African Journal of Marine Science 2007, 29(2): 223–245
Printed in South Africa — All rights reserved

AFRICAN JOURNAL OF
MARINE SCIENCE
EISSN 1814–2338
doi: 10.2989/AJMS.2007.29.2.7.190

A simple implementation of the statistical modelling framework Gadget for cod in Icelandic waters

L Taylor[1]*, J Begley[1], V Kupca[1] and G Stefansson[1, 2]

[1] Marine Research Institute, Reykjavik, Iceland
[2] University of Iceland, Reykjavik, Iceland
* Corresponding author, e-mail: lornaataylor@gmail.com

Gadget is a statistical modelling framework that can be used to assess individual fish stocks and to create multispecies, multi-fleet and multi-area models. The development of a complex multispecies model requires understanding of the modelled single-species population. Simple single-species models are essential in evaluating whether more complex models increase understanding of the population dynamics. In this paper, Gadget is illustrated using a single-species case study with two stock components on one area. Features of the implementation include immature and mature stock components, a maturation process, along with commercial and survey fleets. Parameter estimation is done using maximum likelihood based on a variety of datasets. The estimated parameters relate to growth, maturation, fleet selection, recruitment and the initial population. The data types used to calculate the likelihood include survey indices and biological samples from the commercial catch and surveys. Optimisation of the model, along with a protocol to estimate appropriate weighting of the likelihood components, is described. A range of different assumptions and estimation methods are evaluated including the effect of reduced data availability on parameter estimation by excluding age data from the likelihood components.

Keywords: cod, Gadget, maximum likelihood, statistical multispecies model

Introduction

Fishery science deals with the analysis of the interactions between fisheries and the ecosystem, possibly taking into account economic and social aspects. In order to address this task, data are collected, mainly through routine monitoring programmes.

Data on marine fish populations traditionally come from a variety of sources and contain considerable variability. The data that are routinely collected from the ecosystem are mainly obtained by sampling and measuring fish in various types of surveys and from the fishery, in both cases by recording the catches, effort and taking biological samples from the catches. These biological samples routinely include length measurements of individual fish, expensive age measurements from a subset of these, and stomach content analysis from an even smaller subset.

Much of the variation in the data translates directly into uncertainty when modelling population dynamics, including the assessment of fish stocks or stock projections (Gavaris *et al*. 2000, Patterson *et al*. 2000). In particular, when including species interactions, this may become increasingly difficult because extremely variable stomach content data need to be used (Stefansson 1998, Stefansson and Palsson 1997a).

Many of the species, fisheries and interactions encountered appear to exhibit fundamentally different dynamics, requiring different models for each situation (one may need different growth models or fleet selection for different species/fleets). Similarly, sampling variation in different datasets is notoriously difficult to model and this is reflected in underestimates of uncertainty, as seen for example in Patterson *et al*. (2001). There is thus not only a need to incorporate a variety of different descriptions of processes but also descriptions of a variety of different datasets. Historically, different models of the population dynamics of fish stocks have been able to utilise varying amounts of available data. In recent decades, models have been developed to use the various datasets through likelihood components when describing species, fleets and interactions in models which can be quite disaggregated. This approach involves an internal description of the underlying process including a prediction of the data. The predicted data can then be compared with the observations.

A computer program, Gadget (**G**lobally applicable, **A**rea **D**isaggregated, **G**eneral **E**cosystem **T**oolbox), has been developed to undertake such analyses in as objective a manner as possible, using formal statistical modelling approaches. The model presented in this paper is a statistical age-length structured model, based on the principles laid out in Stefansson and Palsson (1997b) and implemented using the Gadget modelling environment (Begley

2005). This is a simple example of the use of Gadget to describe a species consisting of two stock components, immature and mature fish, in a single area that is fished by one commercial and two survey fleets. Use of the program to include more species and areas is described.

The Gadget environment has been developed to utilise as many different types of fisheries data as possible, using appropriate assumptions on each dataset. One of the principles used here is to model the actual data collected in as raw a form as possible. In particular, rather than massaging age-length keys, length distributions, mean weights and landings into catch-at-age data, each of these data sources can be included in the model separately. For each type of observed data a corresponding fitted dataset can be generated by Gadget, allowing for direct comparisons between the modelled population and the observed population. As a result, whether each model component provides an adequate description of the corresponding dataset can be evaluated. For some reason, issues in combining datasets have largely been ignored in many assessment procedures (e.g. Extended Survivor Analysis [XSA] (Shepherd 1999), most Adaptive Framework [ADAPT] implementations (Gavaris 1988) and even highly detailed time-series analyses (Gudmundsson 1994). Notably, methods that combine all commercial catch datasets into a single set of catches in numbers at age ignore potential changes in selection due to fleet composition. When all survey information is combined across vessels into a single set of survey indices, this ignores possible vessel effects, which might usefully be removed by appropriate models.

The basic model is a forward simulation model, in which, for a given set of parameters, the population can be projected from initial values to obtain a population trajectory for the time period of interest. Given such a trajectory, values of the various data can be predicted and (negative log) likelihood functions evaluated. The process is repeated with different parameter values until the likelihood function is maximised (negative log is minimised).

Models within Gadget can be fully parametric where the specification of the parameters completely determines the model output. Alternative models abound and one possible exception within Gadget, used here, is to assume the total landings are known without error and are directly removed from the fishable biomass at the beginning of a simulation time step.

When fitting such a model to multiple datasets, several technical questions need to be answered. Some of these are common statistical questions, but others are fisheries-specific.

- Is it possible to estimate appropriate weights to be give to different likelihood components?
- Do initial values of parameters affect the solution?
- Are all parameters (uniquely) estimable?
- Do more likelihood components (more datasets) lead to a ‘better’ solution?
- Should the weight for age-length frequency components be based on the total size of the component or the ‘informative’ size?
- Does the use of age data improve the solution?
- Do subsequent optimisations improve the solution?

The examples on weighting issues given here may appear specific to Gadget and this case study, but the issues are, in reality, much more widely applicable to the general class of non-linear models combining data sources. When combining several datasets in a single objective function, the relative weighting of the datasets has the potential to influence the estimation of parameters. Different datasets can provide different perspectives on the population, especially within the context of a model. A method to estimate the weights objectively is therefore an important tool in the use of models of this type. The approach taken in this paper is based on Stefansson (2003), but an alternative approach might be the use of cross-validation to estimate likelihood weights (Wang and Zidek 2005).

In this paper, the term ‘age-length frequencies’ refers to a frequency table describing the observed number of fish measured of each age and length. Correspondingly, the ‘length distribution’ (or ‘age composition’) refers to the observed number of fish recorded at a certain length (or age).

Gadget

The Gadget modelling environment

Models described here are implemented using the computer program Gadget (Begley 2005), a derivative of Bormicon (Stefansson and Palsson 1997b, Stefansson and Palsson 1998, Björnsson and Sigurdsson 2003) and the subsequent Fleksibest (Guldbrandsen Frøysa *et al*. 2002), which uses concepts developed in earlier work such as the multispecies model for the Barents Sea (MULTSPEC; Tjelmeland and Bogstad 1989) and Multispecies Virtual Population Analysis (MVPA; Helgason and Gislason 1979). Gadget can be used to implement a wide variety of multispecies models but only a few features are included in the present case study.

