ICES CM 2004/ FF:20

A stochastic age-length-structured multispecies model applied to North Sea stocks

PETER LEWY

and

MORTEN VINTHER

Danish Institute for Fisheries Research
Department of Marine Fisheries
Charlottenlund Castle
DK-2920 Charlottenlund
Denmark

e-mail: pl@dfu.min.dk, mv@dfu.min.dk

ABSTRACT

A Stochastic Multi-Species model (SMS) including a parameterised size dependent food selection model are formulated and fitted to observations of catches, survey CPUE's and stomach content. The model has been applied to a subset of ICES North Sea Multi Species VPA (MSVPA) data set and the parameters and the variance/covariance matrix have been estimated. For all species except for the prey species sandeel and Norway pout SMS and MSVPA resulted in similar estimates of average fishing mortality and biomass using the same input data. For the latter two species the trends estimated were similar. The predation mortality, M2, was quite different for the two models with a higher inter annual variation in M2 for the MSVPA. The precision of SSB and average F are determined almost entirely by the quality of catch and survey observations and not by the stomach contents observations. Whether M2 was assumed known and constant over years or estimated by the SMS model had almost no effect on the precision of estimated biomass and fishing mortality.

INTRODUCTION

Multispecies stock dynamics including species interactions have been modelled by several authors. Andersen and Ursin (1977) developed a comprehensive model including both primary production and fishing while Gislason and Helgason (1985) simplified their approach and made the basis for ICES multispecies model, MSVPA (Sparre 1991). Multispecies analyses using MSVPA have been carried out by ICES for commercially important stocks in the North Sea and the Baltic Sea since the eighties. In contrast to these deterministic models a stochastic model, Bormicon (1997), was developed. This model and its successor Gadget (Taylor et al.) are models which may include processes as recruitment, mortality, migration, growth, consumption and maturity. Furthermore, Gadget is length-structured, and age-data can supplementary be used.

Compared to Gadget the present model, SMS (Stochastic Multi Species model), includes fewer processes. Only recruitment, fishing mortality and predation are included while e.g. migration, growth and maturity are left out. SMS uses the same data sources as MSVPA, however the food selection model is now parameterised and the parameters involved estimated. Further, in contrast to the fully age-structured MSVPA SMS is a semi age-length structured mode where the stomach content observations and the food selection model are length based. This allows a more realistic food selection model and the use of the originally sample stomach data for the North Sea (e.g. ICES 1989 and ICES 1997), which are stratified by length classless.

Likewise developing stochastic single species models (Lewy and Nielsen 2003; Fournier et al. 1998; Deriso et al. 1985) it is equally important for multispecies models to evaluate the uncertainties of estimated parameters and thereby enabling statistical comparisons of biological hypotheses. In relation to single species analyses it is relevant to test if predation mortalities significantly change over years or not. If not single species models may be sufficient to describe stock dynamics for assessment purposes.

The SMS model has been applied to the North Sea ecosystem including eight species for which four are predators. Stomach content data by length for 1991 (ICES 1997) and commercial and research catch-at-age have been analysed.

MATERIAL AND METHODS

Model structure

SMS is a seasonal stochastic multispecies model for which fishery mortality is described using catch at age data age while predation mortality and food preference are based on stomach contents data by size. The food preference process has been size based because preference depends on size rather than age. Maximum likelihood technique is used to estimate parameters and to weight the various data sources. The likelihood function consists as a sum of four terms for observations of international catch at age, survey CPUE and stomach contents observation, and a stock-recruitment (penalty) function.

Notation

y denotes the year
q denotes the season of the year
s denotes the species
$pred$ denotes the predator species
$prey$ denotes the prey species
of denotes other food
a,b denotes the age
$survey$ denotes a survey

$C(s,a,y,q)$ denotes observed catch in numbers
$\sigma(s,a)$ denotes the standard deviation of log catches
$CPUE(survey,s,a,y,q)$ denotes Catch Per Unit Effort
$Q(survey,s,a)$ denotes the species catchability
$\sigma(survey,s,a)$ denotes the standard deviation of log CPUE
$N(s,a,y,q)$ denotes the stock numbers in the sea
$Z(s,a,y,q)$ denotes total mortality rate
$F(s,a,y,q)$ denotes fishing mortality rate
Y denotes the number of years available
$A(s)$ denotes the number of age groups
NOS denotes the number of seasons
$RS(s)$ denotes the season where the fish is recruited to the fishery
$M2(prey,a,y,q)$ denotes predation mortality estimated in the model
$M1(s,a,y,q)$ denotes residual natural mortality given as input
$SUIT(l_{prey},l_{pred},q)$ denotes food suitability for given prey/length and predator/length groups
l_{prey} denotes a size (length or weight) interval for a prey species
$l_{predator}$ denotes a size (length or weight) interval for a predator species
$\tilde{l}_{prey}$ and $\tilde{l}_{pred}$ denote the midpoints of the size intervals
$\bar{l}_{prey}(a_{prey}(y,q))$ and $\bar{l}_{pred}(b_{pred}(y,q))$ denote the mean size (length or weight) of the age groups
$\eta(pred)$ denotes the parameter expressing the log "mean" of the preferred predator prey size ratio
$\sigma_{pref}(pred)$ denotes the "standard deviation" in the food preference function
$\rho(prey,pred,q)$ denotes vulnerability parameters
$STOM(,l_{prey},l_{pred},y,q)$ denotes the observed average weight proportion of a prey/length group to the total weight of the stomach contents of a predator/length group by year and season.
$AVAILB(l_{pred},y,q)$ denotes total available biomass of a predator of length l_{pred}.
$w(s,a,y,q)$ denotes the mean weight at age
$Food(b_{pred},y,q)$ denotes the food intake

Population dynamics model

Survival of the stocks

The survival of the stock in the sea is described by the usual exponential decay equation

$$N(s,a,y,q+1) = N(s,a,y,q)\exp(-Z(s,a,y,q)) \qquad (1)$$

or

$$N(s,a+1,y+1,q=1) = N(s,a,y,q=\text{last season})\exp(-Z(s,a,y,q=\text{last season}) \qquad (2)$$

Initial stock size, i.e. the stock in the first year and recruitment over years are used as parameters in the model while the remaining stock sizes are considered as functions of the parameters determined by equations (1) and (2).

In a multispecies model including fish predation total mortality, $Z(s,a,y,q)$ is divided into three components, natural mortality exclusive predation (M1), predation mortality (M2) and fishing mortality (F):

$$Z(s,a,y,q) = M1(s,a,q) + M2(s,a,y,q) + F(s,a,y,q) \qquad (3)$$

If *M1* and *M2* are assumed to be known quantities the model described so far is a stochastic single species assessment model (e.g. Lewy et al. 2002) where the dynamic of each of the species can be treated independently of each other.

Fishing mortality

Total fishing mortality, *F(s,a,y,q)*, is modelled as a flexible partly separable model including age, year and seasonal effects (the species index is left out for convenience):

$$F(a,y,q) = F_1(I_1(a),J(y))\,F_2(y)\;F_3(I_2(a),q) \qquad (4)$$

where I_1 and I_2 are different groupings of the range of age groups, 1, 2, …, *A*, and J is a grouping of the time range, 1, 2, …, *Y* defined by

$$I_c(a) = \begin{Bmatrix} 1 & \text{if} & 1 \le a \le A_1^c \\ 2 & \text{if} & A_1^c < a \le A_2^c \\ \ldots & & \\ n_c & \text{if} & A_{n_c-1}^c < a \le A \end{Bmatrix}$$

$$c = 1, 2\,..$$

$$J(y) = \begin{Bmatrix} 1 & \text{if} & 1 \le y \le Y_1 \\ 2 & \text{if} & Y_1 < y \le Y_2 \\ \ldots & & \\ m & \text{if} & Y_{m-1} < y \le Y \end{Bmatrix}$$

The model, (4), indicates that the selection pattern, F_1, may change over years and that the seasonal effect, F_3, may differ for some age groups, but is constant for all years.

Natural Mortality

Natural mortality is divided into two components, predation mortality caused by the predators included in the model (M2) and a residual natural mortality (M1). The residual mortality is assumed to be known and is given as input.

The predation mortality of a prey entity *i* due to predation from predator entity *j* is calculated as suggested by Andersen and Ursin (1977) and Gislason and Helgason (1985)

$$M2(i,y,q) = \sum_j \frac{\overline{N}(j,y,q)\, Food\,(j,y,q)\, SUIT\,(i,j,q)}{AVAILB\,(j,y,q)} \qquad (5)$$

where *Food* intake is taken as input from the literature.

The prey and predator entities may either denote age groups or size.

The available food biomass (*AVAILB*) for predator entity j is defined as the sum of the biomass of available food from all preys, including the biomass of the so-called “other food” (*of*) prey:

$$AVAILB(j,y,q) = \sum_i (\overline{N}(i,y,q)\, w(i,y,q) SUIT(i,j,q)) + B(of) SUIT(of,j,q)$$

As suggested by Andersen and Ursin (1977) the food suitability or preference model, *SUIT*, assumes that food preference can be split into species dependent and size dependent components:

$$SUIT(\tilde{l}_{prey}, \tilde{l}_{pred}, q) = \begin{cases} \rho(prey, pred, q) \exp(-\dfrac{(\ln \frac{\tilde{l}_{pred}}{\tilde{l}_{prey}} - \eta(pred))^2}{2\sigma^2_{1,pref}(pred)}) & if\ \ln \frac{\tilde{l}_{pred}}{\tilde{l}_{prey}} \le \eta(pred) \\ \rho(prey, pred, q) \exp(-\dfrac{(\ln \frac{\tilde{l}_{pred}}{\tilde{l}_{prey}} - \eta(pred))^2}{2\sigma^2_{2,pref}(pred)}) & if\ \ln \frac{\tilde{l}_{pred}}{\tilde{l}_{prey}} \ge \eta(pred) \end{cases} \qquad (6)$$

where ρ are prey and predator species dependent vulnerability parameters, $\eta(pred)$ denotes the relative preferred size of the predator and $\sigma^2_{i,pref}(pred)$ are the "variances" of preferred size and $\tilde{l}_{prey}$ and $\tilde{l}_{pred}$ are the midpoints of the size intervals considered.

Suitability of "other" food, *of*, should be treated differently because that the size of other food prey is not known. Suitability of other food is modelled by assuming that vulnerability is the same for all seasons and that the size dependency of the predator size is exponentially declining with increasing predator size:

$$SUIT(of,\tilde{l}_{pred}) = \rho(of, pred)\exp(-\alpha(pred,q)(\tilde{l}_{pred} - l_{\min}(pred))) \qquad (7)$$

where

$\alpha(pred)$ are parameters and $l_{\min}(pred)$ are known quantities.

One should be aware that $M2$ also is included in the right hand side in eq. (5) ($\overline{N} = \frac{(1-\exp(-M1-M2-F)}{M1+M2+F}N$). As no analytical solution exists this implies that $M2$ has to found numerically. If the time step is sufficiently small, for instance a quarter, the calculation of $M2$ may be facilitated if $M2$ is approximated by

$$M2(i,y,q) = \sum_j \frac{N(j,y,q)\,Food\,(j,y,q)\,SUIT\,(i,j,q)}{AVAILB\,(j,y,q)} \qquad (8)$$

where *AVAILB* is modified correspondingly.

