Advanced configurations in SAM

Anders Nielsen & Olav Breivik

an@aqua.dtu.dk

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
- Options w.r.t. observation correlation should possibly have been categorized as 'basic'

Correlated observations

- Setting this option require two fields `$obsCorStruct` and `$keyCorObs`
- If only the first is set to `"AR"`, then is does not work (interface design flaw?)
- Example:

```
$obsCorStruct
# Covariance structure for each fleet ("ID" independent, "AR" AR(1), or "US" for unstructured).
# Possible values are: "ID" "AR" "US"
 "ID" "US" "AR"

$keyCorObs
# Coupling of correlation parameters can only be specified if the AR(1) structure is chosen above.
# NA's indicate where correlation parameters can be specified (-1 where they cannot).
#V1 V2 V3 V4 V5 V6 V7 V8 V9 V10
  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA
  NA  NA  NA  NA  NA  -1  -1  -1  -1  -1
   0   1   1   1   1   2  -1  -1  -1  -1
```

- The bottom coupling matrix must only be filled in if the AR-structure is used.
- This matrix only has columns for successive paris of ages (one less than number of ages)
- Set to `NA` if unstructured or independent

Correlated observations: (Ir)regular grid AR

- The observation vector $o_y^{(f)}$ for fleet f in year y is assumed $o_y^{(f)} \sim \mathcal{N}(\mu_y^{(f)}, \Sigma)$
- In the regular AR structure the covariance is defined as:

$$\Sigma_{ij} = \rho^{|i-j|}\sqrt{\Sigma_{ii}\Sigma_{jj}}$$

- So correlation only depende on distance between i and j, not which i and j.
- First realize that we can get the same covariance structure by:

$$\Sigma_{ij} = 0.5^{\alpha|i-j|}\sqrt{\Sigma_{ii}\Sigma_{jj}} \quad , \quad \text{where} \quad \alpha > 0$$

- Notice that this implies a regular grid.
- We can extend this structure by defining

$$\Sigma_{ij} = 0.5^{|\theta_i-\theta_j|}\sqrt{\Sigma_{ii}\Sigma_{jj}} \quad , \quad \text{where} \quad \theta_1 = 0 \leq \theta_2 \leq \cdots \leq \theta_A$$

- This corresponds to having the points on an irregular grid.
- If all deltas are the same, then it is a regular AR structure

Correlated observations: Unstructured covariance

- The fully unstructured covariance can be constructured in the following way.

$$\Sigma_{ij} = (D^{-\frac{1}{2}} L L^t D^{-\frac{1}{2}})_{ij} \sqrt{\Sigma_{ii}\Sigma_{jj}}$$

- Here L is a lower triangle matrix (Cholesky of the correlation) and D is the diagonal matrix of (LL^t)

$$L = \begin{pmatrix} 1 & 0 & 0 \\ \theta_1 & 1 & 0 \\ \theta_2 & \theta_3 & 1 \end{pmatrix}$$

- The model parameters are the elements in L and the log-standard deviations
- This is very flexible, but also requires a lot of parameters to be estimated
- Now we have a lot of options (ID, AR, IGAR, US)
- How can we go about choosing an optimal configuration?
- Useful diagnostics: `res <- residuals(fit)`, `plot(res)`, `empirobscorrplot(res)`

Density dependent survey catchability

- Normally survey observations are predicted by:

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

- 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:

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

- The need for this option could be visible from trends over time in residuals
- Mostly considered for stocks with extreme recruitment events

---

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

Catch scaling

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

```
$noScaledYears
# Number of years where catch scaling is applied.
 5

$keyScaledYears
# A vector of the years where catch scaling is applied.
 2001 2002 2003 2004 2005

$keyParScaledYA
# A matrix specifying the couplings of scale parameters (nrow = no scaled years, ncols = no ages).
   0   0   0   0   0   0
   1   1   1   1   1   1
   2   2   2   2   2   2
   3   3   3   3   3   3
   4   4   4   4   4   4
```

