Gadget course

Day 2

Bjarki Þór Elvarsson and Guðmundur Þórðarsson

University of Iceland, Marine Research Institute

22. febrúar 2012

Today:

1. Fitting models to data
   - Available likelihood components
2. Minimization Algorithms in Gadget
3. Data Warehouse
4. Uncertainty Estimates in Gadget
   - Bootstrap

Where are we now...

1. Fitting models to data
   - Available likelihood components
2. Minimization Algorithms in Gadget
3. Data Warehouse
4. Uncertainty Estimates in Gadget
   - Bootstrap

Datasets

A typical model in Gadget tries to emulate various processes such as growth, migration and maturation. In order for these processes to make some biological sense we need to compare our model with real data. Gadget can use a variety of datasets to compare with simulated data from the model. In a typical model parameter estimation relies on datasets such as:

- **Abundance estimates**, such as survey indices or CPUE.
- **Length distribution**, both from surveys and/or commercial operations.
- **Age-length tables**.
- **Tagging data** .
- **Mean length at age**.

Datasets (continued)

Data on a marine resource is sampled for different purposes, i.e. estimate growth or abundance. Therefore data used in modelling exhibits often wildly different characteristics. As an example:

- Abundance estimates are usually estimate for three or four groups on a yearly basis (i.e. very small dataset)
- Age-length keys are however very abundant in data, specially for commercial samples.
- Tagging data is seriously correlated (Fish tagged together stays together).

There also other considerations, such as year effects in the survey samples, year class effects and missing data, that further affect the end result.

Likelihood functions

- The simulated (predicted) data is compared to real data using a likelihood function. The likelihood function is split into likelihood components that depend on the type of the dataset. Typically they are of the form

$$l_i = -\log(f(x - \hat{x}))$$

  where f, x and $\hat{x}$ are the error distribution function, data and prediction respectively.
- We can, in some sense, think of likelihood function as the differences between the model and all datasets. The components then are then the difference between a single dataset and the model.

Combining datasets

Typically, several components enter the likelihood function in any single estimation. Thus the objective function becomes a **weighted** sum of several likelihood components:

$$l = \sum_i w_i l_i$$

The weights, w_i, are necessary for several reasons:

- to prevent some components from dominating the likelihood function.
- to reduce the effect of low quality data.
- as a priori estimates of the variance in each subset of the data.
- increases accuracy (at the expense of introducing some bias).

Assigning weights

- Not a trivial matter, has in the past been the most time consuming part of a Gadget model.
- Usually done using *'Expert judgement'*
- There are however more objective ways;
  - **Iterative re-weighing:** Method introduced in Stefánsson (2003) and implemented for cod in Taylor et al (2007). The approach has now been implemented in R as a general tool for assigning weights to likelihood components in Gadget models.
  - Cross-Validation approach (Wang and Zidek 2005)
  - General heuristics discussed in Francis (2011)

Iterative re-weighing

The general idea behind the iterative re-weighing is to assign the inverse variance of the fitted residuals as component weights. The variances, and hence the final weights, are calculated according the following algorithm:

- Calculate the initial SS given the initial parametrization. Assign the inverse SS as the initial weight for all likelihood components.
- For each likelihood component, do an optimization run with the initial score for that component set to 10000. Then estimate the residual variance using the resulting SS of that component divided by the degrees of freedom (df^*), i.e. $\hat{\sigma}^2 = \frac{SS}{df^*}$.
- After the optimization set the final weight for that all components as the inverse of the estimated variance from the step above (weight $= 1/\hat{\sigma}^2$).

Iterative re-weighing df^*

The effective number of data-points (df^*) is used as a proxy for the degrees of freedom determined from the number of non-zero data-points.

- Viewed as satisfactory proxy when the data-set is large
- For smaller data-sets this could be a gross overestimate.

In particular, if the survey indices are weighed on their own while the yearly recruitment is estimated they could be over-fitted. In general problem such as these can be solved with component grouping, that is in step 2 the likelihood components that should behave similarly, such as survey indices, should be heavily weighted and optimized together.

Available likelihood components

- **BoundLikelihood**, **Understocking** and **MigrationPenalty** assign a penalty if the parameters are out of bounds, more is consumed than is available or a negative amount migrates between areas (which is meaningless).
- **SurveyIndices** compares the development of a stock in the Gadget model to indices calculated from a standardized survey for that stock using:

$$\ell = \sum_{time} \Big(I_t - (\alpha + \beta N_t) \Big)^2 \tag{1}$$

where α and β can either be estimated or specified.