Models implemented with Gadget are parametric statistical models with parameters that can be estimated using maximum likelihood. Internally, such a model tracks the number of fish by age and length within each area, time step and stock unit. The model ecosystem of all such units is projected deterministically forward in time (simulated), resulting in a single realisation of the development of the model ecosystem. A single such simulation results in a variety of information, including stock trends, mean length, etc., for each stock unit. These simulation results can be compared with data through (negative logs of) likelihood functions. A search algorithm can subsequently be used to estimate values of the unknown parameters.

One or more fleets may remove biomass from the population. In Gadget, a fishing fleet is commonly implemented as a predator, done either through the use of a proportional harvesting model (similar to fishing mortalities) or where the total yield in weight is assumed to be known without error (similar to consumption models). In neither case are catches in numbers at age used nor age compositions assumed to be without error.

Denote by N_{almsrt} the number (N) of fish of species s, age a, length l and maturity stage m, in region r and time step t. This group of fish is a subset of the entire collection of fish of a given species. Such a collection will be referred to as a subpopulation (or substock). Mathematically, this is just a collection of numbers, but they are usually unknown and need to be estimated using statistical techniques.

Normally there are only two maturity stages within the model, immature and mature fish (m = 0,1). In many cases only a fixed species, area, time step and maturity stage are considered and in this case such indices are omitted and the notation is simplified to e.g. N_{al}.

Several processes can affect such a group of fish and cause a change in numbers in the group. The change in numbers can be either due to a transfer of fish between groups or mortality. The various processes available as model components in Gadget are listed briefly here, but a more detailed description of most components used in the present case study is given in the following section. Full descriptions are available in Begley (2005).

Consumption
Predation causes mortality in the prey stock and can be used to determine the growth of the predator. Some of the Gadget implementations of predation are described in the Appendix, but these are only implemented for the purpose of fishing in the case study presented.

Fishing
Fishing can be implemented in several different ways in Gadget (see Appendix). The time step is commonly quite short, permitting a model describing a catch directly proportional to biomass (in place of the common catch equation). Alternatively, the fishing fleet can be modelled as any other predator, and this is the approach used here.

Migration
Migration is not used in the present case study. Within Gadget, the migration a subpopulation undertakes on a given time step can be described by matrices,

$$A_{almst} = (a_{almsr_1r_2t})_{r_2=1,\ldots,R}^{r_1=1,\ldots,R}$$

containing the proportion $a_{almsr_1r_2t}$ of the population of the subpopulation that moves from area r_2 to area r_1 and R is the number of areas. Hence, if $u = (u_1,\ldots,u_R)'$ are abundance numbers by area for an age-length-maturity-species group in a subpopulation, the area distribution after the migration has taken place is $A_{almst}u$.

Maturation
Maturation involves shifting fish from a subpopulation of immature fish to a subpopulation of mature fish. This is done using proportions (as in a migration process), which are designed to mimic the resulting proportion of mature fish in each age-length cell. A typical function describing maturity at age and length in Gadget is

$$M(l,a) = \frac{1}{1+e^{-\psi_1(l-l_{50})-\psi_2(a-a_{50})}} \qquad (1)$$

where ψ_1, ψ_2, l_{50} and a_{50} are parameters to be determined (fixed or estimated), and l_{50} and a_{50} are the length- and age-at-50% maturity respectively. Naturally, several other maturity functions could be used.

Individual growth
Within population dynamics models, growth in length may be implemented through growth in weight because the weight increase is more naturally linked to consumption. In this case, the target weight increase ΔW is a function of consumption and fixing a length-weight relationship of the form $W = aL^b$ can give an approximate average length increase through, for example, a Taylor approximation. Alternatively, as used in the present case study, the length growth is according to a specified functional form (the von Bertalanffy growth curve, see e.g. King 1995) and a length-weight relationship is used to implement growth in weight.

Regardless of which method is used to compute ΔW, the length increase ΔL needs to be implemented through an update mechanism that moves fish between length-groups in such a fashion that the average movement corresponds to a length growth of ΔL. Growth in length is therefore modelled through the use of growth update matrices that are described in a parametric manner (see Appendix).

Natural mortality
In addition to predation/fishing, 'other' natural mortality is implemented in the usual fashion (as in Baranov 1918), viz.

$$N_{a,t+1} = e^{-M_a\Delta t}N_{a,t}$$

Ageing
The last time step of a year involves increasing the age by one year, except for the last age-group which is a 'plus group', containing all ages from the oldest age onwards.

Spawning
A (mature) subpopulation may generate offspring and lose biomass. This can also result in spawning mortality. It follows that Gadget can be used for modelling a closed life-cycle, but in this paper recruitment per year is estimated, as is common practice in fisheries.

Data warehouse

Data input to Gadget is normally through a data warehouse, which is used to store data at a minimally aggregated level (i.e. highly disaggregated), as described in Kupca and Sandbeck (2003) and Kupca (2005). In this case study, data from the Marine Research Institute (MRI) database are aggregated onto the subdivisions set out in Taylor (2005). Aggregations are through simple arithmetic summaries (averages or sums, as appropriate). Input data files and data files for use in the likelihood can be extracted from the data warehouse with the user defining the level of aggregation to be output.

Likelihood components

General Gadget likelihood components
Likelihood components on a negative log scale are used in Gadget to compare the modelled population with observed data. The most commonly used components are based on the Gaussian density (sums-of-squared deviations) but multivariate normal likelihoods are also available as well as multinomial likelihoods.

A typical use of a Gaussian likelihood is to compare survey abundance indices with modelled population numbers. The likelihood component for these data is the sum of squares of a log-linear regression with the slope and intercept estimated or with the slope fixed and only the intercept estimated.

$$\ell = \sum_t \left(\ln\left(I_t\right) - \left(\alpha + \beta \ln\left(N_t\right)\right)\right)^2 \tag{2}$$

where I is the observed survey index and N the corresponding population abundance calculated in the model (both at time t). When the slope β is fixed at $\beta = 1$, the intercept is the standard catchability coefficient. Nonlinearity, however, is frequently observed, and is taken care of by relaxing the restriction on the slope.

Many biological samples are counts data and can be used as proportions. These include length distributions, age compositions, age-length frequency tables, etc. Although it is possible to use (negative log) multinomial likelihoods for counts data, it is more common to use a sum-of-squares likelihood function for proportions:

$$\ell = \sum_t \sum_r \sum_a \sum_l \sum_s \left(\rho_{trals} - \pi_{trals}\right)^2 \tag{3}$$

where ρ is the proportion of the data sample for that time(t)/area(or region r)/age(a)/length(l)/substock(s) combination, and π is the proportion of the model population for that time/area/age/length/substock combination.

Naturally, there may be only a single area, species, etc., in which case the corresponding indices can be omitted.

Penalties

Estimation procedures may test parameter values outside feasible ranges. Some such issues can be alleviated, e.g. by working with recruitment only on log-scale, but in other instances bounds need to be set on parameters. Gadget permits the use of penalty functions and parameter ranges for this purpose.

