Fish stock assessment with R

The a4a Initiative

John Doe and friends

2025-01-17

Contents

Chapter 1

cover-image: path to the social sharing image like images/cover.jpg

Placeholder

1.1 License, documentation and development status

1.2 Installing and loading libraries

1.3 How to read this document

1.4 How to get help

1.5 Notation

Chapter 2

Introduction

Placeholder

2.1 The “Assessment for All” Initiative (a4a)

2.2 Multi-stage modelling approach

2.3 Stock Assessment Process

2.4 Stock assessment as a linear model

2.5 Data used in the book

2.5.1 Plaice in area FAO 27, ICES area IV

2.5.2 stock xx

2.5.3 stock xx

Chapter 3

Introduction to Splines

Placeholder

3.1 Understanding Spline Basics

3.2 Generate some artificial data

3.3 Cubic Regression Splines

3.4 Thin plate spline

3.5 Comparison

3.6 The `mgcv` package inside `a4a`

3.7 On the number of knots k

Chapter 4

Modelling Individual Growth and Using Stochastic Slicing to Convert Length-based Data Into Age-based Data

Placeholder

4.1 a4aGr - The growth class

4.2 Adding uncertainty to growth parameters with a multivariate normal distribution}

4.3 Adding uncertainty to growth parameters with a multivariate triangle distribution}

4.4 Adding uncertainty to growth parameters with statistical copulas}

4.5 Converting from length to age based data - the `l2a()` method}

Chapter 5

Modelling Natural Mortality

Placeholder

5.1 `a4aM` - The M class

5.2 Adding uncertainty to natural mortality parameters with a multivariate normal distribution

5.3 Adding uncertainty to natural mortality parameters with statistical copulas

5.4 Computing natural mortality time series - the “m” method

Chapter 6

Stock assessment framework

6.1 Maths description

6.1.1 Observations

The stock assessment model is based on two types of observations: catches, $\hat{C}$, and abundance indices, $\hat{I}$. The model predicts catches at age C_{ay} and indices of abundance I_{ays} for each age a, year y and survey s in the input dataset. It is assumed that the observed catches are normally distributed about the model predictions on the log scale with some observation variance σ^2_{ay}. that is:

$$\log \hat{C}_{ay} \sim \text{Normal}\Big(\log C_{ay}, \sigma^2_{ay} \Big)$$

likewise, the observed survey indices are assumed to be normally distributed about the model predictions on the log scale with some observation variance τ^2_{ays}:

$$\log \hat{I}_{ays} \sim \text{Normal}\Big(\log I_{ays}, \tau^2_{ays} \Big)$$

This leads to the definition of the log-likelihood of the observed catches

$$\ell_C = \sum_{ay} w^{(c)}_{ay} \; \ell_N \Big(\log C_{ay}, \sigma^2_{ay}; \; \log \hat{C}_{ay} \Big)$$

and the log-likelihood of the observed survey indices

$$\ell_I = \sum_{s} \sum_{ay} w^{(s)}_{ays} \; \ell_N \Big(\log I_{ays}, \tau^2_{ays}; \; \log \hat{I}_{ays} \Big)$$

where ℓ_N is the normal log-likelihood function and $w^{(c)}_{ay}$ and $w^{(s)}_{ays}$ are optional weights for the catch and index observations, respectively. The total log-likelihood is then

$$\ell = \ell_C + \ell_I$$

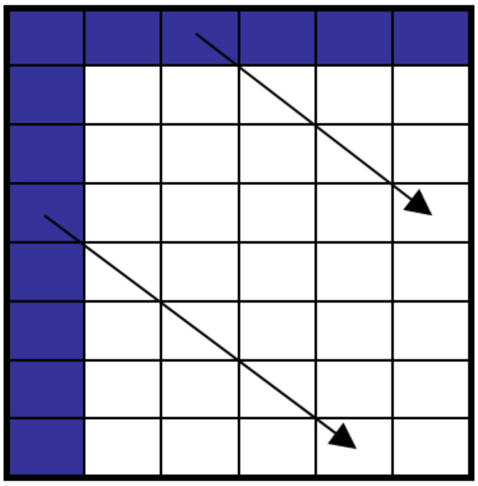
6.1.2 The population dynamics model

To predict catches and survey indices, the model uses the standard population dynamics model

$$N_{a+1,y+1} = N_{ay} \exp\left(-F_{ay} - M_{ay}\right)$$

where N_{ay} is the number of individuals at age a in year y, F_{ay} is the fishing mortality at age a in year y, and M_{ay} is the natural mortality at age a in year y. Any fish that survived beyond the oldest age A in the model are accumulated in the oldest age group and are assumed to be fished at a common rate $F_{A,y}$.

The numbers $N_{a,y}$ are initiated in the first year, $y = 1$ and at the youngest age, $a = 1$, and the matrix of numbers at age are filled in according to the population dynamics model stated above.