Available likelihood components (continued)

- **CatchDistribution** compares distribution data sampled from the model with distribution data sampled from landings or surveys. The user can define a range of likelihood functions. These include:
  - Sum of squares and statified sum of squares::

$$\ell = \sum_{time} \sum_{areas} \sum_{ages} \sum_{lengths} \Big(P_{tral} - \pi_{tral} \Big)^2 \tag{2}$$

  - Multinomial function:

$$\begin{aligned} \ell = 2 \sum_{time} \sum_{areas} \sum_{ages} \Bigg(& \log N_{tra}! - \sum_{lengths} \log N_{tral}! + \\ & \sum_{lengths} \Big(N_{tral} \log \frac{\nu_{tral}}{\sum \nu_{tral}} \Big) \Bigg) \end{aligned} \tag{3}$$

Available likelihood components (continued)

- **CatchDistribution** (cont.)
  - Multivariate normal:

$$\ell = \sum_{time} \sum_{areas} \sum_{ages} \Big(log|\Sigma| + (P_{tra} - \pi_{tra})^T \Sigma^{-1} (P_{tra} - \pi_{tra}) \Big) \qquad (4)$$

  - ... and a few others..
- **CatchStatistics** compares statistical data sampled from the model with statistical data sampled from landings or surveys. Uses either a weigthed or unweigthed sum of squares.
- **StockDistribution** compares distribution data sampled from the model with distribution data sampled from landings or surveys for different stocks within the Gadget model. Uses either a multinomial or a sum of squares.

Available likelihood components (continued)

- **SurveyDistribution** compares the development of a stock in the Gadget model to age-length indices calculated from a survey for that stock.
- **StomachContent** compares consumption data sampled from the model with stomach content data obtained by analysing the stomach contents of various predators. Uses sum of squares.
- **Recaptures** compares recaptures data from tagging experiments within the model with recaptures data obtained from tagging experiments, aggregated according to length at recapture. Uses a Possion likelihood function of the form:

$$\ell = \sum_{time} \sum_{areas} \sum_{lengths} \Big(N_{trl} + \log \nu_{trl}! - N_{trl} \log \nu_{trl} \Big) \tag{5}$$

Available likelihood components (continued)

- **RecStatistics** compares statistical data sampled from tagged subpopulations within the model with statistical data obtained from the fish returned from tagging experiments. Uses weighted sum of squares.
- **CatchInKilos** compares the overall catch from the modelled fleets with landings data. Uses sum of squares.

Where are we now...

1. Fitting models to data
   - Available likelihood components
2. Minimization Algorithms in Gadget
3. Data Warehouse
4. Uncertainty Estimates in Gadget
   - Bootstrap

Parameter estimates

One of the most important use of Gadget is find parameter estimates. It does so is a customary fashion by minimizing an objective function. Gadget uses the likelihood function to compare simulated values to real data. This means that Gadget runs a full simulation for each likelihood score. Currently there are three minimization algorithms implemented in Gadget:

- **Simulated Annealing** a global search algorithm.
- **Hooke and Jeeves** a local search algorithm.
- **BFGS** a quasi–Newton search algorithm.

These algorithm have different properties and usually the estimation procedure uses a combination of all three.

Optinfofile

An example of an optinfofile:

```
[simann]
simanniter 200000 ; number of simulated annealing iterations
simanneps 1e-03 ; minimum epsilon, simann halt criteria
t         30000000   ; simulated annealing initial temperature
rt        0.85  ; temperature reduction factor
nt        2     ; number of loops before temperature adjusted
ns        5     ; number of loops before step length adjusted
vm        1     ; initial value for the maximum step length
cstep     2     ; step length adjustment factor
lratio    0.3   ; lower limit for ratio when adjusting step len
uratio    0.7   ; upper limit for ratio when adjusting step len
check     4     ; number of temperature loops to check
```

Optinfofile

```
[hooke]
hookeiter 20000  ; number of hooke & jeeves iterations
hookeeps  1e-04 ; minimum epsilon, hooke & jeeves halt criteria
rho       0.5   ; value for the resizing multiplier
lambda    0     ; initial value for the step length
[bfgs]
bfgsiter  10000 ; number of bfgs iterations
bfgseps   0.01  ; minimum epsilon, bfgs halt criteria
sigma     0.01  ; armijo convergence criteria
beta      0.3   ; armijo adjustment factor
```

Where are we now...

