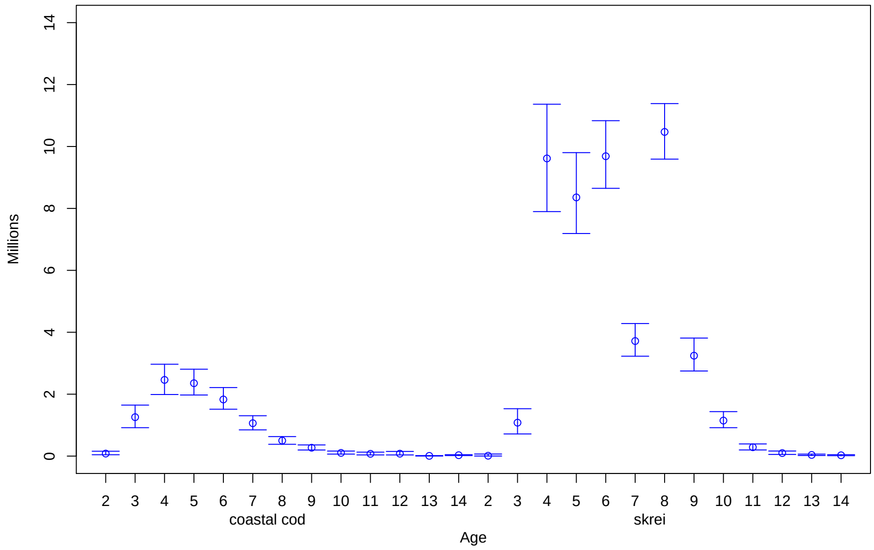
Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Note

Catch-at-age for multiple stocks:

Modelling Skrei and Coastal Cod simultaneously.



Note no SAMBA/46/11

Authors
Hanne Rognebakke
David Hirst
Geir Storvik
Magne Aldrin

Date December 2011

Norwegian Computing Center

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
NO-0314 Oslo, Norway

Besøksadresse
Office address
Gaustadalléen 23
NO-0373 Oslo, Norway

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

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

Title	**Catch-at-age for multiple stocks: Modelling Skrei and Coastal Cod simultaneously.**

Authors

Hanne Rognebakke <hanne.rognebakke@nr.no>
David Hirst <david.hirst@nr.no>
Geir Storvik <geirs@math.uio.no>
Magne Aldrin <magne.aldrin@nr.no>

Date	December 2011
Publication number	SAMBA/46/11

Abstract

The Norwegian Computing Center and the Institute of Marine Research have over years developed a Bayesian hierarchical model to estimate the catch-at-age of fish for a single stock. This has been extended to model multiple stocks, like Atlantic Cod and Coastal Cod. This note presents the multiple stock model, which includes both age reading error and classification error.

Keywords	Bayesian hierarchical models, MCMC, multiple stocks
Target group	Institute of Marine Research
Availability	Open
Project	
Project number	220305
Research field	Statistics for Climate, Environment, Marine Resources and Health
Number of pages	17

Contents

1 Introduction

The Norwegian Computing Center (NR) and the Institute of Marine Research (IMR) have over years developed a Bayesian hierarchical model to estimate the catch-at-age of fish for a single stock, see Hirst et al. (2004, 2005). Norwegian cod stocks are regarded as consisting of two stocks: Atlantic cod (Skrei), found in deep waters, and coastal cod, found nearer the shore. It is rather difficult to distinguish between fish from the two stocks visually, and the classification is usually based on the pattern of growth rings in the otoliths. Even this is not a certain classification, and the fish are divided into certain or uncertain Atlantic cod or coastal cod. The majority of fish do not have their otoliths read, and only have length measurements, with no indication of which stock they come from. The total landings statistics also do not distinguish between the two stocks.

The problem therefore is to estimate the catch-at-age of the two stocks, based on a large number of length measurements, and a smaller number of fish with both length and age measurements. These latter fish also have a classification into certain or uncertain membership of one of the two stocks.

This paper presents the extension of the single stock model that allows uncertainly classified multiple stocks to be modeled.

2 The model

The data are sampled in different years, regions and seasons, and with different gears. We define a cell c as a combination of these. Modelling can be made on haul level or on trip level, and the index u is used to describe the modelling unit of choice.

