fish610.055 Fisheries and ecosystem models

Gunnar Stefansson

10. apríl 2018

Acknowledgements
This tutorial was developed as a part of the EU H2020 project MINOUW.
This project has received funding from the European Commission's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 634495 for the project Science, Technology, and Society Initiative to minimize Unwanted Catches in European Fisheries (MINOUW).
`http://minouw-project.eu/`

Efnisyfirlit

1 From single species through multi-species to ecosystem models

1.1 Classical single species models (assessment models)

Two main classes of single-species models:

- bulk biomass models
- age structured models

These are sometime chosen to match data structure, but not always

1.1.1 Handout

Single-species models can be classified into two main types:

- bulk biomass models
- age structured models

These are sometime chosen to match data structure, but not always. For example, sometimes it is better to fit bulk biomass models even when age data exist and vice versa.

1.2 Fitting methods

- Statistical
- "by hand"

Methods vary, from just making the model output "look like" the trends in the data to complicated nonlinear methods of minimisation.

1.3 Multispecies models

Added complexity:

- multiple species
- biological interactions
- technical interactions

The species may or may not be caught in the same gear(s)
A species may be predators or prey

1.4 Ecosystem models

Typical desired additions:

- hydrography (currents)
- spatial factors

Example: Atlantis
Further reading, best practices: url{`http://www.fao.org/3/a-i0151e.html`}

1.5 Socio- economics

In EAFM we typically also want to take into account:

- social issues (e.g. employment)
- economics (e.g. profitability of the fisheries)

1.6 Including everything

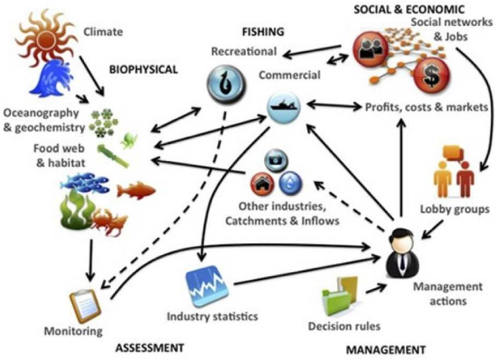


Figure (c) CSIRO, see website link for Atlantis: https://research.csiro.au/atlantis/

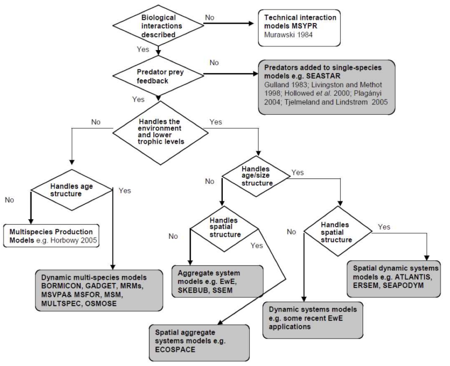
1.7 Classification of models (a)

Further reading: Models for an ecosystem approach to fisheries by Plaganyi, published by FAO in 2007. `http://www.fao.org/tempref/docrep/fao/010/a1149e/a1149e.pdf`

- bulk-biomass vs age/length disaggregated
- temporal vs equilibrium
- strategic vs tactical [long-term yield vs assessments]
- single vs multispecies
- single vs multifleet
- spatial vs non-spatial
- stock component details: age, length, maturity, sex, ...

url{`http://www.fao.org/tempref/docrep/fao/010/a1149e/a1149e.pdf`}
(more in other lectures)

1.8 Classification of models (b)



Plagányi, É.E. 2007. Models for an Ecosystem Approach to Fisheries. FAO Fisheries Technical Paper No. 477. Rome, FAO. 2007. 108p. ISBN 978-92-5-105734-6.

1.8.1 Handout

Further reading: Models for an ecosystem approach to fisheries by Plaganyi, published by FAO in 2007. url{`http://www.fao.org/tempref/docrep/fao/010/a1149e/a1149e.pdf`}

2 Statistical techniques for stock assessments

2.1 Biological models

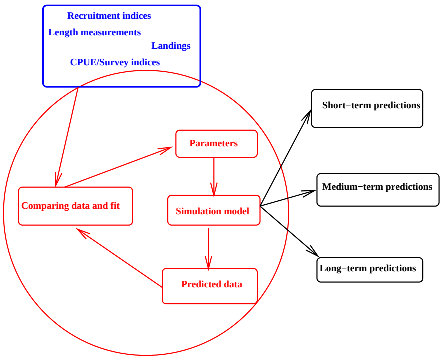
- Conceptual model of biology
- Mathematical description of the model
- Population models: Usually a forward prediction model
- Growth etc: Mathematical description of process
- Remember: **Parsimony!**

2.1.1 Details

When a biological process is to be modeled, the first step is to obtain a conceptual view of the process, followed by a mathematical description, i.e. mathematical formula which describes the process.

2.2 Formal statistical procedures

Model contains parameters
Model predicts data
Parameters are estimated by fitting model to data



2.2.1 Details

A formal statistical procedure always contains a model which describes mathematically how the data have been obtained, i.e. usually specifies the probability distribution. A statistical model always contains unknown parameters (otherwise it is a probabilistic model).

The parameters can subsequently be estimated using techniques which ensure that the model predicts the data as well as possible.

2.3 Population models

Some basic population model types:

- Equilibrium bulk biomass/production model
- Dynamic bulk biomass model
- Dynamic age structured production model
- Back-calculation model

3 Production models

3.1 History

Stock-production models: Typical for whales, shrimp, nephrops and other difficult to age species

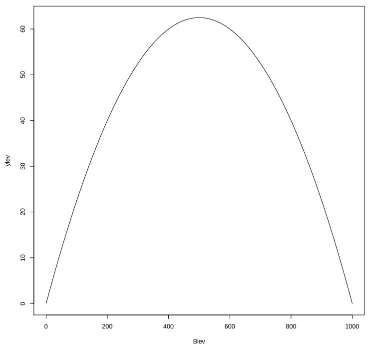
Originally: Assumed equilibrium (oversimplified)
Now: Use dynamic stock-production models

3.1.1 Details

The so called stock-production models have been used to estimate stock sizes for whales, shrimp and other difficult to age species. Originally, such models assumed that the stock was at an equilibrium, but in recent decades such oversimplified models are not used much since catches are known to affect stock size and so dynamic models are needed.