If $M2$ was expressed in terms of stock numbers in the beginning of a quarter rather than the average number, $\overline{N}$, $M2$ could be calculated directly from the right hand side of eq. (5). Both options have been implemented.

In eq. (5) the size dependent predation mortality is calculated. The age-structured predation mortality, $M2(a_{prey},y,q)$ is needed as well due to that the catch model is age structured. $M2$ at age is calculated as follows:

$$M2(a_{prey},y,q) =$$
$$\sum_{pred,b_{pred}} \frac{\overline{N}(b_{pred},y,q)Food\,(b_{pred},y,q)SUIT\,(a_{prey},b_{pred},q)}{\sum_{prey,a_{prey}} \overline{N}(a_{prey},y,q)w((a_{prey},y,q)SUIT\,(a_{prey},b_{pred},q) + B(of,y,q)SUIT\,(of,b_{pred},q)} \qquad (9)$$

Calculation of predation mortality by age requires that suitability by age should be calculated. This can be done by using the same expressions as in equations (6) and (7), but replacing the midpoints of the size groups by mean size by age:

$$SUIT(a_{prey}(q),b_{pred}(q)) = \begin{cases} \rho(prey,pred,q)\exp(-\dfrac{(\ln\frac{\bar{l}_{pred}(b_{pred}(q))}{\bar{l}_{prey}(a_{prey}(q))}-\eta(pred))^2}{2\sigma^2_{1,pref}(pred)})\ \ if\ \ln\frac{\bar{l}_{pred}(b_{pred}(q))}{\bar{l}_{prey}(a_{prey}(q))} \le \eta(pred) \\ \rho(prey,pred,q)\exp(-\dfrac{(\ln\frac{\bar{l}_{pred}(b_{pred}(q))}{\bar{l}_{prey}(a_{prey}(q))}-\eta(pred))^2}{2\sigma^2_{2,pref}(pred)})\ \ if\ \ln\frac{\bar{l}_{pred}(b_{pred}(q))}{\bar{l}_{prey}(a_{prey}(q))} \ge \eta(pred) \end{cases} \quad (10)$$

$$SUIT(of, pred, b_{pred}(q)) = \rho(of, pred, q)\exp(\alpha(pred)(\bar{l}(b_{pred}(q)) - l_{\min(pred)}) \quad (11)$$

where $\bar{l}_{pred}(b_{pred}(q))$ and $\bar{l}_{prey}(a_{prey}(q))$ denote the mean size by age group and quarter. As the suitability parameters appear in the likelihood functions referring to both catch-at-age and stomach content observations the estimated parameters are affected of both sources of information.

Statistical models

Three types of observations are considered: Total international catch at age, research surveys CPUE's and stomach contents observations. For each type a stochastic model is formulated and the the likelihood function calculated. As the three type of observations are independent the total log likelihood is a sum of the contributions from three types of observations. A stock-recruitment (penalty) function is added as a fourth contribution.

Catch-at-age

Catch-at-age observations are considered stochastic variables subject to sampling and process variation. The probability model for these observations is modelled along the lines described by Lewy and Nielsen (2003):

Catch at age is assumed to be log normal distributed with log mean equal to log of the standard catch equation, i.e. $\ln(C(s,a,y,q)) \sim \mathrm{Normal}(E(\ln(C(s,a,y,q)), \sigma^2_{catch}(s,a))$. The variances may be assumed age dependent or the same for a group of ages. Thus, the likelihood function, L_C, associated with the catches is

$$L_{Catch} = \prod_{s,a,y,q} \frac{1}{\sigma_{catch}(s,a)\sqrt{2\pi}} \exp(-(\ln(C(s,a,y,q)) - E(\ln C(s,a,y,q))))^2 / (2\sigma^2_{catch}(s,a))) \quad (12)$$

Where

$$E(\ln(C(s,a,y,q)) = \ln(\frac{F(s,a,y,q)}{Z(s,a,y,q)}(1-\exp(-Z(s,a,y,q))N(s,a,y,q)) \quad (13)$$

The negative log-likelihood for total catches then becomes:

$$l_{Catch} = -\ln(L_C) \propto NOS\ Y \sum_{s,a} A(s)\ln(\sigma_{catch}(s,a)) + \sum_{s,a,y,q} (\ln(C(s,a,y,q)) - E(\ln(C(s,a,y,q))))^2 / (2\sigma^2_{catch}(s,a))) \tag{14}$$

Research survey indices

In analogy the commercial catches the survey indices , *CPUE(survey,s,a,y,q),* are assumed lognormal distributed with mean

$$E(\ln(CPUE(survey,s,a,y,q)) = \ln(Q(survey,s,a)\overline{N}(s,a,y,q)) \tag{15}$$

where Q indicates catchability of the survey and q is the season in which the survey takes place. Catchability may be age dependent or the same for a group of ages. Similarly, the variance of log *CPUE*, $\sigma(survey,a)$, may be estimated by age group or range of ages. The log likelihood is on the same form as equation (14).

$$l_{SURVEY} = -\ln(L_{SURVEY}) \propto Y_{survey,s} \sum_{survey,s,a} A(survey,s)\ln(\sigma_{survey}(survey,s,a)) + \sum_{survey,s,a,y} (\ln(CPUE(survey,s,a,y) - E(\ln(CPUE(survey,s,a,y)))^2 / (2\sigma^2_{survey}(survey,s,a))) \tag{16}$$

Stomach contents

The observations considered for modelling predator food preference are the average proportions by weight in the stomach averaged over the entire North Sea and obtained from stomach samples (e.g. ICES 1989; ICES 1997). The observations for given prey and predator species, $STOM(l_{prey}, l_{pred}, y, q)$, are grouped by size groups, l_{prey} and l_{pred} and are assumed to be stochastic variables subject to sampling and process variations. The individual size may be either weight or length.

For given predator/size the observations, proportions of the prey species/size groups included, may be correlated of two reasons: 1. The proportions are summing to one ($\sum_{prey,l_{prey}} STOM(l_{prey}, l_{pred}, y, q) = 1$). 2. Specific prey items may co-occur in the stomachs in a systematic pattern due to the combined effect of the preference of the predator and the spatial distribution of the predator and prey species. Thus, for a given predator the probability distribution of the stomach observations including all prey/length groups in principle needs to be a multivariate distribution including the correlation between observations.

To different distributions have been considered: A multivariate log normal and a Dirichlet distribution. The parameterisation of the Dirichlet distribution is described in the Appendix. For the log normal distribution the expected values and the variance/covariance matrix need to be specified. For a preliminary simple case considered here the correlations are, however, ignored and set to zero, which means that observations are assumed to be independent lognormal variables. If each of the stomach proportions are small, the true correlations probably may be small and hence the assumption of independency may be a reasonable approximation.

The expected value of the stomach observations is modelled on the basis the theory developed by Andersen and Ursin (1977) and the modifications made by Gislason and Helgason (1985). Assuming that the seasonal stomach sampling has taken place uniformly over the seasons the median of log observations is set to:

$$MEDIAN(STOM(l_{prey}, l_{pred}, y, q)) = \text{proportion of available food} =$$

$$\frac{\overline{N}(l_{prey}, y, q) w(\tilde{l}_{prey})(SUIT(\tilde{l}_{prey}, \tilde{l}_{pred}, q)}{\sum_{prey, l_{prey}} \overline{N}(l_{prey}, y, q) w(\tilde{l}_{prey}) SUIT(\tilde{l}_{prey}, \tilde{l}_{pred}, q) + B(of, y, q) * SUIT(of, \tilde{l}_{pred}, q)} \qquad (17)$$

where l_{prey} and l_{pred} are indices for specified size groups and $\tilde{l}_{prey}$ and $\tilde{l}_{pred}$ are the midpoints of the corresponding size intervals, and where the suitability function is defined by eqs. (6) and (7).

In spite of that both the catch models and the deterministic MSVPA are age-structured models the predation model considered here has been chosen to be size based. This is simply due to that the predation process is determined by the size rather than by the age of the predators and their prey and further more, the stomach contents observations are given by size classes (and have been transformed to observation by age group for use in MSVPA).

As we prefer to keep the age-structured catch model, eq. (14), including stock numbers by age, stock numbers by size in the predation model, (14), need to be expressed by age as well. This is simply done by aid of length/age keys, which are assumed to be known without errors: Let $\beta(l_{prey}, y, q \mid a_{prey})$ denote the proportion of fish in size group l_{prey} of an a_{prey} year old fish $(\sum_{l_{prey}} \beta(l_{prey}, y, q \mid a_{prey}) = 1)$. If the age/length key samples are collected evenly over the season we approximate $\overline{N}(l)$ at size by $\overline{N}(a)$ at age:

$$\overline{N}(l_{prey}, y, q) = \sum_{a_{prey}} \beta(l_{prey}, y, q \mid a_{prey}) \overline{N}(a_{prey}, y, q)$$

Unpublished analyses of the present authors of stomach contents data for North Sea predators (ICES (1997)) indicate that the relationship between the variance and the median may formulated in the following way:

$$VAR(STOM(l_{prey}, l_{pred}, y, q)) =$$

$$\frac{1}{V_{pred} U_{l_{pred}, q}} MEDIAN(STOM(l_{prey}, l_{pred}, y, q))(1 - MEDIAN(STOM(l_{prey}, l_{pred}, y, q))) \qquad (18)$$

where $U_{l_{pred},q}$ is the number of hauls containing the predator/length in the stomach sampling programme and where V_{pred} are parameters.

As $STOM(l_{prey}, l_{pred}, y, q)$ is assumed to follows lognormal distributions with median and variance of the (untransformed) variables determined by eqs. (17) and (18), $\ln(STOM(l_{prey}, l_{pred}, y, q))$ follows a normal distribution with mean $\ln(Med)$ and the variance, γ^2, equal to

$$\gamma^2(l_{prey}, l_{pred}, y, q) = \ln\left(0.5 + \sqrt{0.25 + \frac{\frac{1}{MED(STOM(l_{prey}, l_{pred}, y, q))} - 1}{V_{pred} U_{lpred,q}}}\right) \qquad (19)$$

As finally observations for different predator/size groups are assumed independent the likelihood then becomes

$$l_{stom} \approx -\ln(L_{stom}) = \sum_{pred, l_{pred}, prey, l_{prey}, y, q} \left\{ \ln(\gamma(l_{prey}, l_{pred}, y, q)) + \frac{(\ln(STOM(l_{prey}, l_{pred}, y, q)) - \ln(MEDIAN(STOM(l_{prey}, l_{pred}, y, q))))^2}{2\gamma^2(l_{prey}, l_{pred}, y, q)} \right\} \qquad (20)$$

The empirical correlation between two prey types found in the stomach of a given predator can be found using basic data. Therefore, an alternative to the assumption of independent prey observations could be to use multivariate log normal distributions either with a specified correlation structure included or simply with known correlations equal to the empirical determined correlations.