2.1 The model for stock and proportion at age

The model for one stock considers the numbers at age in unit u, cell c to be multinomial with parameter $\boldsymbol{p}_{c,u}$ of length A where A is the number of age groups considered plausible for the species. Thus $p_{c,u}(a)$, $a = 1, \ldots, A$, is the proportion of fish in age group a in the unit.

The model for two stocks extends this to a vector of length $2A$, where the first A give the proportion in age groups $1, \ldots, A$ for the first stock, and the second A give the proportion at age in the second stock. This simple extension allows the single stock model to be used almost unchanged. A more detailed description of the model and the simulation algorithm is given in Rognebakke et al. (2011).

The proportions are reparameterised as

$$p_{c,u}(a) = \frac{\exp(\alpha^a_{c,u})}{\sum_{a'} \exp(\alpha^{a'}_{c,u})}.$$

The parameters $\alpha^a_{c,u}$ are modelled in terms of various covariates as

$$\alpha^a_{c,u} = \alpha^{const,a} + \alpha^{year,a}_y + \alpha^{season,a}_s + \alpha^{gear,a}_g + x^{hsz}_u \alpha^{hsz,a}_u + \zeta^{region,a}_r + \zeta^{cell,a}_c + \zeta^{unit,a}_{c,u}.$$

The α terms and $\{\zeta^{region,a}_r\}$ are the main effects for year, season, gear and region. The α terms are fixed effects. It is assumed that there will always be some data for all levels of the fixed effects that are of interest. For identifiability, we assume

$$\sum_a \alpha^{const,a} = 0,$$
$$\sum_a \alpha^{year,a}_y = 0, \ \forall y,$$
$$\sum_y \alpha^{year,a}_y = 0, \ \forall a,$$

and likewise for the season, gear and haulsize effects.

Further, $\{\zeta^{region,a}_r\}$ are spatially smoothed random effects. It is necessary to estimate the proportions in areas with no data, and our approach is to introduce some spatial smoothing. This is accomplished by assuming $\{\zeta^{region,a}_r\}$ follow a Gaussian conditional autoregressive distribution (CAR) model (Besag, 1974).

The terms $\{\zeta^{cell,a}_c\}$ and $\{\zeta^{unit,a}_{c,u}\}$ are independent random effects and model the differences between the fit from the main-effects-only model and the true cell means, and the differences between units within a cell, respectively. We assume these effects are iid Gaussian distributed with zero mean and common precisions, τ^{cell}_{age} and τ^{unit}_{age}, and again with a sum-constraint over ages.

The only parameters common to the two stocks are the precisions of the unit, cell and area effects.

2.2 The models for length-given-age and weight-given-length

It is necessary to have separate length-age and weight-length models, since the two stocks (may) have different growth patterns. Hence, the models for the two stocks have no common parameters in the length-given-age and weight-given-length models.

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

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

where $\epsilon_{u,f}^{fish} \overset{iid}{\sim} \mathcal{N}\left(0, \tau_{l,fish}^{-1}\right)$ is the random within-unit variation in length-given-age.

The intercept is given by

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

Here the β terms and ϵ_r^{region} are main effects. The ϵ terms are random effects. They are all modelled similarly to the corresponding terms in the age model.

The slope β_1 is common to all cells and units.

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

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

where we have fixed the parameters $\theta = 0.5$ and $c = 1$, while γ is estimated. The function is standardised to lie between 0 and 1. Note that age here is a continuous variable, ie age in years and months or years and seasons. The age in the age model is the year class.

The weight-given-length model is similar to the log-linear length-given-age model.

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

where $\nu_{u,f}^{fish} \overset{iid}{\sim} \mathcal{N}\left(0, \tau_{w,fish}^{-1}\right)$ 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.

2.3 Age reading errors

Age reading errors were included in the single stock model in the following way: Suppose there are A age groups. Then we can create an $A \times A$ matrix $\boldsymbol{E}$, where the columns give the conditional probability of the observed fish age, given the true age, i.e.

$$E_{i,j} = \Pr(a^{obs} = i | a = j).$$

Assuming the age errors are independent, we simply replace the vector of true age probabilities on a boat $\boldsymbol{p}_b$ by the probability of observed ages $\boldsymbol{E}\boldsymbol{p}_b$. We assume $\boldsymbol{E}$ is known.

This method can be used directly in the two-stock analysis, except that there are now two age-error matrices, say $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$, one for each stock. These could be equal, but they do not need to be.

