Gadget course

Day 2 - the devil is in the details

Guðmundur Þórðarsson and Bjarki Þór Elvarsson

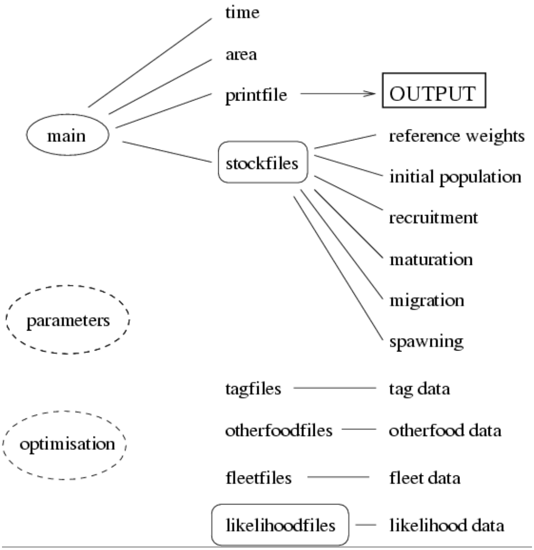
Marine Research Institute, University of Iceland

23. febrúar 2012

ToC:

1. What we learned so far
2. Data Input files
3. Changes to the simulation model
   - The likelihood file
   - The print file
4. Running a minimization
5. Getting the results
   - Likelihoodprinters plots
   - Results plots
6. Prognosis

Gadget file structure



Data Input Files

We have only looked very briefly at the data input files for `Gadget` . The file we have looked at so far or the:

- tuskrefw.dat

Can not really be termed as data input files but rather as a supplementary file. So far the only actual data file is the `fleet.data` file.
Now we will add three types of data files to the model:

- Disaggregated survey indices (10 cm)
- Length distributions (2 cm) interval
- Age-structured data

Disaggregated survey indices: `surveyindices.dat`

```
;year step area lengthgr no
1985 1 allareas len10-19 107.9
1985 1 allareas len20-29 552.7
1985 1 allareas len30-39 918.5
1985 1 allareas len40-49 1208.2
1985 1 allareas len50-59 801.7
1985 1 allareas len60-69 317
1985 1 allareas len70-79 55.9
1985 1 allareas len80+ 16.5
1986 1 allareas len10-19 110.8
1986 1 allareas len20-29 378.8
...
2011 1 allareas len50-59 952.6
2011 1 allareas len60-69 226.2
2011 1 allareas len70-79 13.7
2011 1 allareas len80+ 12.9
```

Note that `Gadget` does not tolerate zeros in the the numbers.

Length distributions: catchledist.dat, surveyledist.dat

catchledist.dat: These are length distributions from the commercial fleet divided by quarters by year. surveyledist.dat: is the length distributions from the March survey (set at step 1).

```
; from catchledist.dat
....
2001 1 allareas allages len100-101 1
2001 1 allareas allages len34-35 1
2001 1 allareas allages len36-37 1
2001 1 allareas allages len38-39 8
2001 1 allareas allages len40-41 19
2001 1 allareas allages len42-43 13
2001 1 allareas allages len44-45 25
2001 1 allareas allages len46-47 32
2001 1 allareas allages len48-49 40
2001 1 allareas allages len50-51 32
2001 1 allareas allages len52-53 33
2001 1 allareas allages len54-55 40
2001 1 allareas allages len56-57 31
2001 1 allareas allages len58-59 34
.....
```

Note that Gadget does not care about zeros in these data files.

Age structured data: catchalkeys.dat, surveyalkeys.dat

Same applies as to the length distribution data, below is an example of the structure form the surveyalkeys.dat file:

```
;year step area, age, length no
2011 1 allareas 6 len50-51 2
2011 1 allareas 7 len50-51 22
2011 1 allareas 8 len50-51 25
2011 1 allareas 9 len50-51 5
2011 1 allareas 10 len50-51 3
2011 1 allareas 5 len52-53 1
2011 1 allareas 7 len52-53 6
2011 1 allareas 8 len52-53 14
2011 1 allareas 9 len52-53 15
2011 1 allareas 10 len52-53 1
2011 1 allareas 11 len52-53 1
2011 1 allareas 12 len52-53 1
2011 1 allareas 6 len54-55 1
```