One example of an infeasible situation is when there is insufficient prey to satisfy the consumption equations for the predators. A particular instance is when the available biomass of a species is less than the catch subsequently taken by the fleet. In this case, the predator is only allowed a scaled version of the required biomass and a penalty is added for this 'understocking' situation.

A likelihood component describing 'understocking' is given by the equation

$$\ell = \sum_t \sum_r \left(\sum_p U_{trp} \right)^g \tag{4}$$

where U is the amount by which the required predation exceeds the biomass available to the predators (understocking) during a model simulation. Normally the power is fixed at g = 2.

Estimation issues

As noted by several authors (e.g. Methot 1989), considerable attention needs to be given to the weights attributed to each negative log-likelihood component. Weights are calculated based on the method proposed by Stefansson (1998) and Stefansson (2003). Initial parameters are chosen arbitrarily, the only constraints being that there should be no understocking (i.e. sufficient fish) and if survey index slopes are to be estimated, the initial slopes are greater than zero (i.e. the relationship between the survey and population is increasing). With these parameters and all component weights set to one, the sum of squares is calculated for each component (the component likelihood score), the inverse of these scores will be referred to as the 'inverse SS'. The total likelihood score with the inverse SS and initial parameters therefore equals the number of components. Results from an optimising run using the inverse SS are compared with those from the iterative reweighting scheme of Stefansson (2003) (referred to as the standard model).

The reweighting scheme requires a separate optimising run for each likelihood component. Each component is taken in sequence, first the inverse SS of the component is multiplied by 10 000, and an optimising run is done to minimise the negative log-likelihood function with the result for this particular negative log-likelihood component taken as a measure of how well the model can best fit to this dataset. The number of terms (degrees of freedom) in the component is then divided into the best (minimum) value and used as a variance estimate, to become (after inversion) the final weight for this component. Whereas the number of terms is easily determinable for some likelihood components, such as survey indices, it is less clear how many should be assigned to data types such as age-length frequencies for which many values are expected to be zero. In these examples, two possibilities are considered for age-length frequencies: the degrees of freedom were either taken to be the full size of the dataset or the number of potentially informative length cells for each year/step/age cell, i.e. the 'informative' size. For all other data types, the total size of the dataset is used. Because there is only one datum for each survey index in each year, given annual recruitment parameters, it is possible for the model to fit perfectly to these components. To circumvent this, the equivalent spring and autumn indices are weighted simultaneously.

Optimisation in the iterative reweighting scheme involves the sequential use of Simulated Annealing and Hooke and Jeeves. Once these have been undertaken, the weights have been estimated. For the subsequent parameter estimation runs, optimisation consists of two optimisation runs, each a combination of Simulated Annealing followed by Hooke and Jeeves. The aim of the first run is to move the parameters into the vicinity of the solution and the second is a more precise run. The two-run approach also allows for a second global search (Simulated Annealing) if the first optimisation moved the solution into a local minimum. For model comparison, the second Hooke and Jeeves run must always find a solution with a step length of 1e–5.

A Case Study

Cod in Icelandic waters

The primary case study in this paper models two stock components of cod, i.e. mature and immature cod, in Icelandic waters which is considered a single area. This study illustrates biological assumptions that need to be made and encompasses several statistical estimation issues.

A 'standard run' is chosen, against which alternatives are compared. This run fixes the slopes in the log-log relationship between indices and abundance to unity and uses the complete iterative reweighting scheme described earlier. Here, the reweighting is based on the use of estimated variances using simply the total number of data points in each dataset. Parameter estimation is started from an arbitrary set of initial values. All likelihood components are used, including age data in the form of age-length frequencies.

A range of alternative model formulations must be evaluated. The purpose of these is to compare a variety of different assumptions and estimation methods. As mentioned above, the slopes can be either fixed to unity or estimated.

Weights can be allocated to the likelihood components in several ways. First, the proposed iterative reweighting scheme can be used, but a simple scheme that uses the inverse of the initial likelihood component scores as weights can also be used. This latter scheme ensures that all initial scores are on a similar scale and is therefore a plausible approach to avoid individual datasets dominating the minimisation. Next, when considering the age-length frequencies, it is noted that these contain many zeros and it is therefore reasonable to reduce the effective number of degrees of freedom to the 'informative' sample size. Starting parameter values can be either from some arbitrary initial values, or from a judiciously chosen starting point from an earlier estimation run. Finally, it is of interest to consider the effect of not using any age information directly, because this corresponds to the common situation where no age readings are available. These various model formulations are described in Table 1.

Further, when using the estimation protocol described earlier (see Estimation issues), several questions need to be addressed: (1) Is it possible to estimate appropriate weights objectively? (2) Do the starting parameters affect the solution? (3) Are age data required in the objective function? (4) Can the power in the survey index be estimated and does it affect the results? (5) Should the weighting of age-length frequency (ALF) components reflect the total component size or the informative size?

In addition, if growth, maturation, selection patterns and the structure of the initial population are considered known, which data types are required in the objective function to estimate recruitment and a scaling multiplier for the initial population? To address this, the parameters estimated from the standard model are used as input values for three optimisation runs with three alternative objective functions: (1) with only survey indices, (2) with indices and length distributions and (3) with indices, length distributions and age-length frequencies.

Data collection

Biological data are collected by the MRI in Reykjavik as a part of a standard monitoring programme and data on landings are collected by the Directorate of Fisheries. In both cases, all data are stored in Oracle databases. For ease of extraction for use in Gadget, these data have been aggregated and stored in the data warehouse described earlier. The data of interest to the current case study are:

- Landings data by species, fleet and month
- Age, length and maturity measurements from survey and biological sampling of the catch
- Survey indices

Data and likelihood functions

As indicated in Stefansson (1998), a statistical multispecies model can be based on a variety of different data summaries, but care must be taken not to use the same information twice, e.g. both through mean length at age and as length distributions with age-length frequencies. These data should also be direct observations rather than processed or modelled data.

It is important to aggregate the data to a level at which coverage of the selected length/age interval is adequate, as zeros are taken to mean zero. Time steps with low sampling or sampling restricted to a limited range of length-/age-groups may result in zeros that are an artefact of sampling and unrepresentative of the population. This is true for all

Table 1: Conditions in the standard model and the alternatives (Alts). The slopes of the regressions can either be fixed (F) or estimated (E) for Length-groups 1 and 2; weights can be inverse (I), using all weights with the age-length frequency (ALF) scaled by the total size (T), using all weights with the ALF scaled by the informative size (R) or with the ALF weight set to zero (0); the initial parameters can be arbitrary (A), optimised values from the inverse SS run (B), optimised values from the standard run (C), optimised values from the inverse SS run for Alt1 (D) (i.e. the equivalent model with the slopes fixed) or arbitrary with the growth parameter k fixed (G); and the components can be all (L) or with no age data (K)

Parameter	Standard	Inverse	Alt 1	Alt 2	Alt 3	Alt 4	Alt 5	Alt 6	Alt 7	Alt 8
Slopes	F	F	F	E	E	F	F	E	E	E
Weight	T	I	R	T	R	0	0	T	R	T
Initial	A	A	A	A	A	A	G	B	D	C
Components	L	L	L	L	L	K	K	L	L	L

likelihood components. Another consideration is that using a finer resolution can result in noisier data.