Stock-recruitment relation included as a penalty function

In order to enable estimation of recruitment in the last year for the cases where survey CPUE and catch at recruitment age are missing (e.g. for North Sea saithe) a stock-recruitment relationship $R(s, y) = R_s(SSB(s, y) \mid \beta(s), \gamma(s))$ penalty function is included in the likelihood function. Presently the Ricker and Beverton & Holt functions are available together with a constant recruitment model. Assuming that recruitment takes place at the middle of the year and that recruitment is lognormal distributed the parameters $\beta(s)$ and $\gamma(s)$ are estimated from the log likelihood contributions previously defined and the log penalty contribution, l_{s-r}, equal to

$$l_{s-r} = \ln(L_{s-r}) \propto Y \sum_{s} \sigma_{s-r}(s) + \sum_{s,y} ((\ln(N(s, a=0, y, q=NOS/2)) - E(\ln(R(s, y))))^2 / 2\sigma^2_{s-r}(s)) \qquad (21)$$

where

$E(\ln(R(s,y))) = \beta(s)SSB(s,y)\exp(-\gamma(s)SSB(s,y))$ for the Ricker case,

where

$E(\ln(R(s,y))) = \dfrac{\beta(s)SSB(s,y)}{1+\gamma(s)SSB(s,y)}$ for the Beverton & Holst case,

and where

$E(\ln(R(s,y))) = \beta(s)$ in case of constant recruitment.

The application of the penalty function, eq. (21), may be interpreted as an Bayesian approach.

Total likelihood function and parameterisation

The total negative log likelihood function, *l,* is now found as the sum of the four terms:

$$l = l_{CATCH} + l_{SURVEY} + l_{STOM} + l_{S-R}$$

To ensure unique parameters restrictions on these have to be set. The parameters, which have to be specified, are the vulnerability parameters $\rho(prey, pred, q)$ and the unknown biomass of "other food", $B(of, y, q)$. Eqs. (9), (17), (6) and (7) indicate that for each year and quarter it is only possible to determine vulnerability parameters, $\rho(prey, pred, q)$ relative to $B(of, y, q)\rho(of, pred)$. Hence, for each year and quarter one of these parameters need to be specified. As we do not know the biomass of other food we arbitrarily choose to set $B(of, y, q)\rho(of, pred) = 1\,\text{mill}$ tonnes for each predator, year and quarter.

Similarly, to ensure unique fishing mortality parameters two parameters have to be set. We have chosen to fix $F_2(y=1)=1$ and $F_3(I_2(a), q = \text{last season}) = 1/NOS$.

The parameters in the model are:

Initial stock size in the first year : $N(s, a \geq 1, y=1, q=1)$,

Recruitment : $N(s, a=0, y, q=RS(s)), y = 1,\ldots,Y$

Fishing mortality parameters: $F_1(s, I_1(a), J(y))\, F_2(s,y)\;\; F_3(s, I_2(a), q)$

Vulnerability: $\rho(prey, pred)$, $\rho(of, pred) = 1$

Preferred relative size: $\eta(pred)$

Variances: $\sigma^2_{pref}(pred)$

Stock-recruitment parameters: $\beta(s)$, $\gamma(s)$

Survey catchability: $Q(survey, s, a)$

Other food size dependency: $\alpha(pred, q)$

The parameters have been estimated by minimizing the negative log likelihood, *l*, and the variance/covariance matrix has been approximated by the inverse Hessian matrix.

Implementation

The package AD Model Builder, ADMB, (Otter Research Ltd 2003) was used to calculate maximum likelihood parameters and the Hessian matrix. ADMB was used to carry out Markov Chain Monte Carlo simulations (Gilks et al. 1996), MCMC, to estimate the posterior distributions of the parameters.

Estimation of parameters in Phases

To obtain the maximum likelihood estimates of parameters in the high dimensional parameter space it has been necessary to estimate parameters in a number of phases, i.e. several parameters are estimated in the first step and more and more parameters are added to the model in the following phases.

For the North Sea case the model 500-1000 parameters are included of which approximately 10% are predation mortality parameters. If predation mortality, *M*2, is constant over years (and known) the total natural mortality is known and SMS performs a number of "single species" assessments. Even with a high number of species, the optimisation is relatively fast, as the parameters for the individual species are independent. The "single species" estimates are used as initial values in the multispecies model to facilitate fast and efficient parameter estimation. The estimation is done in four phases:

1. Estimate realistic stock numbers and recruitment parameters from catch at age data assuming a fixed annual F at 1.0 and fixed natural mortality e.g the values used by the ICES assessments. .
2. Use parameter estimates from step 1 as initial values and estimate all "single species" parameters in the full "single species" model using a fixed natural mortality,
3. Use parameter estimates from step 2 as fixed parameter values in a full multispecies model and estimate M2 and predation parameters.
4. Re-estimate all model parameters in the full multispecies model using initial parameter values from step 2 and 3.

Technically, step 1 is not necessary for convergence, however it reduces the running time. The fixation of "single species" parameters in phase three is optional. Fixed parameters speed up the optimisation and does not change the final results after phase 4.

Test runs and comparisons of results with MSVPA

Simulated data sets with a varying degree complexity and "sampling noise" have been used to investigate the characteristics of the estimators (not presented here). This test procedure provided evidence that the model works as intended for simulated data.

SMS results using sub-set of the North Sea MSVPA data set

To test the model on "real" data the model was applied for a subset set of data used for the MSVPA "key-run" made in 2002 (ICES 2002). The so-called "other predators" included in the MSVPA are not yet implemented in SMS and these species were left out of dataset. The remaining species included are presented in Table 1.

The North Sea average of stomach contents data used as observations in SMS were calculated from the basic MSVPA data (ICES 1997) for year 1991 only. For use in a comparable MSVPA these length base stomach data were processed further into stomach contents data by age group using the methodology described in ICES (1997).

Mean length per age group and length-weight relations were estimated from quarterly IBTS data 1991-1997 and commercial catches (sandeel). Data for all years were pooled and one set of quarterly age-size key were derived and used as an approximation for all years in the model.

Specifications of the separable fishing mortality models are presented in Table 2 while survey specifications are given in Table 3.

To show the effect of estimating predation mortality within the model versus to use "known" natural mortality, SMS was run with and without predation estimation.

Comparison with MSVPA

The input data for SMS are mainly based on MSVPA data and the formulation of predation mortality for both models are based on the same theory. The results of the two models should therefore be comparable. Hence results from SMS and MSVPA using the same data sources as for SMS were compared. The 4M package (Vinther et al. xxx) was used to obtain the MSVPA results.

RESULTS

A minimum of the negative of likelihood was found for which it also was possible to obtain the inverse Hessian matrix. There was a problem with the size selection parameter, $\sigma_{pref}(pred)$, for saithe, which was found to lie on the boundary. Instead the parameter was fixed to 5. The value 5 was chosen to be within the range of $\sigma_{pref}(pred)$ estimated for the other predators and to give a realistic estimate of the preferred predator-prey size ratio, $\eta(pred)$ for saithe.

The negative log likelihood contributions by species and type of observations are presented in Table 4. The number of observations differs in the various groups and the contributions cannot be compared directly. It is however clear, that catch and that catch and survey observations has the best fit (low negative log likelihood) for cod, whiting and haddock. The sum of log likelihood contributions from stomach contents observations is positive for all predators.

The size preference parameters are presented in Table 5. A symmetric size preference model was chosen for all species except for whiting. The parameters for preferred size predator/prey ratio are estimated with a coefficient of variation (CV) in the range 6-30%. Haddock had the largest CV due to relatively few observations available. The preference curves based on the estimated parameters are presented in Figure 1. Whiting prefers a prey which body weight is 171 times smaller than it self. Cod prefer slightly smaller preys (a factor of 260) and saithe prefer a prey 420 times smaller. The smallest prey is preferred by haddock, which prefers a prey 2600 times smaller than it self. Size selection range seems unrealistic wide for all predators except whiting.

The relative vulnerability parameters are shown in Table 6. The precision of the estimated vulnerability parameters is highest for predator cod, followed by whiting and saithe. Haddock eats relatively few fish preys and the vulnerability parameter for its two preys, sandeel and Norway pout are estimated with a CV of more than 125%. In general the precision of the estimated vulnerability parameters is low. Cod and whiting prefer cod, whiting and haddock and only to a lesser extent herring, sandeel and Norway pout. Sandeel seems to be the least preferable of the species considered.

Whether natural mortality, M, remains constant over time for a given age group is a crucial question in stock assessment. This is due to that if M is constant over years each stock assessments can be done separately as single species assessments. (Of course M has to be estimated regularly in the multispecies model). Annual predation mortality, $M2$ and the 95% confidence intervals indicate that the $M2$ values significantly different over the period 1975-2000 (Figure 2), as the 95% confidence interval is not overlapping for the whole time period. A formal statistical test is not possible, as the individual $M2$ values are not independent. $M2$ seems to decline since 1980.

Figure 3 and 4 present the precision of SSB and average F for the last 5 model years. The smallest CVs are those for cod, haddock and whiting, followed by saithe and herring. The small sized and short-lived prey species sandeel and Norway pout have the largest CV. Further, the estimates of F and SSB are – not surprisingly - correlated. An example is shown for the last model year for a low (cod) and a high CV species (sandeel) in Figure 5, where the correlation is calculated using MCMC with 200.000 chains thinned by a factor 200 resulting in 1000 uncorrelated chains. Uniform priors were used for all parameters except for the recruitment, for which the Ricker function was assumed. The correlation is visualised using contour plots of the two dimensional density of the simultaneous distribution of F and SSB. It appears that the correlation is larger for cod than for sandeel.

The sum of log likelihood contributions from stomach contents observations is positive for all predators (Table 4). Individual contribution from each stomach content observation is shown for cod on Figure 6. There is a clear negative relationship between the likelihood contribution and stomach weight proportion in the stomach indicating a better model fit for larger weight proportions. A residual plot (Figure 7) confirms that large relative stomach contents has the smallest residuals. For a species like cod, for which “other food” part amounts at least 50% relatively small residuals are obtained from this component, whereas the prey species included in the model obtain larger residuals, due to their low proportion in the diet.

MSVPA and SMS comparison

SMS and MSVPA are based on the same theory apart from that suitability has been modelled in the SMS (eqs. (6) and (7)) and SMS is a stochastic model. To results from the two models they have been tested on the same data set. Regarding SSB the results for the two models are also quite similar for all the species except for the short-lived prey species sandeel and Norway pout (Figure 8). The difference is larger for the two prey species but have the same trends.

Predation mortality (M2) for the youngest ages, estimated by SMS and MSVPA however, are different but has in general the same temporal trend (Figure 9). M2 differs for many species - age groups combinations with a factors up to 3. MSVPA M2 estimates are larger for the 0-group and the 2-group for five out of the six species, while there is no clear pattern for the 1-group. In most cases the between years variation of M2 is larger for the MSVPA than for SMS.

The effect of estimating M2 in SMS

Besides of running the full SMS with species interactions included, SMS has the option to run with fixed (known) natural mortality. This corresponds to making single species stochastic, stock assessment. Table 4 and Table 7 show the likelihood contributions from the two models. It is clear that the addition of predation mortality has a very little effect on the “single species” (catch, CPUE & S/R) likelihood contributions. This further indicates that the precision of estimates should be similar for the two models, which is confirmed by the comparison of the CV of the estimated SSB for the two models (Figure 10).

