

# Introduction: Stock assessment modeling

Olav N.R. Breivik

*Thanks to Anders Nielsen for borrowing course material*



Norsk Regnesentral  
ANVENDT DATAFORSKNING

# Overview

- What is an assessment model
- What is a statistical assessment model
- What is a statistical state-space assessment model
  - SAM is a state-space assessment model
  - We will implement the official NEA cod model during the day!

# Fish stock assessment

## Want to know:

- How many fish are in the ocean?
- What is the fishing selectivity?

## Data:

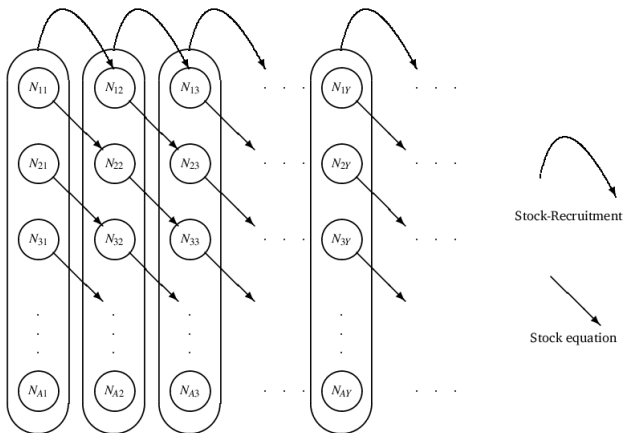
- $C_{a,y}$ : Yearly catch-at-age (Typically calculated with ECA in Norway)
- $I_{a,y}$ : Abundance indices-at-age

## NEA haddock case study:

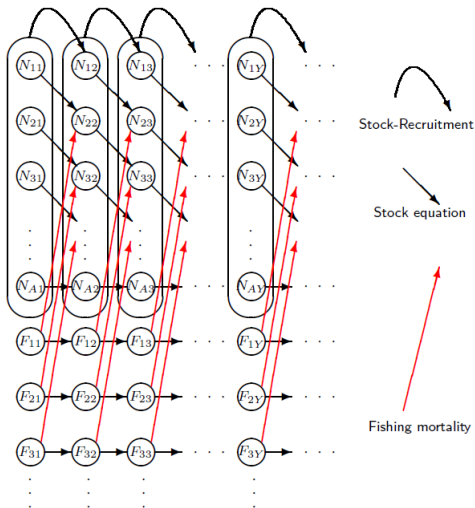
	Year 1950–2023									
Age e.g. 3–13+										

	Year 1994–2024									
Age e.g. 3–10+										

# Population dynamic model



# Population dynamic model



# Population dynamic model

**Stock equation:** The number of fish in a cohort is expected to follow:

$$\frac{dN_t}{dt} = -(F_t + M_t)N_t$$

If  $F$  and  $M$  are assumed constant within each year we get:

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

- Typically included a plus age group

**Stock–recruitment:** Obviously connected — Difficult to estimate

- Often assumed random walk
- Several options we will elaborate tomorrow

## Data:

- $C_{a,y}$ : Yearly catch-at-age (Typically calculated with ECA in Norway)
- $I_{a,y}$ : Abundance indices-at-age

NEA haddock case study:

	Year 1950–2023									
Age e.g. 3–13+										

	Year 1994–2024									
Age e.g. 3–10+										

# Observation equations

**Catch equation:** The number of fish in a cohort after one year can be separated into:

$$N_{a,y} = \underbrace{N_{a+1,y+1}}_{\text{survived}} + \underbrace{C_{a,y} + D_{a,y}}_{\text{died}}$$

The part of the cohort that was caught during the year is:

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

**Index equation:** The number of fish-at-age expected in index:

$$I_{a,y} = Q_a \underbrace{e^{-(F_{a,y} + M_{a,y})\tau}}_{\text{survived in year}} N_{a,y}$$

where  $\tau \in (0, 1)$  is time of survey



# Deterministic example

Imagine that a cohort consists of  $N_0 = 1000$  fish in year 0. We know that  $M = 0.2$  and  $F = 0.4$ .

```
1 N0 <- 1000; F <- 0.4; M <- 0.2
```

- We can calculate catch in year 0 and abundance in start of year 1:

```
1 C0 <- F / (F+M) * (1-exp(-F-M)) * N0  
2 N1 <- N0 * exp(-F-M)
```

- We can calculate catch in year 1 and abundance in start of year 2:

```
1 C1 <- F / (F+M) * (1-exp(-F-M)) * N1  
2 N2 <- N1 * exp(-F-M)
```

- **Exercise 1a:** Calculate catch and abundance the 10 first years
- **Exercise 1b:** How many percentage dies because of fishery each year?

See `exampleDeterministic.R` to get you started

# Deterministic model

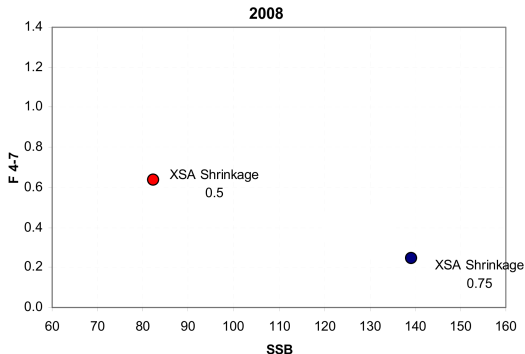
- A deterministic model ignores observation noise
- Examples:
  - Virtual population analysis (VPA)
  - The XSA-model
- These models works roughly speaking by
  - 0: Guess the number of survivors  $N_{A+1,y}$  and  $N_{a,Y+1}$
  - 1: Back calculate all  $N_{a,y}$  by subtracting catch and natural mortality
  - 2: Use surveys to adjust all  $N_{a,y}$  and update survivors accordingly
  - 3: Repeat 1-3 until survivors converge

We will not do this deterministic approach! But that's how it was done before statistical models took over.

