

# Basic configurations

Olav N.R Breivik



**NR**  Norsk Regnesentral  
ANVENDT DATAFORSKNING

## Recap of standard SAM run:

```
1 library(stockassessment)
2 cn <- read.ices("cn.dat")
3 cw <- read.ices("cw.dat")
4 dw <- read.ices("dw.dat")
5 lw <- read.ices("lw.dat")
6 mo <- read.ices("mo.dat")
7 nm <- read.ices("nm.dat")
8 pf <- read.ices("pf.dat")
9 pm <- read.ices("pm.dat")
10 sw <- read.ices("sw.dat")
11 lf <- read.ices("lf.dat")
12 surveys <- read.ices("survey.dat")
13
14 dat <- setup.sam.data(surveys=surveys,
15                       residual.fleet=cn,
16                       prop.mature=mo,
17                       stock.mean.weight=sw,
18                       catch.mean.weight=cw,
19                       dis.mean.weight=dw,
20                       land.mean.weight=lw,
21                       prop.f=pf,
22                       prop.m=pm,
23                       natural.mortality=nm,
24                       land.frac=lf)
25
26 conf<- defcon(dat)
27 par <- defpar(dat, conf)
28 fit <- sam.fit(dat, conf, par)
29 saveConf(fit$conf, file = "conf.cfg")
30
31 #Modify conf.cfg and load it
32 conf2 = loadConf(dat, file = "conf.cfg")
33 par2 <- defpar(dat, conf2)
34 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:

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

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

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

were

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

Observe:

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

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

Configurations automatically generated based on data:

- minAge

```
1 $minAge
2 # The minimum age class in the assessment
3     3
```

- maxAge

```
1 $maxAge
2 # The maximum age class in the assessment
3     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:

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

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

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

were

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

Observe:

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

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

Example with random walk:

```
1 $stockRecruitmentModelCode
2 0
```

# Recruitment functions

- 0: plain random walk
- 1: Ricker
- 2: Beverton-Holt
- 3: Piece-wise constant

$$\log N_{1,y} \sim N(\log N_{1,y-1}, \sigma_r^2)$$

$$\log N_{1,y} \sim N(f_r(SSB_{y-1}), \sigma_r^2)$$

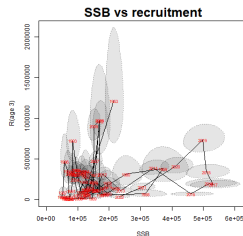
$$\log N_{1,y} \sim N(f_b(SSB_{y-1}), \sigma_r^2)$$

$$\log N_{1,y} \sim N(\mu_y, \sigma_r^2)$$

Example with random walk:

```
1 $stockRecruitmentModelCode  
2 0
```

- Recruitment is difficult to predict
- More options included that are typically used in MSY analysis, see the course material.



# Recruitment and survival 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:

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

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

```
1 $keyVarLogN
2 # Coupling of the recruitment and survival process variance parameters
3   0 1 1 1 1 1 1 1 1 1 1
```



# Recruitment and survival process

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

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

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

```
1 $keyVarLogN
2 # Coupling of the recruitment and survival process variance parameters
3 0 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:

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

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

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

# Fishing mortality process

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

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

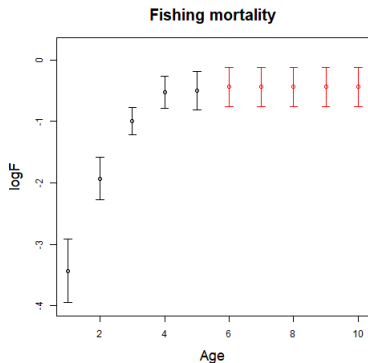
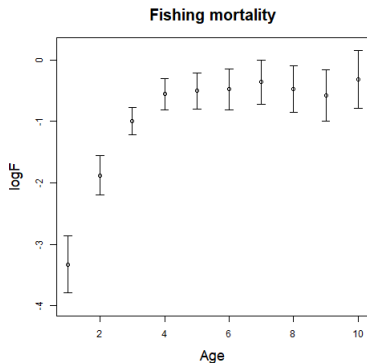
$$\log N_{A,y} = \log(N_{A-1,y-1} e^{-F_{A-1,y-1} - M_{A-1,y-1}} + N_{A,y-1} e^{-F_{A,y-1} - M_{A,y-1}}) + \eta_{A,y}$$

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

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

# Fishing mortality process

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

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

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