Final on data input files

We have now covered the data input files used in the tusk model. As mentioned earlier there are many more types of data input files Gadget can use. One type of data file should be mentioned here as it can be quite important and that is the lengthgivenstddev file which can give Gadget information on growth in the absence of age structured data.

```
;year step area number meanle stddev
1995 1 allareas 5 11 45     4.656
1995 1 allareas 6 42 46.476 5.024
1995 1 allareas 7 64 51.406 5.364
1995 1 allareas 8 49 53.592 5.74
1995 1 allareas 9 26 57.538 6.116
1995 1 allareas 10 20 61.2   6.375
1995 1 allareas 11  4 66.25  6.366
1995 1 allareas 12  2 72.5   5.908
1995 2 allareas 6  4 54.75  5.024
```

Making Gadget aware of datafiles

The `Gadget` model population and fleets are linked to the data through the `likelihood` file.

In many ways this file is the fundamentally most challanging file in a `Gadget` model. The reasons are many but the one that pops into mind at first glance is the weights of the various likelihood components.

The 'likelihood components' refers to the different datasets that are part of the overall objective function.

We need to modify our 'simulation model' from Monday to bbe able to change it to an 'estimation model'.

- Add fleet 'survey' to the FLEET-file
- Add survey catches to the fleet.data file
- Change the LIKELIHOOD file (the bridge)
- Change the PRINT file to get measure of the fit between data and model

The likelihood file (I)

The first part of the likelihood file is normally the penalty and the understocking. This can be viewed as components to keep the model in line.

- The bounds component keeps the H&J algorithim in line so that it does not search outside of the bounds specified in the parameter file.
- The understocking component makes sure that we can not have more catches than are available in the ocean according to the model

```
[component]
name bounds
weight 10
type penalty
datafile penaltyfile
;
[component]
name understocking
weight 1
type understocking
```

The likelihood file (II)

In the tusk model we divide the disaggregated survey indices into three likelihood components, namely: si2039 (juveniles), si4069 (fishable stock) and si70110 (OAPs).

Each component has different weight and different lenaggfile.

```
[component]
name si2039
weight  8.477335
type surveyindices
datafile Data/surveyindices.dat
sitype lengths
areaaggfile AggFiles/allarea.agg
lenaggfile AggFiles/si2039len.agg
stocknames tusk
fittype fixedslopeloglinearfit
slope 1
;
[component]
name si4069
weight 14.349347
....
;
[component]
name si70110
weight  4.664863
....
```

Aggregation files for survey indices

```
si2039len.agg
;   name minl maxl
len20-29 19.5 29.5
len30-39 29.5 39.5

si4069len.agg
;   name minl maxl
len40-49 39.5 49.5
len50-59 49.5 59.5
len60-69 59.5 69.5

si70110    len.agg
;   name minl maxl
len70-79 69.5 79.5
len80+ 79.5 110.5
```

The likelihood file (III)

```
[component]
name ldist.catch
weight 0.09065913
type catchdistribution
datafile Data/catchledist.dat
function multinomial
overconsumption 1
minimumprobability 20
areaaggfile AggFiles/allarea.agg
ageaggfile AggFiles/allage.agg
lenaggfile AggFiles/len2.agg
fleetnames comm
stocknames tusk
;
[component]
name ldist.survey
weight 0.03107802
type catchdistribution
datafile Data/surveyledist.dat
function multinomial
overconsumption 1
minimumprobability 20
areaaggfile AggFiles/allarea.agg
ageaggfile AggFiles/allage.agg
lenaggfile AggFiles/len2.agg
fleetnames survey
stocknames tusk
```

The likelihood file (IV)

```
[component]
name alkeys.catch
weight 0.36706690
type catchdistribution
datafile Data/catchalkeys.dat
function multinomial
overconsumption 1
minimumprobability 20
areaaggfile AggFiles/allarea.agg
ageaggfile AggFiles/age.agg
lenaggfile AggFiles/len2.agg
fleetnames comm
stocknames tusk
;
[component]
name alkeys.survey
weight 0.21764955
type catchdistribution
datafile Data/surveyalkeys.dat
function multinomial
overconsumption 1
minimumprobability 20
areaaggfile AggFiles/allarea.agg
ageaggfile AggFiles/age.agg
lenaggfile AggFiles/len2.agg
fleetnames survey
stocknames tusk
```

