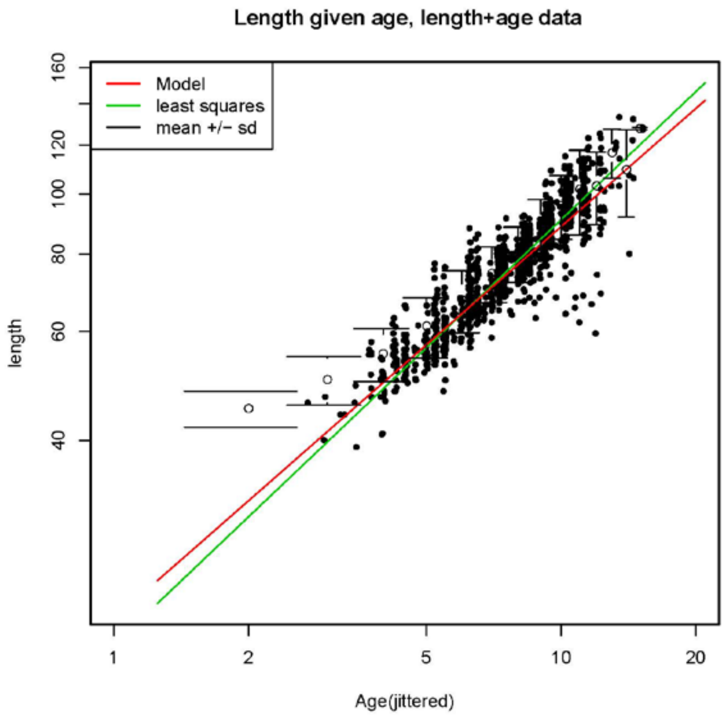
Norsk Regnesentral
Norwegian Computing Center
(+47) 22 85 25 00
nr@nr.no · www.nr.no

Note

Catch-at-age – Version 2.0:

New features and validation of program



Note no SAMBA/57/11

Authors Hanne Rognebakke
David Hirst

Date November 2011

Norsk Regnesentral

Norsk Regnesentral (Norwegian Computing Center, NR) is a private, independent, non-profit foundation established in 1952. NR carries out contract research and development projects in the areas of information and communication technology and applied statistical modelling. The clients are a broad range of industrial, commercial and public service organizations in the national as well as the international market. Our scientific and technical capabilities are further developed in co-operation with The Research Council of Norway and key customers. The results of our projects may take the form of reports, software, prototypes, and short courses. A proof of the confidence and appreciation our clients have for us is given by the fact that most of our new contracts are signed with previous customers.

Norsk Regnesentral
Norwegian Computing Center
Postboks 114, Blindern
N-0314 Oslo, Norway

Besøksadresse
office address
Gaustadalléen 23A, B
N-0371 Oslo, Norway

Telefon · telephone
(+47) 22 85 25 00
Telefaks · telefax
(+47) 22 69 76 60

Bankkonto · bank account
8200.01.48888
Foretaksnr. · enterprise no.
NO 952125001 VAT

Internett · internet
www.nr.no
E-post · e-mail
nr@nr.no

Title	**Catch-at-age – Version 2.0: New features and validation of program**

Authors	**Hanne Rognebakke** hanne.rognebakke@nr.no
	David Hirst david.hirst@nr.no

Date	November 2011
Publication number	SAMBA/57/11

Abstract

NR and the Institute of Marine Research have over years developed a Bayesian hierarchical model to estimate the catch-at-age of fish. The model is implemented in C with an R interface. A new version of the program (version 2.0) is presented here, with a description of new features, validation of the program and comparison with the previous version (version 1.0).

Keywords	Bayesian hierarchical models, MCMC
Target group	Institute of Marine Research
Availability	Open
Project number	220305
Research field	Statistics for Climate, Environment, Marine Resources and Health
Number of pages	34

Contents

1 Introduction

Estimates of catch-at-age are a critical input in most age structured stock assessment processes for commercial fish species throughout the world. NR and the Institute of Marine Research have over years developed a Bayesian hierarchical model to estimate the catch-at-age of fish (see Hirst et al (2004, 2005)).

The last version of the program was sent to IMR in March 2009. This version will be referred to as version 1.0. The new version from 2011 will be referred to as version 2.0. This note is mainly a description of the new features in the model with emphasis on the differences between version1.0 and 2.0. A more detailed description of the model and the simulation algorithm is found in Rognebakke et al (2011).

1.1 Overview

Chapter 2 gives a summary of the model. The changes between version 1.0 and 2.0 are emphasised. The differences in the simulation algorithm between the two versions are described in Chapter 3. Some new features are available in the version 2.0, which are also described in Chapter 3. Chapter 4 gives a thorough validation of the program. In Chapter 5 results from the two versions are compared.

2 The model

The central concept of the model is that we can estimate the proportion-at-age of fish caught, and then assuming we know the total weight landed we can convert the proportions-at-age to numbers-at-age by estimating the mean weight of a fish. We therefore develop models for proportion at age, length-given-age (so we can utilise data where length but not age is measured) and weight-given-length (so we can estimate mean weight). We assume there are four other covariates: year, season, gear and area. We will call a combination of these covariates a cell. It would be simple to include more covariates if required. We assume that the total weight landed is known for each cell. It is well known that between-unit variation in catch composition can be very large and it is crucial to take this into account. This is built into the model by cell- and unit-specific random effects.

The models are described below. Depending on the observation scheme, modelling can be made either on haul level or on trip level. The index u is used to describe the modelling unit of choice. In all cases, the most general models are described, alternatives can be obtained by removing different terms.

2.1 Age model

Let $p_u(a)$ be the probability of a randomly drawn fish from sampling unit u to belong to age group *a*. Age groups are indexed by $a \in \{1,...,A\}$, where *A* will contain fish of age *A* or older. We assume a multinomial logistic-type model

$$p_{c,u}(a) = \frac{\exp(\alpha_u{}^a)}{\sum_{a'} \exp(\alpha_u{}^{a'})},$$

where

$$\alpha_u^a = \alpha^{const,a} + \alpha_y^{year,a} + \alpha_s^{season,a} + \alpha_g^{gear,a} + x_u^{hsz}\alpha_u^{hsz,a} + \zeta_r^{region,a} + \zeta_c^{cell,a} + \zeta_{c,u}^{unit,a}.$$

(1)

The main effects are the following terms; constant $\{\alpha^{const,a}\}$, year $\{\alpha_y^{year,a}\}$, season $\{\alpha_s^{season,a}\}$, gear $\{\alpha_g^{gear,a}\}$, region $\{\zeta_r^{region,a}\}$ and haulsize $\{\alpha_u^{hsz,a}\}$. The terms $\{\zeta_c^{cell,a}\}$ and $\{\zeta_{c,u}^{unit,a}\}$ are cell and unit-specific effects. The *α*-terms are fixed effects while the *ζ*-terms are random effects. x_u^{hsz} is the haulsize in unit *u* and can be measured in either numbers or weight. For identifiability, we assume

$$\sum_a \alpha^{const,a} = 0,$$

$$\sum_a \alpha_y^{year,a} = 0, \quad \forall y,$$

$$\sum_y \alpha_y^{year,a} = 0, \quad \forall a,$$

and likewise for the season, gear and haulsize effects.

If region is included in the model, we assume $\{\zeta_{r_i}^{region,a}\}$ to follow a conditional autoregressive model

$$\left[\zeta_{r_i}^{region,a} \mid \zeta_{r_{i'\neq i}}^{region,a}\right] = N\left(\phi_{age,r} n_i^{-1} \sum_{i'\in\delta(i)} \zeta_{r_i}^{region,a}, \left(\tau_{age}^{region}\left[\phi_{age,r} \cdot n_i + 1 - \phi_{age,r}\right]\right)^{-1}\right),$$

where n_i is the number of neighbours of region *i*, while $\delta(i)$ is the set of neighbours of region *i*. Further, $\phi_{age,r}$ is the AR-parameter, which is assumed uniformly distributed between 0 and 1, and τ_{age}^{region} is the precision parameter. Previously in version 1.0, the variance was given by $\left(\tau_{age}^{region}\phi_{age,r}\right)^{-1}$, which is reasonable if $\phi_{age,r}$ is close to one. However, if $\phi_{age,r}$ is close to zero, which should correspond to independence, the different region effects have different variances. We have therefore used a different model for the variance, which corresponds to making the precision matrix a sum of a spatial and an independent part.

$\{\zeta_c^{cell,a}\}$ are random effects that accounts for interactions between the main effects, while $\{\zeta_{c,u}^{unit,a}\}$ are random effects that accounts for the within-unit correlation. For the random effects we assume

$$\zeta_c^{cell,a} \overset{iid}{\sim} N\left(0, \tau_{age}^{cell\,-1}\right)$$

$$\zeta_{c,u}^{unit,a} \overset{iid}{\sim} N\left(0, \tau_{age}^{unit\,-1}\right),$$

again with a sum-constraint over ages.

In cases of modelling two species, e.g. coastal cod and Atlantic cod (skrei), the same model is used, but with $a \in \{1,\ldots A, A+1,\ldots 2A\}$. The first *A* age groups then corresponds to coastal cod, and the last *A* age groups corresponds to Atlantic cod.