The data used in the current analyses consist of measurements from commercial catches and two groundfish surveys. Not all months have measurements for all fleets and, in particular, the groundfish surveys each occur only in a single month. Each of these datasets is linked to the model through a likelihood component. Given the problems inherent in mixing likelihood components corresponding to different probability distributions and, in particular, the known problems with a multinomial distribution being inadequate for length distributions (Hrafnkelsson and Stefansson 2004), only normal distributions are used and each negative log-likelihood component is simply a (weighted) sum of squares. For a particular data type, such a sum of squares will be referred to as a negative log-likelihood component or 'likelihood component' when unambiguous. The observed data used in the likelihood components will be referred to as 'likelihood data'.

Data used in the case study are:

Input data

- Landings data obtained from the database of the Directorate of Fisheries

Biological sampling likelihood data

- Length distributions, aggregated on 2cm intervals from: sea and harbour sampling of commercial catches by month from 1984 to 2003; the spring groundfish survey (Palsson *et al*. 1989) from 1985 to 2003; the autumn groundfish survey (Sigurdsson *et al*. 1997) from 1995 to 2003.
- Age-length frequencies, aggregated on 4cm intervals from: commercial catches by month from 1984 to 2003; the spring groundfish survey from 1989 to 2003; the autumn groundfish survey from 1995 to 2003. Age data from the spring survey have only been used from years with random otolith sampling.
- Age compositions from the spring groundfish survey from 1989 to 2003 and from the autumn groundfish survey from 1995 to 2003. Age data from the spring survey have only been used from years with random otolith sampling.

These likelihood data are of the same class — biological sampling — with all ages aggregated for the length distributions and all lengths aggregated for the age compositions. The age-length frequencies and age compositions are used directly because otolith sampling is random (not stratified by length). The age compositions are not used for parameter optimisation, but the fit of the model to these data is used for model comparison. The sum-of-squares likelihood function is given by Equation 3.

Stock composition likelihood data

- Proportion mature at length in 2cm length-classes from the spring groundfish survey (Palsson *et al*. 1989) from 1985 to 2003.

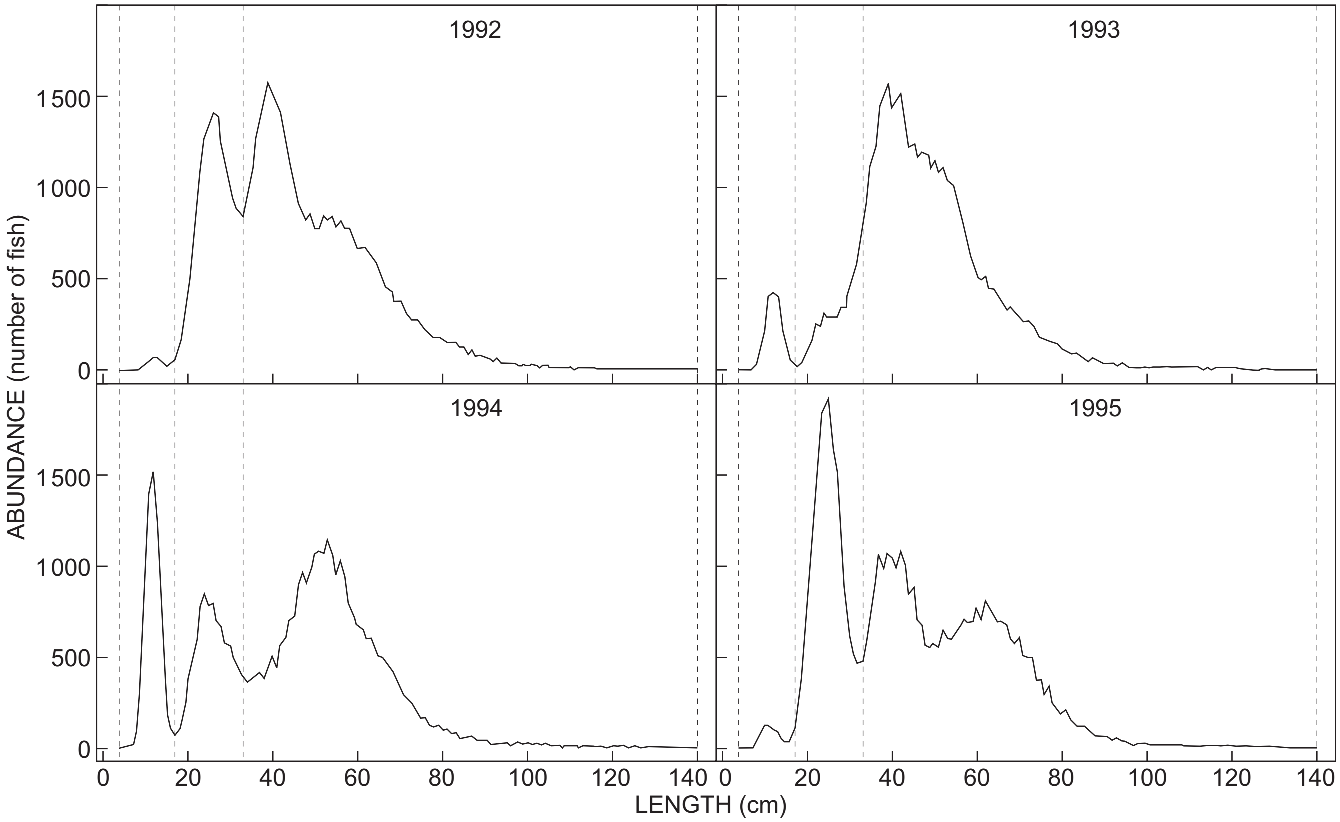

Figure 1: Length distributions from the spring groundfish survey, 1992–1995, with vertical lines indicating the length-groups within which the data are aggregated to calculate the survey indices

This groundfish survey is undertaken at a time when the maturity stage is easily determined and is unaffected by the potential biases owing to targeting of the mature stock component by the commercial fleet. Although the maturation process in Gadget can be age and length dependent, in this case study only length is taken into account and the data are aggregated over all ages.

The sum-of-squares likelihood function for this component is given by Equation 3.

Survey index likelihood data

Three indices are calculated from each survey by splitting the length distribution into three groups, with the division based on the 'typical' structure of the length distribution for the survey over all years. These length-groups represent Age 1, Age 2 and Age 3 and older. Figure 1 illustrates the length-groups over which data are aggregated for the spring survey for a selection of years.

Age data are not used as indices as there is greater potential for age sampling to introduce sampling bias. Whereas calculating number at age by combining age and length data might reduce sampling bias, it involves processing likelihood data. Disaggregation of the length distribution does not need to be based on an assumed age structure but equivalent indices allow data from the two surveys to be linked. This is explained more fully earlier (see Estimation issues). An additional consideration is that disaggregating indices by length rather than age allows more flexibility when modelling populations where ages are poorly determined or unavailable.