The main difference with respect to the results of a “single” or “multispecies” SMS is differences in the mortality of the younger ages and thereby the estimated recruitment. That is clearly seen in the stock-recruitment plots (Figure 11 and 12) for the two models. Compared to single species case the multispecies model results in much more variable recruitment estimates for cod whiting and Norway pout, while the opposite is the case for haddock and Sandeel. The range of SSB is quite similar for the two models, except for sandeel and Norway pout.

Discussion

REFERENCES

Andersen, K. P., and Ursin, E. 1977. A Multispecies Extension to the Beverton and Holt Theory of Fishing, with account of Phosphorus Circulation and Primary Production. Meddr. Danm. Fisk.- og Havunders. **7** 319-435.

Deriso, R.B., Quinn II T.J., and Neal, P.R. 1985. Catch-Age Analysis with Auxiliary Information. Can. J. Fish. Aquat. Sci. **42**: 815-824.

Fournier, D.A., Hampton, J., and Sibert, J.R. 1998. MULTIFAN-CL: A length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, Thunnus alalunga. Can. J. Fish. Aquat. Sci. **55**: 2105-2116.

Gislason, H., and Helgason, T. 1985. Species interaction in assessment of fish stocks with special application to the North Sea. Dana **5**: 1-44.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. 1996. Markov Chain Monte Carlo in practise. Chapman & Hall, London, UK. 486 pp.

ICES 1997. Database report of the stomach sampling project 1991. ICES Coop. Res. Rep. No. 219.

ICES 2002. Report of the Workshop on MSVPA in the North Sea, 8-12 April 2002. ICES CM 2002/D:04

Lewy, P., and Nielsen, A. 2003. Modelling stochastic fish stock dynamics using Markov Chain Monte Carlo. ICES J. Mar. Sci., 60: 743-752.

Nielsen, A., and Lewy, P. 2002. Comparison of the frequentist properties of Bayes and the maximum likelihood estimators in an age-structured fish stock assessment model. Can. J. Fish. Aquat. Sci. **59**: 136-143.

Otter Research Ltd 2003. AD Model Builder Documentation. Otter Research Ltd, Sydney, British Columbia. http:\\otter-rsch.com.

Sparre, P. 1991. Introduction to multispecies virtual population analysis. ICES J. Mar. Sci. Symp. **193**: 12-21.

Taylor, L., Begley, J., Kupca, V., and Stefansson, G. 2004. ICES, FF:24.

Vinther, M., and Lewy, P. 2002. Analysis of the variation of the diet of North Sea demersal roundfish species. In prep.

APPENDIX. THE DIRICHLET DISTRIBUTION USED TO DESCRIBE STOMACH CONTENTS DISTRIBUTIONS

An alternative to the independent lognormal distribution used to describe the stomach contents distribution, eq. (20), could be the Dirichlet distribution. The latter distribution has the advantage that it at least accounts for the correlation due to that observations sum to one. Applying the stomach contents model (eq. (17), where the median has been replaced by the expected value, *E*) the density distribution of the Dirichlet can be formulated as follows:

Assume for a given predator, *pred/l_{pred}* year and quarter that the observed diet, $STOM(l_{prey}, l_{pred}, y, q)$ ($\sum_{prey, l_{prey}} STOM(l_{prey}, l_{pred}, y, q) = 1$) follows a Dirichlet distribution with parameters $p(l_{pred}, l_{prey}, y, q)$. The with density function is

$$f_{l_{pred}}(y,q) = f((STOM(l_{prey}, l_{pred}, y, q))_{prey, l_{prey}} \mid (p(l_{pred}, l_{prey}, y, q))_{prey, l_{prey}})$$
$$= \frac{\Gamma(p(l_{pred}, y, q)}{\prod_{prey, l_{prey}} \Gamma(p(l_{pred}, l_{prey}, y, q))} \prod_{prey, l_{prey}} STOM(l_{prey}, l_{pred}, y, q)^{p(l_{pred}, l_{prey}, y, q) - 1} \quad \text{(A1)}$$

where

$$p(l_{pred}, y, q) = \sum_{prey, l_{prey}} p(l_{pred}, l_{prey}, y, q)$$

herher

As the predation model is expressed through the mean values of each of the prey/prey length combinations the parameters $p(l_{pred}, l_{prey}, y, q)$ needs to be expressed as functions of the predation parameters through the mean values. As the mean in the Dirichlet distribution for a given predator/length group of a specified prey is

$$E(l_{pred}, l_{prey}, y, q) = E(STOM(l_{prey}, l_{pred}, y, q)) = \frac{p(l_{pred}, l_{prey}, y, q)}{p(l_{pred}, y, q)}$$

we simply put

$$p(l_{pred}, l_{prey}, y, q) = p(l_{pred}, y, q) E(STOM(l_{prey}, l_{pred}, y, q)) \quad \text{(A2)}$$

By inserting eq. (A2) in the Dirichlet distribution, (A1), the parameters now include the predation parameters and $p(l_{pred}, y, q)$.

The variance and covariance is known to be

$$VAR(l_{pred}, l_{prey}, y, q) = \frac{1}{p(l_{pred}, y, q) + 1} E(l_{pred}, l_{prey}, y, q)(1 - E(l_{pred}, l_{prey}, y, q)) \quad \text{(A3)}$$

$$COV(l_{pred}, l_{prey1,} l_{prey2}, y, q) = -\frac{1}{p(l_{pred}, y, q) + 1} E(l_{pred}, l_{prey1,}, y, q) E(l_{pred}, l_{prey2,}, y, q) \qquad \text{(A4)}$$

The parameter $p(l_{pred}, y, q)$ in eq. (A3) is modelled using eq. (18):

$$p(l_{pred}, y, q) = V(l_{pred}, y, q) U(l_{pred}, y, q) - 1$$

Hence, the parameters, *p*, is replaced by the parameters, *V*.

Assuming that the diet observations are independent for the predator/length groups the negative log likelihood function including all predators/length groups are:

$$l_{STOM} = -\ln(L_{STOM}) = -\ln(\prod_{pred, l_{pred}, y, q} f_{l_{pred}}(y, q)) \qquad \text{(A5)}$$

where

$$p(l_{pred}, l_{prey}, y, q) = (V(l_{pred}, y, q) U(l_{pred}, y, q) - 1) E(STOM(l_{prey}, l_{pred}, y, q)) \qquad \text{(A6)}$$

and where *E(STOM)* equals the right hand side of eq. (17).

Figure 1. Size preference by predator as estimated by SMS

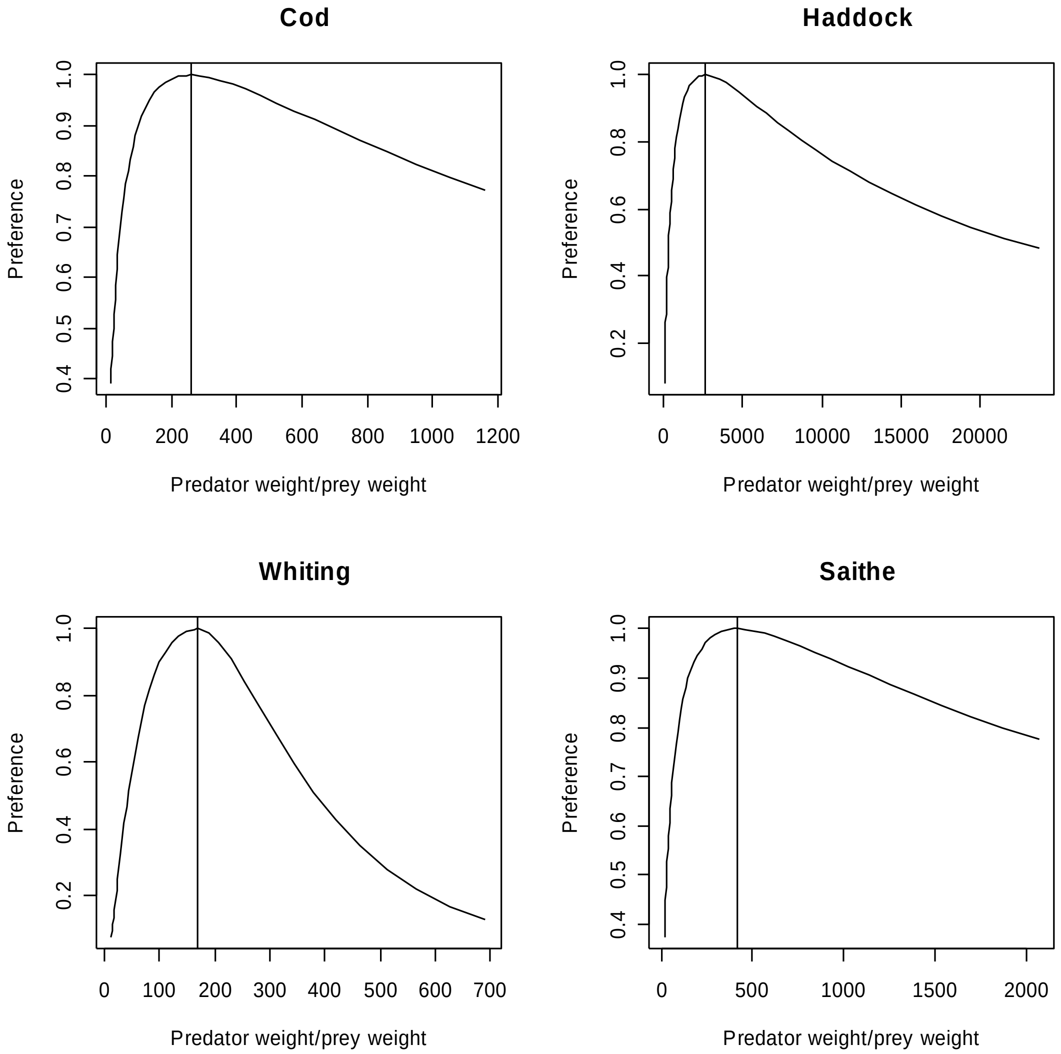

Figure 2. Annual predation mortality (M2) and the 95% confidence intervals for age group 1.