3.2 An equilibrium model

A simple equilibrium production model:
$Y = rB(1 - \frac{B}{K})$



Equilibrium yield and biomass

3.2.1 Details

A traditional model of the possible production of a stock is given by the equilibrium production model.

Definition 3.1. Equilibrium production model

$$Y = rB(1 - \frac{B}{K})$$

B=biomass
K=maximum biomass corresponding to no fishing
r=rate of population increase
Y=production

If the production is described using the equilibrium production model, then this is also the equilibrium yield, i.e. the yield that can be obtained at this level of biomass.

3.2.2 Examples

Example 3.1. The production, or equilibrium yield, curve can easily be plotted given values of r and K, e.g. with the following R code.

```
r<-0.25
K<-1000
Blev<-(0:100)*10
ylev<-r*Blev*(1-Blev/K)
plot(Blev,ylev,'l')
```

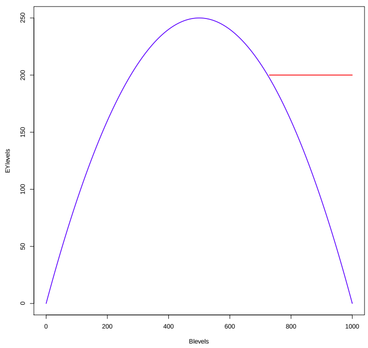
3.3 Simple forward projections

Given B the production is $rB(1-B/K)$.

If the catch is Y the stock biomass will become

$$B+rB(1-B/K)-Y.$$

Notably, the stock increases if Y is less than the production and vice versa.



Equilibrium yield with a constant catch line.

3.3.1 Examples

Example 3.2. Suppose a stock starts out at $B=K=1000$ tonnes and gets fished with constant catch, $Y=100$ tonnes.

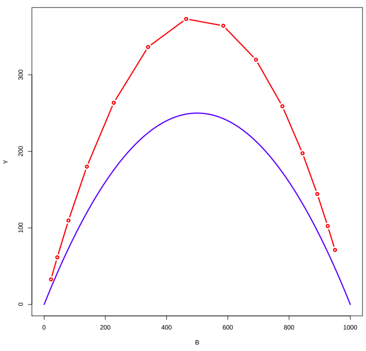
The following R code will track the development of the stock for the first few years.
http://tutor-web.net/fish/fish5108statass/lecture20/stockprod.r

3.4 The problem with equilibrium models

Over-fished=catch above sustainable yield

Observed biomass and catch all above the underlying curve

Note: Will overestimate production if assume equilibrium



The bias issue when estimating data from an equilibrium model. The solid line is the true model and the dot/dashed line is from generated data when fishing is 1.5 times the equilibrium catch.

3.4.1 Details

Suppose data are collected on biomass and yield while a stock is being over-fished, i.e. the annual catch is more than the annual sustainable yield. In this case the stock will be reduced each year.

The problem is that even in the best scenario, when perfect information is available on historical yields and biomass, the corresponding data points, (B_y, Y_y), are all above the equilibrium curve.

In this case, the approach of assuming that the data are measurements of points on the equilibrium curve is clearly extremely dangerous and will lead to serious overestimation of the production for any given level of biomass.

3.4.2 Examples

Example 3.3. The following R code simulates the development of a stock which is being over-fished each year. The stock is fished using a simple rule which is to take double the sustainable yield (SY).

http://tutor-web.net/fish/fish5108statass/lecture20/stockprod-1.r

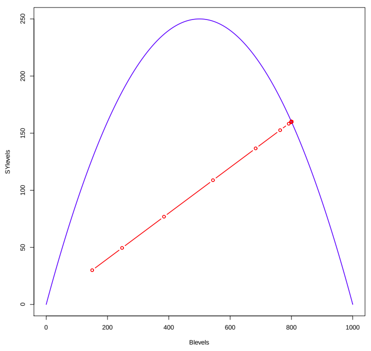
3.5 A dynamic biomass model

A simple forward projection model:

$$B_{y+1} = B_y + rB_y(1 - B_y/K) - Y_y$$

where the catch is specified according to some rule, e.g. proportional harvest:

$$Y_y = pB_y.$$



Equilibrium yield and curve describing the trajectory of catch and biomass under a constant harvest strategy.

3.5.1 Details

The simplest model for describing the development of a stock is probably the simple bulk biomass model.

Definition 3.2. Simple bulk biomass model

$$B_{y+1} = B_y + rB_y(1 - B_y/K) - Y_y.$$

Y_y=annual catches taken from the stock

In order to investigate the behavior of the simple bulk biomass model, the catch can be simulated according to some simple rule, e.g. as a proportion of the stock size,$Y_y = pB_y$.

3.5.2 Examples

Example 3.4. http://tutor-web.net/fish/fish5108statass/lecture20/stockloop.r

3.6 Unknown parameters

Model: $B_{y+1} = B_y + rB_y(1 - B_y/K) - Y_y$
Unknown parameters: r, K, and B_0.

- Given model and parameters, the trajectory can be generated.
- Assumes catches are known constants (no error)
- Parsimony: May (need to) assume $B_0 = K$.

3.6.1 Details

Even the simplest population dynamics models contain parameters which are unknown. In the equilibrium surplus production models these parameters are r and K.

In the dynamic production model, $B_{y+1} = B_y + rB_y(1 - B_y/K) - Y_y$, the unknown parameters are r, K, and the initial population size, B_0. This approach assumes that the catches are simply known constants (i.e. without error). An important aspect of this approach is that for a given set of parameters the entire population trajectory can be generated.

