Basic configurations

Olav N.R Breivik

Recap of standard SAM run:

```
library(stockassessment)
cn <- read.ices("cn.dat")
cw <- read.ices("cw.dat")
dw <- read.ices("dw.dat")
lw <- read.ices("lw.dat")
mo <- read.ices("mo.dat")
nm <- read.ices("nm.dat")
pf <- read.ices("pf.dat")
pm <- read.ices("pm.dat")
sw <- read.ices("sw.dat")
lf <- read.ices("lf.dat")
surveys <- read.ices("survey.dat")

dat <- setup.sam.data(surveys=surveys,
                    residual.fleet=cn,
                    prop.mature=mo,
                    stock.mean.weight=sw,
                    catch.mean.weight=cw,
                    dis.mean.weight=dw,
                    land.mean.weight=lw,
                    prop.f=pf,
                    prop.m=pm,
                    natural.mortality=nm,
                    land.frac=lf)

conf<- defcon(dat)
par <- defpar(dat,conf)
fit <- sam.fit(dat, conf, par)
saveConf(fit$conf,file = "conf.cfg")

#Modify conf.cfg and load it
conf2 = loadConf(dat,file = "conf.cfg")
par2 <- defpar(dat,conf2)
fitNew <- sam.fit(dat, conf2, par2)
```

Basic configurations

We will now elaborate basic configurations for:

- Recruitment function
- Survival process
- Fishing mortality process
- Survey catchability
- Observation uncertainty
- Prediction-variance relation
- Correlated observations
- Age plus groups in surveys and catch

State space stock assessment

We assume the standard stock equations:

$$\begin{aligned}
\log N_{1,y} &= \log R(\mathbf{N}_{y-1}) + \eta_{1,y} \\
\log N_{a,y} &= \log N_{a-1,y-1} - F_{a-1,y-1} - M_{a-1,y-1} + \eta_{a,y} \\
\log N_{A,y} &= \log(N_{A-1,y-1} e^{-F_{A-1,y-1} - M_{A-1,y-1}} + N_{A,y-1} e^{-F_{A,y-1} - M_{A,y-1}}) + \eta_{A,y}
\end{aligned}$$

were

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

Observe:

$$\begin{aligned}
\log C_{a,y} &\sim N(\mu_{C_{a,y}}, \sigma^2_{c,a}) \\
\log I^{(s)}_{a,y} &\sim N(\mu^{(s)}_{I_{a,y}}, \sigma^2_{s,a})
\end{aligned}$$

Configurations automatically generated based on data:

- minAge

```
$minAge
# The minimium age class in the assessment
     3
```

- maxAge

```
$maxAge
# The maximum age class in the assessment
     11
```

These provide the maximum and minimum age in the assessment data.

Note: These should not be changed without changing the input data accordingly

Recruitment function

We assume the standard stock equations:

$$
\begin{aligned}
\log N_{1,y} &= \log R(\mathbf{N}_{y-1}) + \eta_{1,y} \\
\log N_{a,y} &= \log N_{a-1,y-1} - F_{a-1,y-1} - M_{a-1,y-1} + \eta_{a,y} \\
\log N_{A,y} &= \log(N_{A-1,y-1} e^{-F_{A-1,y-1} - M_{A-1,y-1}} + N_{A,y-1} e^{-F_{A,y-1} - M_{A,y-1}}) + \eta_{A,y}
\end{aligned}
$$

were

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

Observe:

$$
\begin{aligned}
\log C_{a,y} &\sim N(\mu_{C_{a,y}}, \sigma^2_{c,a}) \\
\log I^{(s)}_{a,y} &\sim N(\mu^{(s)}_{I_{a,y}}, \sigma^2_{s,a})
\end{aligned}
$$

Example with random walk:

```
$stockRecruitmentModelCode
     0
```

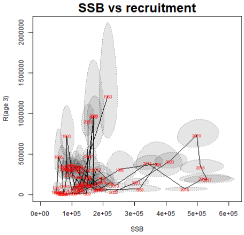
Recruitment functions

- 0: plain random walk $\log N_{1,y} \sim N(\log N_{1,y-1}, \sigma_r^2)$
- 1: Ricker $\log N_{1,y} \sim N(f_r(SSB_{y-1}), \sigma_r^2)$
- 2: Beverton-Holt $\log N_{1,y} \sim N(f_b(SSB_{y-1}), \sigma_r^2)$
- 3: Piece-wise constant $\log N_{1,y} \sim N(\mu_y, \sigma_r^2)$

Example with random walk:

```
$stockRecruitmentModelCode
      0
```

- Recruitment is difficult to predict
- More options included that are typically used in MSY analysis.