2.2 Length-given-age model

Let $l_{u,f}$ be log-length measurements of fish *f* from unit *u* and $a_{u,f}$ the corresponding age. Then

$$l_{u,f} = \beta_{0,u} + \beta_1 g(a_{u,f};\theta_g) + \varepsilon_{u,f}^{fish},$$

where $\varepsilon_{u,f}^{fish} \overset{iid}{\sim} N(0, \tau_{\lg a}^{fish^{-1}})$ is the random within-unit variation in length-given-age. We assume

$$\beta_{0,u} = \beta^{const} + \beta_y^{year} + \beta_s^{season} + \beta_g^{gear} + x_u^{hsz}\beta_u^{hsz} + \varepsilon_r^{region} + \varepsilon_c^{cell} + \varepsilon_{c,u}^{unit}.$$

Here the β's are main coefficients and the ε's are random effects. They are all modelled similarly to the corresponding terms in the age model.

Note that the age in the length-given-age model, $a_{u,f}$, should be as close as possible to the actual age of the fish *f* rather than the age-class *a* modelled in the age model. We use $a_{u,f} = a + season/4$, if there are 4 seasons in the data set numbered from 1 to 4, and age-class $a = 1, 2$ refer to fish of ages 1, 2 etc.

In the simplest case, $g(\cdot)$ is the log-function (in which case θ_g is empty). Otherwise we use a non-linear age-length model, given by

$$g(a_{u,f};\theta_g) = \log\left[1 - \theta\exp\left(-\gamma \cdot a_{u,f}^c\right)\right],$$

where θ, γ and c are parameters that could be estimated. However, there seems to be hardly any information regarding θ in the data, and simulations indicate that there are many parameter sets that define the same function. Hence, we have fixed two parameters; $\theta = 0.5$ and $c = 1$. In order to avoid identifiability problems with respect to $\beta_{0,u}$ and β_1, we have linearly transformed $g(\cdot)$ such that $g(a_{\min}) = 0$ and $g(a_{\max}) = 1$. In version 1.0 the Schnute-Richards model was used in the non-linear case, but there was often problems with convergence when estimating the parameters since many parameter sets defined the same function.

In case of modelling two species, separate age-length relationships are considered for each species, i.e. separate parameters $\beta_{0,u}$ and β_1, and g-functions.

2.3 Weight-given-length model

Defining $w_{u,f}$ to be the log-weight of fish *f* from unit *u*, we assume

$$w_{u,f} = \delta_{0,u} + \delta_1 l_{u,f} + \nu_{u,f}^{fish},$$

where $\nu_{u,f}^{fish} \overset{iid}{\sim} N\left(0, \tau_{wgl}^{fish^{-1}}\right)$ is the random within-unit variation in weight-given-length, and where $\delta_{0,u}$ and δ_1 are modelled similarly to $\beta_{0,u}$ and β_1.

In case of modelling two species, separate weight-length relationships are considered for each species.

2.4 Age uncertainty

If ages are read by errors, we assume the knowledge of an $A \times A$ transition matrix $\boldsymbol{E}$, where the columns give the conditional probability of the observed fish age, given the true age. Hence, elements $\boldsymbol{E}_{ij}$ contain the probability of observing age i given that the true age is j. The method is also used in the two-stock analysis, except that there are now two age-error matrices, $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$, one for each stock. These could be equal, but they do not need to be.

2.5 Species uncertainty

When modelling two stocks, there could be uncertainty in classifying the different species. This is due to the shape and pattern of the otoliths, rather than to the person who interpreted them. This uncertainty can be included in the model by regarding the classification as equivalent to classification into age groups. The only difference is that there are two different types of classification for both stocks, type 1 (which is "certain" and easy to classify) and type 2 (which is "uncertain" and harder to classify). If we make the assumption that a type 1 fish is never confused with a type 2 fish, then the new error matrix takes is given by

$$\begin{array}{c} \\ \\ \text{Type1 CC} \\ \text{Type2 CC} \\ \text{Type1 AC} \\ \text{Type2 AC} \end{array} \begin{array}{cccc} \multicolumn{2}{c}{\text{Coastal cod}} & \multicolumn{2}{c}{\text{Atlantic cod}} \\ \text{Type1 CC} & \text{Type2 CC} & \text{Type1 AC} & \text{Type2 AC} \end{array}$$

$$\begin{pmatrix} pclass_1^C E_1^{CC} & 0 & (1-pclass_1^A)E_1^{AC} & 0 \\ 0 & pclass_2^C E_2^{CC} & 0 & (1-pclass_2^A)E_2^{AC} \\ (1-pclass_1^C)E_1^{CA} & 0 & pclass_1^A E_1^{AA} & 0 \\ 0 & (1-pclass_2^C)E_2^{CA} & 0 & pclass_2^A E_2^{AA} \end{pmatrix}$$

where $pclass_1^C$ is the probability that a type 1 coastal cod will be correctly classified, $pclass_2^C$ is the probability that a type 2 coastal cod will be correctly classified, $pclass_1^A$ is the probability that a type 1 Atlantic cod will be correctly classified, and $pclass_2^A$ is the probability that a type 2 Atlantic cod will be correctly classified. E_1^{CC} is the age error matrix for coastal cod that are classified as type 1 coastal cod, E_2^{CC} is the age error matrix for coastal cod that are classified as type 2 coastal cod, E_1^{CA} is for coastal cod that are misclassified as type 1 Atlantic cod, and so on. Hence, the columns give the conditional probability of the observed type, given the true species. We can allow the age error matrices to be different for the certain and uncertain types.

2.6 Estimation of catch-at-age

The model described above is for proportion at age in individual units, whereas we want to estimate the total numbers caught at age in each cell. The total catch in a cell, w_c, is given in weight. In order to calculate T_c, the total catch in each cell in number of fish, we need to calculate the mean weight of fish caught in the cell, $E_c[w]$.Then, the number at age in a cell is given as

$$T_c(a) = \frac{w_c}{E_c[w]} E_c[p(a)].$$

We can also estimate $P_c(l \mid a)$, the probability of a fish of a given age a being in a length interval l. Then the number at age in a cell for a given age and length interval is given by

$$T_c(a,l) = \frac{w_c}{E_c[w]} E_c[p(a)] P_c(l \mid a).$$

3 Inference and MCMC

Inference on the unknown parameters are obtained by using a Bayesian framework. We use an MCMC algorithm to obtain samples from the posterior distribution. In this Chapter we discuss the changes in the simulation algorithm from the old version to the current version. Changes have been made when estimating the age model and the length-given-age model, while estimating the weight-given-length model is unchanged. There is a small change in the program when predicting catch-at-age.

3.1 Age and length-given-age model

Use of observed and simulated data

The main difference between the old version and the current version is the way that the data is used in estimating the age and length-given-age model. There are three types of data; age-length data, age-given-length data and length-only data. Previously all the data was used to estimate the models. For the data where only the length was observed, the missing age was simulated, and then used similar to the observed age data.

In the version 2.0 the missing ages are still simulated. All the data is used when estimating the parameters in the age model. But when estimating the parameters in the length-given-age model, only units where there are some observed ages are included. Hence, the length-only data is not used here. The exception is when estimating the unit effect $\varepsilon_{c,u}^{unit}$, where all the data is used.

Running the program

A new feature in the current version is that all parameters from the last iteration are saved in a file, and it is possible to continue a simulation later on.

In the current version there is a split option, which first estimates the model using only the observed age-length data. Then the parameter set from the last iteration is used as starting values for a new estimation, which includes all the data. The idea is to get better starting values using only age-length data, where the simulation is fast, in order to obtain faster convergence when using the full data set.

Parameter delta.age

If there are a large number of age groups with hardly any fish, so that for some levels of the covariates there may be no fish of that age at all, it can be difficult to estimate the covariates. One way to improve the estimation is to add a small amount delta.age to the probability of each age group, in each haul. An appropriate value is 0.005. This amount is then subtracted when the hauls are simulated to estimate the catch-at-age. An example is for the cod 2004 data, using age groups 1-20. The ages above 10 are very rare in this data. The effect on the season covariate can be seen in Figure 3.1.