Two sets of indices are calculated:

- Three survey indices from the spring survey (Palsson *et al*. 1989) (Figure 2) with one datum for each group for the years 1985–2003.
- Three survey indices from the autumn survey (Sigurdsson *et al*. 1997) (Figure 2) with one datum for each group for the years 1996–2003. The first year of the survey contained fewer stations and is not included in the index.

In this case study, Equation 2 is used with the slope always fixed to 1 for the third length-group, because for fish of this length the survey index is considered to be linearly related to the population abundance. When $\beta = 1$, the intercept has the interpretation of log-catchability.

In Figures 2 and 3 it can be seen that for Length-groups 1 and 2, the surveys provide similar information for the overlapping years (correlations of 0.86 and 0.94 for Length-groups 1 and 2 respectively), but if the entire time-series is considered, there is conflicting information from the third length-group (correlation of –0.10). The Length-group 3 data are in two groups, within both of which there is a log-linear relationship between the surveys (i.e. 1994–1997 plus 1999 and 2000–2002 plus 1998). This may indicate a change over time of the relative catchability of larger (or older) fish in the spring and autumn surveys.

For both the spring and autumn surveys, there is a strong positive relationship between Length-group 1 and 2 indices, as indicated in Figure 4, with correlations of 0.90 and 0.86 for the spring and autumn surveys respectively. These results support the assumption that the Group 1 and 2 indices represent cohorts.

It should be noted that a particularly simple sum of squares is used as the negative log-likelihood in all cases. There are several reasons for this, all somewhat technical. The use of sums of squares in this manner is numerically equivalent to assuming that the corresponding (possibly transformed) data are Gaussian. It is known that considerable deviations from this assumption are tolerated, still giving reasonable point estimates. However, this is not a general rule. Thus, although a multinomial distribution may appear a more reasonable assumption for length distributions (MacDonald and Pitcher 1979), it is well known that this assumption is seriously flawed, not only through over-dispersion (MacCall 2003) but also in correlation structure (Hrafnkelsson and Stefansson 2002). The net effect of

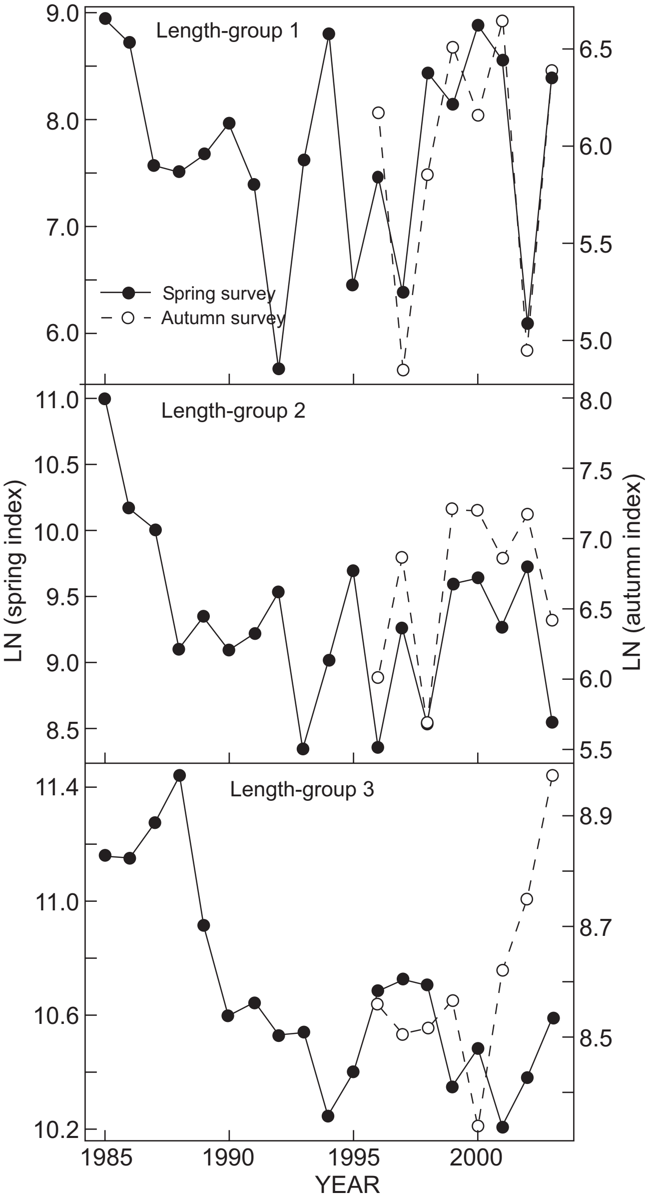

Figure 2: Survey indices for the spring and autumn surveys for each length-group

these incorrect assumptions is unknown and hence simple sums of squares are used instead in this paper. It is clear that these methods may be improved (e.g. by using multivariate Gaussian assumptions, cf. Elvarsson 2005), and this is a promising route of future work.

Parameter settings and initial values

The model runs from 1984 to 2003 on one area with two cod components — immature and mature. The immature are aged 1–10 and the mature 3–12, with the final age being a plus group. Immature fish mature according to the maturation function and any not mature by the end of their 10th year move into the mature component. A single commercial fleet operates along with two surveys.

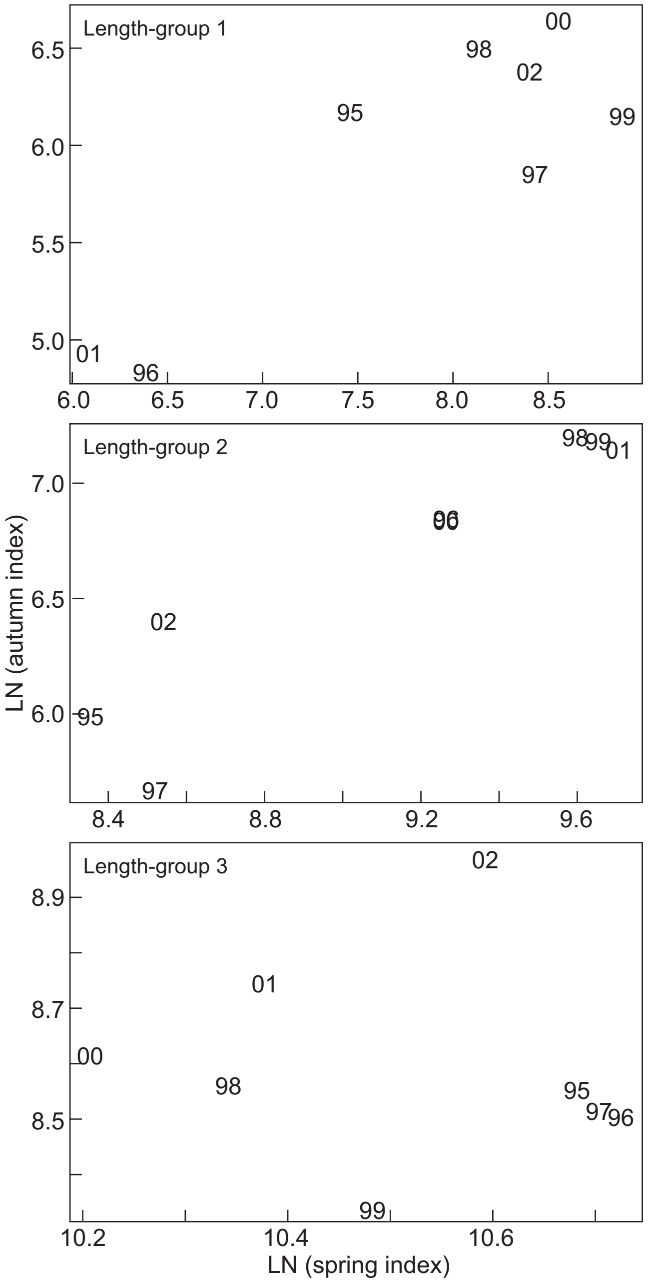

Figure 3: Scatter plots of the autumn vs the spring survey indices for each length-group on a log scale. Each point is indicated by the year of the survey