```
1 $keyVarF
2 # Coupling of process variance parameters for log(F)-process
3   0   1   1   1   1   1   1   1   1   1   1
4  -1  -1  -1  -1  -1  -1  -1  -1  -1  -1  -1
```

# Fishing mortality variance

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

```
1 $keyVarF
2 # Coupling of process variance parameters for log(F)-process
3   0   1   1   1   1   1   1   1   1   1   1
4   -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  $\sigma_{2,F} = \sigma_{3,F} = \dots = \sigma_{11,F}$

# Fishing mortality correlation

We assume the standard 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:

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

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

```
1 $corFlag
2 # Correlation of fishing mortality across ages (0 independent, 1 compound symmetry,
3 # 2 AR(1),
4     2
```

# Fishing mortality correlation

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

```
1 $corFlag  
2 # Correlation of fishing mortality across ages (0 independent, 1 compound symmetry,  
3 # 2 AR(1),  
4     2
```

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

- 0: Independent - no correlation between  $F$  across age
- 1: Compound symmetry - equal correlation across all ages
- 2: AR1 - similar ages are more correlated



# Survey catchability

We assume the standard 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

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

Observe:

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

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

```
1 $keyLogFpar
2 # Coupling of the survey catchability parameters
3   -1  -1  -1  -1  -1  -1  -1  -1  -1  -1  -1
4   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:

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

```
1 $keyLogFpar
2 # Coupling of the survey catchability parameters
3   -1  -1  -1  -1  -1  -1  -1  -1  -1  -1  -1
4   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:

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

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

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

were

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

Observe:

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

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

```
1 $keyVarObs
2 # Coupling of the variance parameters for the observations.
3   0   1   2   2   2   2   2   2   2   2   2
4   3   3   3   3   3   3   3   3  -1  -1  -1
```

# Observation variance

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

Observation equations:

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

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

```
1 $keyVarObs
2 # Coupling of the variance parameters for the observations.
3     0   1   2   2   2   2   2   2   2   2   2
4     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  $\sigma_{c,3} = \dots \sigma_{c,11}$
- One for the survey:  $\sigma_{s,1} = \dots \sigma_{s,8}$

# Link between observation mean and variance

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

```
1 $predVarObsLink
2 # Coupling of parameters used in a prediction-variance link for observations.
3 0 1 2 2 2 2 2 2 2 2
4 3 3 3 3 3 3 NA NA NA NA
5 -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 (2021), 78(36), 3659–3657. <https://doi.org/10.1093/icesjms/fsab205>

## Original Article

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

Olav Nikolai Breivik<sup>1,\*</sup>, Anders Nielsen<sup>2</sup>, and Casper W. Berg<sup>2</sup>

# Observation correlation

We assume the standard 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

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

Observe:

$$\log \mathbf{C}_y \sim \mathcal{N}(\boldsymbol{\mu}_{C_y}, \boldsymbol{\Sigma}_c)$$

$$\log \mathbf{I}_y^{(s)} \sim \mathcal{N}(\boldsymbol{\mu}_y^{(s)}, \boldsymbol{\Sigma}_s)$$

```
1 $obsCorStruct
2 # Possible values are: "ID" "AR" "US"
3 "ID" "US" "AR"
```

```
1 $keyCorObs
2 #3-4 4-5 5-6 6-7 7-8 8-9 9-10 10-11 11-12 12-13
3 NA NA NA NA NA NA NA NA NA NA
4 NA NA NA NA NA -1 -1 -1 -1 -1
5 0 1 1 1 1 2 -1 -1 -1 -1
```

# Correlated observations

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

```
1 $obsCorStruct
2 #Covariance structure for each fleet ("ID" independent, "AR" AR(1), or "US" for
   unstructured).
3 # Possible values are: "ID" "AR" "US"
4 "ID" "US" "AR"
5
6 $keyCorObs
7 # 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).
8 #V1 V2 V3 V4 V5 V6 V7 V8 V9 V10
9 NA NA NA NA NA NA NA NA NA NA
10 NA NA NA NA NA -1 -1 -1 -1 -1
11 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 pairs of ages (one less than number of ages)
- Set to NA if unstructured or independent

# Independent observations

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

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

where

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



# (Ir)regular grid AR

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

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

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

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

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

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

- 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), empirobcorrplot(res)`

# Plus groups

Plus group age can vary between fleets

```
1 $maxAgePlusGroup  
2 # Is last age group considered a plus group for each fleet (1 yes, or 0 no)  
3 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*

```
1 $fbarRange  
2 # lowest and highest age included in Fbar  
3 4 7
```

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

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