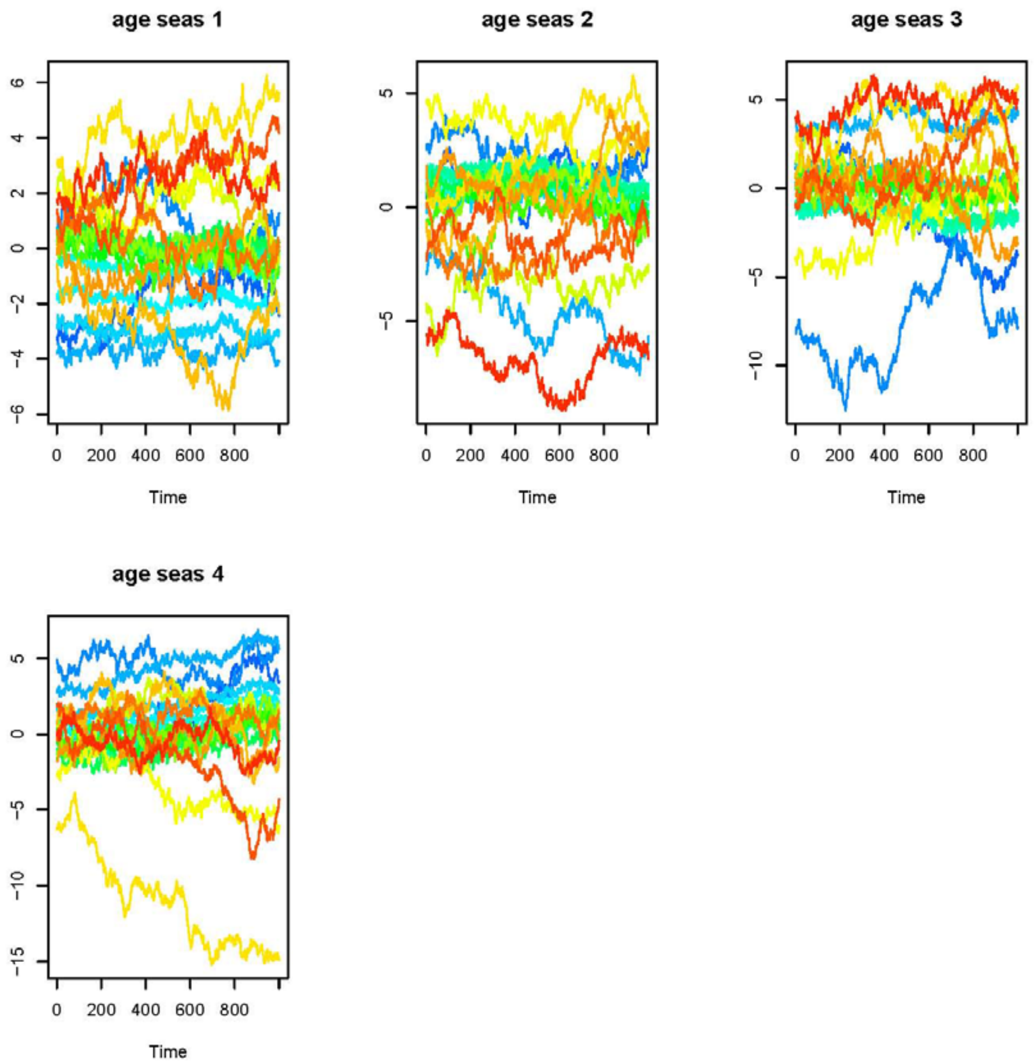


Figure 3.1: Estimation of season effect, cod 2004, 20 age groups, all covariates, delta.age=0. The estimates are very unstable.

When a delta.age value of 0.005 is used, the resulting estimates are shown in Figure 3.2. They are much more stable than in the previous plot.

The effect on the estimation of catch-at-age combined over all cells is fairly small, though the variance does decrease somewhat. The effect on the estimates for a single cell could be much bigger however. The combined cell estimates are shown in Table 3.1 and Table 3.2.

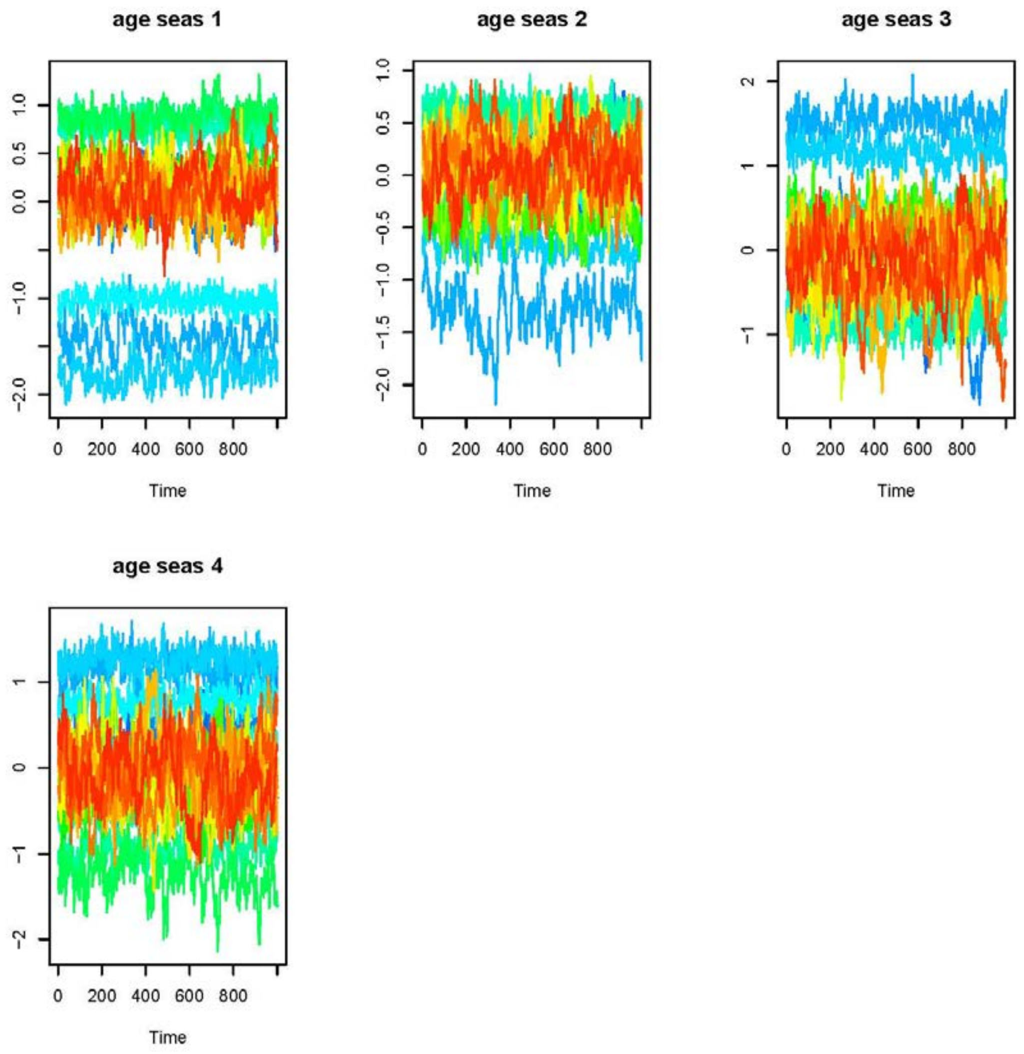


Figure 3.2: Estimation of season effect, cod 2004, 20 age groups, all covariates, delta.age=0.005. The estimates are much more stable than for delta.age=0.

Table 3.1: Estimates of catch at age for delta.age = 0.

	Mean	Sd	2.5%	97.5%
1	10.80	14.72	1.16	49.02
2	13.61	10.83	1.31	40.71
3	526.40	179.96	291.03	929.97
4	3488.73	1185.20	2114.95	5870.99
5	6930.11	1647.13	4751.63	10842.35
6	14927.43	2440.97	11635.60	19248.43
7	16544.27	1681.40	13698.39	20105.56
8	8577.27	840.24	7110.35	10349.07
9	3460.06	361.53	2780.02	4207.97
10	985.56	127.11	771.16	1263.79
11	159.60	37.24	105.63	250.59
12	100.29	47.12	42.05	231.42
13	55.90	75.94	14.44	307.97
14+	31.37	8.38	19.82	51.17

Table 3.2: Estimates of catch at age for delta.age = 0.005.

	Mean	Sd	2.5%	97.5%
1	71.19	43.87	19.97	177.69
2	79.49	53.19	16.35	202.47
3	602.81	184.24	368.56	1073.51
4	3378.90	1078.41	2165.11	5566.59
5	6740.63	1508.22	4866.44	9911.95
6	14300.33	2199.63	11192.79	18715.33
7	15901.83	1577.68	13171.20	19124.70
8	8195.68	767.78	6832.41	9826.07
9	3383.34	330.50	2766.44	4060.56
10	963.83	130.40	759.91	1246.83
11	185.23	60.63	103.71	343.96
12	128.11	48.27	58.70	242.75
13	79.26	37.61	32.52	179.34
14+	427.73	101.17	271.19	656.27

3.2 Catch-at-age

When predicting the catch-at-age in each cell we need to estimate the mean weight and the probability at age. We estimate $E_c[p(a)]$ from Monte Carlo simulation over a given number of units. In version 1.0 the same number of units were used for all cells. In version 2.0 the number of units is input to the program and is weighted for each cell with the catch in the cell. This gives a large number of Monte Carlo simulations for cells with large catch and few simulations for cells with almost no catch. The default value is set to 1000.

In prediction of catch-at-age it is possible to specify a length interval, which is used for producing a gadget file. However, computing the catch-at-age for many length intervals is more time consuming. So if this is not needed, it is now possible to put this equal to 0 (default), which means that there is only one length interval encompassing all the possible lengths.

4 Validation of the program

The program has been tested by simulation. This Chapter presents results from the simulations.

The methodology was as follows:

1. Simulate data from the model.
2. Find true catch-at-age for the parameters.
3. Estimate catch-at-age.
4. Find 80% intervals for each age group.
5. Find 80% bootstrap intervals for each age group.
6. Calculate summary statistics: bias, coverage of intervals (yes/no), width of intervals.
7. Repeat a large number of times.

More details and results are given for some of the simulations.