Length growth is defined by Equation App.2 with L_∞ fixed and k estimated within the model. The beta-binomial parameter β is fixed to 1 000 with maximum length-group growth n = 10. Weight growth is according to Equation App.3 with different length-weight relationships for the immature and mature components. The values of c and b were calculated from survey data and subsequently fixed.

The maturation process is based on tracking the maturity function given in Equation 1. As maturation in this case study is only a function of length, the age parameters ψ_2 and a_{50} are set to zero; l_{50}, the length-at-50% maturity, and the rate ψ_1 are estimated within the model.

The number at age in the initial population (Ages 2–11) is estimated within the model. To reduce the number of parameters there are no fish in the Age 12+ group in the first year. In some examples, the age structure of the initial population is fixed and an estimated multiplier scales the population abundance. The mean and standard deviation of length of the initial populations were calculated from the spring survey and output from the data warehouse. The length-weight relationships used for weight growth were also used for the initial population.

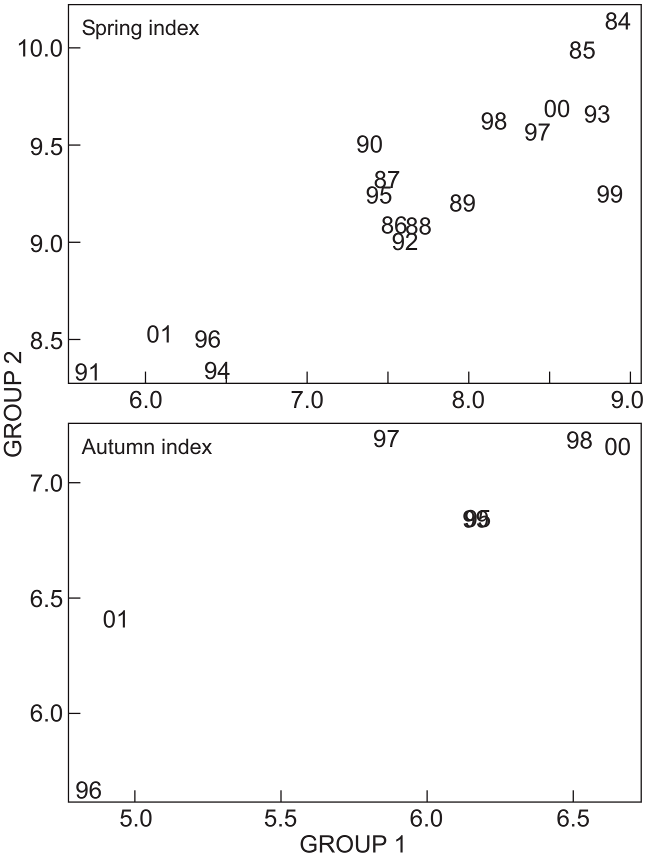

Figure 4: Scatter plots of the Length-group 2 index against the corresponding Length-group 1 index for each survey on a log scale. Each point is labelled with the cohort year

The number of Age 1 recruits for each year (1984–2003) is estimated within the model. Mean and standard deviation of length of the Age 1 recruits is calculated from the spring survey for each year and output from the data warehouse, with the mean length for 1985–1989 used as the mean length for 1984.

Natural mortality at age is given by a fixed vector: 0.5 and 0.35 for Ages 1 and 2, 0.2 for Ages 3–9 and 0.3, 0.5 and 0.7 for Ages 10–12.

Selection by the surveys and commercial catch are described by the suitability function in Equation App.6, with $\phi_3 = 0$, $\phi_4 = 1$ and ϕ_1 and ϕ_2 estimated for each fleet separately.

Results

Parameter estimation and model evaluation of the standard model are discussed in some detail, followed by comparisons of the alternative models.

Standard model

The minimum sums of squares from the iterative reweighting scheme described earlier (see Estimation issues) are given in Table 2 for all negative log-likelihood components. All the models with all survey regression slopes fixed to 1 are based on these initial runs.

Table 2 illustrates the conflicting information, within this Gadget implementation, from the different data sources. For example, the minimum sum of squares of error (sse) from weighting the spring survey length distribution (LD$_s$) is 0.045, whereas the same component has a score of 0.154 (a value 3.4 times greater) when the catch length distribution is heavily weighted. In general, within the same type of component (e.g. length distribution) there is greater consistency between the surveys than between either survey and the catch. Weighting the spring survey

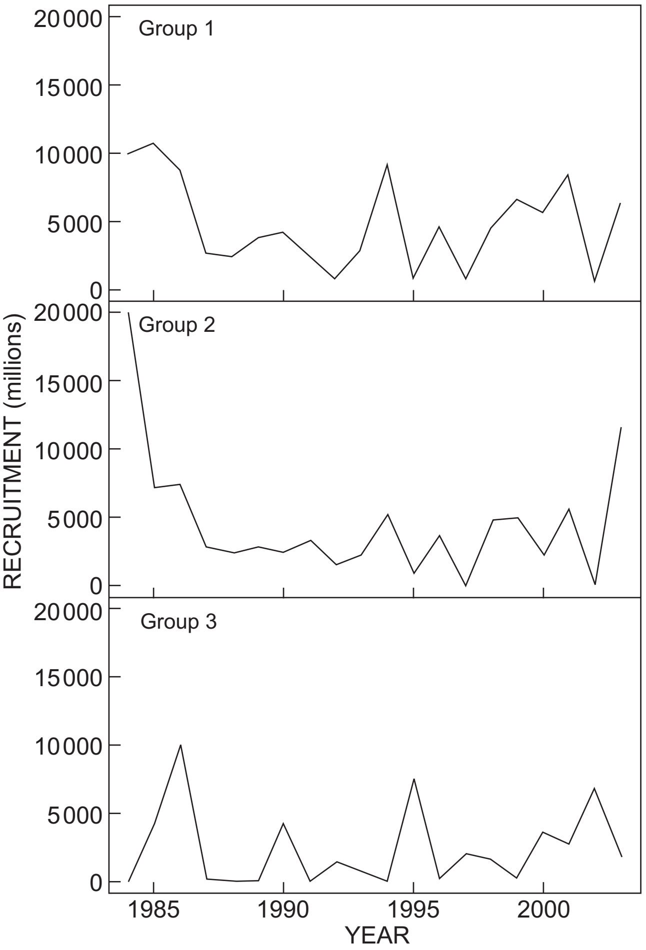

Figure 5: Recruitment estimated from the iterative reweighting procedure with Length-groups 1–3 of the survey data heavily weighted