1. Fitting models to data
   - Available likelihood components
2. Minimization Algorithms in Gadget
3. Data Warehouse
4. Uncertainty Estimates in Gadget
   - Bootstrap

Introduction

Gadget models only go so far were it not for data. As in any statistical analysis proper data handling is important to ensure that the model describes reality. Classically statistical analysis is split up into (at least) three dependent stages:

1. **Acquisition:** surveys and fleet samples...
2. **Assimilation:** data validation, merging datasets, formatting..
3. **Analysis:** model building, testing hypothesis, projections..

Typically, step 2 is the most time consuming and not handled consistently.

DST[2]

- DST[1] was a 4-year fisheries project aimed at modeling and simulation of marine ecosystem that used Gadget as its foundation for all modeling work.
- Gadget needs for its run fair amounts of ecosystem data. These are typically available from the institutional databases.
- These databases are usually very heterogeneous between institutions and often lack robustness and consistency.
- To enable easier and potentially faster access to these databases, with the stress being put on the creation of Gadget input data files as specialized database, the Data Warehouse (DW), was developed.

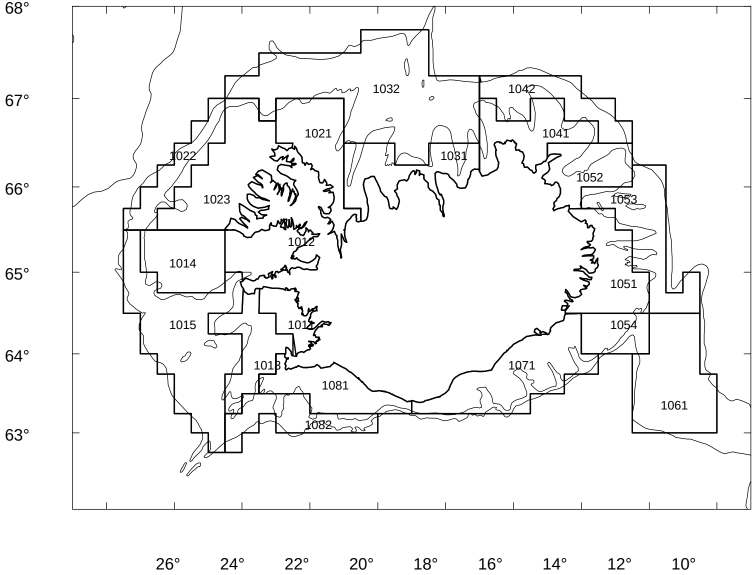
[1]The acronym stands for *Development of structurally detailed statistically testable models of marine populations*

Data Warehouse

- The DW is designed to be a centralised repository for all Gadget data. The interface to the DW is consistent across institutions and can import data from other databases using ASCII–files[2].
- The DW is designed to contain all significant data used by ecosystem models in a minimally aggregated form. This means that the scale of the data is not too fine as that would introduce unwanted statistical properties (such as diel variation) into the data. It is however not too coarse either to allow the user to model all relevant processes.
- Typically data is aggregated on a monthly basis according by geographical areas. These areas are defined in such a way that all significant correlation is contained within the area.

[2] Lowest common denominator

Gadget areas



Where are we now...

1 Fitting models to data
- Available likelihood components

2 Minimization Algorithms in Gadget

3 Data Warehouse

4 Uncertainty Estimates in Gadget
- Bootstrap

Inferences with Gadget

We have now seen how to estimate model parameters with Gadget. What about uncertainty? And more importantly what can be done with it. Possible uses include:

- Find the best time and space resolution for the model.
- Compare different models
- Introduce randomness into model forcast

The following slides introduce current work on uncertainty estimates in Gadget.

Uncertainty estimates in Multi-species models

- Variance estimates of parameters in nonlinear models have commonly been derived from the inverted Hessian matrix at the optimum, when the method of least squares (or maximum likelihood) is employed for parameter estimation.
- The Jacobian matrix (i.e. variance–covariance matrix) of the residuals can also be used.
- However, several conditions need to be satisfied for statistical inference, e.g. confidence statements to hold in the finite-sample case.

  1. the model needs to be correct.
  2. observations need to be normally distributed.
  3. variance assumptions i.e. homoscedasticity and knowledge of the ratios of variances in individual data sets, need to be appropriate.

- Alternative data distributions can in principle be used as extensively developed in the theory of generalised linear models.