4.1 Model with one covariate

The model:

- 6 age groups, with approximate probabilities 0.31, 0.02, 0.31, 0.02, 0.31, 0.02.
- Covariate simulated at random in each run.
- Log-linear length-given-age model.
- 10 hauls with age and length data, 80 fish in each haul.
- 40 hauls with length-only data, 80 fish in each haul.
- As near as possible to equal numbers of hauls for each level of the covariates: 4 seasons, 5 gears, 5 areas.
- Equal total catch weight in each cell.

The results are similar for each covariate, and only results for season are shown, in Figure 4.1 to Figure 4.3. Figure 4.1 shows the coverage of the intervals. Both with and without the length only data the program gives coverage close to the correct value of 80%. The bootstrap is much poorer, probably due to the small sample size. Figure 4.2 shows that the interval width gets much smaller when the length-only data is added to the age data. Figure 4.3 shows that all methods are nearly unbiased.

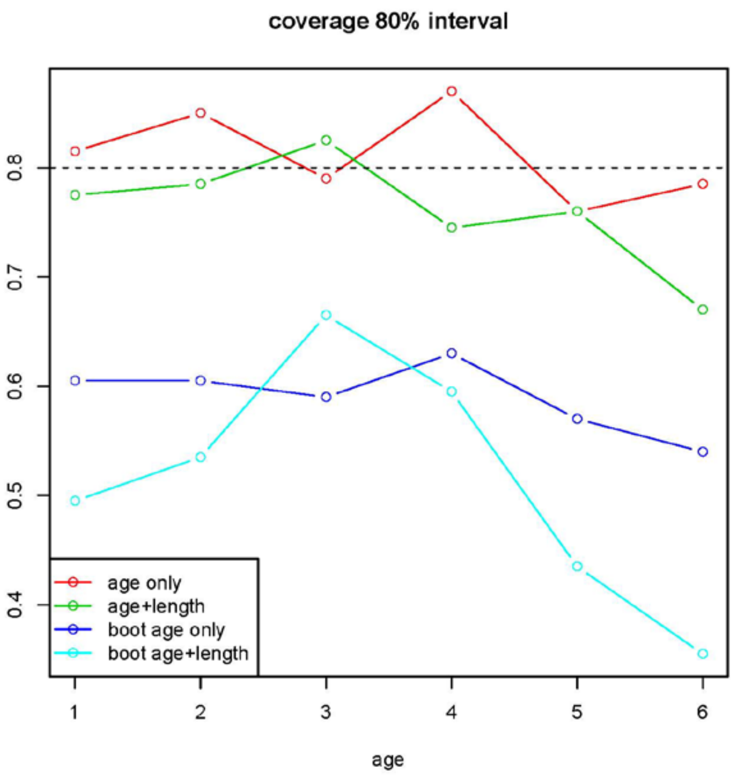


Figure 4.1: Coverage of nominal 80% intervals, only 1 covariate (season).

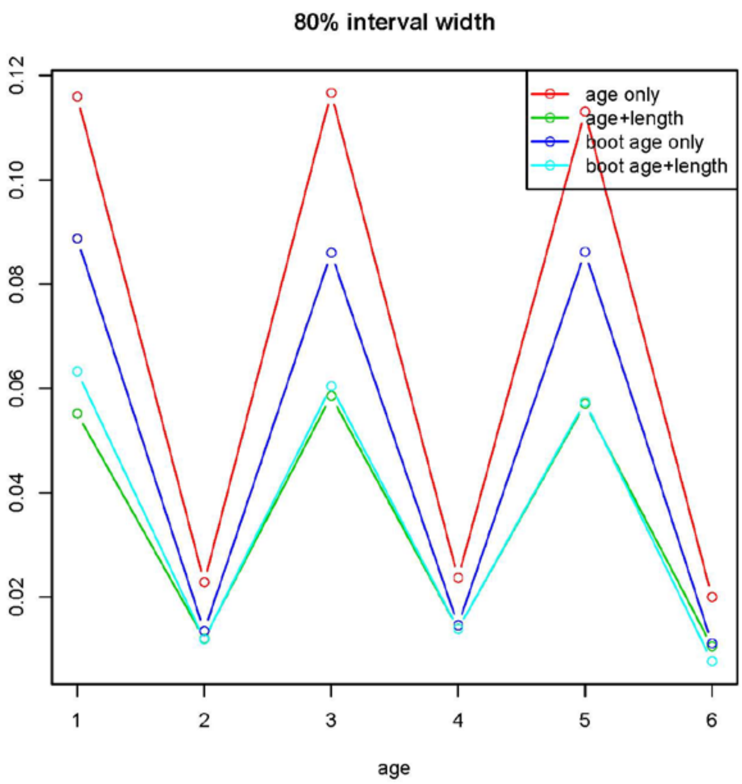


Figure 4.2: Width of 80% intervals, only 1 covariate (season).

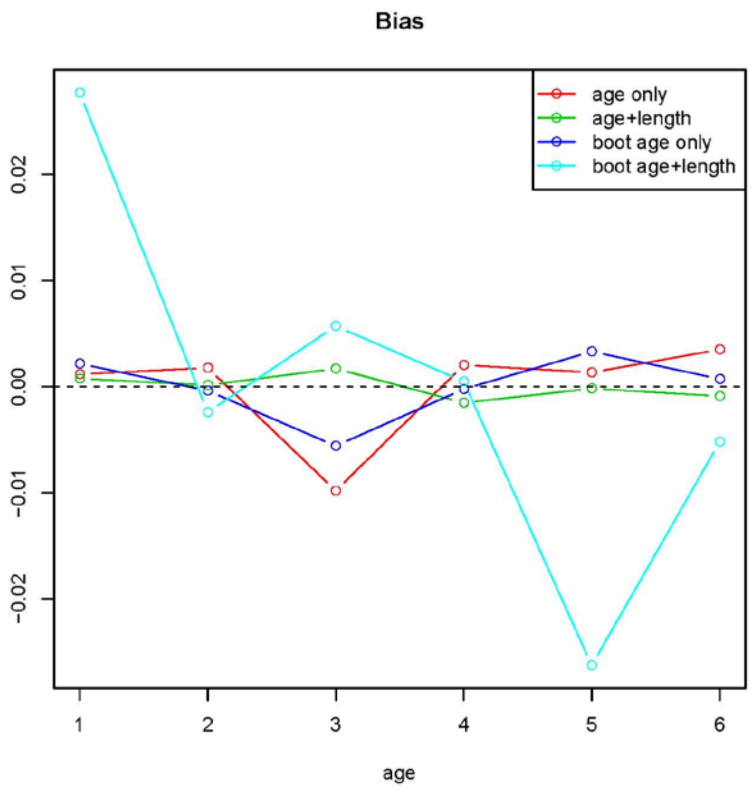


Figure 4.3: Bias of different methods, only 1 covariate (season).

4.2 Model with all covariates (season, gear, region and cell)

The model is the same as in Section 4.1, but with 100 and 400 age and length-only hauls. Many cells have no data so the bootstrap is done within cells with data, then scaled up to the correct total weight. All the covariates are included simultaneously.

Figure 4.4 and Figure 4.5 show the coverage of the intervals and interval widths, respectively. The results are similar as in the previous section, although the bootstrap does particularly badly with the length only data, probably because the age-length-key is very badly estimated within cells.

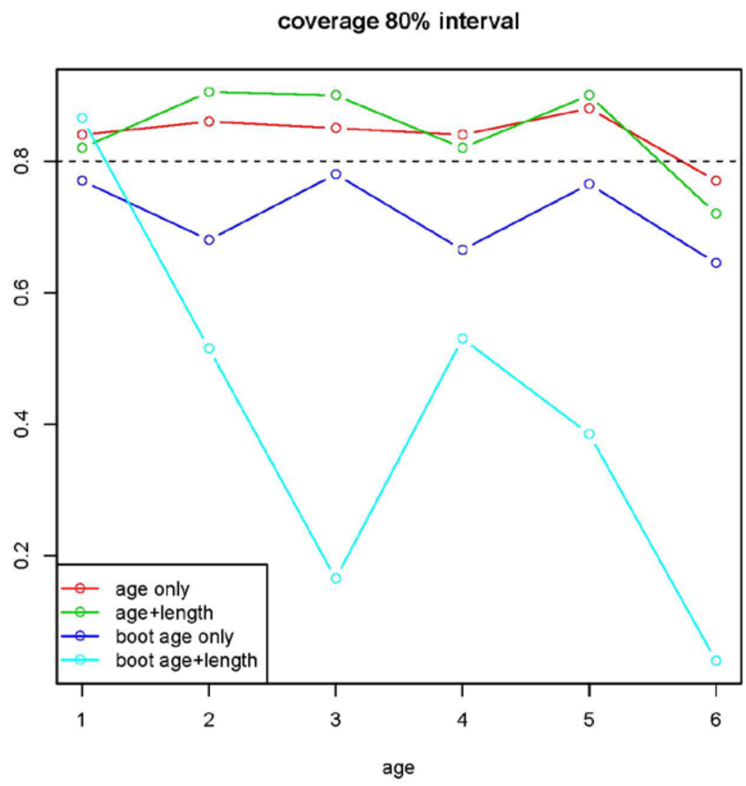


Figure 4.4: Coverage of intervals, all covariates.