3.6.2 Assignment

Assignment 3.1. Take each figure in this section and redraw it in R using the commands behind each figure. Next, rerun the commands for each figure using different assumptions: Investigate the effects of varying K, r, B_0 and the harvest rate p (or the constant catch, as the case may be). Draw conclusions concerning the sustainability of different harvesting regimes.

4 Fitting criteria

4.1 The issue of fitting to data

Have some internal model of stock development
Model may be bulk or age-based
Need to compare to data-not contort the data!
Data may be bulk or age-based
Note several possible combinations: model or data may be complex or simple

4.1.1 Details

Suppose there is some given (internal) model of stock development. This model may be a bulk (biomass production) model or a highly disaggregated (e.g. age-based) model.

Such a model will inevitably contain several unknown parameters which can only be estimated based on some measurements. Thus, there is a need to make some formal comparison with data.

There is no a priori specification of whether the data need to be disaggregated or bulk-biomass. It is important that actual data be used with as little transformation as possible. For example, catch and effort data should NOT be transformed to catch and CPUE data.

In particular, there are several possible combinations of disaggregated or complex models and bulk or disaggregated data. For example a model may be age-based but fitted to length data and so forth.

4.2 Linking biomass to an index

Annual index of total abundance:

$$I_y = qB_y e^{\varepsilon_y}$$

Assume ε comes from Gaussian distribution
q is catchability

4.2.1 Details

Recall that a statistical model always links some mathematical model to data. For example, if I_y is an index of abundance in year y and B_y is the model-based biomass in year y, then one could use $\hat{I}_y = qB_y$ as a predicted index. In this simple model it is assumed that there is a constant catchability which describes how the index relates to stock biomass.

If the model is a simple dynamic production model, $B_{y+1} = B_y + rB_y(1 - B_y/K) - Y_y$, the primary unknown parameters are r, K, and B_0 but in addition q needs to be estimated.

For a given set of parameters several potential measures of the quality of the model can be defined. The simplest is without doubt a direct sum of squares:

$$\sum_y \left(I_y - \hat{I}_y\right)^2$$

though it may be more appropriate to use log-transformed data:

$$\sum_y \left(ln(I_y) - ln(\hat{I}_y)\right)^2$$

The actual indices are usually assumed to be related to the predicted indices on a log-scale and one may specify a formal distribution, such as log-normal for the measurements. Thus, one would commonly assume that $ln(I_y) = ln(\hat{I}_y) + \varepsilon_y$ and often assume further that the deviations come from a Gaussian distribution.

Continuing in this manner leads to the full statistical model for the data.

Definition 4.1. Annual index of total abundance:

$$I_y = qB_y e^{\varepsilon_y}$$

4.3 Fitting in a VPA setting

- VPA/cohort analysis gives historical information
- Survey indices provide time series proportional to abundance
- Effort data provide time series proportional to F_y
- Can connect through regression (i.e. revise the terminal year)
- Can reduce to single parameter (F_{term})
- Better to formally estimate using a statistical approach, e.g. minimize

  $$\sum (E_y - (1/q)F_y)^2$$

  over the terminal F_{term}.
- Can also write up formal methods (e.g. ADAPT) for estimating an entire parameter set, including selection pattern s_a.

4.3.1 Details

In order to move forward from simple VPA or cohort analysis (which typically assumes a given fishing mortality in the last year), towards objective methods of assessment, some model is needed whereby fishing mortality can be estimated.

The simplest such relationship is probably to consider effort data, which might be related to fishing mortality through a simple linear relationship:

$$F = qE.$$

The available data are now E_y and the model predicted fishing mortality is F_y, for any value of F in the last year.

A natural way forward is to estimate the fishing mortality in the last year to be the value which gives the best linear fit to the effort data.

In this case the nonlinearity lies in the VPA (or cohort analysis) which is used to compute the historical fishing mortality.

A serious problem with this particular method is that the basic model, $F = qE$ reverses the usual statistical model framework, i.e. the data E is used to predict the model F and not vice versa. In this simple version, the problem can easily be alleviated by reversing the direction, i.e. writing $E = (1/q)F$ and minimizing the sum of squares

$$\sum (E_y - (1/q)F_y)^2$$

Several points should be noted:

- VPA/cohort analysis mainly gives historical information
- Survey indices provide a time series proportional to abundance
- Effort data provide a time series proportional to F_y

- Can connect through regression (i.e. revise the terminal year)
- Can reduce to single parameter (F_{term}).
- Better to formally estimate using a statistical approach, e.g. minimize

$$\sum (E_y - (1/q)F_y)^2$$

  over the terminal F_{term}
- Can also write up formal methods (e.g. ADAPT) for estimating entire parameter set, including e.g. selection pattern s_a, annual recruitment etc.

4.3.2 Examples

Example 4.1. It is possible to carry out a Virtual Population Analysis for the Icelandic summer spawning herring. The input for the VP-Analysis (i.e. mortality coefficients for the last year) is adjusted in such a way that the time sequence of the stock assessment is as consistent as possible with the acoustic surveys.

The model is an age-based model and uses catches in numbers at age. However the model fitting mechanism is of the same form as before, i.e. a sum of squares is minimized in order to find the estimate.

4.3.3 Handout

The simplest case of estimating stock sizes in a back-calculation setting is when a selection pattern is assumed to be known and fishing mortality on the oldest ages are assumed to be equal (or proportional to) to the fishing mortality on some younger ages. Here, only one value remains unknown, the terminal fishing mortality F_{term}.

In principle, any data set can be used to estimate the terminal fishing mortality. This includes survey indices, an effort time series, commercial CPUE data and so forth.

4.4 ADAPT

Simplified ADAPT:

- Start with VPA or cohort analysis
- Use regression to predict indices - get SSE
- Find the best possible regression by varying F_{term}
- NB: The index-values are on the y-axis!!

$$SSE = \sum_{a,y} w_a [\ln U_{ay} - (\alpha_a + \beta_a \ln N_{ay})]^2$$

Often we set $w_a = 1$ for most ages. Usually we set $\beta_a = 1$ for most ages.

Full ADAPT: Use nonlinear minimization and estimate initial age composition, annual F_y and recruitment along with q_a.