Recruitment and survival process

We assume the stock equations:

$$
\begin{aligned}
\log N_{1,y} &= \log R(\mathbf{N}_{y-1}) + \eta_{1,y} \\
\log N_{a,y} &= \log N_{a-1,y-1} - F_{a-1,y-1} - M_{a-1,y-1} + \eta_{a,y} \\
\log N_{A,y} &= \log(N_{A-1,y-1} e^{-F_{A-1,y-1} - M_{A-1,y-1}} + N_{A,y-1} e^{-F_{A,y-1} - M_{A,y-1}}) + \eta_{A,y}
\end{aligned}
$$

were

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

Observe:

$$
\begin{aligned}
\log C_{a,y} &\sim N(\mu_{C_{a,y}}, \sigma^2_{c,a}) \\
\log I^{(s)}_{a,y} &\sim N(\mu^{(s)}_{I_{a,y}}, \sigma^2_{s,a})
\end{aligned}
$$

```
$keyVarLogN
# Coupling of the recruitment and survival process variance parameters
     0 1 1 1 1 1 1 1 1 1 1
```

Recruitment and survival process

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

```
$keyVarLogN
# Coupling of the recruitment and survival process variance parameters
     0 1 1 1 1 1 1 1 1 1 1
```

- Typically a separate variance parameter for recruitment and survival
- Typically a common variance parameter for survival

Fishing mortality process

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:

$$\log C_{a,y} \sim N(\mu_{C_{a,y}}, \sigma^2_{c,a})$$
$$\log I^{(s)}_{a,y} \sim N(\mu^{(s)}_{I_{a,y}}, \sigma^2_{s,a})$$

```
$keyLogFsta
# Coupling of the fishing mortality states processes for each age
     0    1    2    3    4    5    5    5    5    5    5
    -1   -1   -1   -1   -1   -1   -1   -1   -1   -1   -1
```

Fishing mortality process

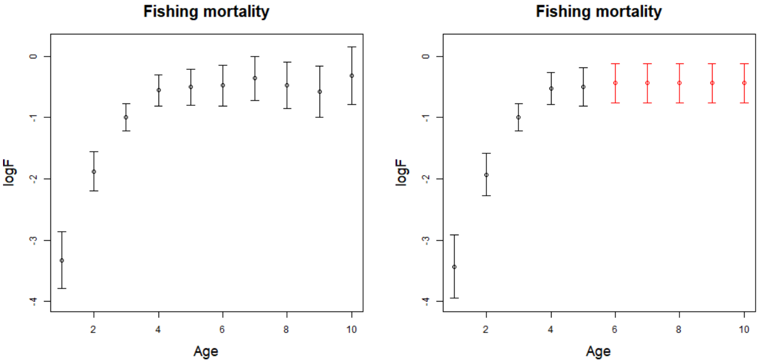
$$\log N_{1,y} = \log R(\mathbf{N}_{y-1}) + \eta_{1,y}$$
$$\log N_{a,y} = \log N_{a-1,y-1} - F_{a-1,y-1} - M_{a-1,y-1} + \eta_{a,y}$$
$$\log N_{A,y} = \log(N_{A-1,y-1} e^{-F_{A-1,y-1} - M_{A-1,y-1}} + N_{A,y-1} e^{-F_{A,y-1} - M_{A,y-1}}) + \eta_{A,y}$$

```
$keyLogFsta
# Coupling of the fishing mortality states processes for each age
     0   1   2   3   4   5   5   5   5   5   5
    -1  -1  -1  -1  -1  -1  -1  -1  -1  -1  -1
```

- The configuration *keyLogFsta* defines how we couple the fishing mortality states.
- Here $(F_{1,y}, ..., F_{11,y}) = (F_{1,y}, ..., F_{5,y}, F_{6,y}, ..., F_{6,y})$
- Often a flat selectivity pattern for older ages is selected

Fishing mortality process

```
$keyLogFsta
# Coupling of the fishing mortality states processes for each age
     0   1   2   3   4   5   5   5   5   5   5
    -1  -1  -1  -1  -1  -1  -1  -1  -1  -1  -1
```



- Defines dimension of latent **F**
- Does not effect number of model parameters

Fishing mortality variance

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

$$\log \mathbf{F}_y = \log \mathbf{F}_{y-1} + \epsilon_y^F.$$

Observe:

$$\log C_{a,y} \sim N(\mu_{C_{a,y}}, \sigma^2_{c,a})$$
$$\log I^{(s)}_{a,y} \sim N(\mu^{(s)}_{I_{a,y}}, \sigma^2_{s,a})$$

```
$keyVarF
# Coupling of process variance parameters for log(F)-process
     0    1    1    1    1    1    1    1    1    1    1
     -1   -1   -1   -1   -1   -1   -1   -1   -1   -1   -1
```