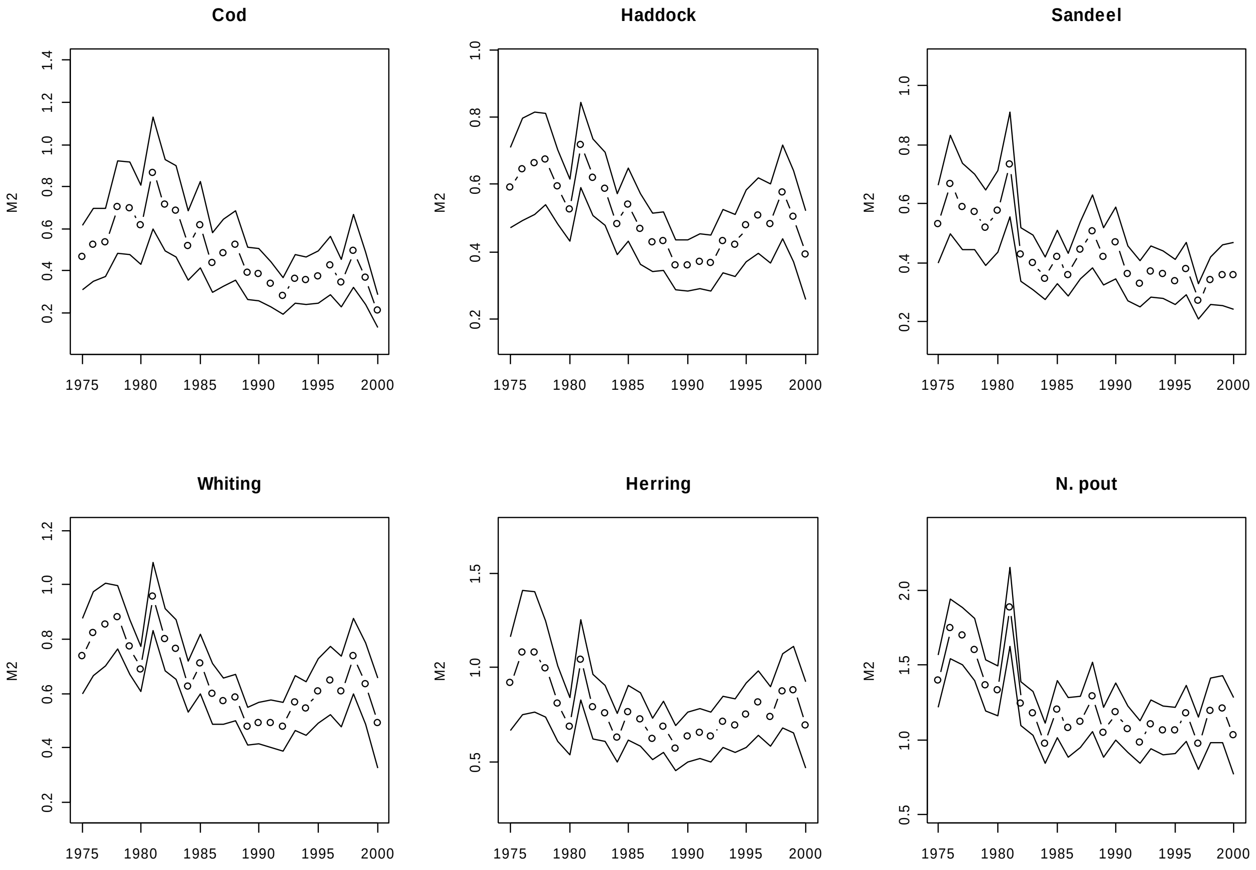

Figure 3. Coefficient of variation (%) of spawning stock biomass

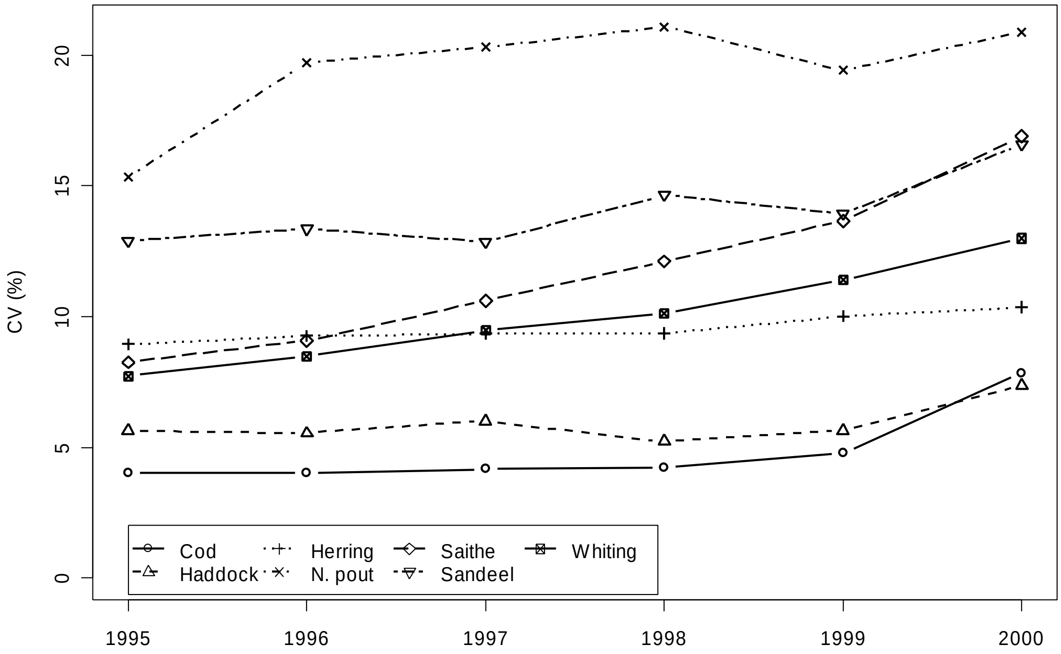

Figure 4. Coefficient of variation (%) of average fishing mortality

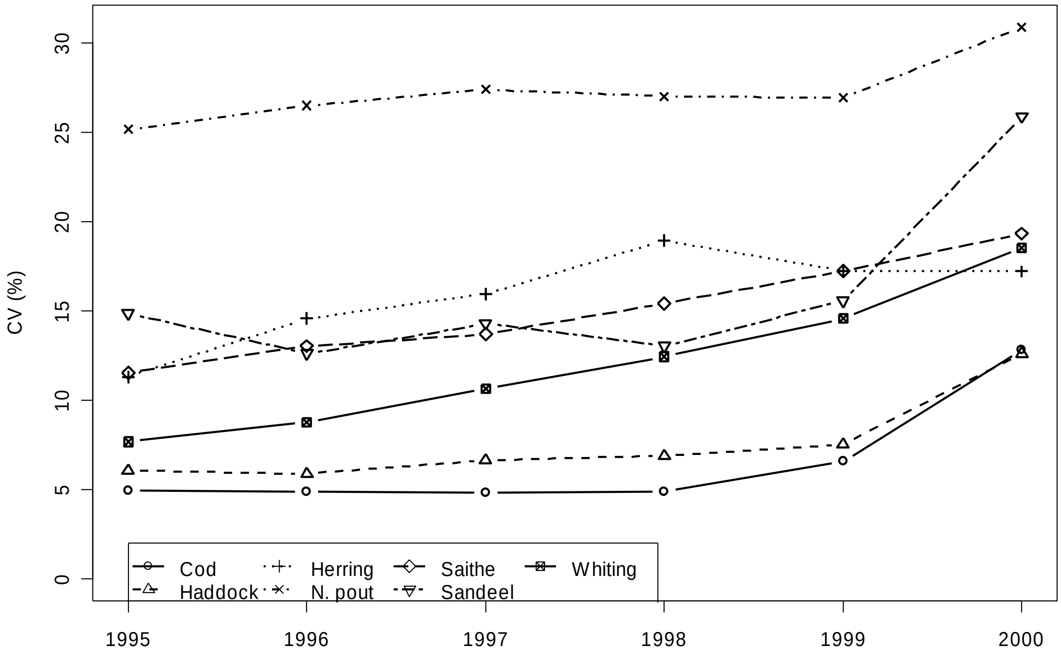

Figure 5. Posterior density of SSB, average F and their simultaneous distribution in the last year estimated from 200000 Markov Chain Monte Carlo simulations.

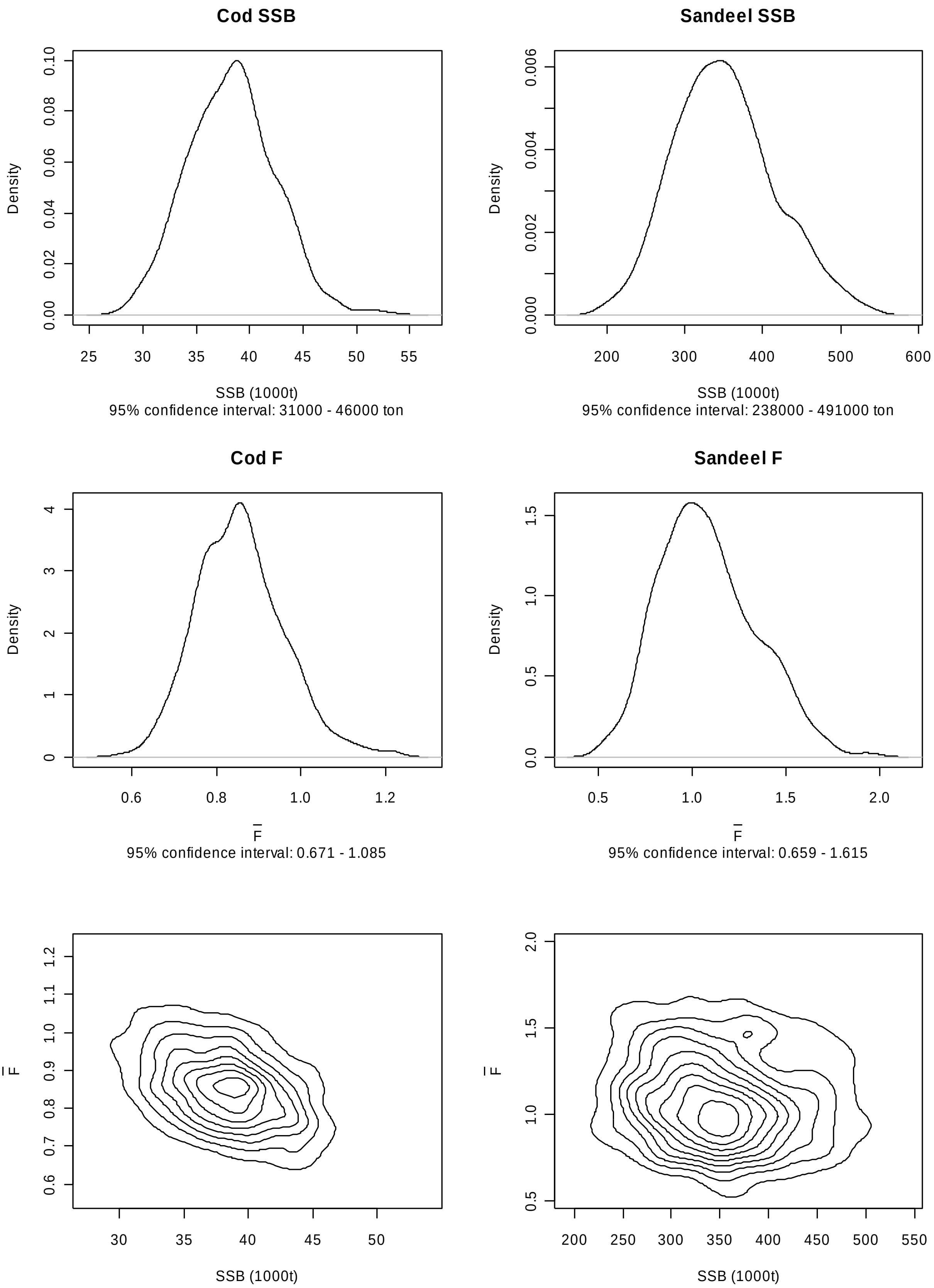

Figure 6. log likelihood contributions for individual stomach contents observations