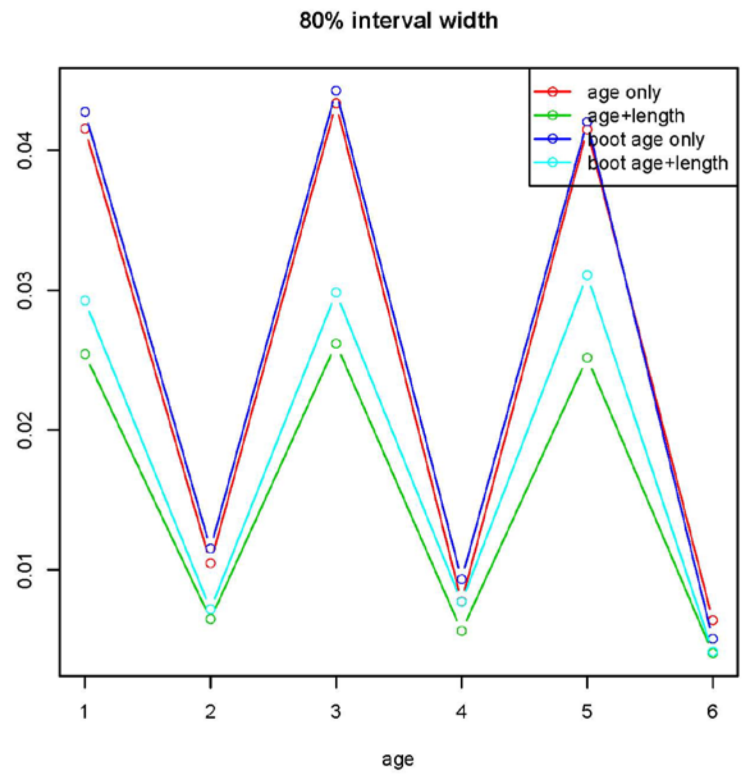


Figure 4.5: Width of intervals, all covariates.

4.3 Non-linear model

The model is the same as in Section 4.1, but the non-linear age-length model is used instead of the log-function. Figure 4.6 shows the coverage of intervals.

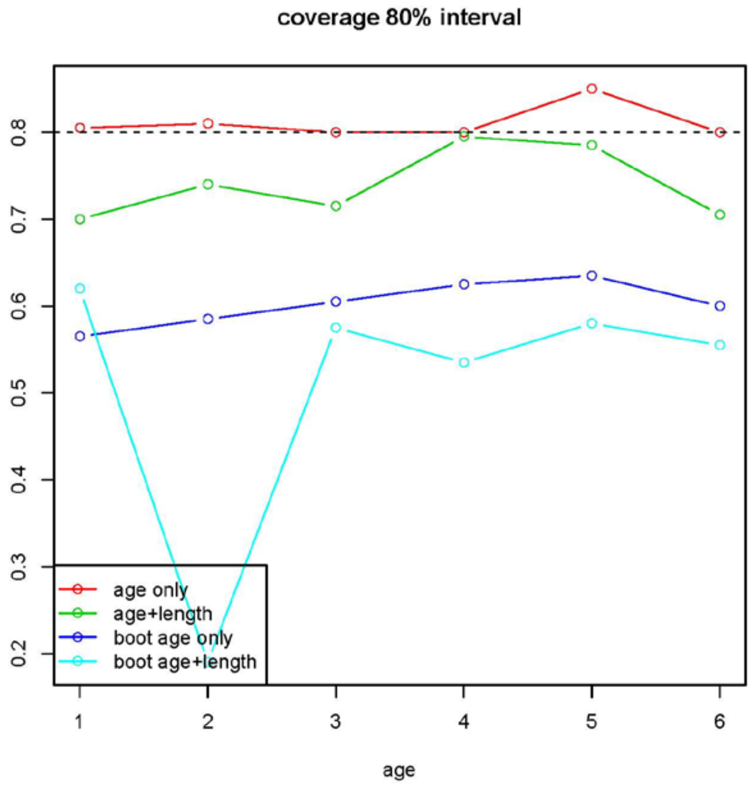


Figure 4.6: Coverage of intervals, non-linear model, only 1 covariate (season).

4.4 Age uncertainty

The model is the same as in Section 4.1, but now we include an age error matrix. The resulting coverage of intervals is shown in Figure 4.7.

4.5 Species uncertainty

The model is the same as in Section 4.1, but now we use the two stock analysis. The uncertainty in classification is given by $pclass_1^C = 1$, $pclass_2^C = 0.7$, $pclass_1^A = 1$ and $pclass_2^A = 0.7$. Figure 4.8 shows the coverage of intervals.

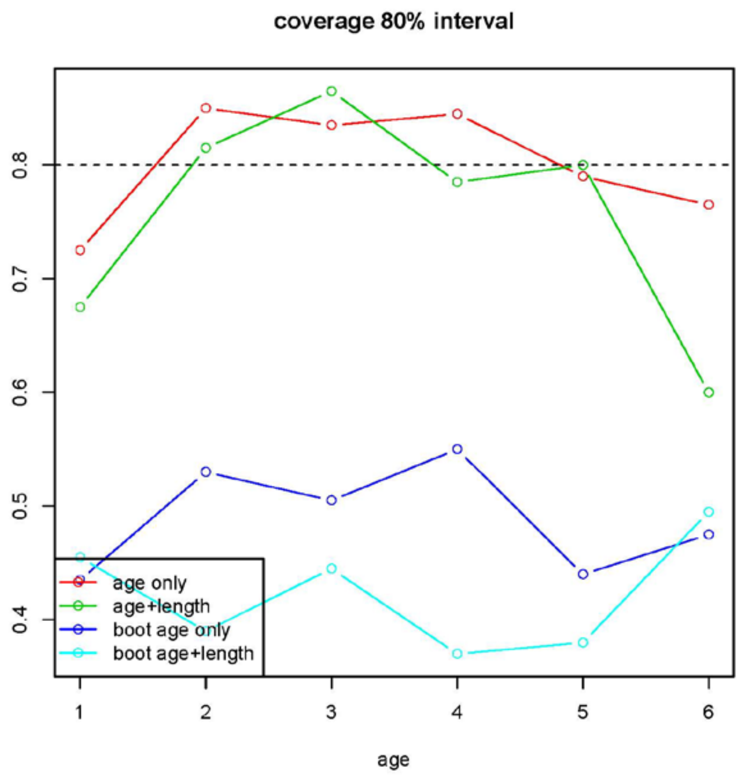


Figure 4.7: Coverage of intervals, age uncertainty, only 1 covariate (season).

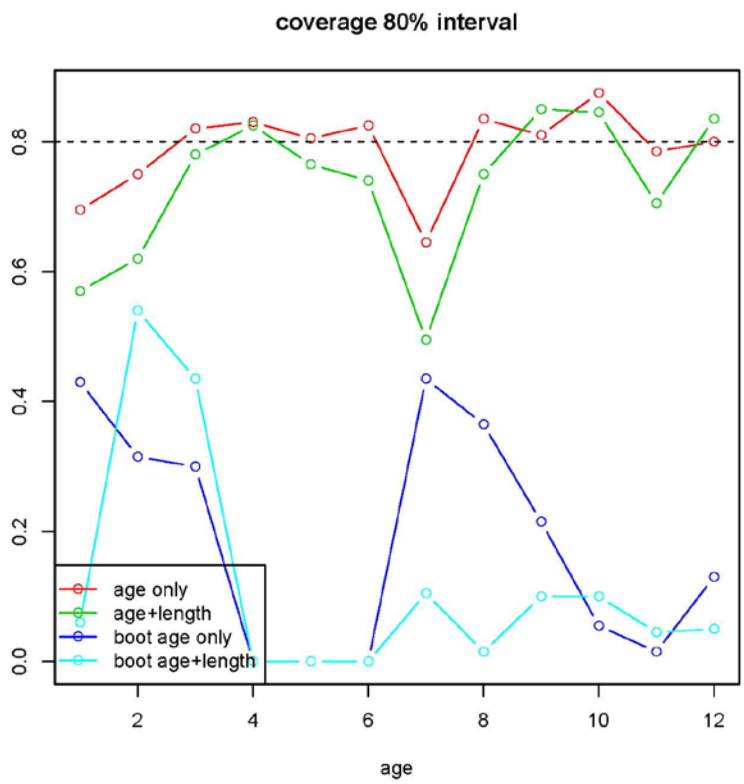


Figure 4.8: Coverage of intervals, species uncertainty, only 1 covariate (season).

5 Comparison of versions

This Chapter presents results from predicting the total catch-at-age using both version 1.0 and version 2.0.

Data that are used for comparison are cod from the years 2004, 2005 and 2006. We have used age-length data and length-only data from amigo and reference fleet. We have used 1000 samples with 1000 burn-in samples.

5.1 Cod 2006

The cod 2006 data consists of 800 hauls with together 43611 fish. Of these 12925 have age measurements, while 30686 have missing ages. The minimum observed age is 2 and the maximum observed age is 15. The model is set to have ages from 1 to 16.

The model is first fitted using the covariates season, gear and region. Results from predicting the total catch-at-age using version 1.0 and 2.0 are shown in Figure 5.1. There are differences between the two versions. For some age groups the mean value using version 2.0 is outside the 95% prediction interval using version 1.0. Also the prediction intervals are generally larger when using version 2.0.

Then the model is fitted without any covariates. The results from predicting the total catch-at-age are shown in Figure 5.2.

In addition we have included results from fitting the model with only data where both age and length are measured. This is shown in Figure 5.3. Here the results are almost identical. This should be the case since the model has not changed when having only observed ages.