4.4.1 Details

The Adaptive framework (ADAPT) for stock assessments are a straight-forward nonlinear statistical estimation technique.

An internal stock-projection model is used. This could be backwards computation but more often forward projection using cohort analysis is used.

Unknown parameters will typically include recruitment, selection parameters and overall annual mortality rates along with the initial age-disaggregated population size and catchabilities (or similar parameters).

For a given set of parameters, the stock can be used to predict e.q. survey indices.

The prediction error can now be minimized to estimate unknown parameters.

This approach requires a number of assumptions, e.g. that the catch data is known. The following describes the approach a little more fully.

The collection of catch data concerning various important fish stocks are of such quality that it is possible to assume that the data are reasonably well measured, i.e. contains little measurement error. This results in the assumption that if VP-Analysis is employed with the correct input it will provide a sound stock estimate. This estimate could in turn be used to predict indices from survey data and therefore it is feasible to verify whether a given stock estimate is in accordance with a time series of survey data.

One possible way to conduct such a comparison is through stating that for given mortality coefficients in the last year and a given relationship with indices, the deviation in the forecast concerning indices is given by:

$$SSE = \sum_{a,y} w_a \left[\ln U_{ay} - (\alpha_a + \beta_a \ln N_{ay})\right]^2$$

In a simplified version of ADAPT, using cohort analysis for back-calculation, the unknown coefficients in the model are only α_a , β_a and the overall fishing mortality rate for the last year. The selection pattern for the last year can be given on the basis of the patterns of previous years (i.e. restricted or set to the average), and fishing mortality on the oldest fish is fixed (i.e. restricted to an average of younger age groups) so it becomes unnecessary to estimate all mortality coefficients and only the multiplier (F_y) needs to be estimated.

In the simplified version, for a given fishing mortality multiplier, the best estimate of the coefficients α_a and β_a will be obtained from a simple linear regression. Thus, it becomes simple to compute SSE for each value on F_y. Different values may be tested for F_y until a low value for SSE has been established. Thus, an estimation of the fishing mortality for the last year has been obtained.

The SSE-equation contains scaling coefficients (weights), w_a, which need to indicate the relative precision for the main age groups in the groundfish survey data. These numbers can be estimated by examining how low the sums of squares can be for each age group.

It should be noted that the equation assumes different coefficients α_a and β_a for each age group. By way of simplification, it could be stated that $\beta_a = 1$ for all age groups, but little can be said for α_a, which then becomes the catchability coefficient (denoted q_a).

4.5 Abundance index (research or log-books)

- Log-books: catch per towing hour
  - use mean weight to get number of fish per towing hour
  - use age determination to get number per age group
- Groundfish survey: direct count
  - use age determination to get number per age group
  - Other research cruises...

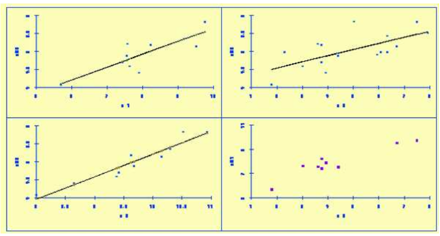
4.5.1 Details

In the above discussion it has been assumed that survey information is available. A survey index could in principle be age-disaggregated, a bulk index, or any other measure which can in some way be linked to stock size or fishing mortality.

In principle a survey is not needed, as it is possible to use commercial catch-per-unit-effort data, effort data, or other measures. On the other hand, it is quite likely that data from commercial fleets will have time trends as far as relationships with stock size are concerned.

4.6 Nature of relationships

- Poor relationship between N and U
  - Is VPA bad?
  - Uncertain indices?
  - Variable M?
- Good relationship between N and U
  - VPA and indices OK
  - M stable
- $N-U$ bad but $U-U$ good
  - Variable M and/or VPA bad



4.6.1 Details

Various different methods for the estimation of current stock size have been introduced in the above sections. No single method is perfect, but there are certain indications that some methods are much better than others.

What we can learn from these methods is that if there is any cause for doubt regarding the estimate of stock size, it is desirable to try other methods and data sets in order to examine whether the results are verifiable when the problem is viewed through a different pair of spectacles.

5 Formal statistical stock assessments in dynamic bulk production model

5.1 Aggregate data

Have only total catch and abundance data, Y_y and I_y.
No error in catches:

$$B_{y+1} = B_y + rB_y(1 - B_y/K) - Y_y$$

Errors in indices:

$$I_y = qB_y e^{\varepsilon_y}$$

or:

$$\hat{I}_y = qB_y$$

and

$$\ln I_y = \ln \hat{I}_y + \varepsilon_y$$

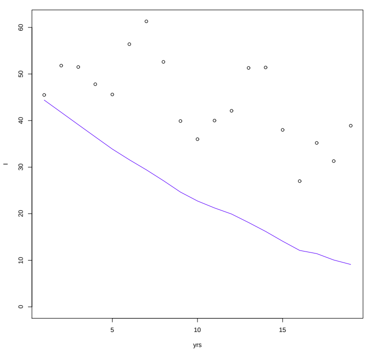
5.1.1 Examples

Example 5.1. Nephrops in Icelandic waters, 1980-1998:

http://tutor-web.net/fish/fish5108statass/lecture40/nephrons-80-98.r

5.2 Initial values for bulk production models

1. Define the model and parameters to be estimated.
2. Initialize parameters.
3. Evaluate the model fit.
4. Optimize the model fit.



5.2.1 Details

It is usually quite important to carefully select the initial values and parameter transformations when fitting stock-production models. There are several reasons for this.

Firstly, if the initial biomass (B_0) and productivity (r) are too low, then forward projections will result in zero or negative biomass values and in turns problems with fitting. Similarly, if parameters are estimated on the original scale, a minimization algorithm will not know that the parameters typically need to be positive and may test negative values. One solution is to use the logarithms of the parameters for optimization.