The print file (I)

We do want two things from our Gadget run:

- Diagnostics of the fit to the dat
- the results (estimates)

Here are the first two printers in the PRINT file. The first one is the same from the simulation model. The second is new and gives the likelihood value for all the components (Have never looked at this)

```
[component]
type            stockstdprinter
stockname       tusk
printfile       out/tusk.std
yearsandsteps   all all
;
[component]
type            likelihoodsummaryprinter
printfile       out/likelihoodsummary
;
```

likelihoodsummaryprinter

```
; Gadget version 2.1.06 running on hafsili Mon Feb 20 22:01:15 2012
; Summary likelihood information from the current run
; year-step-area-component-weight-likelihood value
all   all   allareas       si2039   8.477     3.189979
all   all   allareas       si4069   14.35    6.3084728
all   all   allareas      si70110   4.665    19.093046
1984    1   allareas  ldist.catch 0.09065913    113.08829
1984    4   allareas  ldist.catch 0.09065913    127.28275
1986    1   allareas  ldist.catch 0.09065913    167.62647
1986    4   allareas  ldist.catch 0.09065913    139.56037
1987    1   allareas  ldist.catch 0.09065913    147.29708
1987    4   allareas  ldist.catch 0.09065913    171.50817
1991    1   allareas  ldist.catch 0.09065913    263.33502
1991    2   allareas  ldist.catch 0.09065913    212.40015
```

This may actually be quite useful for spotting dataproblems :-)

The print file (II)

These printers are possibly the most important ones as they are the 'first line of defence' i.e. we check if the model fits these components and if not we normally discard the run.

```
[component]
type       likelihoodprinter
likelihood  si2039
printfile  out/likelihoodsi2039
;
[component]
type     likelihoodprinter
likelihood  si4069
printfile  out/likelihoodsi4069
;
[component]
type     likelihoodprinter
likelihood  si70110
printfile  out/likelihoodsi70110
```

Likelihoodprinter

This gives us the estimated numbers in each length interval and the regression information (parameters not in the parameter file).

```
; year-step-area-label-number
1985    1   allareas   len40-49     11520478
1985    1   allareas   len50-59    5838413.6
1985    1   allareas   len60-69    2658219.5
1986    1   allareas   len40-49     11459825
1986    1   allareas   len50-59    6485150.1
...
; Regression information for area allareas
; len40-49 intercept -9.2447159 slope 1 sse 1.0669309
; len50-59 intercept -9.175146 slope 1 sse 1.6219904
; len60-69 intercept -9.5781428 slope 1 sse 3.6195514
```

The print file (III)

These printers are not as important as the others but nevertheless can be quite informative. They tell us how well Gadget is fitting the length distributions. We can do similar things for the age data.

```
[component]
type      likelihoodprinter
likelihood  ldist.survey
printfile  out/likelihoodsLDsurvey
;
[component]
type      likelihoodprinter
likelihood  ldist.catch
printfile  out/likelihoodsLDcatch
```

Likelihoodprinter

This gives us the estimated numbers in each length interval and the regression information (parameters not in the parameter file).

```
; Gadget version 2.1.06 running on hafsili Mon Feb 20 22:01:15 2012
; Likelihood output file for the likelihood component ldist.catch
; year-step-area-age-length-number
1984    1   allareas    allages   len10-11    0.8138476
1984    1   allareas    allages   len12-13    3.6590701
1984    1   allareas    allages   len14-15    12.133467
1984    1   allareas    allages   len16-17    32.936613
1984    1   allareas    allages   len18-19    77.288214
1984    1   allareas    allages   len20-21    162.92477
1984    1   allareas    allages   len22-23    319.36918
```

Running a minimization

Typically we run a minimisation by typing in the command line:

```
gadget -l -i params.in -opt optinfofile
```

The main difference from the simulation model is the '-l' which means that a 'likelihood' or an optimisation srun wil be performed.

The '-opt' means that Gadget should use the settings specified int the 'optinfofile' for the minimisation algorithims.