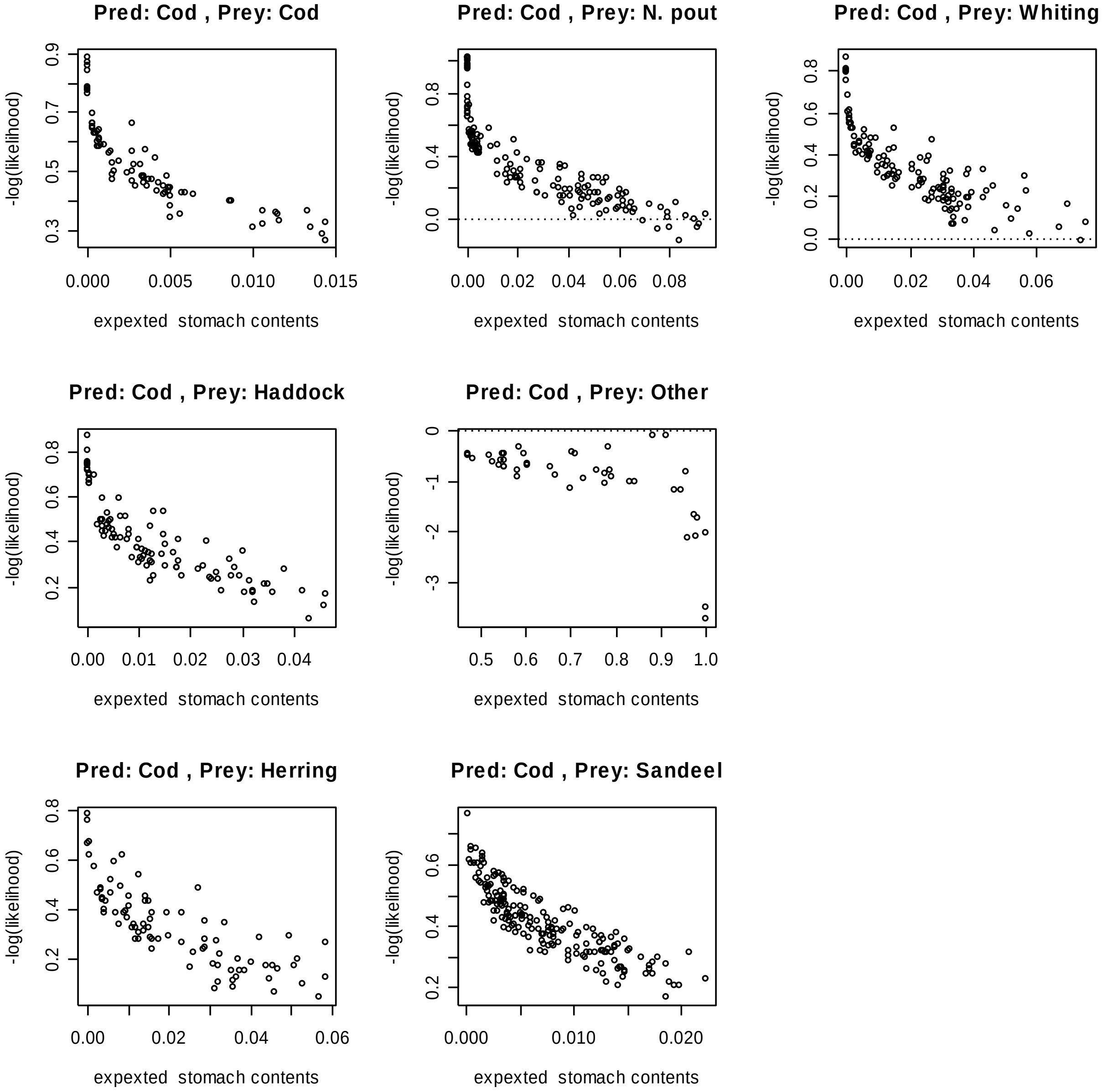

Figure 7. Residuals of stomach contents data for Cod as predator

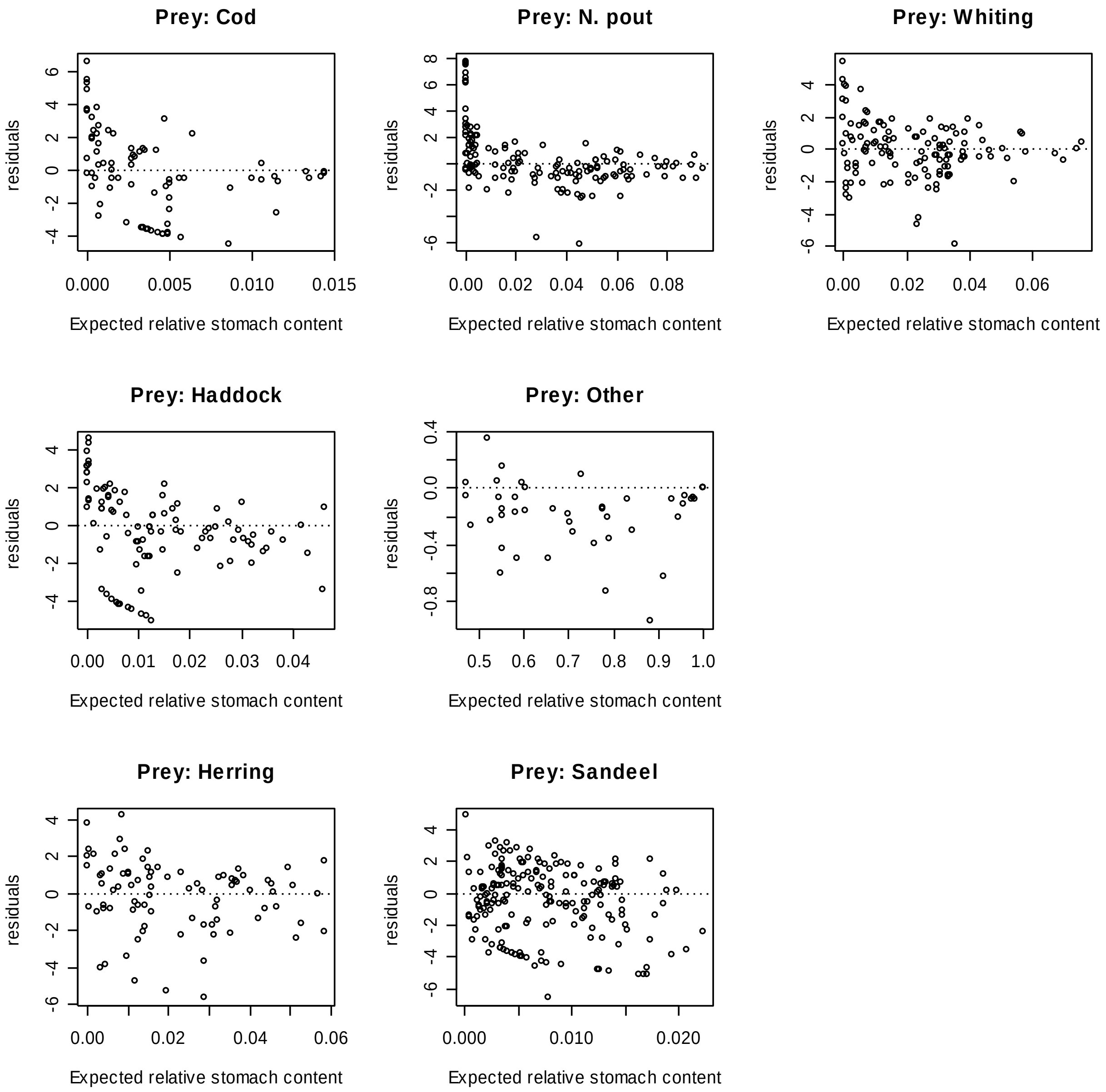

Figure 8. Spawning stock biomass by species and year estimated by SMS (dotted line) and MSVPA (solid line).

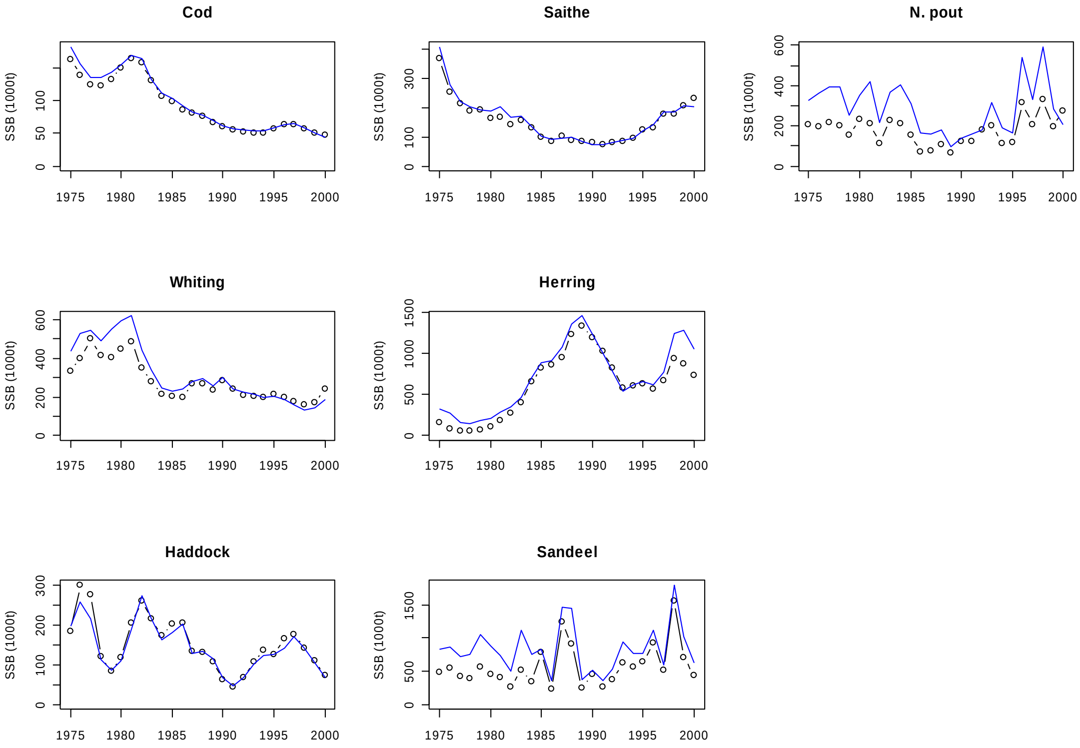

Figure 9. Predation mortality (M2) by species and age estimated by SMS (dotted line) and MSVPA (solid line). M2 is annual values except for age group 0, which include only the second half of the year.