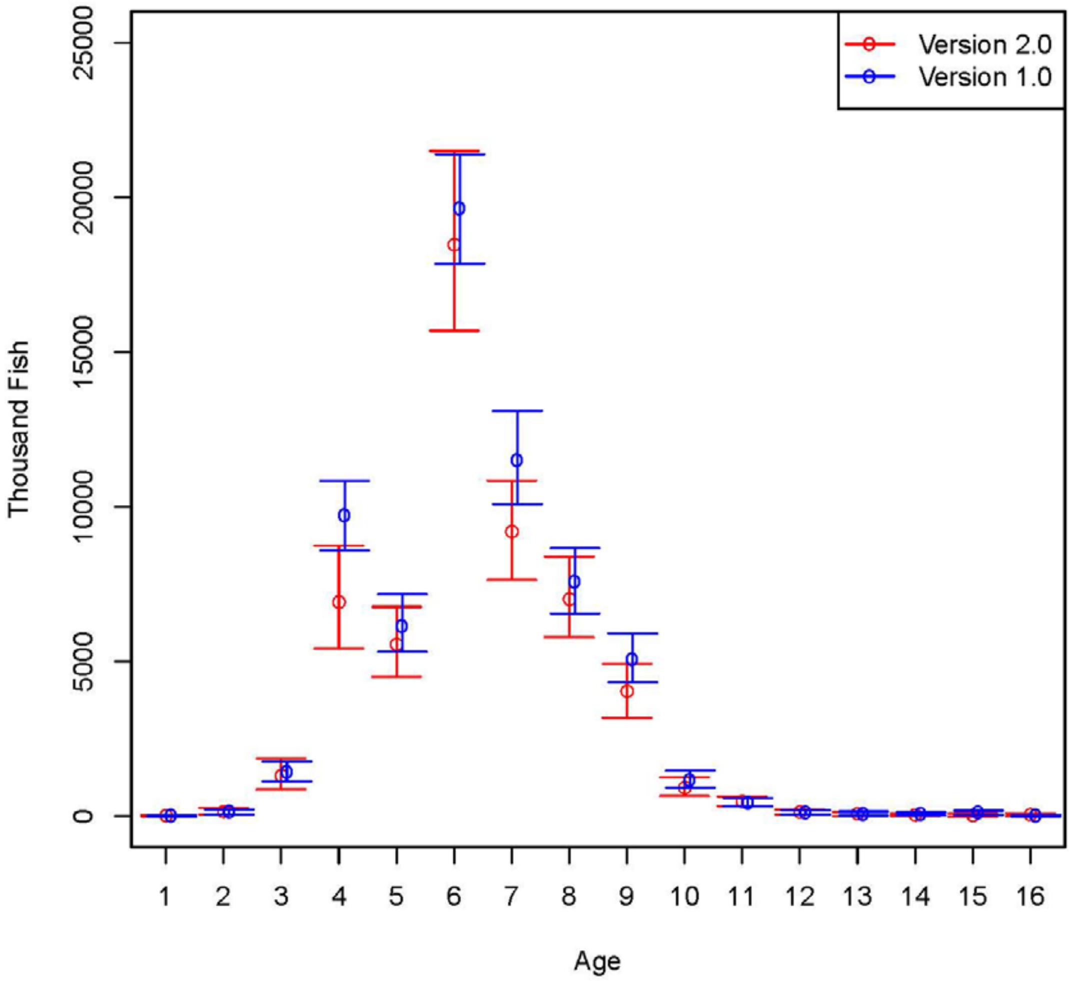


Figure 5.1: Total catch-at-age for cod 2006, with season, gear and region in fitted model. The circles are the mean values of the MCMC samples, while the intervals are 95% prediction intervals.

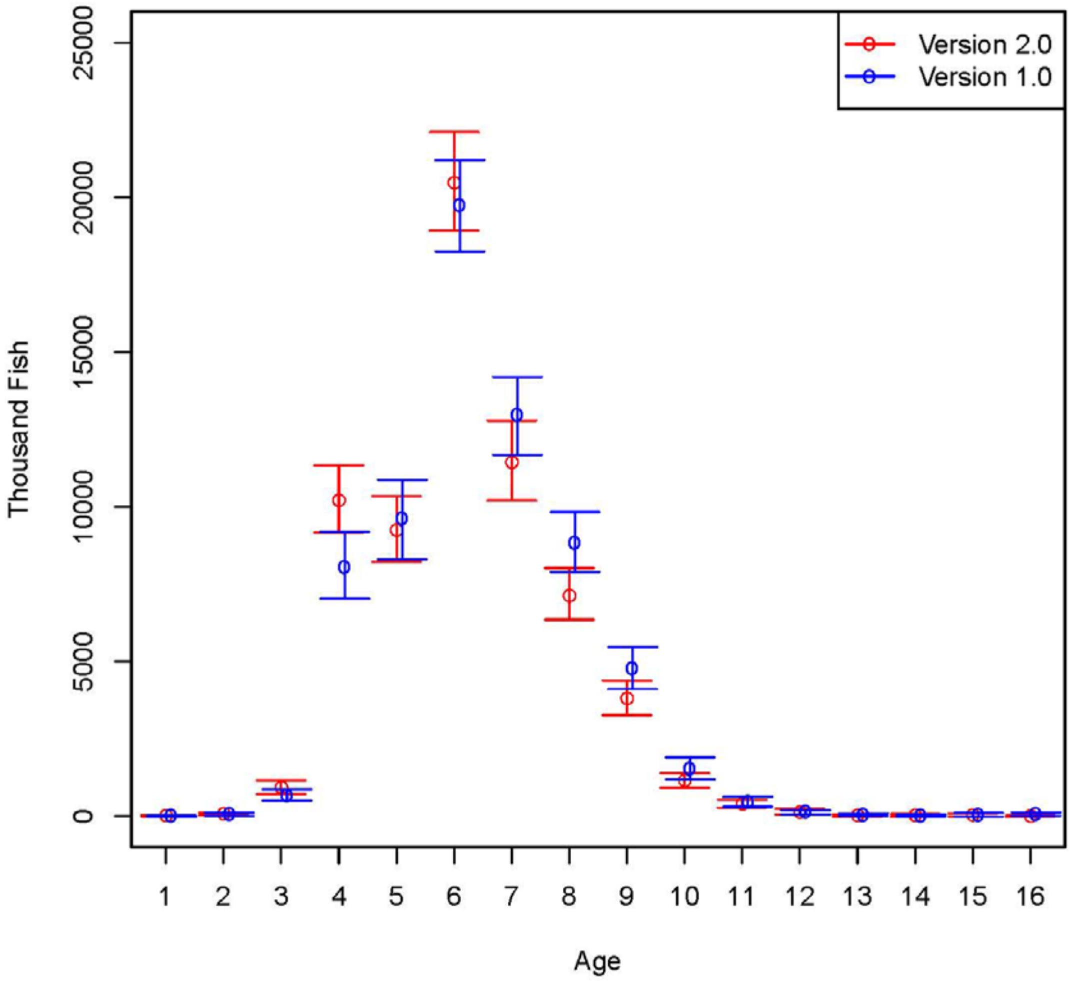


Figure 5.2: Total catch-at-age for cod 2006, with no covariates in fitted model. The circles are the mean values of the MCMC samples, while the intervals are 95% prediction intervals.