2.4 Uncertainty in classification

It is difficult to distinguish between fish from the two stocks visually, and classification, which is usually based on reading the otoliths, divides the fish into certain or uncertain coastal cod or Atlantic cod. Uncertainty in classification is mostly due to the shape and pattern of the otoliths, rather than to the person who interpreted them. The majority of fish only have length measurements, giving no indication of which stock they come from. The total landings statistics also do not distinguish between the two stocks.

The uncertainty in classification is included in the model by regarding the classification to species 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 the form

$$\begin{array}{c} \\ \\ \text{Type1C} \\ \text{Type2C} \\ \text{Type1A} \\ \text{Type2A} \end{array} \begin{array}{c} \begin{array}{cc} \text{Coastal cod (C)} & \text{Atlantic cod (A)} \end{array} \\ \begin{array}{cccc} \text{Type1C} & \text{Type2C} & \text{Type1A} & \text{Type2A} \end{array} \\ \begin{pmatrix} pclass_1^C E_1^{CC} & 0 & (1-pclass_1^A)E_1^{AC} & 0 \\ 0 & pclass_2^C E_2^{CC} & 0 & (1-pclass_2^A)E_2^{AC} \\ (1-pclass_1^C)E_1^{CA} & 0 & pclass_1^A E_1^{AA} & 0 \\ 0 & (1-pclass_2^C)E_2^{CA} & 0 & pclass_2^A E_2^{AA} \end{pmatrix} \end{array}.$$

The probabilities $pclass_1^C$ and $pclass_2^C$ are the probabilities that a type 1 and type 2 coastal cod, respectively, will be correctly classified. Similarily, $pclass_1^A$ and $pclass_2^A$ are the probabilities that a type 1 and type 2 Atlantic cod, respectively, will be correctly classified. E_1^{CC} and E_2^{CC} are the age error matrices for coastal cod that are classified as type 1 and type 2 coastal cod, respectively, while E_1^{CA} and E_2^{CA} are the age error matrices for coastal cod that are misclassified as type 1 and type 2 Atlantic cod, respectively, 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.

3 Prediction

The total catch is known through fisheries reports, and is given in weight for each cell. The number of fish is equal to the total weight divided by the average weight, while the number of fish in an age group is equal to the number of fish multiplied by the proportion of fish in that age group. Usually the number of fish is so large that the average weight can be replaced by expected weight. The prediction procedure is as follows:

- Fit the model to get full posterior distributions for all parameters.
- Simulate a large number of units in each cell using one sample from the joint posterior to get a proportion at age and mean weight for the cell.
- Divide the total catch by the mean weight to get a total number of fish for each cell, and multiply by the proportion at age in the cell to get numbers at age.
- Repeat for a large number of samples from the joint posterior to get a posterior distribution of numbers at age in each cell.

4 Example

We will use cod data from 2008 to show the difference between modelling skrei using the single stock model and the two stock model. The covariates season, gear, region and cell are included in the model. We have used 1000 samples with 1000 burnin. Section 4.1 shows results when we assume there is no classification error, while Section 4.2 shows results for two different levels of classification error.

4.1 Results when no classification error

First, we show the results when using data where both age and length are observed. This includes 426 hauls with a total of 13818 fish. Figure 1 shows estimates of catch-at-age for skrei for the 2008 data set where both age and length are observed. These results are compared with the single stock analysis.

Then, we show results when including data with only length measurements. This includes 1879 hauls with a total of 97625 fish. The resulting estimates of catch-at-age is shown in Figure 2.

Both results generally show a reduction of estimates of the catch-at-age for skrei, especially for the smaller age groups. Hence, a lot of the younger fish that in the single stock analysis contribute to the catch-at-age estimate of skrei, will in the two stock analysis be catch-at-age estimates of coastal cod instead. When using the two stock analysis, we also get an estimate of the catch-at-age for coastal cod. The results are shown in Figure 3 and 4, when using only age-length data and when including length-only data, respectively.