# Pros and cons with deterministic model

- + Very fast to compute
- + Fairly simple to explain the path from data to stock numbers
- Difficult to explain why it converges, and what a solution mean
- These algorithms contain many ad-hoc settings (shrinkage, tapered time weights, ...) that makes them less objective
- No framework for comparing models (different settings)
- No quantification of uncertainties within model
- The assumptions are difficult to identify and verify
- With no clearly defined model more ad-hoc methods are needed to make predictions

# Example: F-shrinkage for Eastern Baltic Cod



- These differences are not small and theoretical
- There are no objective way to choose between these two deterministic approaches
- Things would be simpler if we had a statistical model

# Full parametric statistical assessment model

**Stock equation:** The number of fish in a cohort is expected to follow:

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

where  $Z_{a,y} = F_{a,y} + M_{a,y}$ .

Need to make some simplifications to not have a too flexible model:

- Lets assume:  $F_{a,y} = F_a F_y$

Observations equations:

$$\log(C_{a,y}) \sim \mathcal{N} \left( \log \left( \frac{F_{a,y}}{Z_{a,y}} (1 - e^{-Z_{a,y}}) N_{a,y} \right), \sigma_c^2 \right)$$

$$\log(I_{a,y}) \sim \mathcal{N} \left( \log \left( Q_a e^{-Z_{a,y} T} N_{a,y} \right), \sigma_s^2 \right)$$

# A full parametric statistical model

- A RTMB example is given `basicfsa.R`
- Data in `fsa.RData`
- The code is very short, so lets go trough it together.

# Pros and cons with full parametric model

- + Very fast to compute
- + Fairly simple to explain the path from data to stock numbers
- + Have framework for comparing models (different settings)
- + Have quantification of uncertainties within model
- + Can verify assumptions
- Difficult to include time-varying fishing selectivity without over fitting
- Ad-hoc methods are needed to make predictions

# State-space assessment model

- We studied a RTMB example is given `basicfsa.R`
- **Q1:** What is needed when modifying this implementation to a state-space assessment model?
- We will now introduce state-space models
  - Let's switch to presentation: `randomEffects.pdf`
  - We will be back soon



# State-space assessment model

**Stock equation:** The number of fish in a cohort is expected to follow:

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

We log-transform and get:

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

- Let  $\log N_{a,y}$  be a latent variable
- Let  $\log F_{a,y}$  be a latent variable

# State-space assessment model

In SAM we assume that:

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

Where  $\eta_a \sim \mathcal{N}(0, \sigma_a)$

- $M_{a,y}$  is given as data
- $F_{a,y}$  is estimated as a random walk

# State-space assessment model

In SAM we assume that:

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

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

Where  $\eta_a \sim \mathcal{N}(0, \sigma_a)$

- $M_{a,y}$  is given as data
- $F_{a,y}$  is estimated as a random walk

# State-space assessment model

In SAM we assume that:

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

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

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

Where  $\eta_a \sim \mathcal{N}(0, \sigma_a)$

- $M_{a,y}$  is given as data
- $F_{a,y}$  is estimated as a random walk

# Exercise

- We introduced a full parametric assessment model, see `basicfsa.R`
- We can modify it into a state space assessment model by
  - Letting  $\log F$  and  $\log N$  be latent processes
- **Exercise 3:** Modify `basicfsa.R` into a state space assessment model

## Hints:

- Need to include  $\log N$  and  $\log F$  as parameters
- Need to include process parameters in  $\log N$  and  $\log F$
- Need to calculate the likelihood contribution by the latent process
- Need to inform RTMB that we should marginalize over  $\log N$  and  $\log F$

```
1 obj <- MakeADFun(nll, par, random = c("logF", "logN"))
```

- See `basicfsaStateSpace.R` to get you started

- That is SAM in a nut shell. Not very difficult!
- There are many model extension in SAM:
  - Correlated observations
  - Densitydependent catchability
  - Prediction-variance relation
  - Use of external data on uncertainty
  - Much more ...
    - we will go through everything in detail the next two days
- Validation is done simple in SAM:
  - Residuals
  - Retropective analysis
  - Leave-out runs
  - AIC comparison

# Pros and cons with state space model

- + Very fast to compute
- + Fairly simple to explain the path from data to stock numbers
- + Have framework for comparing models (different settings)
- + Have quantification of uncertainties within model
- + Can verify assumptions
- + Easy to include time-varying fishing selectivity
- + No ad-hoc methods are needed to make predictions
- Difficult to understand?
  - Easy to implement with RTMB

# NEA cod assessment model

- Now we have implemented a state-space assessment model
- Only one small modification needed to make the NEA cod assessment model!
  - Correlated observation
  - See `SAM-NEAcod.R` for an implementation