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

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

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

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:

```
$keyBiomassTreat
# To be defined only if a biomass survey is used (0 SSB index, 1 catch index, 2 FSB index, 3 total catch,
# 4 total landings, 5 TSB index, 6 TSN index, and 10 Fbar idx).
 -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: $\log(\text{SSB}_y) \sim \mathcal{N}(\log(q^{(s)}\widehat{\text{SSB}_y}), (\sigma^{(s)})^2)$

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

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

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

Observational likelihoods

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 Logisitc 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

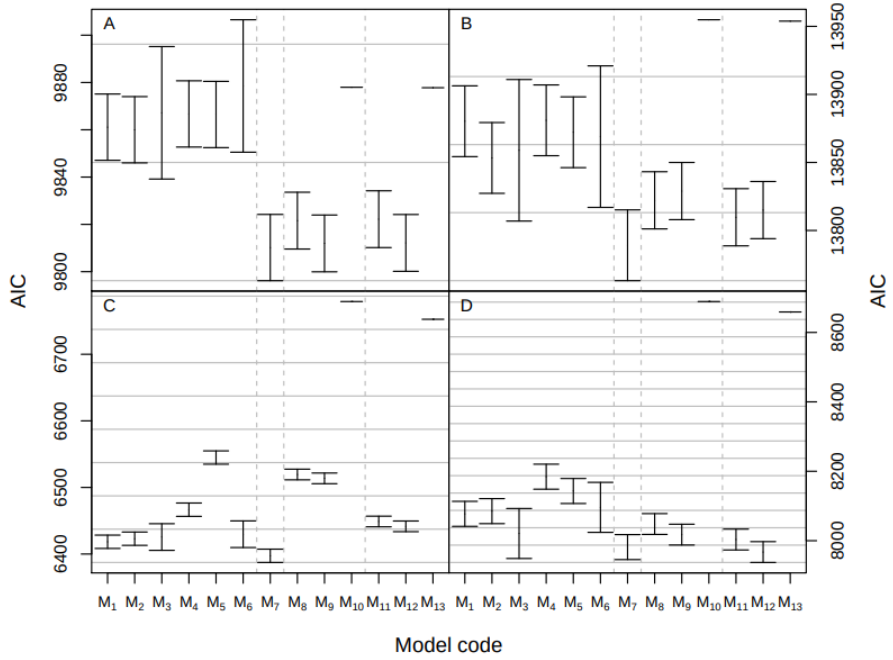


Figure 3: AIC intervals for models M_1 to M_{13} (Table 1) in the case studies: Blue Whiting (A), North-East Arctic Haddock (B), North Sea Cod (C), and Northern Shelf Haddock (D). The horizontal grey lines indicate AIC differences of 50 starting at the lowest lower bound of the models. Vertical dashed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

```
$constRecBreaks
# 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.
# 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)
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)

Link between observation mean and variance

- When using the log-normal the relationship between mean μ and variance v is $v = \alpha\mu^2$
- This relationship may not be correct, so instead the power β can be estimated in $v = \alpha\mu^\beta$

```
$predVarObsLink
# Coupling of parameters used in a prediction-variance link for observations.
   0   1   2   2   2   2   2   2   2   2   2
   3   3   3   3   3   3  NA  NA  NA  NA  NA
  -1  -1  -1  -1  -1  -1  -1  NA  NA  NA  NA
```

- In the configuration above this is configured for the first two fleets
- Used in situations where the size of e.g. catches vary greatly in the time period
- Residuals can also be inspected (plot residual versus predicted)

Prediction–variance relation in a state-space fish stock assessment model

Olav Nikolai Breivik [1,*], Anders Nielsen [2], and Casper W. Berg [2]

[1] *Norwegian Computing Center, Gaustadaleen 23A, 0373 Oslo, Norway*
[2] *National Institute of Aquatic Resources, Technical University of Denmark, Kemitorvet, 2800 Kgs. Lyngby, Denmark*

* *Corresponding author: tel: +47 22852558; e-mail: olavbr@nr.no*

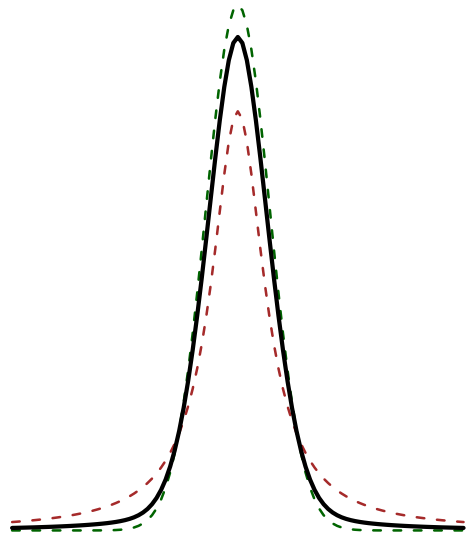
Breivik, O. N., Nielsen, A., and Berg, C. W. Prediction–variance relation in a state-space fish stock assessment model. – ICES Journal of Marine Science, 78: 3650–3657.