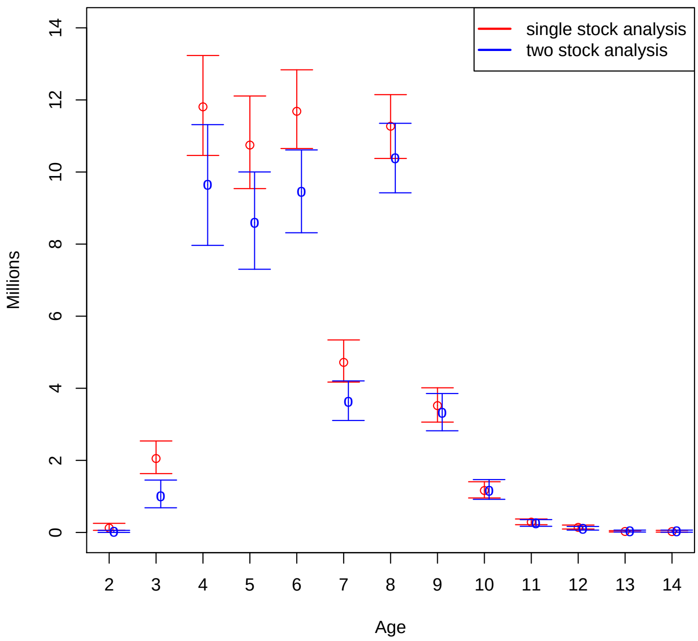


Figure 1. Estimates of catch-at-age of skrei using the single stock anaylsis and the two stock analysis, when using age-length data. The circles are the mean values of the MCMC samples, while the intervals are 95% prediction intervals.

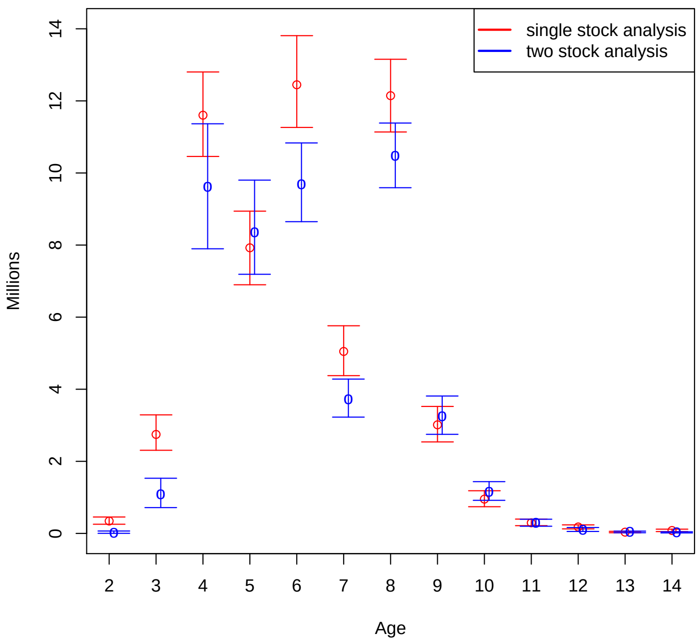


Figure 2. Estimates of catch-at-age of skrei using the single stock anaylsis and the two stock analysis, when using both age-length data and length-only data. The circles are the mean values of the MCMC samples, while the intervals are 95% prediction intervals.