Fishing mortality variance

$$\log \mathbf{F}_y = \log \mathbf{F}_{y-1} + \epsilon_y^F.$$

```
$keyVarF
# Coupling of process variance parameters for log(F)-process
     0   1   1   1   1   1   1   1   1   1   1
    -1  -1  -1  -1  -1  -1  -1  -1  -1  -1  -1
```

- The configuration *keyVarF* defines how we couple the variances of the fishing mortality increments.
- Here two variances parameters are included: $\sigma_{1,F}$ and $\sigma_{2,F} = \sigma_{3,F} = \cdots = \sigma_{11,F}$

Fishing mortality correlation

We assume the standard stock equations:

$$
\begin{aligned}
\log N_{1,y} &= \log R(\mathbf{N}_{y-1}) + \eta_{1,y} \\
\log N_{a,y} &= \log N_{a-1,y-1} - F_{a-1,y-1} - M_{a-1,y-1} + \eta_{a,y} \\
\log N_{A,y} &= \log(N_{A-1,y-1} e^{-F_{A-1,y-1} - M_{A-1,y-1}} + N_{A,y-1} e^{-F_{A,y-1} - M_{A,y-1}}) + \eta_{A,y}
\end{aligned}
$$

were

$$\log \mathbf{F}_y = \log \mathbf{F}_{y-1} + \epsilon_y^F.$$

Observe:

$$
\begin{aligned}
\log C_{a,y} &\sim N(\mu_{C_{a,y}}, \sigma^2_{c,a}) \\
\log I^{(s)}_{a,y} &\sim N(\mu^{(s)}_{I_{a,y}}, \sigma^2_{s,a})
\end{aligned}
$$

```
$corFlag
# Correlation of fishing mortality across ages (0 independent, 1 compound symmetry,
# 2 AR(1),
    2
```

Fishing mortality correlation

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

```
$corFlag
# Correlation of fishing mortality across ages (0 independent, 1 compound symmetry,
# 2 AR(1),
     2
```

The configuration *corflag* defines the correltion structure of the fishing mortality increments.

- 0: Independent - no correlation between F across age
- 1: Compound symmetry - equal correlation across all ages
- 2: AR1 - ages close to each other are more correlated

Survey catchability

We assume the standard stock equations:

$$
\begin{aligned}
\log N_{1,y} &= \log R(\mathbf{N}_{y-1}) + \eta_{1,y} \\
\log N_{a,y} &= \log N_{a-1,y-1} - F_{a-1,y-1} - M_{a-1,y-1} + \eta_{a,y} \\
\log N_{A,y} &= \log(N_{A-1,y-1} e^{-F_{A-1,y-1} - M_{A-1,y-1}} + N_{A,y-1} e^{-F_{A,y-1} - M_{A,y-1}}) + \eta_{A,y}
\end{aligned}
$$

were

$$\log \mathbf{F}_y = \log \mathbf{F}_{y-1} + \epsilon_y^F.$$

Observe:

$$
\begin{aligned}
\log C_{a,y} &\sim N(\mu_{C_{a,y}}, \sigma^2_{c,a}) \\
\log I^{(s)}_{a,y} &\sim N(\log Q^{(s)}_a N_{a,y}, \sigma^2_{s,a,y})
\end{aligned}
$$

```
$keyLogFpar
# Coupling of the survey catchability parameters
     -1  -1  -1  -1  -1  -1  -1  -1  -1  -1  -1
      0   1   1   1   1   2   2   2  -1  -1  -1
```

Survey catchability

- The catchability parameter is the proportionality constant between stock size and the observed index.
- *keyLogFpar* defines how the catchability parameters are coupled across age groups.

Survey observation equation:

$$\log I_{a,y}^{(s)} \sim N(\log Q_a^{(s)} N_{a,y}, \sigma_{s,a,y}^2)$$

```
$keyLogFpar
# Coupling of the survey catchability parameters
     -1   -1   -1   -1   -1   -1   -1   -1   -1   -1   -1
      0    1    1    1    1    2    2    2   -1   -1   -1
```

Here it is included a unique catchability for the youngest age, a common catchability for the four next ages, and a common catchability for the older ages.