A simple approach to obtaining initial values of the right magnitude is to select scenarios which correspond to data without negative biomasses etc. For example, if the initial biomass is chosen equal to the total historical catch and a very low productivity is used, then the corresponding modeled stock biomass will correspond to a depletion model, but one which can explain the catches (though it may give a poor fit to the indices).

Similarly, if the initial values of catchability are set to some ratio of the survey index to catch, then this will lead to appropriate units for q.

5.2.2 Examples

Example 5.2. The initialization of a dynamic bulk biomass model involves setting initial values for all parameters. A procedure is then needed to forward project the population and subsequently to evaluate the quality of the fit to the data. The full example of R commands to project the stock forward and plot the predicted index from initial parameter values is as follows:

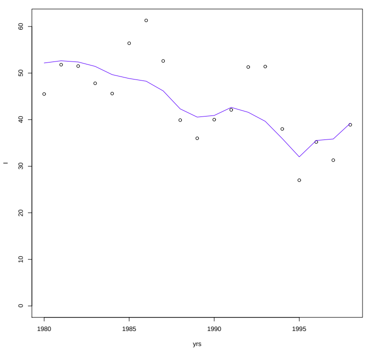
http://tutor-web.net/fish/fish5108statass/lecture40/bulkprodmodel.r

5.2.3 Handout

The estimation procedure is usually separated into a few parts, namely initialization of parameters, projection based on those parameters, evaluation the fit to data, and search for parameters which give the best fit to the data.

5.3 Estimation in bulk production models

1. Define the model and parameters to be estimated
2. Initialize parameters
3. Evaluate the model fit
4. Optimize the model fit



5.3.1 Details

Having obtained initial values, the next step is to project the future stock and evaluate the model fit using actual data. A minimization algorithm is then used to estimate the parameter values.

5.3.2 Examples

Example 5.3. The following procedure will project forward the population and evaluate the quality of the fit to the data:

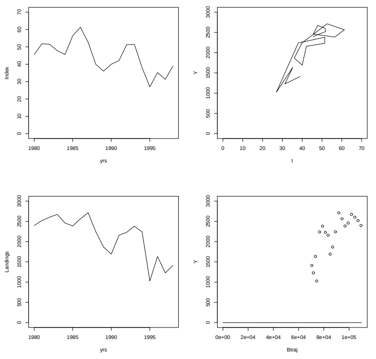
http://tutor-web.net/fish/fish5108statass/lecture40/forwardprojection.r

5.3.3 Handout

The estimation procedure is usually separated into a few parts, namely initialization of parameters, projection based on those parameters, evaluation the fit to data and then search for parameters which give the best fit to the data.

5.4 Revising the model

Need to verify output



5.4.1 Examples

Example 5.4. Modify the model to follow the assumption $B_0 = K$.
http://tutor-web.net/fish/fish5108statass/lecture40/bo-k.r

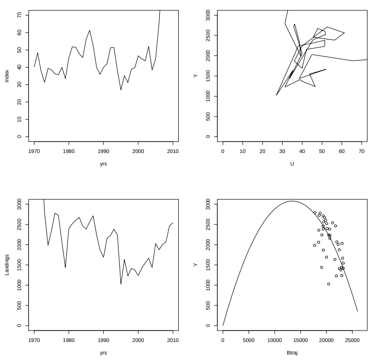
5.5 More nephrops case studies

Longer time series are available

Series can be spatially disaggregated

See file nephrops.dat

Caveat: Fleet changes may have occurred during the time period - invalidates the analysis.



5.5.1 Examples

Example 5.5. Data on *Nephrops norvegicus* from the MRI 2011 annual report (see http://www.hafro.is).

Read in the data, fit, and plot
http://tutor-web.net/fish/fish5108statass/lecture40/nephronsfull.r

6 Case studies of stock-production models

6.1 History

6.1.1 Details

Recall that stock-production models have been used to estimate stock sizes of difficult to age species including whales and shrimp.
Initially, these models assumed that the stock was at equilibrium which is obviously an oversimplification.
Such models have been replaced with dynamic models.

6.2 Prawn models

6.2.1 Details

The simple bulk biomass model (see definition 2.2 of tutorial 8) can be modified in various ways to suit real situations. Recruitment indices for deep-sea prawn are, for example, available and should contain such information. It is known that cod feed to a considerable extent on deep-sea prawn and this factor should be entered straight into the model. Natural mortality is, on the other hand, less well known and it is difficult to estimate the increase in biomass between years.

Definition 6.1. Prawn model:

$$B_{y+1} = \alpha B_y - Y_y + \beta R_y - \delta D_y$$

R_y= recruitment index
D_y=index of cod predation on prawn
α= production excluding recruitment and mortality excluding predation by cod

The system is described in such a way that biomass at the beginning of the year is multiplied by a coefficient, catch is subtracted, recruitment is added and predation is subtracted.

Unknown coefficients in the model are q, α, β , δ , q and B_0 . Given these coefficients, it is possible to compute B_y for all the years. Estimation of the coefficients is arrived at by establishing the values that match existing data on catch per unit effort most closely.

This model can then be used to examine biomass changes from one year to the next. For a given size of the cod stock, the effect of catches on the stock can be examined.

6.2.2 Examples

Example 6.1. The following data provides an example of the type of data needed for a prawn model:

http://tutor-web.net/fish/fish5108statass/lecture50/prawnmodel.dat

The data can be entered into a spreadsheet and the calculations completed. When using the spreadsheet, we begin by guessing the values of unknown coefficients. These values are then used to update the stock and make predictions as to catch per unit effort. Then the quadratic deviation of the projection is computed. Finally, different values for the coefficients are tested in order to examine which values give the lowest quadratic sum.

http://tutor-web.net/fish/fish5108statass/lecture50/prawnmodelexpanded.dat

As can be seen from the last column, an estimate is obtained of the amount of prawn consumed by cod on an annual basis. This is denoted as δD.

It is possible to plot measured and projected catch per unit effort. Although it is seen that the predictions match the data quite closely, it is also clear that the number of parameters is high compared to the number of data points.