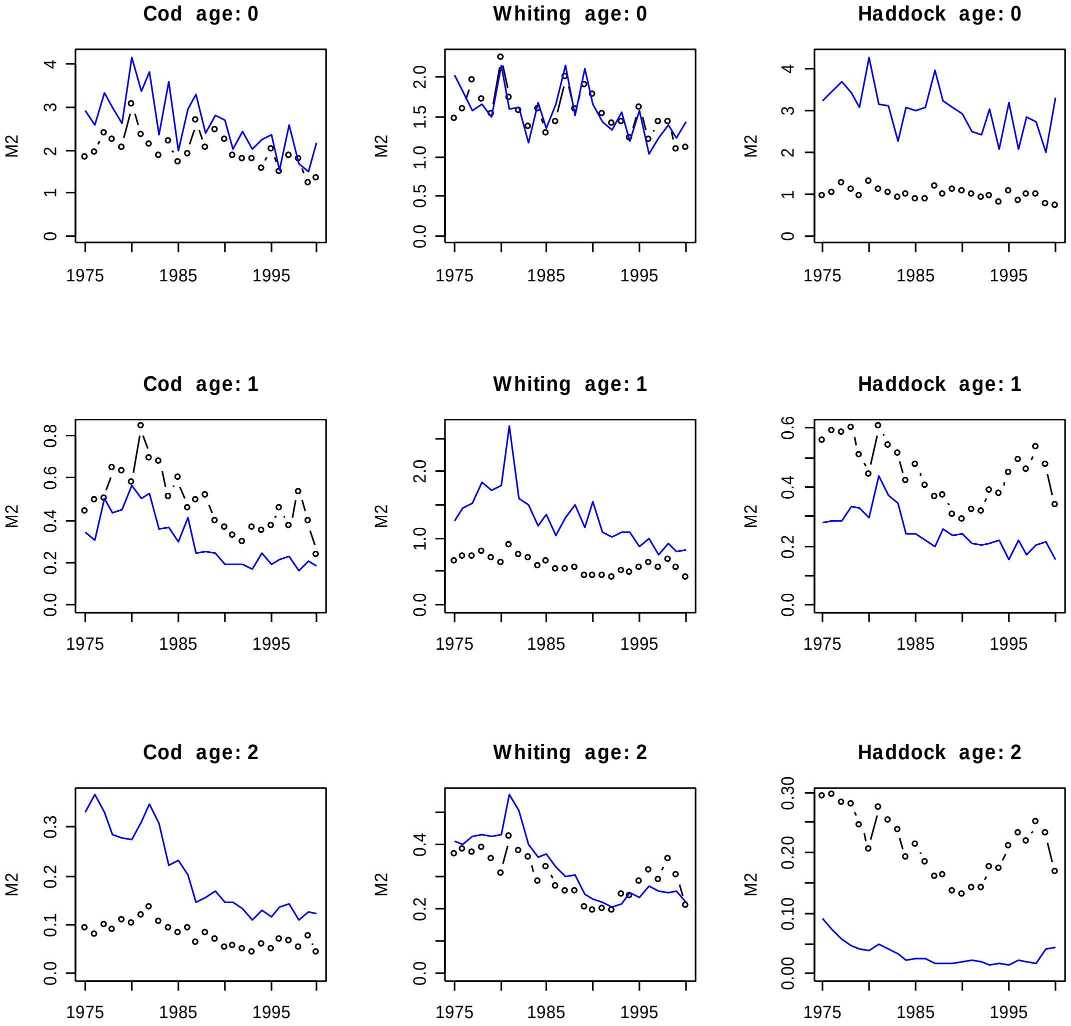

Figure 9 (continued). Predation mortality (M2) by species and age estimated by SMS (dotted line) and MSVPA (solid line). M2 is annual values except for age group 0 which include only the second half of the year.

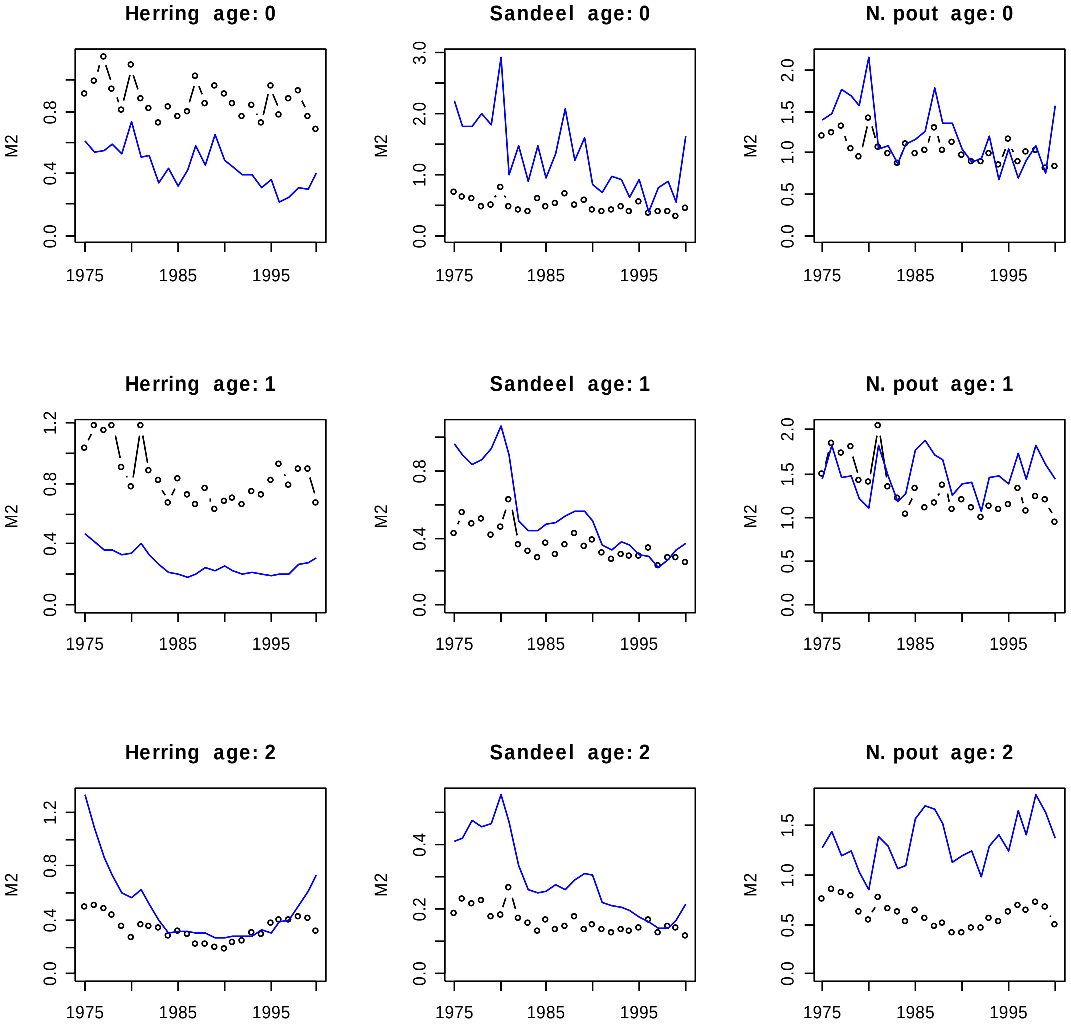

Figure 10. Coefficient of variation of SSB estimated from SMS model including species interactions (solid line) and single species analyses

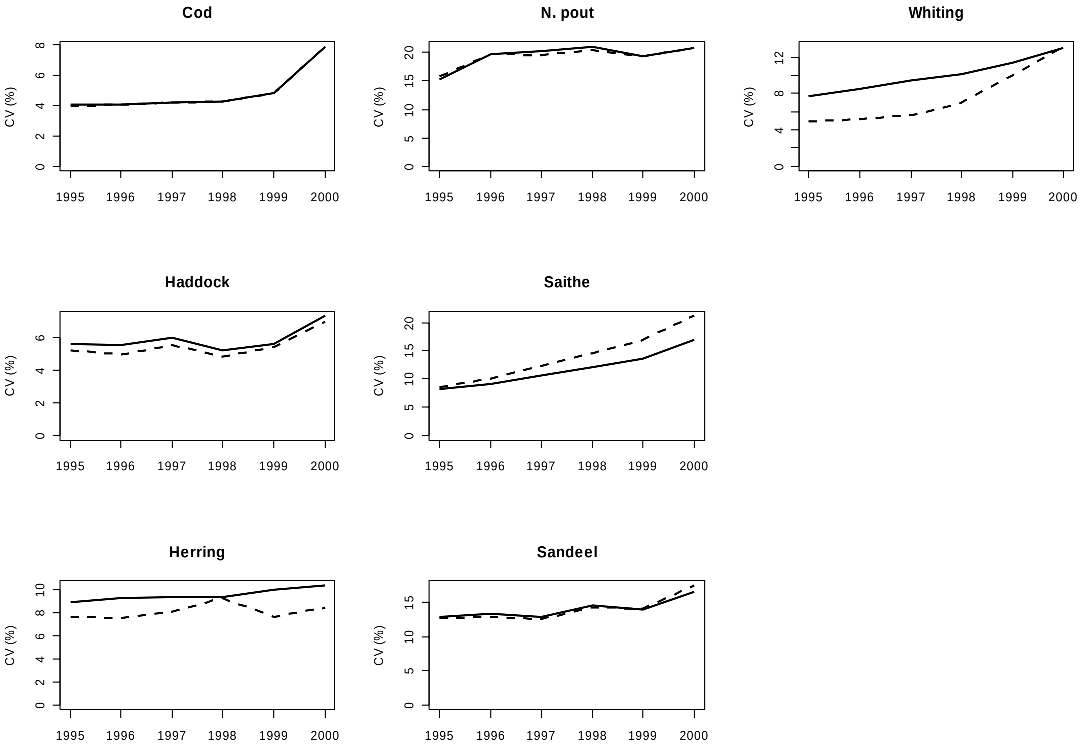

Figure 11. Recruitment numbers and spawning stock biomass for the period 1975-2000 estimated by SMS including species interactions