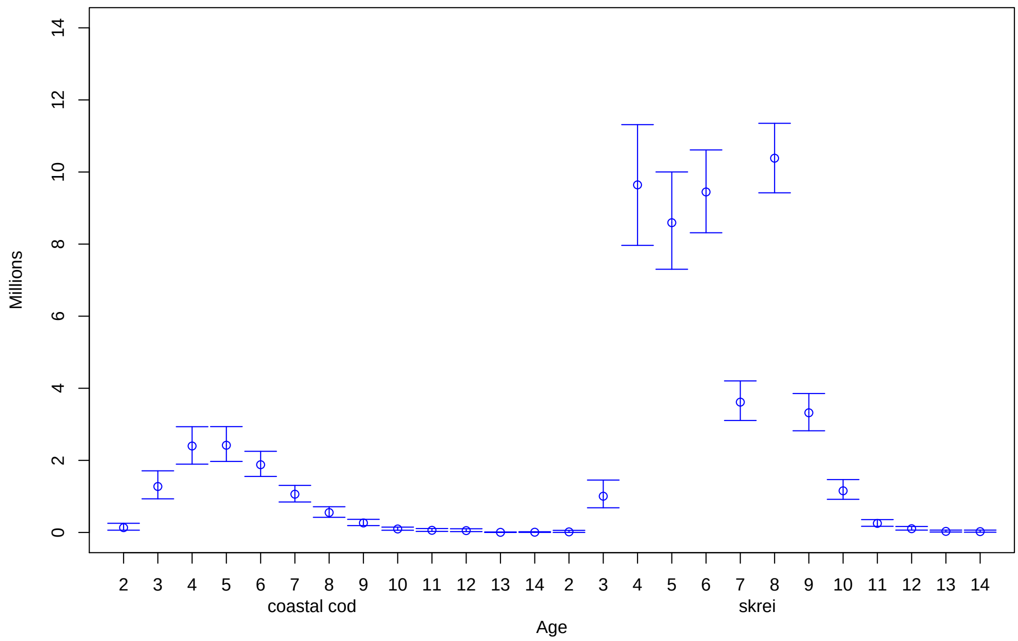


Figure 3. Estimates of catch-at-age of coastal cod and skrei using the two stock anaylsis, when using age-length data. The circles are the mean values of the MCMC samples, while the intervals are 95% prediction intervals.

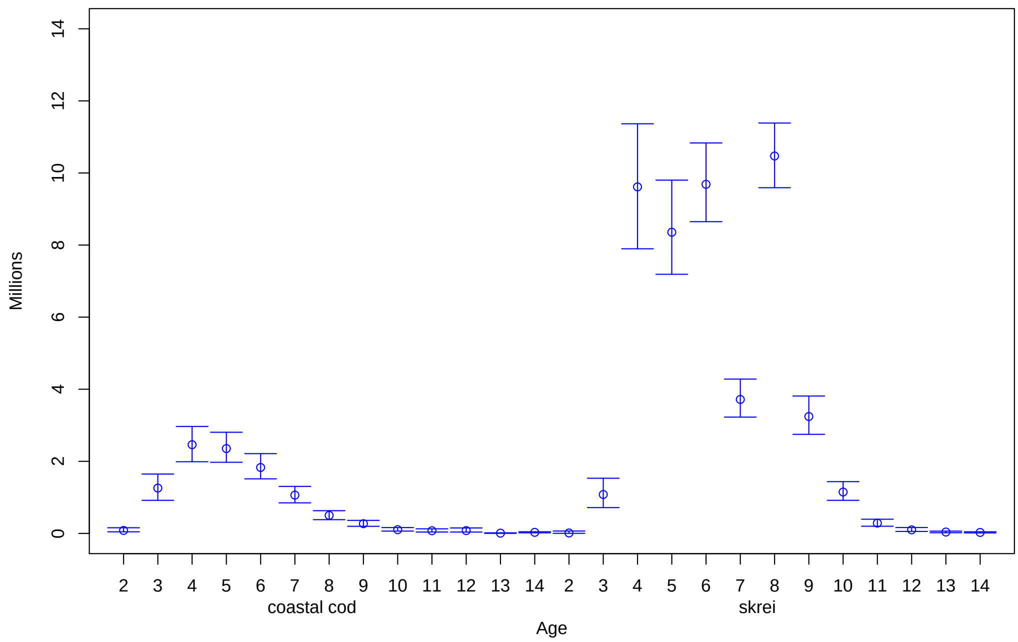


Figure 4. Estimates of catch-at-age of coastal cod and skrei using the two stock anaylsis, when using age-length and length-only data. The circles are the mean values of the MCMC samples, while the intervals are 95% prediction intervals.

4.2 Results when including classification error

The effect of including the uncertainty in the classification in stock is explored. Two levels of uncertainty are used.

- **Level 1:**
  70% probability of type 2 coastal cod being classified correctly,
  100% probability of type 1 coastal cod being classified correctly.
  Same probabilities for skrei.

- **Level 2:**
  70% probability of type 2 coastal cod being classified correctly,
  98% probability of type 1 coastal cod being classified correctly.
  Same probabilities for skrei.

A table of the classification number by age and type is shown below.

		type			
		Type1C	Type2C	Type2A	Type1A
age	2	54	1	1	1
	3	364	19	13	139
	4	525	84	138	1148
	5	599	124	228	1169
	6	453	174	337	2044
	7	345	84	164	927
	8	146	67	224	3031
	9	86	19	43	857
	10	33	9	20	265
	11	17	4	11	67
	12	15	3	5	26
	13	0	0	0	8
	14	2	1	0	4

Figures 5 and 6 show estimates of catch-at-age for both coastal cod and skrei, when using age-length data only and when using both age-length and length-only data, respectively. Both levels of uncertainty are included, and also the case when having no uncertainty in the classification.