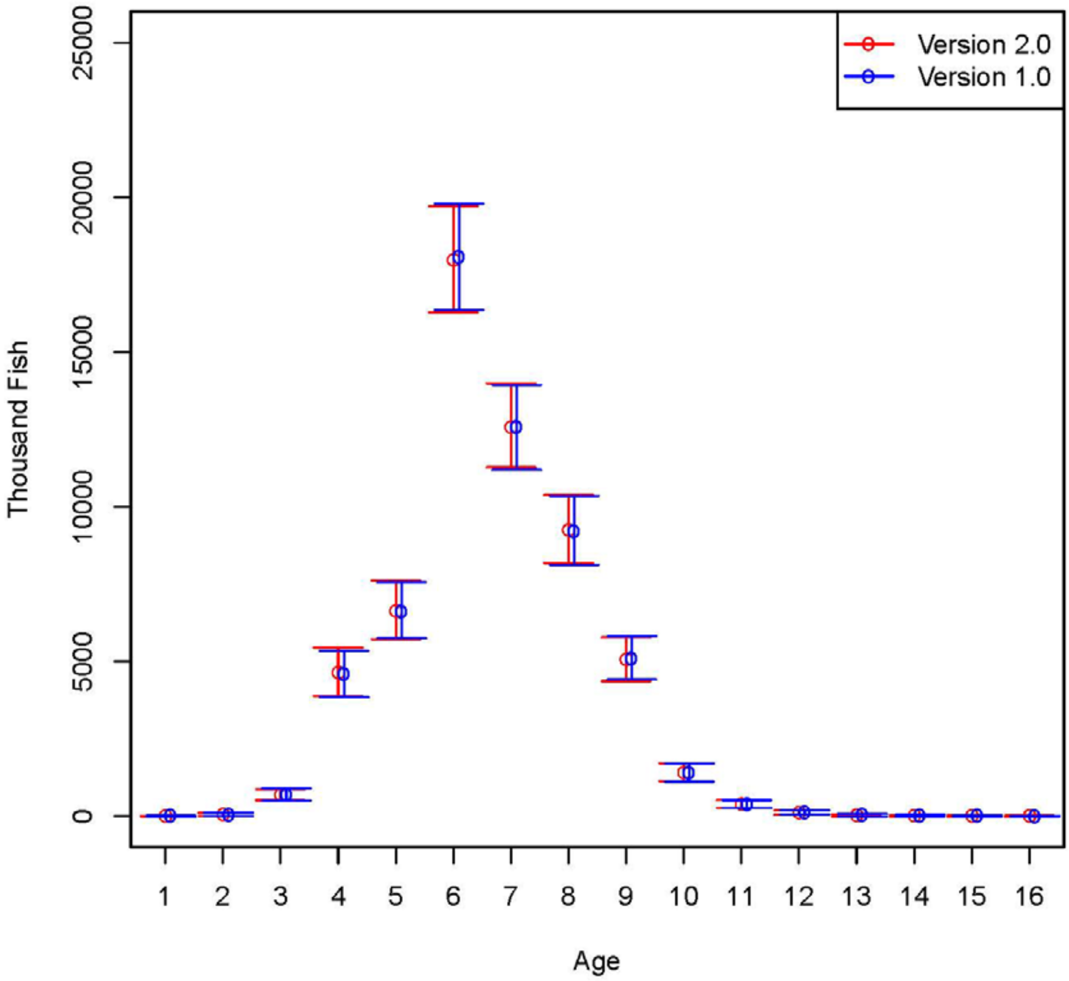


Figure 5.3: Total catch-at-age for cod 2006 with only Amigo data (all ages observed), with no covariates in fitted model. The circles are the mean values of the MCMC samples, while the intervals are 95% prediction intervals.

5.2 Cod 2005

The cod 2005 data consists of 721 hauls with together 41094 fish. Of these are 14630 age measured, while 26464 have missing ages. The minimum observed age is 2 and the maximum observed age is 15. The model is set to have ages from 1 to 16.

The model is first fitted using the covariates season, gear and region. Results from predicting the total catch-at-age using version 1.0 and 2.0 are shown in Figure 5.4. There are some differences between the two versions, but they are within the Monte Carlo variation. Here the prediction intervals are also larger when using version 2.0.

Then the model is fitted without any covariates. Results from predicting the total catch-at-age are shown in Figure 5.5. The differences between the two versions are larger than when fitting the model with all the covariates.

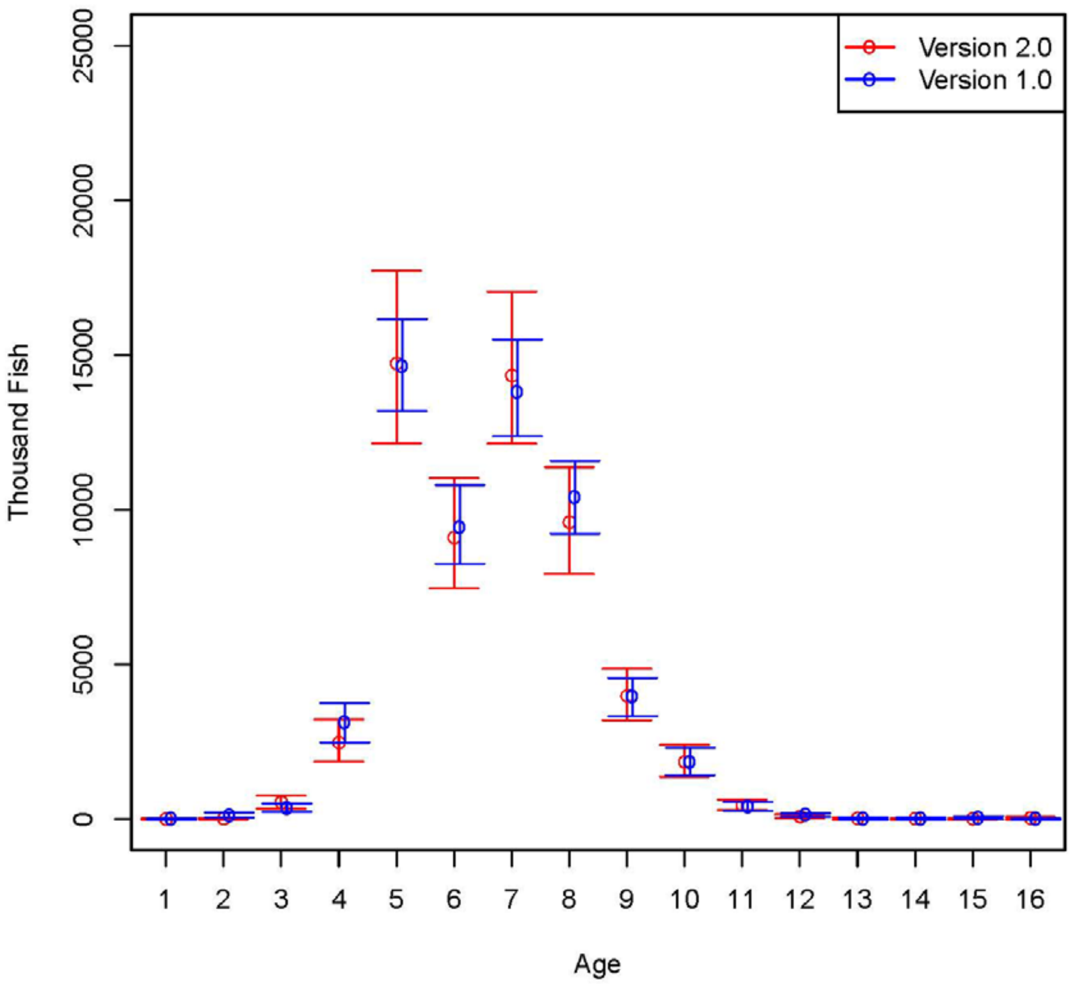


Figure 5.4: Total catch-at-age for cod 2005, with season, gear and region in fitted model. The circles are the mean values of the MCMC samples, while the intervals are 95% prediction intervals.

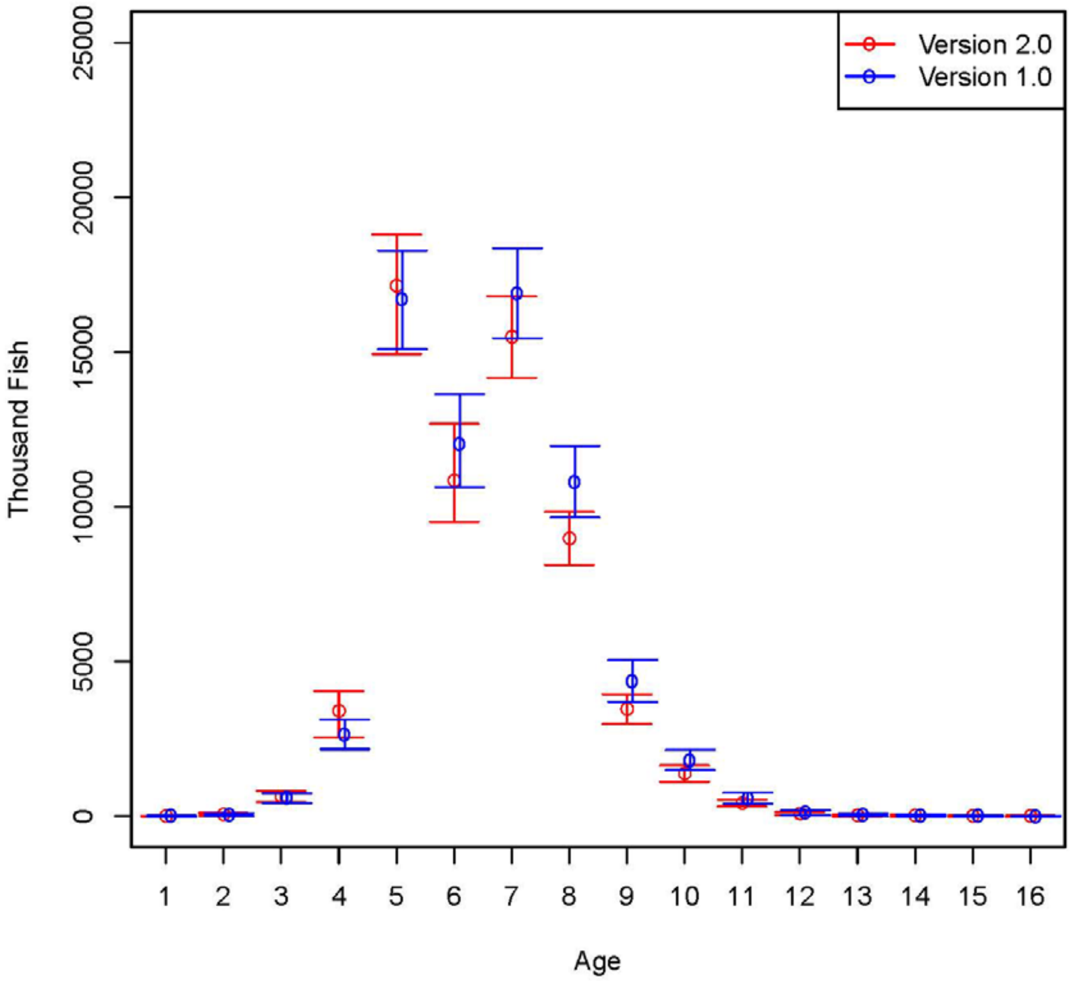


Figure 5.5: Total catch-at-age for cod 2005, with no covariates in fitted model. The circles are the mean values of the MCMC samples, while the intervals are 95% prediction intervals.