Observation variance

We assume the standard stock equations:

$$
\begin{aligned}
\log N_{1,y} &= \log R(\mathbf{N}_{y-1}) + \eta_{1,y} \\
\log N_{a,y} &= \log N_{a-1,y-1} - F_{a-1,y-1} - M_{a-1,y-1} + \eta_{a,y} \\
\log N_{A,y} &= \log(N_{A-1,y-1} e^{-F_{A-1,y-1} - M_{A-1,y-1}} + N_{A,y-1} e^{-F_{A,y-1} - M_{A,y-1}}) + \eta_{A,y}
\end{aligned}
$$

were

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

Observe:

$$
\begin{aligned}
\log C_{a,y} &\sim N(\mu_{C_{a,y}}, \sigma^2_{c,a}) \\
\log I^{(s)}_{a,y} &\sim N(\mu^{(s)}_{I_{a,y}}, \sigma^2_{s,a})
\end{aligned}
$$

```
$keyVarObs
# Coupling of the variance parameters for the observations.
     0    1    2    2    2    2    2    2    2    2    2
     3    3    3    3    3    3    3    3   -1   -1   -1
```

Observation variance

- Observations include noise
- *keyVarObs* couples the observation variance parameters

Observation equations:

$$\log C_{a,y} \sim N(\mu_{C_{a,y}}, \sigma^2_{c,a})$$
$$\log I^{(s)}_{a,y} \sim N(\mu^{(s)}_{I_{a,y}}, \sigma^2_{s,a})$$

```
$keyVarObs
# Coupling of the variance parameters for the observations.
     0    1    2    2    2    2    2    2    2    2    2
     3    3    3    3    3    3    3    3   -1   -1   -1
```

Four variance parameters included

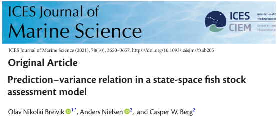
- Three for catch: $\sigma_{c,1}$, $\sigma_{c,2}$ and $\sigma_{c,3} = \cdots \sigma_{c,11}$
- One for the survey: $\sigma_{s,1} = \cdots \sigma_{s,8}$

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)
- Go trough `SAM-predVar.pdf`


ICES Journal of Marine Science

ICES Journal of Marine Science (2021), 78(10), 3650–3657. https://doi.org/10.1093/icesjms/fsab205

Original Article

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

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

Observation correlation

We assume the standard stock equations:

$$
\begin{aligned}
\log N_{1,y} &= \log R(\mathbf{N}_{y-1}) + \eta_{1,y} \\
\log N_{a,y} &= \log N_{a-1,y-1} - F_{a-1,y-1} - M_{a-1,y-1} + \eta_{a,y} \\
\log N_{A,y} &= \log(N_{A-1,y-1} e^{-F_{A-1,y-1}-M_{A-1,y-1}} + N_{A,y-1} e^{-F_{A,y-1}-M_{A,y-1}}) + \eta_{A,y}
\end{aligned}
$$

were

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

Observe:

$$
\begin{aligned}
\log \mathbf{C}_y &\sim \mathcal{N}(\boldsymbol{\mu}_{C_y}, \boldsymbol{\Sigma_c}) \\
\log \mathbf{I}_y^{(s)} &\sim \mathcal{N}(\boldsymbol{\mu}_y^{(s)}, \boldsymbol{\Sigma_s})
\end{aligned}
$$

```
$obsCorStruct
# Possible values are: "ID" "AR" "US"
"ID" "US" "AR"
```

```
$keyCorObs
#3-4 4-5 5-6 6-7 7-8 8-9 9-10 10-11 11-12 12-13
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
```

Correlated observations

- Setting this option require two fields `$obsCorStruct` <u>and</u> `$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

Independent observations

From the example on the previousl slide, we then assume:

$$\log \mathbf{C}_y \sim \mathcal{N}\left(\boldsymbol{\mu}_{C_y}, \boldsymbol{\Sigma}\right) \tag{1}$$

where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2_{C,a_1} & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2_{C,a_2} & 0 & \cdots & 0 \\ 0 & 0 & \sigma^2_{C,a_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2_{C,A} \end{pmatrix}$$

(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

Unstructured covariance

- The fully unstructured covariance can be constructed 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)`

Plus groups

Plus group age can very between fleets

```
$maxAgePlusGroup
# Is last age group considered a plus group for each fleet (1 yes, or 0 no)
1 1
```

Here both catch and survey consist of observations where the oldest age group contains that age and above.

The plus group implies the following update to the stock equation:

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

- If A_f is less than A, then the stock number used to predict observations at age A_f for the fleet is the sum of the stock numbers from A_f to A.

F-bar range

Age range in $\bar{F}$ is controlled by *fbarRange*

```
$fbarRange
# lowest and higest age included in Fbar
4 7
```

Here ages 4 to 7 are included in $\bar{F}$, i.e.,:

$$\bar{F}_y = \frac{1}{4}\sum_{a=4}^{7} F_{a,y}$$