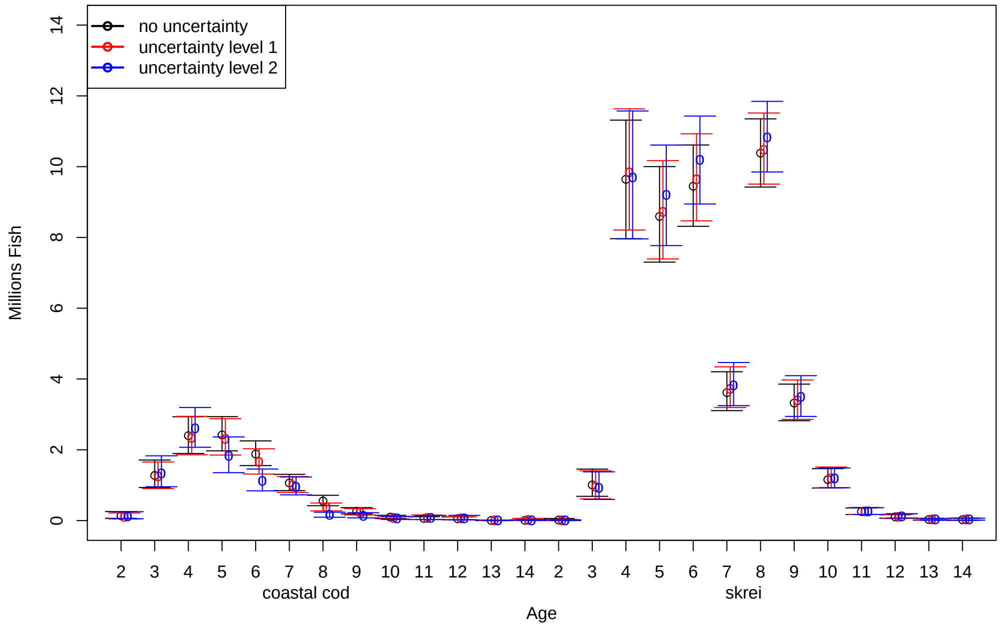


Figure 5. Estimates of catch-at-age of coastal cod and skrei including classification error in the two stock analysis, when using only age-length data. The circles are the mean values of the MCMC samples, while the intervals are 95% prediction intervals.

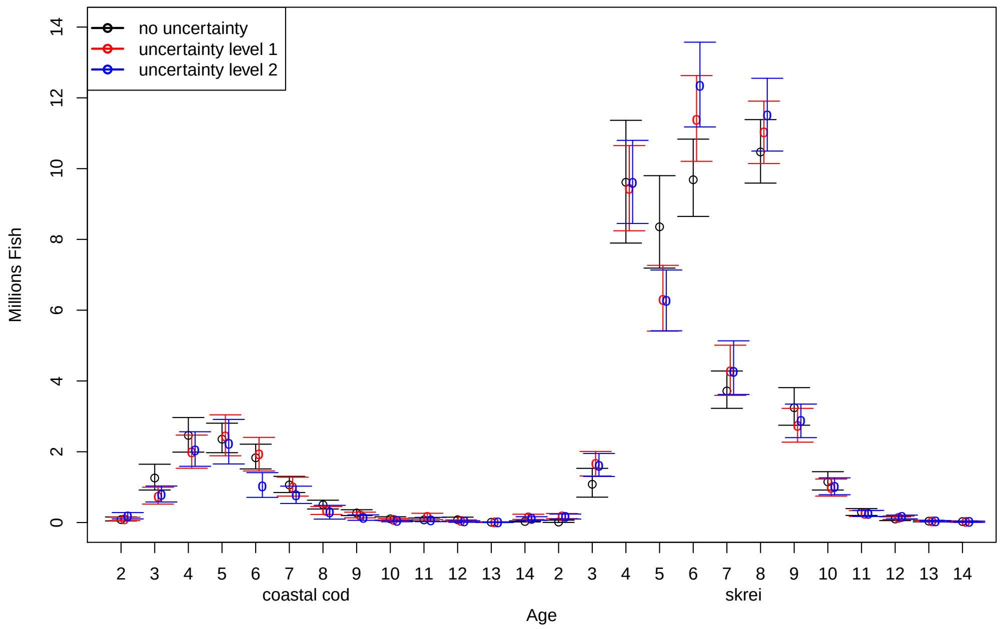


Figure 6. Estimates of catch-at-age of coastal cod and skrei including classification error in the two stock analysis, when using both age-length data and length-only data. The circles are the mean values of the MCMC samples, while the intervals are 95% prediction intervals.