This model was introduced in (Stefánsson et.al. 1994)). The results of stomach content investigations in Icelandic waters have been presented in various articles by Pálsson et.al.

Example 6.2. Herring models
Acoustic surveys provide an estimation of the size of the herring stock at a certain point in time.

Measurements of stock size do, however, not give any information on yield potential. This requires additional information on how the stock reacts to catches. If, for example, renewal within the stock is slow, hardly anything can be caught from it without depleting it by an amount which almost equals the catch. If, on the other hand, renewal is fast (high natural mortality rates, good recruitment and fast individual growth) it will be possible to catch a bigger proportion from the stock each year.

Stock-production models like the ones described above have therefore been designed for herring stocks. These models are then estimated in such a way that is most consistent with counts or catch per unit effort.

Biomass models have been used for many species, including prawns.

6.3 Marine mammals

6.3.1 Details

The most important information on the size of the whale stock is obtained by sighting surveys (counting). It can be assumed, that the total number is known for the year in which the counting took place.

Elaborate statistical methods are then used to link the resulting counts to total abundance for the stocks.

Stock-production models have been designed for whale stocks. Parameters of the models are estimated in order to obtain the best possible fit to abundance measurements and other historic data.

6.4 Redfish

6.4.1 Details

Redfish are notoriously difficult to assess due to ageing issues. As a result, simple models of a production-type have been used for several redfish stocks. However,these are complicated by their long life span. Thus, acoustic surveys for the *Sebastes mentella* stock of redfish have recently begun in order to use stock-production models that mimic that used for whales.

7 Models with internal age structure

7.1 Introduction

Internal age structure
Incorporate growth information
Link to any available data

- Length measurements
- Age data
- Survey indices

7.1.1 Details

The term "age-based dynamic production model"has been used to encompass a fairly wide range of models which have several common features. These models incorporate all the positive features of all earlier models of single-species population dynamics. The following section describes these models and illustrates how most current fisheries models form a single class of models. Some of these do not explicitly model the production as a function of the adult population and the term "statistical assessment model"is therefore sometimes preferred.

These statistical models are flexible yet parsimonious in parameters and can be linked to any available fisheries data but they do not require the existence of a particular data set. They use an internal age structure in order to encapsulate recruitment variation. And, they commonly incorporate growth information, at least in the form of length distributions.

Typical data sets accounted for are

- Annual landings
- Survey or CPUE indices in numbers or weight
- Length measurements
- Age data

This section only considers models which do not take growth explicitly into account and thus do not link directly to length measurements. If length measurements are used in these particular methods, this is through some transformation. For example, a survey length distribution is typically used to obtain an index of recruitment, even an index for 2-3 age groups (through some form of cohort slicing).

7.2 Biomass or numbers

Choose whether population model is in biomass or numbers

7.2.1 Details

The choice of a numbers or biomass model is not trivial.

Note 7.1. A biomass-based model directly incorporates the growth of the individuals and recruitment in biomass as well as survival into a single production component.

Note 7.2. A numbers-based model allows for a separation of the recruitment in numbers and thus deals with natural mortality separately which tends to be poorly estimated.

Note 7.3. Intermediate models may include one or two true age groups and a single lumped plus group.

7.2.2 Handout

The various dynamic models of fish populations are naturally linked through the use of "the plus group". First, the dynamic bulk biomass and numbers models (ignoring fishing):

$$B_{t+1} = B_t + rB_t\,(1 - B_t/K)$$

and

$$N_{t+1} = N_t + rN_t\,(1 - N_t/K)$$

are the same models, though one described the change in numbers and the other describes the change in biomass. There is a difference in interpretation of course, since the production term in the numbers model describes a change in weight which includes both recruitment in numbers and the weight increase of individuals in the population.

Naturally one can rewrite the numbers model as consisting of two steps:

$$R_t = rN_t\,(1 - N_t/K)$$

$$N_{t+1} = N_t + R_t$$

which makes the model more like a fisheries model with a recruitment term.

Inserting a mortality on the adult population and using a Beverton-Holt formulation for the recruitment term transforms this model into:

$$\begin{aligned} R_t &= \frac{\alpha N_t}{1+N_t/K} \\ N_{t+1} &= N_t e^{-Z_t} + R_t \end{aligned}$$

This particular formulation assumes that the catches do not include the recruiting year-class. Alternatively, one might assume that the catches are in part from the recruiting year-class (applying a different Z-value to R in the equation) or that recruitment enters the fishery at the beginning of the year with:

$$\begin{aligned} R_t &= \frac{\alpha N_t}{1+N_t/K} \\ N_{t+1} &= (N_t + R_t)\, e^{-Z_t} \end{aligned}$$

7.3 Forwards or backwards

Should computations be forwards or backwards in time?

7.3.1 Details

The traditional VPA and cohort analyses project the populations backwards in time. This has the benefit of VPA convergence, which provides stable estimates of historical numbers in each age group.

A forward projection is in many ways a more natural approach. This starts with some initial population either in numbers or biomass. The population is brought forward one year at a time using a set of projection equations.

A forward projection is needed in order to incorporate a plus group.

7.4 Models of catches

Want to include errors in catches Unlike many bulk biomass and VPA models Can use $F_{ay} = F_y s_a$ to reduce parameters

7.4.1 Details

The forward mechanism allows several parametrizations which are more natural than those used in VPA. In particular it is very easy to restrict the number of parameters, e.g. by assuming that fishing mortality is separable ($F_{ay} = F_y s_a$) every year (not just the last).

7.5 The plus group

Traditional biomass model

$$B_{y+1} = B_y + rB_y(1 - B_y/K) - Y_y$$

Numbers model with true ages $1,\ldots,A$:

$$N_{A+1,y+1} = \left(N_{A,y} + N_{A+1,y}\right) e^{-Z_A}$$

7.5.1 Details

A traditional biomass model has a simple aggregate measure of total biomass:

$$B_{y+1} = B_y + rB_y(1 - B_y/K) - Y_y$$

A model of population numbers with true ages $1, \ldots, A$ may use "age"$A+1$ as a “plus group”. Each year this group gets reduced by mortality but a new age group enters the plus group.

$$N_{A+1,y+1} = \left(N_{A,y} + N_{A+1,y}\right) e^{-Z_A}$$

Examples of this approach include the case where there are very many true ages so the plus group is negligible but also cases where there is only a single true age and the plus group becomes very similar to the bulk biomass model.