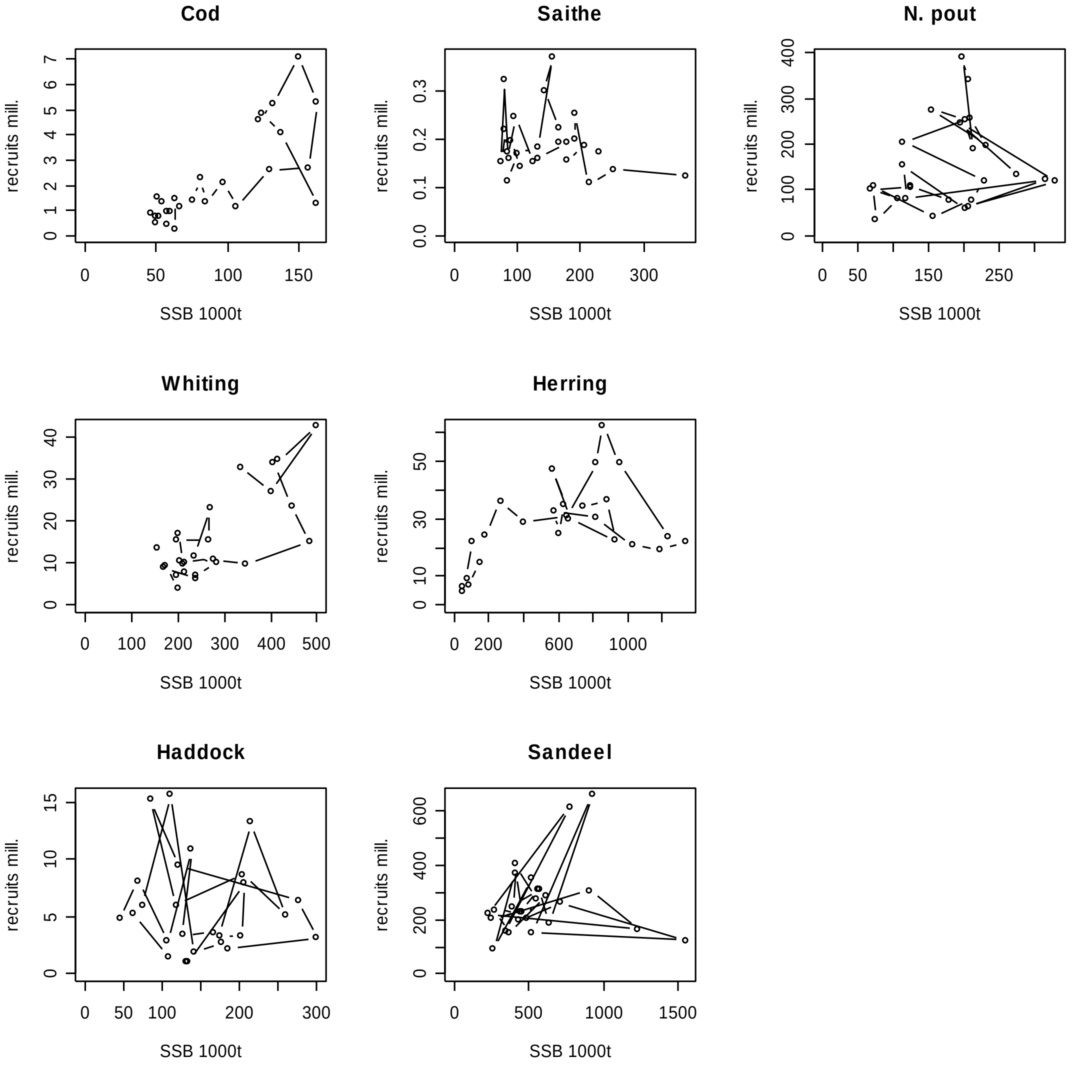

Figure 12. Recruitment numbers and spawning stock biomass for the period 1975-2000 estimated by single species SMS using fixed natural mortality from the ICES single species assessment.

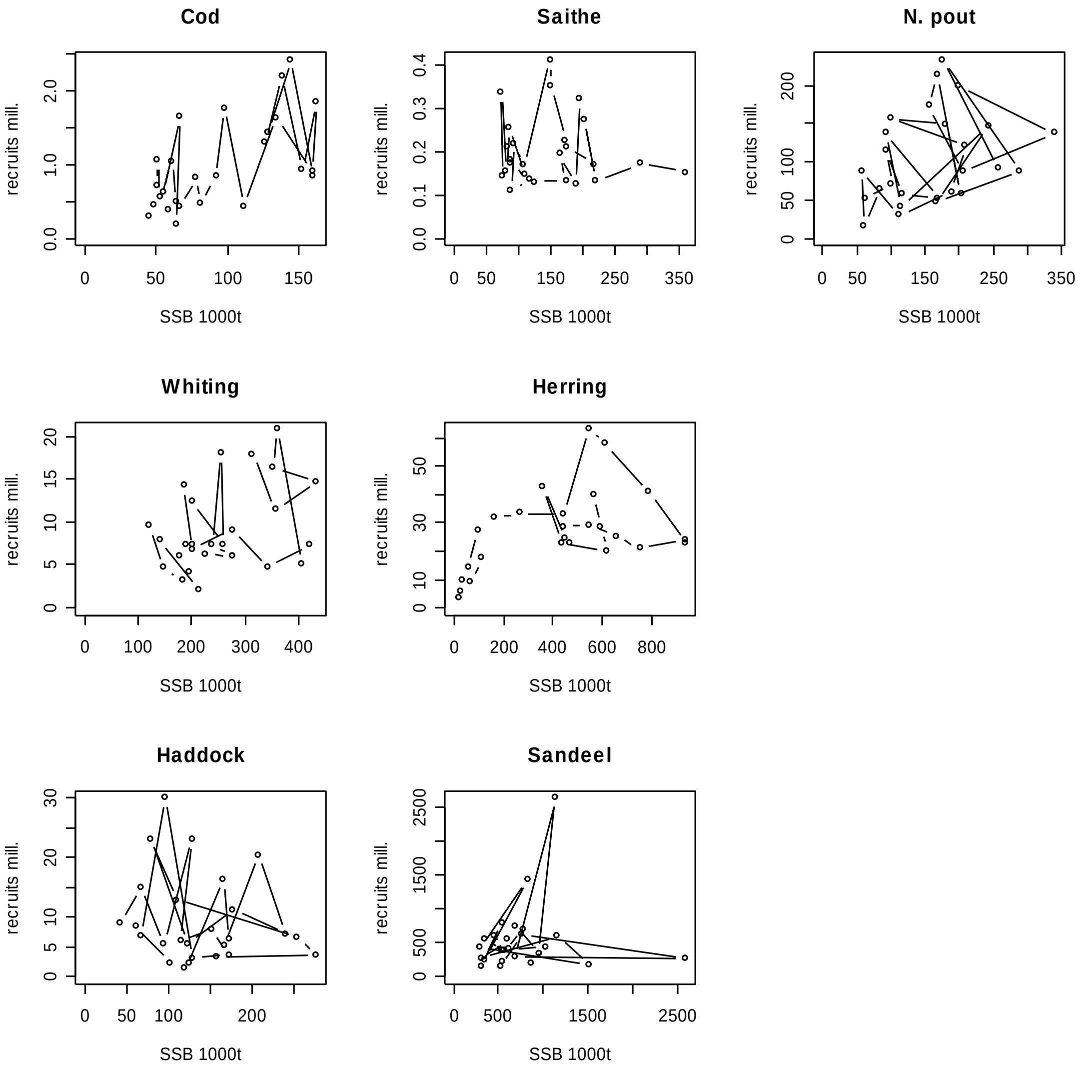

Table 1. Species and specification of options

Species	Max age	Predator	Prey	SSB recruitment relation
Cod	10+	yes	yes	Ricker
Whiting	8+	yes	yes	Ricker
Haddock	10+	yes	yes	Geom. mean
Saithe	10+	yes	no	Ricker
Herring	7+	no	yes	Ricker
Sandeel	4+	no	yes	Geom. mean
Norway pout	3	no	yes	Geom. mean

Table 2. Specifications of Fishing mortality model

Species	First age where catches are used	Last age with age dependent selectivity	Year range with constant age selectivity	Age groups for the same seasonal selectivety	Age groups for variance of catch
Cod	1	7	1975-2000	1, 2, 3-10	1, 2-6, 7-10
Whiting	0	7	1975-2000	0, 1-8	0, 1, 2-8
Haddock	0	7	1975-2000	0, 1, 2-10	0, 1, 2-6, 7-10
Saithe	2	8	1975-2000	2, 3-10	2, 3-10
Herring	0	5	1975-1977 1978-1982 1983-2000	0, 1, 2-7	0, 1, 2-7
Sandeel	0	2	1975-2000	0, 1, 2-4	0, 1-3, 4
Norway pout	0	2	1975-2000	0, 1-4	0, 1-2, 3

Table 3. Survey data and specifications.

Species	Survey	Ages	Age groups for catchability	Age groups for variance
Cod	English Groundfish, 1977-2000	1-5	1, 2, 3, 4-5	1-2, 3-5
	IBTS Q1, 1977-2000	1-5	1, 2, 3, 4-5	1, 2, 3-5
	Scottish Groundfish, 1982-2000	1-6	1, 2, 3, 4-6	1, 2-6
Whiting	English Groundfish, 1991-2000	1-1	1	1
	IBTSQ1, 1991-2000	1-4	1-4	1, 2-3, 4
	Scottish Groundfish, 1991-1997	1-6	1, 2, 3-6	1, 2-6
Haddock	English Groundfish, 1977-2000	0-5	0, 1-5	0-5
	IBTS Q1, 1975-2000	0-5	0, 1-5	0-5
Saithe	English Groundfish, 1977-2000	2-3	2, 3	2, 3
	Norweigian acoustic, 1996-2000	3-7	3, 4, 5-7	3, 4-7
	Scottish Groundfish, 1982-2000	2-3	2, 3	2, 3
Herring	Acoustic, 1989-2000	2-7	2, 3, 4, 5, 6, 7	2-3, 4-7
	IBTSA, 1983-2000	2-5	2, 3, 4-5	2-5
	IBTSY79, 1979-2000	1-1	1	1
Sandeel	Northern NS 1 half, 1976-2000	1-1	1	1

	Northern NS 2 half, 1976-2000	0-0	0	0
	Southern NS, 1 half, 1982-2000	1-4	1, 2, 3-4	1, 2-4
N. pout	English Groundfish, 1982-2000	0-3	0, 1, 2-3	0, 1-3
	IBTS Q1, 1975-2000	1-3	1, 2, 3	1-3
	Scottish Groundfish, 1982-2000	1-3	1, 2-3	1-3

Table 4. Sum of negative –log likelihood contributions by species and type of observations

Multi species

	Catch	CPUE	Stock-rec.	Stomachs	all
Cod	-289.6	-96.4	-2.1	586.8	198.8
Whiting	-201.3	-50.3	-8.0	499.5	239.9
Haddock	-136.0	-116.4	5.4	76.5	-170.4
Saithe	98.9	53.0	-20.3	276.3	407.9
Herring	234.2	-82.6	-18.5	0.0	133.2
Sandeel	70.9	26.0	-8.2	0.0	88.7
N. pout	176.0	58.4	0.4	0.0	234.8
All	-46.9	-208.1	-51.3	1439.1	1132.8

Tabele 5. Parameter estimates and CV of size preference parameters.

Species	$\eta(pred)$ log "mean" of the preferred predator prey size ratio	CV (%) of $\eta(pred)$	$\sigma_{pref}(pred)$ "standard deviation" in the food preference function	CV (%) $\sigma_{pref}(pred)$
Cod	5.56	7	4.36	21
Whiting	5.14	6	1.35 & 0.47	16 & 137
Haddock	7.87	30	3.32	71
Saithe	6.04	10	5.0	-

Table 6. Estimated vulnerability parameters and their coefficient of variation in percent (in brackets). The (relative) values are only comparable separately for each predator.

	Prey species					
PREDATOR	Cod	Whiting	Haddock	Herring	Sandeel	N. pout
COD Relative to N. pout	1089 (27) 15.1	427 (25) 5.9	378 (25) 5.3	58 (26) 0.8	19 (16) 0.3	72 (19) 1
WHITING Relative to N. pout	105 (49) 4.8	103 (47) 4.7	123 (51) 5.6	31 (50) 1.4	4 (35) 0.2	22 (42) 1
HADDOCK Relative to N. pout	-	-	-	-	323 (125) 0.5	673 (131) 1
SAITHE Relative to N. pout	-	1085 (45) 1.2	1096 (56) 1.2	1427 (51) 1.5	62 (47) 0.1	943 (37) 1

Table 7. Sum of negative –log likelihood contributions by species and type of observations for a run with fixed natural mortality

Single species

	Catch	CPUE	Stock-rec.	Stomachs	all
Cod	-284.7	-100.8	-3.7	0.0	-389.2
Whiting	-201.1	-51.0	-6.6	0.0	-258.7
Haddock	-138.6	-115.4	6.7	0.0	-247.3
Saithe	97.0	51.9	-19.1	0.0	129.8
Herring	254.2	-95.5	-19.9	0.0	138.9
Sandeel	65.9	25.1	-2.3	0.0	88.7
N. pout	174.5	54.3	-2.7	0.0	226.1
All	-32.8	-231.5	-47.5	0.0	-311.7