Defining $R_y = N_{1,y}$, the numbers at age can be written (ignoring the plus group) as:

Catches in numbers by age and year are defined in terms of the three quantities: natural mortality, fishing mortality and recruitment; using a modified form of the well known Baranov catch equation:

$$C_{ay} = \frac{F_{ay}}{F_{ay} + M_{ay}} \left(1 - e^{-(F_{ay}+M_{ay})}\right) R_y e^{-\sum(F_{ay}+M_{ay})}$$

Survey indices by age and year are defined in terms of the same three quantities with the addition of survey catchability:

$$I_{ays} = Q_{ays} R_y e^{-\sum(F_{ay}+M_{ay})}$$

Observed catches and observed survey indices are assumed to be log-normally distributed, or equivalently, normally distributed on the log-scale, with specific observation variance:

$$\log \hat{C}_{ay} \sim \text{Normal}\Big(\log C_{ay}, \sigma^2_{ay}\Big)$$

$$\log \hat{I}_{ays} \sim \text{Normal}\Big(\log I_{ays}, \tau^2_{ays}\Big)$$

The log-likelihood can now be defined as the sum of the log-likelihood of the observed catches:

$$\ell_C = \sum_{ay} w_{ay}^{(c)} \; \ell_N \Big(\log C_{ay}, \sigma_{ay}^2; \; \log \hat{C}_{ay} \Big)$$

and the log-likelihood of the observed survey indices as:

$$\ell_I = \sum_s \sum_{ay} w_{ays}^{(s)} \; \ell_N \Big(\log I_{ays}, \tau_{ays}^2; \; \log \hat{I}_{ays} \Big)$$

giving the total log-likelihood

$$\ell = \ell_C + \ell_I$$

which is defined in terms of the strictly positive quantites, M_{ay}, F_{ay}, Q_{ays} and R_y, and the observation variances σ_{ay} and τ_{ays}. As such, the log-likelihood is over-parameterised as there are many more parameters than observations. In order to reduce the number of parameters, M_{ay} is assumed known (as is common).

%==
%THE FOLLOWING NEEDS REVISION, need to bring in N1 submod %================================

The remaining parameters are written in terms of a linear combination of covariates x_{ayk}, e.g.

$$\log F_{ay} = \sum_k \beta_k x_{ayk}$$

where k is the number of parameters to be estimated and is sufficiently small. Using this tecnique the quantities $\log F$, $\log Q$, $\log \sigma$ and $\log \tau$ %log initial age structure % this is not present in the above (in bold in the equations above) can be described by a reduced number of parameters. The following section has more discussion on the use of linear models in a4a.

%==

The a4a statistical catch-at-age model can addionally allow for a functional relationship that links predicted recruitment $\tilde{R}$ based on spawning stock biomass and modelled recruitment R, to be imposed as a fixed variance random effect. [NEEDS REVISION, sentence not clear]

Options for the relationship are the hard coded models Ricker, Beverton Holt, smooth hockeystick or geometric mean. This is implemented by including a third component in the log-likelihood:

$$\ell_{SR} = \sum_y \ell_N \Big(\log \tilde{R}_y(a, b), \phi_y^2; \; \log R_y \Big)$$

giving the total log-likelihood

$$\ell = \ell_C + \ell_I + \ell_{SR}$$

Using the (time varying) Ricker model as an example, predicted recruitment is

$$\tilde{R}_y(a_y, b_y) = a_y S_{y-1} e^{-b_y S_{y-1}}$$

where S is spawning stock biomass derived from the model parameters F and R, and the fixed quantites M and mean weights by year and age. It is assumed that R is log-normally distributed, or equivalently, normally distributed on the log-scale about the (log) recruitment predicted by the SR model $\tilde{R}$, with known variance ϕ^2, i.e.

$$\log R_y \sim \text{Normal}\Big(\log \tilde{R}_y, \phi_y^2\Big)$$

which leads to the definition of ℓ_{SR} given above. In all cases a and b are strictly positive, and with the quantities F, R, etc. linear models are used to parameterise $\log a$ and/or $\log b$, where relevant.

%===
%THE FOLLOWING NEEDS REVISION, this is not just the default I guess, % it's always present since R predictions will be a mix of S/R and γ %==

By default, recruitment R as apposed to the reruitment predicted from a stock recruitment model $\tilde{R}$, is specified as a linear model with a parameter for each year, i.e.

$$\log R_y = \gamma_y$$

This is to allow modelled recruitment R_y to be shrunk towards the stock recruitment model. However, if it is considered appropriate that recruitment can be determined exactly by a relationship with covariates, it is possible, to instead define $\log R$ in terms of a linear model in the same way as $\log F$, $\log Q$, $\log \sigma$ and $\log \tau$. %But this is pretty much the same as taking a geometric mean, with a model on log a, and making the variance very small.