Methods

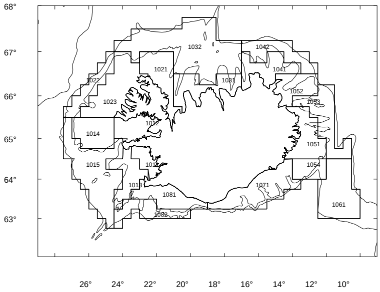
- Hessian based methods have seen to fail as seen in several examples in fishery science (Patterson et. al., 2001).
- One can therefore not assume a priori that Hessian–based methods yield reasonable results.
- Disregarding correlation structure, as an example, when present can potentially lead to incorrect conclusions, with serious consequences (Myers and Cadigan, 1995).
- Multimodal likelihood functions have also been seen in real applications (i.e. IWC fin whale model).

Bootstrap

- Alternative methods include bootstrap methods. The simplest bootstrap method assumes that the data are independent measurements without correlation. Doing this assumes that all individually measured fish are independent which is invalid for several reasons (e.g Hrafnkelsson and Stefansson, 2004).
- Semi-parametric approaches have also been developed to sample from model residuals, which assumes that all relevant correlations have been modelled.
- Resampling entire fish samples (as is done by Singh et al., 2011) can potentially be used to account for intra–haul correlation.
- However, considering samples as units may not be quite enough, since fish at close geographic locations will also tend to be similar due to a fine–scale spatial structure which can not be easily modelled (e.g. Stefansson and Palsson, 1997a).

Bootstrap in statistical fisheries model

- When modelling population dynamics, the pertinent spatial and temporal scales will depend on the research question posed.
- Aggregations are made in order to reduce correlations between the elementary data units.



- Aggregation by month and Gadget–area should eliminate intra–haul correlations (Pennington and Volstad, 1994) and those correlations between age–groups (Myers and Cadigan, 1995) which are related to local shoals or small feeding patches.

Bootstrap based on geographical units

The bootstrapping approach consists of the following:

- The base data are stored in the DW
- To bootstrap the data, the list of subdivisions required for the model is sampled (with replacement) and stored. For a multi–area model resampling of subdivisions is conducted within each area of the model.
- The list of resampled (with replacement) subdivisions is then used to extract data.
- For a single bootstrap Gadget model, the same list of resampled subdivisions is used to extract each likelihood dataset i.e. length distributions, survey indices and age–length frequencies are extracted from the same spatial definition.
- The full dataset is extracted and 1000 bootstrap samples generated.

Estimation protocol(s)

The whole purpose is to obtain an estimate of the distribution of parameter estimates using the bootstrap datasets.

- How does the bootstrap approach compare to a Hessian – based approach?
- For each dataset the whole estimation procedure is repeated, in particular iterative re–weighing is repeated for all datasets (think 1000 x 2 x 8 hours). Can we reduce the time required by using fixed weights for all datasets?
- What is the effect of the number of bootstrap samples? Can we make do with 100 samples?

To answer these question we will use a model for Cod.

Cod model – fitted parameters

The model used here was initially described in Taylor et. al (2003). The parameters estimated in it are:

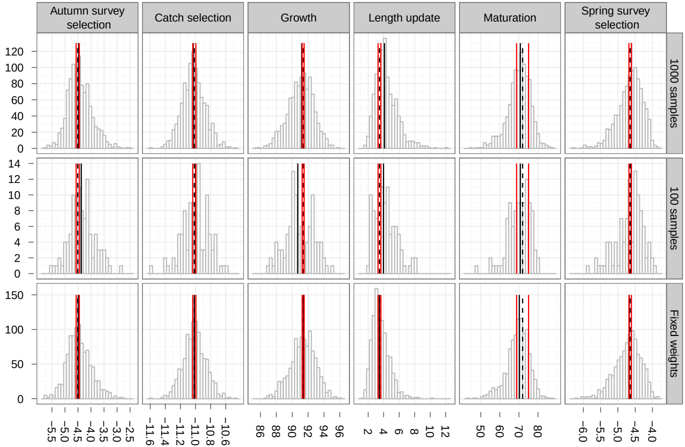
- one parameter for the Von Bertalanffy growth function
- one parameter for the length update (i.e. Beta-Binomial growth transition matrix)
- one parameter for each fleet selection pattern (commercial catch, March survey and October survey). Exponential function.
- two parameters defining the maturation ogive, l_{50} and a_{50} (which are correlated)
- the number of recruits (abundance of age 1) for each year (1984 — 2003)
- the abundance at ages 2 — 11 at the start of the model in 1984