Getting the results

Like stated earlier we control the output through the printfile (PRINTFILE). Several types of routines and plots have been developed to visualize the results have been developed. Some of these are available through the `gadgetR`-package.

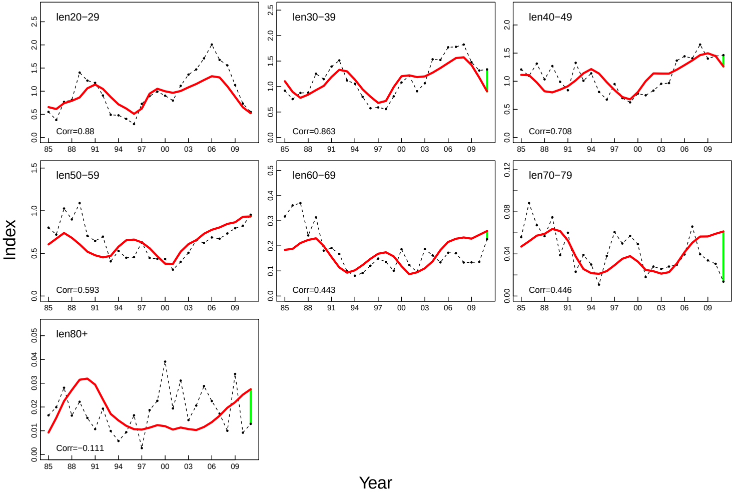
- Likelihoodprinters plots
- Results plots

Likelihoodprinters plots

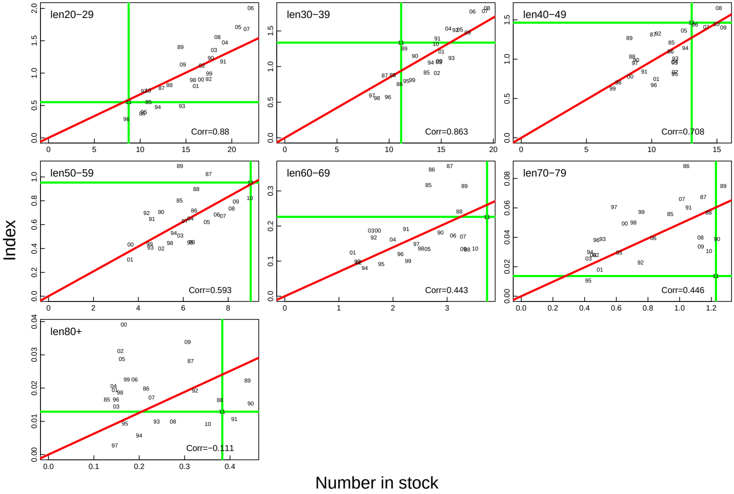
We can divide these into several plots

- Fit to survey indices
  - Timeseries
  - XY-Scatter, fit vs observed
  - Bubble plots
- Fit to length distributions (not implemented in `gadgetR`)