%===
% WE NEED TO ADD SOMETHING ABOUT HOW THE PLUSGROUP IS MODELLED %===

6.2 Classes Description

Figure 6.1 illustrates the input/output model of the statistical stock assessment method based on catch-at-age data.

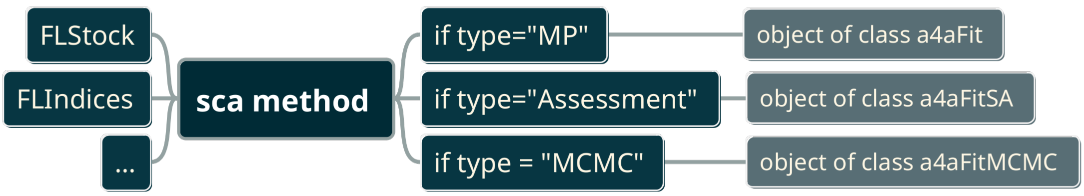


Figure 6.1: Figure: The fit process

The `type` argument in the `sca` function defines the fitting method—either maximum likelihood or MCMC—and specifies whether the variance-covariance matrix of the parameters is calculated and returned in the case of maximum likelihood. The resulting object belongs to a specific class, depending on the selected options. The data structure used to store and report the fitting process follows an object-oriented paradigm (e.g., the S4 system in R) and is hierarchically organized. The primary class, `a4aFit`, is extended by `a4aFitSA` and `a4aFitMCMC` for specific configurations.

Table 6.1: Fit Types and Associated Classes

Type of Fit	Fit Method	Variance-Covariance Matrix	Output Object Class
MP	Maximum Likelihood	No	`a4aFit`
Assessment	Maximum Likelihood	Yes	`a4aFitSA`
MCMC	MCMC	Yes	`a4aFitMCMC`

Type `MP`, representing "Management Procedure," returns an `a4aFit` class object designed for use in MSEs (Management Strategy Evaluations) with full feedback models (REF PUNT and KELL). Inverting the Jacobian to compute the variance-covariance matrix is computationally intensive in maximum likelihood models, and as MSE analyses often involve thousands of iterations, using `type="MP"` significantly speeds up the process. This option is advantageous for scenarios requiring multiple model fits. However, the lack of immediate feedback on model convergence is a drawback, as convergence is assessed by inverting the Jacobian. A failed inversion indicates non-convergence. Table 6.2 describes the composition of the `a4aFit` class.

Table 6.2: `a4aFit` Class Description

Class	Slots	Slot's Class	Description
a4aFit	call	call	Code used to run the analysis
	catch.n	FLQuant	Catch numbers at age and year
	clock	numeric	Time to run the analysis
	desc	character	Description of the stock and/or analysis
	fitSumm	array	Summary statistics of the fit (e.g., number of data points)
	harvest	FLQuant	Fishing mortality at age and year
	index	FLQuants	Indices of abundance (age/biomass, by year)
	name	character	Stock name
	range	numeric	Age and year range of the data
	stock.n	FLQuant	Population in numbers (age and year)

The `a4aFitSA` and `a4aFitMCMC` classes extend `a4aFit`, retaining all its slots while adding a `pars` slot of class `SCAPars`. Table 6.3 outlines these classes.

Table 6.3: `a4aFitSA` and `a4aFitMCMC` Class Description

Class	Slots	Slot's Class	Description
a4aFitSA	All `a4aFit`		Inherited from `a4aFit`
	pars	SCAPars	Parameter information
a4aFitMCMC	All `a4aFit`		Inherited from `a4aFit`
	pars	SCAPars	Parameter information

The `SCAPars` class stores details about submodel parameters, such as formulas and distributions, and includes three slots: `stkmodel` for stock model parameters, `qmodel` for catchability parameters, and `vmodel` for variance parameters. Table 6.4 describes this class.

Table 6.4: `SCAPars` Class Description

Class	Slots	Slot's Class	Description
SCAPars	stkmodel	a4aStkParams	Details of fishing mortality, stock recruitment, and initial stock numbers
	qmodel	submodel	Details of catchability parameters
	vmodel	submodel	Details of variance parameters

The `stkmodel` slot encompasses parameters for fishing mortality, stock recruitment, and initial stock numbers. Due to potential correlations among these parameters, their variance-covariance matrix is reported collectively. Table 6.5 describes the slots of the `a4aStkParams` class.