Cod model – likelihood data

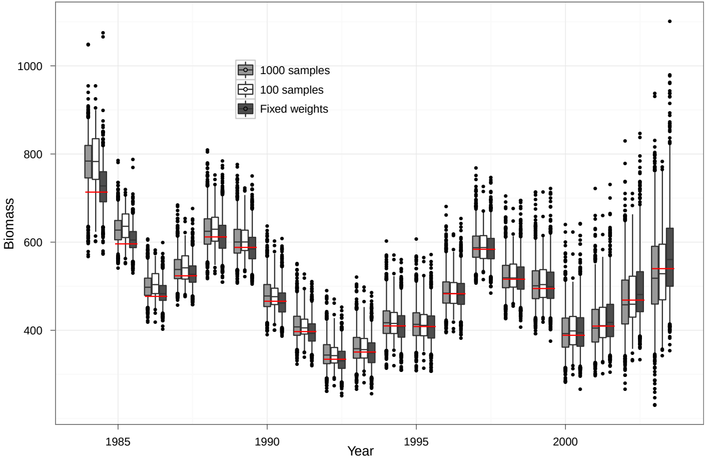
The likelihood data used are:

- Length distributions from the March survey (1985 — 2003), October survey (1995 — 2003) and sampling from the commercial fishery (1984 — 2003).
- Age–length frequencies from the March survey (1989 — 2003), October survey (1995 — 2003) and sampling from the commercial fishery (1984 — 2003).
- Survey indices from the March survey (1985 — 2003) and October survey (1995 — 2003). These are calculated from the length distributions and are disaggregated ("sliced") into three groups which correspond roughly to age 1, age 2 and age3+.
- The ratio of immature:mature by length group from the March survey (1985 — 2003).

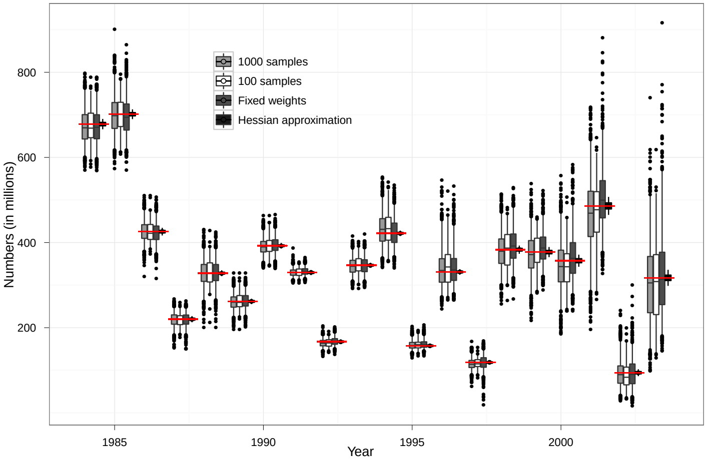
Results


Autumn survey selection
Catch selection
Growth
Length update
Maturation
Spring survey selection
1000 samples
100 samples
Fixed weights
120
100
80
60
40
20
0
14
12
10
8
6
4
2
0
150
100
50
0
-5.5 -5.0 -4.5 -4.0 -3.5 -3.0 -2.5
-11.6 -11.4 -11.2 -11.0 -10.8 -10.6
86 88 90 92 94 96
2 4 6 8 10 12
50 60 70 80
-6.0 -5.5 -5.0 -4.5 -4.0

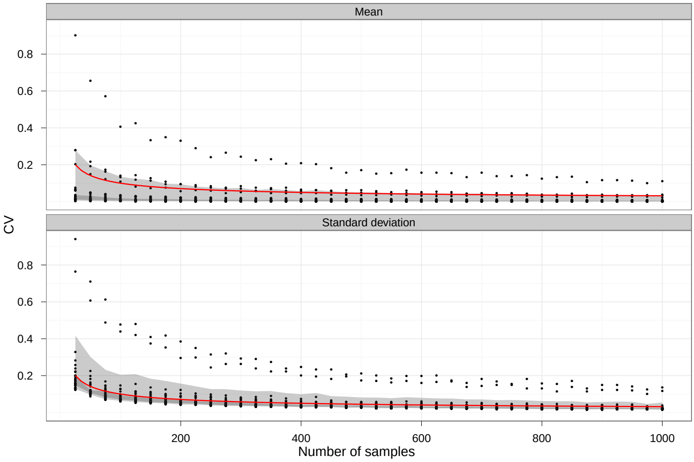
Results


1000 samples
100 samples
Fixed weights
Biomass
1000
800
600
400
Year
1985
1990
1995
2000

Results


1000 samples
100 samples
Fixed weights
Hessian approximation
Numbers (in millions)
800
600
400
200
1985
1990
1995
2000
Year

Results


Mean
0.8
0.6
0.4
0.2
CV
Standard deviation
0.8
0.6
0.4
0.2
200
400
600
800
1000
Number of samples

Discussion