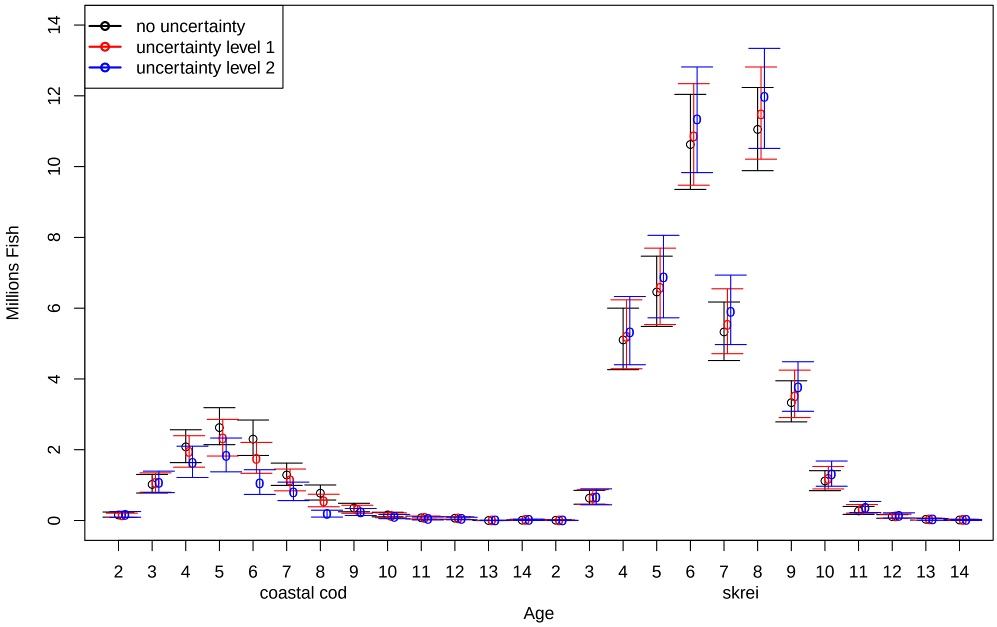


Figure 7. Estimates of catch-at-age of coastal cod and skrei including classification error in the two stock analysis, when using only age-length data. No covariates are included in the model. The circles are the mean values of the MCMC samples, while the intervals are 95% prediction intervals.

It can be seen from Figures 5 and 6 that including uncertainty in the classification can change the estimates of catch-at-age considerably. The details of the changes are difficult to anticipate because the number of uncertain classifications varies with stock and age. In general the effect of increasing uncertainty is to increase the skrei catch estimates at the expense of the coastal cod. Also when including length-only data, the effect of having classification error increases. Reducing the probability of type 1 classifications being correct to 98% (uncertainty level 2), can also give a large effect on the estimates.

If the coastal cod are largely restricted to a few cells, the difference between uncertainty levels would be smaller when including all the covariates in the model than when having no covariates in the model. This is illustrated in Figure 7, which shows estimates of catch-at-age for coastal cod and skrei when using age-length data only and no covariates.

5 Conclusions

The catch-at-age model has been extended to model multiple stocks, and it also allows uncertainty in the classification of species.

The conclusions are as follows:

- The effect of including two stocks in one analysis has a considerable effect on the estimation of skrei.
- A two stock analysis provides a previously unavailable estimate of coastal cod catch-at-age.
- In general, the effect of increasing uncertainty is to increase skrei catch estimates at the expense of the coastal cod.
- Including a measure of uncertainty in the classification has a potentially dramatic effect on the estimates. If the type 1 ("certain") classifications are allowed to have some error, even if the error is as low as 2%, it can also give a large effect on the estimates.

References

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society. Series B (Methodological)*, 32(2):192–236.

Hirst, D., Aanes, S., Storvik, G., Huseby, R. B., and Tvete, I. F. (2004). Estimating catch-at-age from market sampling data using a Bayesian hierarchical model. *Applied Statistics*, 53(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*.