Table 2: Minimum sums of squares from iterative optimisations with each column representing a run with that component heavily weighted. The values in bold are the sums of squares of the heavily weighted component in each optimisation. These components are from the biological sampling likelihood components (LD — length distribution, ALF — age-length frequencies, AC — age composition, with s, c and a indicating whether spring survey, catch sampling or autumn survey respectively), the ratio of immature to mature fish (mat) and the survey indices (I). For the indices, I$_i$, i = 1,2,3 are the length-groups and the s and a subscripts indicate the spring and autumn surveys. The corresponding length-groups of the surveys are weighted simultaneously

Component	Mat	LD$_s$	LD$_c$	LD$_a$	ALF$_s$	ALF$_c$	ALF$_a$	I$_1$	I$_2$	I$_3$
Mat	**32.47**	65.73	37.17	78.84	41.94	44.68	42.06	36.71	34.96	47.51
LD$_s$	0.64	**0.04**	0.15	0.36	0.16	0.13	0.40	0.18	0.27	0.74
LD$_c$	12.84	4.49	**0.89**	12.06	7.98	1.94	9.63	10.05	5.53	12.28
LD$_a$	0.30	0.03	0.04	**0.02**	0.03	0.04	0.02	0.09	0.08	0.23
ALF$_s$	0.75	0.15	0.60	0.44	**0.06**	0.10	0.18	0.32	0.20	1.13
ALF$_c$	19.68	6.78	13.72	13.86	6.35	**2.85**	7.47	10.12	5.79	25.30
ALF$_a$	0.49	0.07	0.40	0.26	0.04	0.06	**0.03**	0.23	0.07	0.67
AC$_s$	2.60	0.36	1.12	1.32	0.09	0.31	0.91	0.69	0.91	3.57
AC$_c$	68.85	19.12	38.23	71.17	27.06	5.60	34.84	22.57	23.12	106.60
AC$_a$	1.90	0.07	0.66	0.51	0.05	0.15	0.04	0.37	0.28	2.32
I$_{1s}$	77.85	15.41	19.49	56.01	24.46	23.84	40.46	**1.09**	32.15	91.44
I$_{2s}$	43.84	16.48	17.84	41.18	28.41	27.70	41.57	2.08	**0.03**	40.95
I$_{3s}$	3.93	15.05	14.80	11.11	20.90	29.09	12.83	0.75	1.49	**0.35**
I$_{1a}$	10.13	2.16	1.43	1.50	2.02	1.68	1.90	**0.08**	9.57	19.07
I$_{2a}$	12.46	1.70	1.30	1.35	1.56	1.78	1.66	0.88	**0.03**	6.36
I$_{3a}$	1.01	1.26	0.37	1.77	1.74	1.39	2.12	0.06	0.16	**0.05**

does not result in a large increase in the sse for the corresponding autumn survey component (0.029 compared with 0.023). Because the autumn survey does not cover the entire time period of the spring survey, the reverse comparison is invalid.

Table 3: Component weights: the inverse minimum sums of squares, the scaled weights calculated from the iterative reweighting procedure and the ratio of the sum of squares from the optimised model (sse_f) to the minimum from the weighting run (sse_m). Abbreviatons as in Table 1

Component	Inverse SS	Scaled weights	sse_f:sse_m
Mat	0.027	51	1.08
LD_s	3.256	28 930	2.26
LD_c	0.220	14 790	1.33
LD_a	10.049	24 658	0.90
ALF_s	5.411	107 860	1.24
ALF_c	0.223	17 538	1.07
ALF_a	9.328	112 387	1.07
I_{1s}	0.070	17	5.07
I_{2s}	0.850	54	1.94
I_{3s}	0.360	293	34.02
I_{1a}	0.206	553	67.03
I_{2a}	0.291	97	6.46
I_{3a}	3.685	157	2.54

The difference in information deriving from the different components is particularly large for the survey index data. With the emphasis on Length-group 1 or 2, the fit to Length-group 3 is relatively good. But there are clear inconsistencies between Length-groups 1 and 2.

Being a plus group extending over at least 10 age-groups, Length-group 3 is less informative with respect to the recruitment pattern. This results in the scores for Length-groups 1 and 2 being particularly high when the emphasis is on Group 3. Figure 5 illustrates the difference in predicted recruitment when emphasis is put on the different survey length-groups. Groups 1 and 2 display similarities (with a correlation of 0.73), but recruitment predicted from Group 3 (the plus group) is unsurprisingly quite different. It should be noted that Length-group 1 provides no information for 1984 because the first survey datum is for 1985.

The initial inverse SS and the weights determined from the iterative procedure are given in Table 3 along with the ratio of the sum of squares from the final model to the minimum for each component. As a result of incorporating information from the different data sources, the final sse for each component is typically greater than the minimum from the iterative reweighting.

The estimated parameters are shown in Figure 6 with plots of the fitted log-linear regression in Figure 7. With a