Table 6.5: `a4aStkParams` Class Description

Class	Slots	Slot's Class	Description
a4aStkParams	centering	FLPar	Centering parameters
	coefficients	FLPar	Model coefficients
	desc	character	Description
	distr	character	Distributions
	fMod	formula	Fishing mortality model
	link	function	Link function
	linkinv	function	Inverse link function
	m	FLQuant	Mortality parameters
	mat	FLQuant	Maturity parameters
	n1Mod	formula	Initial stock numbers model
	name	character	Stock name
	range	numeric	Age and year range
	srMod	formula	Stock-recruitment model
	units	character	Units of measurement
	vcov	array	Variance-covariance matrix

Class	Slots	Slot's Class	Description
	wt	FLQuant	Weights

The `qmodel` and `vmodel` slots share the `submodel` class, which describes single submodels. Table 6.6 provides details.

Table 6.6: `submodel` Class Description

Class	Slots	Slot's Class	Description
submodel	centering	FLPar	Centering parameters
	coefficients	FLPar	Model coefficients
	desc	character	Description
	distr	character	Distributions
	formula	formula	Submodel formula
	link	function	Link function
	linkinv	function	Inverse link function
	name	character	Stock name
	range	numeric	Age and year range
	vcov	array	Variance-covariance matrix

Chapter 7

Submodel structure

Placeholder

7.1 Submodel building blocks and fundamental `R` formulas

7.2 The major effects available for modelling

7.3 The submodel class and methods

Chapter 8

Fitting

Placeholder

8.1 Fishing mortality submodel (F_{ay})

8.1.1 Separable model

8.1.2 Constant selectivity for contiguous ages or years

8.1.3 Time blocks selectivity

8.1.4 Time changing selectivity

8.1.5 Closed form selection pattern

8.1.6 More models [TO BE MOVED TO ITS OWN SECTION?]

8.2 Abundance indices catchability submodel (Q_{ays})

8.2.1 Catchability submodel for age based indices

8.2.2 Catchability submodel for age aggregated biomass indices

8.2.3 Catchability submodel for single age indices

8.3 Stock-recruitment submodel (R_y)

8.4 Observation variance submodel ($\{\sigma^2_{ay}, \tau^2_{ays}\}$)

8.5 Initial year abundance submodel ($N_{a,y=1}$)

8.6 Data weigthing

8.7 Working with covariates

8.8 Assessing `ADMB`files

8.9 Missing observations in the catch matrix or index

Chapter 9

Diagnostics

Placeholder

9.1 Residuals

9.2 Predictive skill

9.3 Aggreagted catch in weight

9.4 Fit summary, information and cross-validation metrics

9.5 The package a4adiags

9.6 Residuals and submodels misspecifiction

9.6.1 The "mean" model

9.6.2 The age effects

9.6.3 The fishing mortality year model

9.6.4 The initial year population abundance model, aka N1

9.6.5 The stock recruitment submodel

9.6.6 The variance submodel

Chapter 10

Hindcast

alternative setups and metrics and what it means (EJ to share an initial structure)

Chapter 11

Reference Points

Placeholder

11.1 Yield per recruit reference points

11.2 Stock recruitment relationship based reference points

11.2.1 Stock recruitment after fitting the stock assessment model

11.2.2 Stock recruitment during fitting the stock assessment model

11.3 Economics reference points

11.4 Computing user specific reference points

Chapter 12

Projections and harvest control rules

Placeholder

12.1 Initial condition assumptions [CHECK fwdWindow]

12.2 Scenarios

12.2.1 Relative scenarios

12.2.2 Limits

12.2.3 Complex scenario

Chapter 13

Predict and simulate

Placeholder

13.1 Basic functions}

13.1.1 simulate()

13.1.2 genFLQuant()

13.2 submodels

13.3 Predict

13.4 Simulate

Chapter 14

The statistical catch-at-age stock assessment framework with MCMC

Placeholder

14.1 Diagnostics with CODA

Chapter 15

Propagate uncertainty into stock assessment

Placeholder

Chapter 16

Modelling fleet selectivity

Chapter 17

Modelling spatial effects

```
ridx01 <- stk0@stock.n[1]*0.7*0.001
ridx01 <- log(ridx01*rlnorm(ridx01))
ridx02 <- stk0@stock.n[1]*0.3*0.001
ridx02 <- log(ridx02*rlnorm(ridx02))

srmod <- ~ geomean(a~ridx01+ridx02, CV=0.5)
cvar <- FLQuants(ridx01 = ridx01, ridx02 = ridx02)
fit01 <- sca(stock,tun.sel[c(1)],fmodel=fmod,qmodel=qmod, srmodel=srmod, covar=cvar)
coef(fit01)
srmod <- ~ geomean(CV=0.1)
fit02 <- sca(stock,tun.sel[c(1)],fmodel=fmod,qmodel=qmod, srmodel=srmod)
coef(fit02)
```

- check situation where the two areas are negatively correlated
- the two covariates need to be at the same scale, in the sense of representing the same process
- other examples (ICES, ask in the plenary if we can have access to the data)
- dan ghotel ask for spatial workshop data

Chapter 18

Sections to be added!?

- Assessing the coverage of confidence intervals??