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).

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

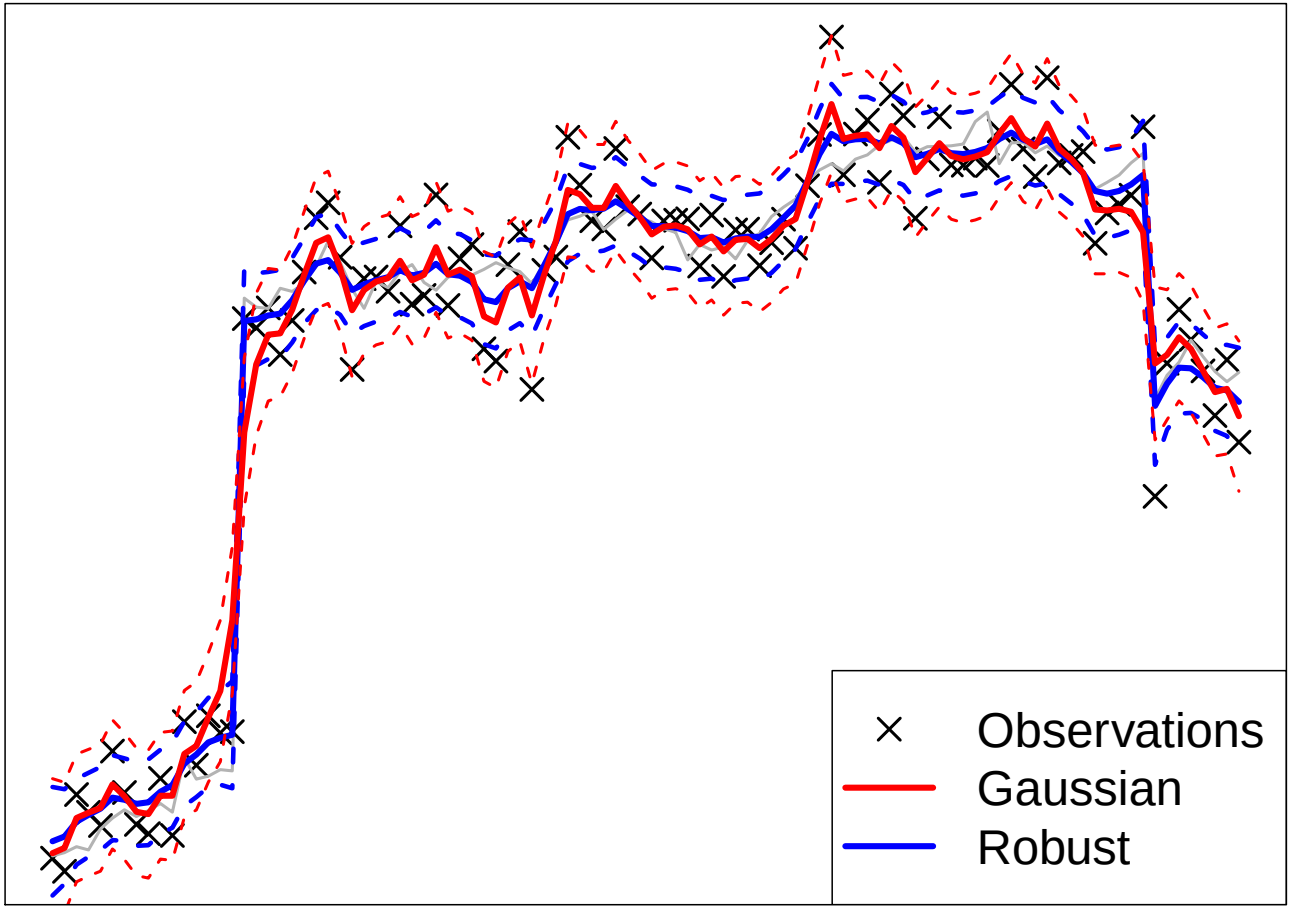
- Robust distributions can also be used for incerments in N, and F processes

```
$fracMixN
# The fraction of t(3) distribution used in logF increment distribution
 0

$fracMixF
# The fraction of t(3) distribution used in logN increment distribution (for each age group)
 0
```

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

Robust increments — does exactly what we want



Estimating observation noise for selected observations

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

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

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

$matureModel
# 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)
 1
$keyMatureMean
# Coupling of mature process mean parameters (not used if matureModel==0)
 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 bilogical 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:

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

- Primitive, but works (demonstrate)