5.3 Cod 2004

The cod 2004 data consists of 928 hauls with together 52174 fish. Of these are 15205 age measured, while 36969 have missing ages. The minimum observed age is 2 and the maximum observed age is 15. The model is set to have ages from 1 to 16.

The model is first fitted using the covariates season, gear and region. Results from predicting the total catch-at-age using version 1.0 and version 2.0 are shown in Figure 5.6. Then the model is fitted without any covariates. Results from predicting the total catch-at-age are shown in Figure 5.7.

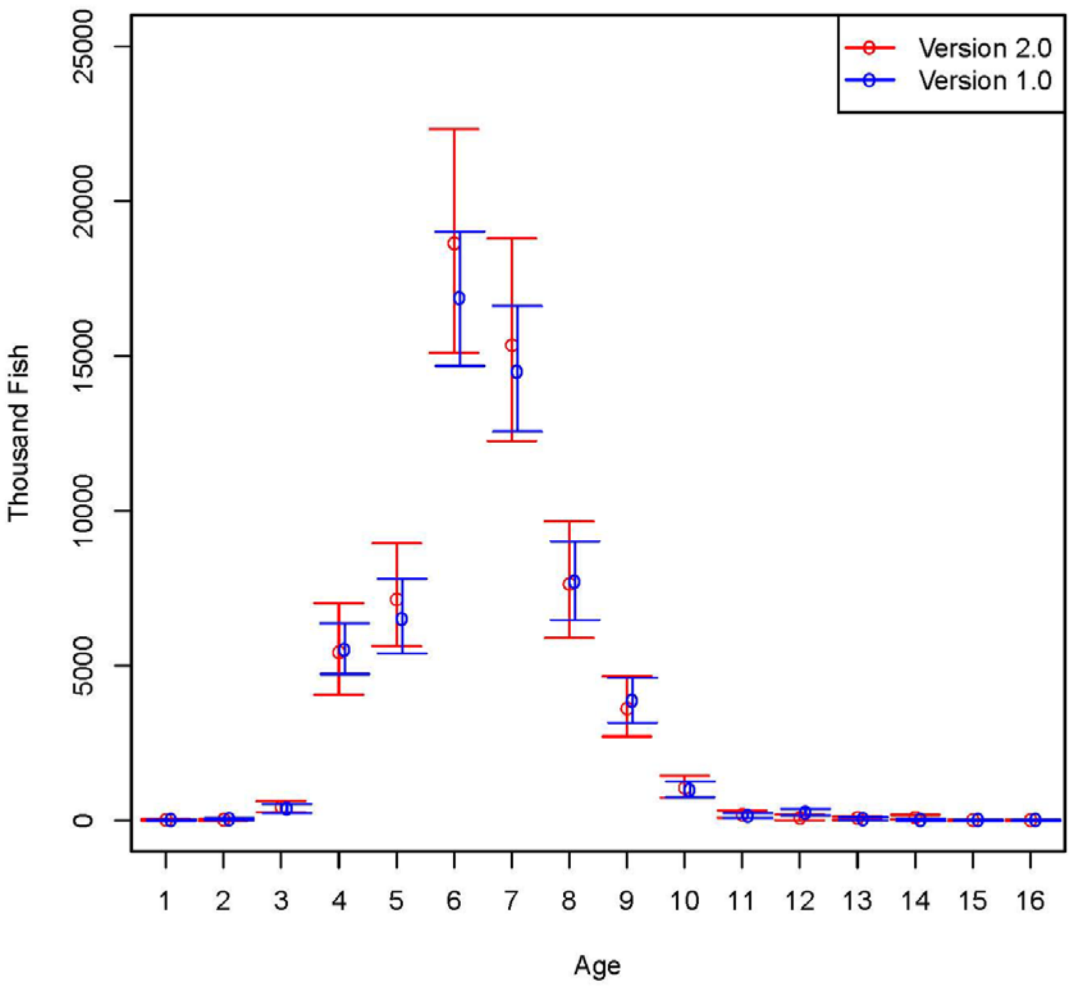


Figure 5.6: Total catch-at-age for cod 2004 with season, gear and region covariates in fitted model. The circles are the mean values of the MCMC samples, while the intervals are 95% prediction intervals.

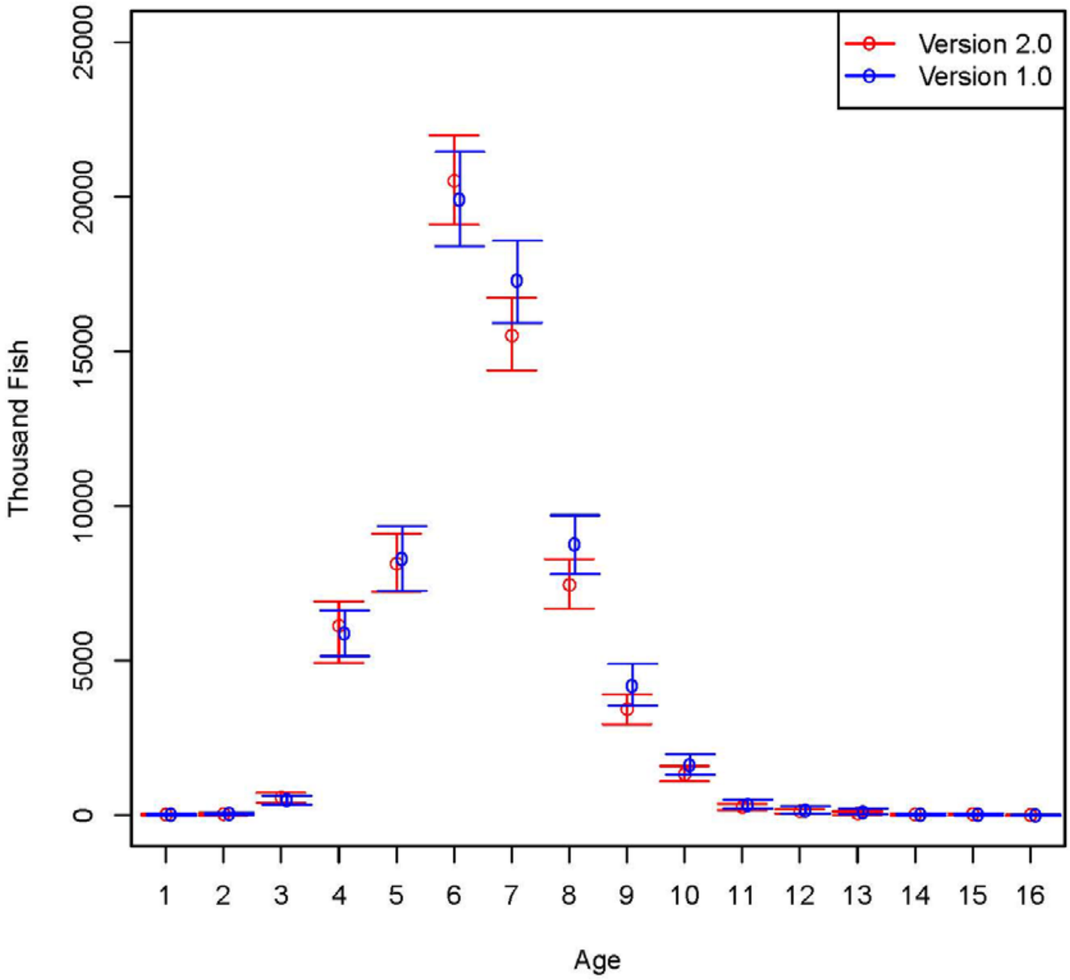


Figure 5.7: Total catch-at-age for cod 2004 with no covariates in fitted model. The circles are the mean values of the MCMC samples, while the intervals are 95% prediction intervals.

6 Conclusions

There are some changes to Catch-at-age version 2.0 from version 1.0. The main difference between the two versions is how the data is used when estimating the age and the length-given-age model. The change has made the simulation algorithm more robust to different data sets.

The program has been extensively tested by simulations. It gives satisfactory results for all the combinations of covariates and features of the model (like age error etc.) that have been tested.

If there is no length-only data, then the results are almost identical for the two versions. If there are lots of length-only data, there are some differences. For one of the data sets used in the comparison, the difference is larger than the 95% prediction interval for several age groups. The differences are larger when fitting the model without any covariates.

7 References

Hirst, D., Aanes, S., Storvik, G., Huseby, R. B., and Tvete, I. F. (2004). Estimating catch-at-age from market samplilng data using a Bayesian hierarchical model. Applied Statistics, 52(1):1-14.

Hirst, D., Storvik, G., Aldrin, M., Aanes, S., and Huseby, R. B. (2005). Estimating catch-at-age by combining data from different sources. Canadian J. of Fisheries and Aquatic Sciences, 62(6):1377-1385.

Rognebakke, H., Hirst, D., and Storvik, G. (2011). Catch-at-age – Version 2.0: Technical Report, NR Note SAMBA/44/11.