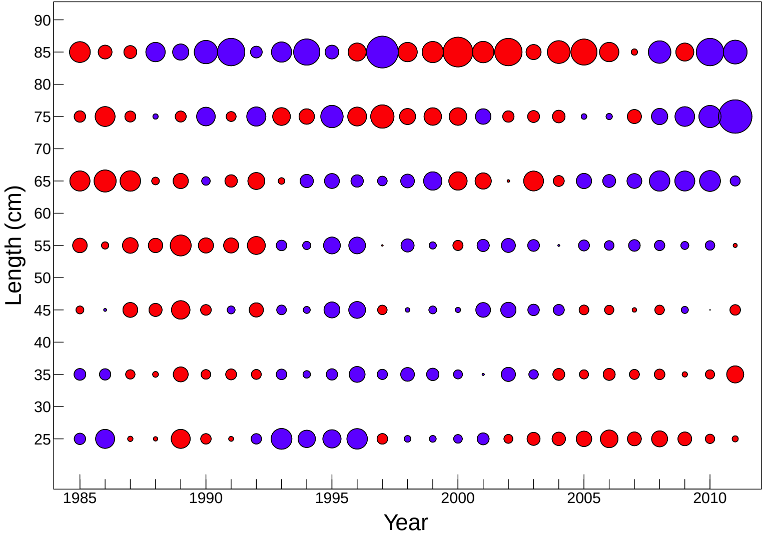
Timeseries plot



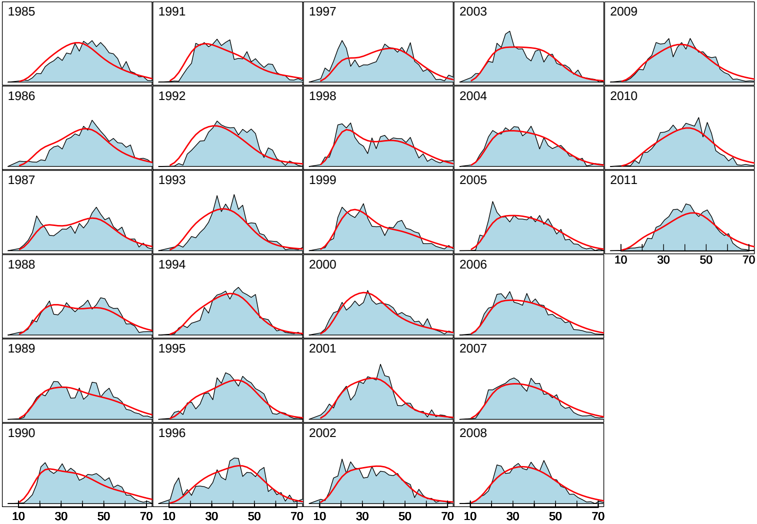
XY plot


len20−29
Corr=0.88
len30−39
Corr=0.863
len40−49
Corr=0.708
len50−59
Corr=0.593
len60−69
Corr=0.443
len70−79
Corr=0.446
len80+
Corr=−0.111
Index
Number in stock

Bubble plot

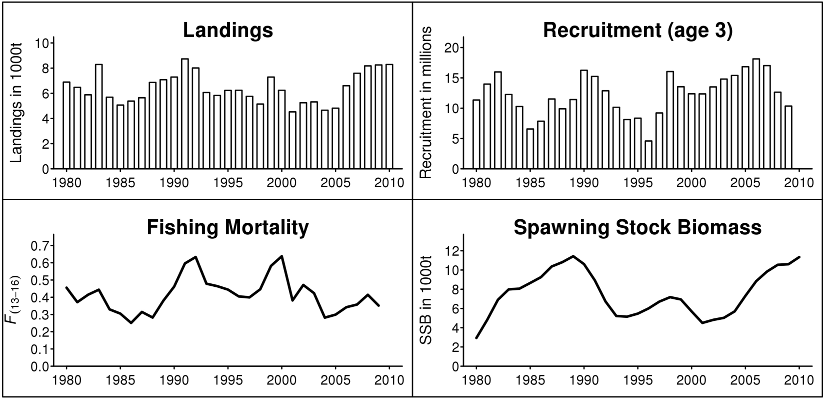

90
85
80
75
70
65
60
55
50
45
40
35
30
25
Length (cm)
1985
1990
1995
2000
2005
2010
Year

Fit to length distributions



Results plots

Just the 'standard' ICES plot.



Prognosis I

We can run prognosis either on F or with constant catches. All we have to do is to define some 'future'-fleet in the FLEET-file:

```
[fleetcomponent]
linearfleet     future
livesonareas     1
multiplicative   aaa
suitability
tusk     function ExpsuitfuncL50    #alphacomm   #L50comm
amount           Data/fleet.predict
```

We could specify a single F we are interested in in the 'multiplicative' and run it. But often we do want to try a few different Fs such as $F = 0$, F_{Max}, $F_{0.1}$ etc. Thats when we use a shell-script (or maybe through R).

Prognosis II

The shell-script `runEffort.sh`:

```
sed -e "s/aa/0/g" < ../predc1.dat.orig > ../predc1.dat

for fmort in 0 0.16 0.29 0
do
    rm -f  FLEET
    sed -e "s/aaa/$fmort/g" < FLEET.orig > FLEET
    gadget  -s -i params.26may11 -o tmp.out
    cp out/tusk.std out/prognosis/tusk.std.$fmort
done
```