Although this seems like a very appealing method, two immediate issues arise. Firstly, a mean weight "at age"needs to be attached to the plus group. For a given year the plus group has a certain age composition and therefore a mean weight can be estimated. However, this will change dynamically as can best be seen by considering a large cohort entering the plus group. This cohort may dominate the plus group in numbers and individuals in this cohort will of course gain weight in time. Thus, the mean weight in the plus group changes with time. A solution to this may be to use available data on the mean weight in the plus group, if such data is available on an annual basis.

A second problem with the plus group is that an incorrect value of natural mortality may cause a lot of non-existing fish to aggregate within the plus group.

7.6 Recruitment

Numbers model:
Recruitment is the number of individuals in an incoming year-class. Growth of older individuals, natural mortality and weight of recruits is separate.
Typical model: $R = \alpha S/(1 + S/K)$.

Production model:
Recruitment is the growth in biomass or surplus production. Incorporates growth of individuals, natural mortality and recruiting year-class in weight.
Typical model: $rB(1 - B/K)$.

7.7 Initial population size

In a forward projection one needs to set an initial population size (all ages, in first year)
Can use equilibrium assumption $\Rightarrow$ only one parameter:

$$N_{0,a} = N_{0,0} e^{-M \cdot a}$$

(where we start the ages and years from 0 so $N_{y,0} = R_y$)
Can try to estimate all ages $\Rightarrow$ more parameters but also more potential to fit data for initial year (if available).

7.7.1 Details

When using a model which projects forward in time some assumption needs to be used for the initial population size. This applies to all ages in the first year, i.e. a single total biomass in a bulk biomass model, but all true ages as well as the plus group in a numbers model.

It is possible to use an equilibrium assumption to reduce the number of parameters required. For example, in a numbers model an assumption of steady-state and no historical fishing will mean that the numbers at age in the first year are given by $N_{a,0} = N_{a-1,0}e^{-M_a}$ where M_a is usually an assumed number and this reduced the first-year parameter set to only the historical number of recruits.

Naturally one can also try to estimate the numbers at age for all ages in the first year. This will require more parameters but the approach also has more potential to fit the data better for the initial year (if such data is available, e.g. in the form of a length distribution with distinct peaks).

7.8 Initializing an estimation procedure

7.8.1 Details

When a parametric statistical approach is used to fit a population dynamics model, the first step is to input data and set the initial values of the parameters.

7.8.2 Examples

Example 7.1. Some R code for setting initial values in a statistical age-based population model is given below. The data are for nephrops in Icelandic waters from 1980.

http://tutor-web.net/fish/fish5108statass/lecture50/aspminit.r

7.9 Projecting a stock in numbers forward in time

7.9.1 Details

When a stock in numbers is to be projected forwards in time from known parameters, the formulas are usually the standard catch and stock equations, with provisions for the recruiting year-class and the plus group as indicated previously.

7.9.2 Examples

Example 7.2. The following R code will project stock biomass forward in time. This uses a stock-numbers model and assumes some true ages (at least one) with a plus group. Note that this only assumes estimation of fishing mortality and recruitment. The initial stock size is set to an equilibrium stock based on a constant recruitment.

The code can be downloaded from:
http://tutor-web.net/fish/fish5108statass/lecture60/aspm.r

7.10 Evaluating a model fit

Usually an evaluation of a fit is coded as a function which returns a fit measure for a given set of parameters.

7.10.1 Details

In order to fit a model a method is needed for comparing model output to data. Such a method typically computes the sums of squares between each data set and the corresponding fitted values. In particular, when a model is used for fishing mortality, the catches are usually predicted and similarly, survey indices are typically predicted from proportionality with the biomass.

7.10.2 Examples

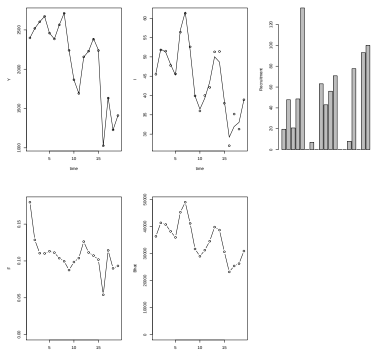
Example 7.3. A typical fit-evaluation function in R takes a vector of parameters as input. Each element (or group of elements) of the vector corresponds to a specific population parameter (or group). In the following example, the first parameter is the cathcability. This is followed by a vector of annual fishing mortalities and a vector of annual recruitment values.

http://tutor-web.net/fish/fish5108statass/lecture60/ssefcn.r

7.11 A complete run

Need more data, e.g. on recruitment in order to reduce the number of parameters compared to number of data points.

Would prefer production to be linked to stock size.



7.11.1 Details

A complete model-fitting run consists of first initializing the parameters, followed by function definitions, and finally calling the nonlinear minimizers.

7.11.2 Examples

Example 7.4. When fitting a model to the nephrops data and plotting the results the R commands below can be used.

One normally first sets up a "run file", containing commands which can be run using a single "source"command in R. This should automatically read in any other R command files, which read data, initialize variables, estimate the stock size and plot results.

http://tutor-web.net/fish/fish5108statass/lecture60/aspmrunplot.r

References ISBN:
ISBN:

8 Finicky details

8.1 The weighting factors

$$SSE_Y + SSE_I$$

or weighted?

Logged data:

$$\lambda \sum_t \left(ln\left(x_t\right) - ln\left(\hat{x}_t\right) \right)^2$$

where

$$\lambda = \frac{1}{V\left[ln\left(x_t\right)\right]} = \frac{1}{\sigma^2_{ln(x_t)}}$$

if possible...

8.1.1 Details

As indicated above, there may be several data sources. Notably, there will usually be landings information and one or more survey indices. The corresponding sums of squares are commonly added together to form a single objective function to be minimized. This is only rarely acceptable and may lead to seriously incorrect results.

8.1.2 Handout

Consider the estimation of weighted factors (weights) to be assigned to sums of squares which will be minimized in order to estimate parameters in a population dynamics model for a fish species. Assume that all of the terms correspond to logged data. Each term is thus of the form

$$\lambda \sum_t \left(ln\left(x_t\right) - ln\left(\hat{x}_t\right) \right)^2$$

where the x_t's may be annual landings, indices, or any other data which can be predicted from the model and λ is the weighted factor.

The "correct"weighted factor from a statistical viewpoint is the inverse variance,

$$\lambda = \frac{1}{V\left[ln\left(x_t\right)\right]} = \frac{1}{\sigma^2_{ln(x_t)}}$$

and it should be noted that in the case of low variability, the standard deviations of the logged quantities are close to the coefficient of variation of the original numbers. Since the latter are convenient to think of and the former are computationally convenient, the term $CV(x)$ is commonly used when $\sigma_{ln(x_t)}$ is meant.

In the case of the landings data, these are sometimes known quite precisely in which case it would be natural to assume e.g. $CV(Y) = 0.1$, reflecting a belief that 95% of the annual catch estimates are within 20% (two standard deviations) of their true value [1].

[1] One notes that this approach tacitly assumes a symmetry in the errors

In the case of an abundance index, e.g. a CPUE index, U_t is intended to reflect adult biomass, one could set up a simple time series model to estimate the uncertainty in the index or one could fit a polynomial in time and use the residuals for the variance estimation (cf IWC work, IWC 19xx).

In either case one must keep in mind that the resulting variance estimate only estimates the variance in the data set itself. It does not include the variance with regard to the model uncertainty which may be much higher. These approaches do, however, give fairly objective guidelines on the choice of weights.

Traditional text books on statistics only skim over the choice of weights and there is a reason for this: A fundamental assumption in statistical models is that the model is correct! When this assumption holds the parameter estimates are unbiased regardless of the choice of weights. Unfortunately this assumption rarely holds for any model and although it is often "approximately correct", it is commonly seriously violated in fisheries.

If there are model errors, then the model will typically not be able to fit all data sets simultaneously. In this case the choice of a high weight on one data set will “drive the model” to fit that data set and give poor fits to other sets. It is therefore imperative that several choices of weights be investigated in order to evaluate the level of inconsistency observed.

8.2 Caveats

Typically too many unknowns!

8.2.1 Details

Although the annual fishing mortality is typically quite well determined by the annual yields, few other parameters are well defined. In particular, attempting to determine annual recruitment in the presence of a single total abundance index is doomed to be dubious.

8.3 Adding a production term

8.3.1 Details

It makes little sense to talk of a production model unless there is a production term. In an age-structured model the production term is normally in the form of a stock-recruitment relationship.

8.4 Smoothing

Add penalty terms?

$$\sum (F_{t+1} - F_t)^2$$

or

$$\sum \left(R_{t+1} - \hat{R}_t\right)^2$$

etc

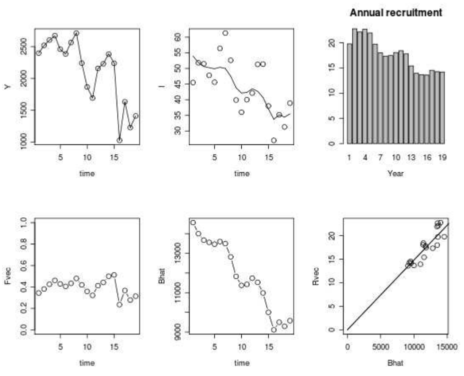
8.4.1 Details

Inserting time-series-like components as sums of squares will smooth the time series of fishing mortalities and/or recruitments. This will usually result in more stable estimates.

The exact way in which this is done is somewhat important.

The method used here has been criticized.

8.5 A typical run



Example of a complete run

8.5.1 Examples

Example 8.1. A collection of R routines for conducting a simple age-structured assessment with a production component is given below.

First, consider the initialization phase.
http://tutor-web.net/fish/fish5108statass/lecture70/aspminit.r

Next, a function to project the stock forward one year is needed.
http://tutor-web.net/fish/fish5108statass/lecture70/aspm.r
http://tutor-web.net/fish/fish5108statass/lecture70/aspmrun.r
http://tutor-web.net/fish/fish5108statass/lecture70/ssefcn.r

These methods can be improved quite a bit. To get more precise estimates one needs more data, however!

9 Some case studies

9.1 Stochastic simulations

Simulate a population with errors in data

9.1.1 Examples

Example 9.1. Example R code for the assessment:

init.r
http://tutor-web.net/fish/fish5108statass/lecture80/init.r

predict.r
http://tutor-web.net/fish/fish5108statass/lecture80/predict.r

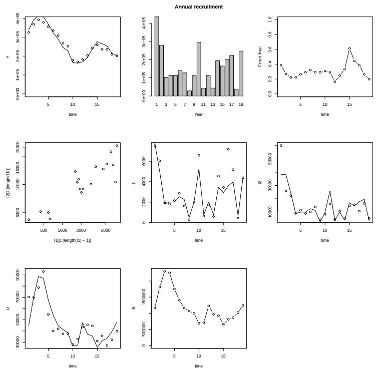
simpop.r
http://tutor-web.net/fish/fish5108statass/lecture80/simpop.r

ssefcn.r
http://tutor-web.net/fish/fish5108statass/lecture80/ssefcn.r

statass.r
http://tutor-web.net/fish/fish5108statass/lecture80/ssefcn.r

9.2 Cod in Icelandic waters

9.2.1 Details



Consider a simple example based on cod in Icelandic waters.

Data: Total landings and indices based on a groundfish survey by extracting "ages"1 and 2 from the scaled length distributions as well as the plus group (all in numbers).

Note that the landings are almost entirely of ages 3+ so the approximate age indices are pre-recruitment indices.