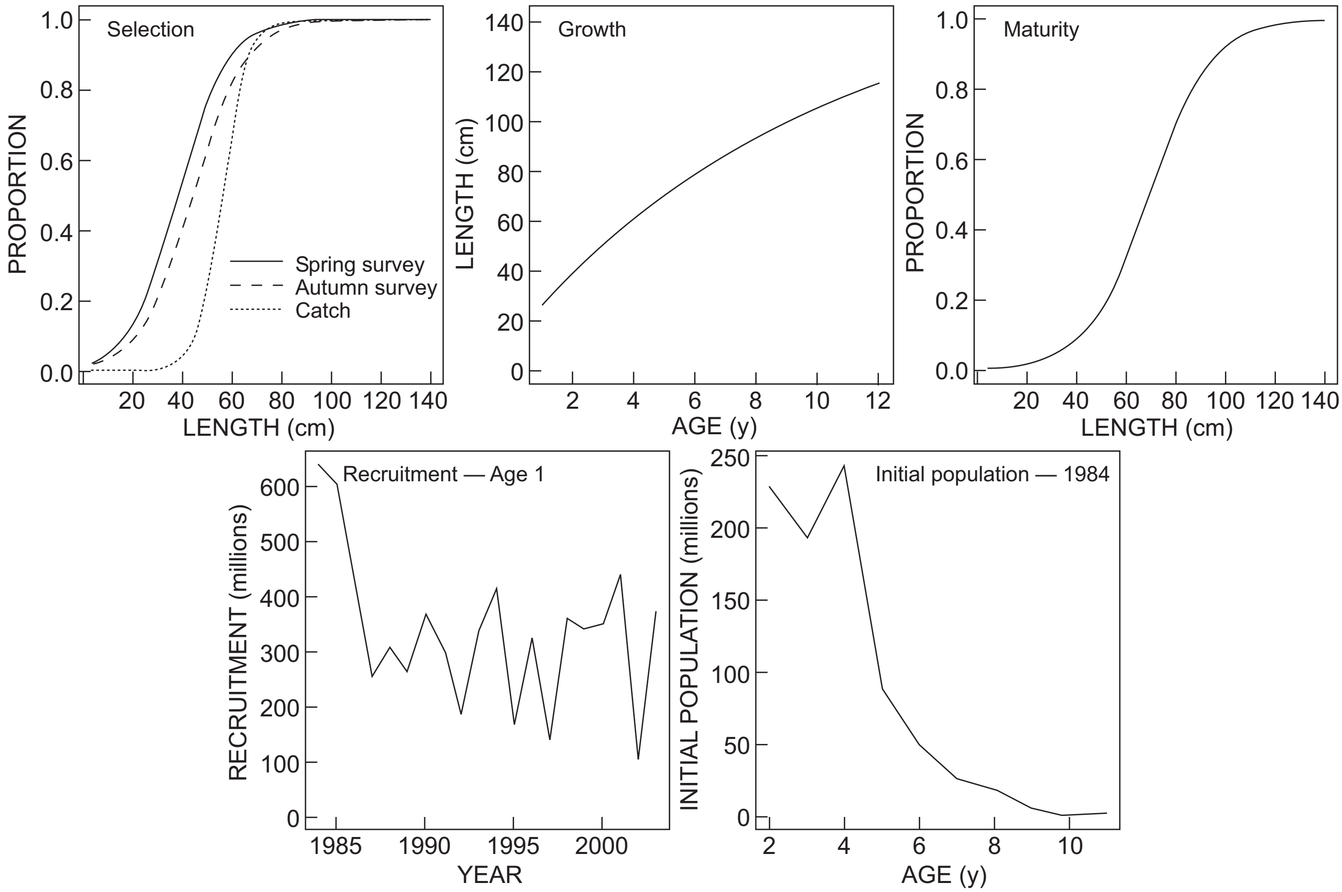

Figure 6: Parameter estimates from the final optimisation: selection patterns of the spring survey, autumn survey and catch, growth, maturation, number of Age 1 recruiting into the model and the initial population, Ages 2–11

linear relationship assumed between the indices and population abundance, the catchability (i.e. the intercept) of the length-group increases with length for both surveys. In addition, catchability is lower for the autumn survey, which has fewer stations, than the spring survey.

Evaluating the fit of the model to the data

An advantage of a statistical model is that the fit of the model to the data can be assessed in detail. The negative log-likelihood score by year, step and component can be plotted, as in Figure 8 for the spring survey data and Figure 9 for the commercial catch data, to determine whether there are any patterns, e.g. trends in time. The use of autumn survey data from 1995 onwards, given the degree of consistency between data from the spring and autumn surveys, may contribute to the downward trend in the survey age-length frequency and age composition scores (Figure 8). On the other hand, apart from the very high score for 1985, there is an increasing trend for the length distribution residuals. The high scores for the spring survey age data, and to a lesser extent the catch age data in 1990 and 1991, indicate problems in modelling the age structure in these years, although the corresponding

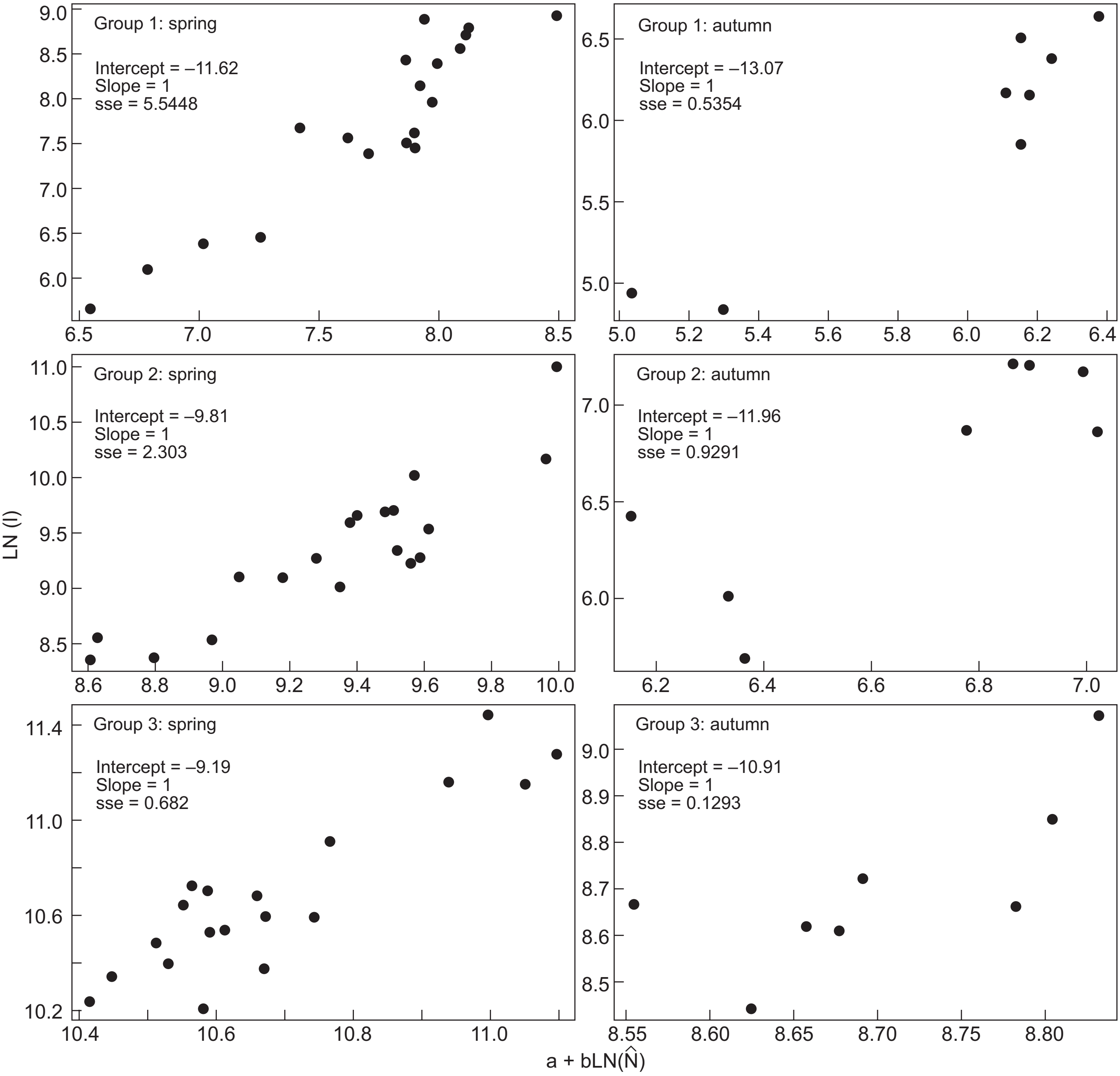

Figure 7: Plots of the log-linear regression as estimated in Gadget, with one plot for each negative log-likelihood component and the corresponding slope (fixed to 1 in this implementation) and sse for each survey index. Ln(I) is the index and a + b[ln($\hat{N}$)] the fitted regression to the modelled population

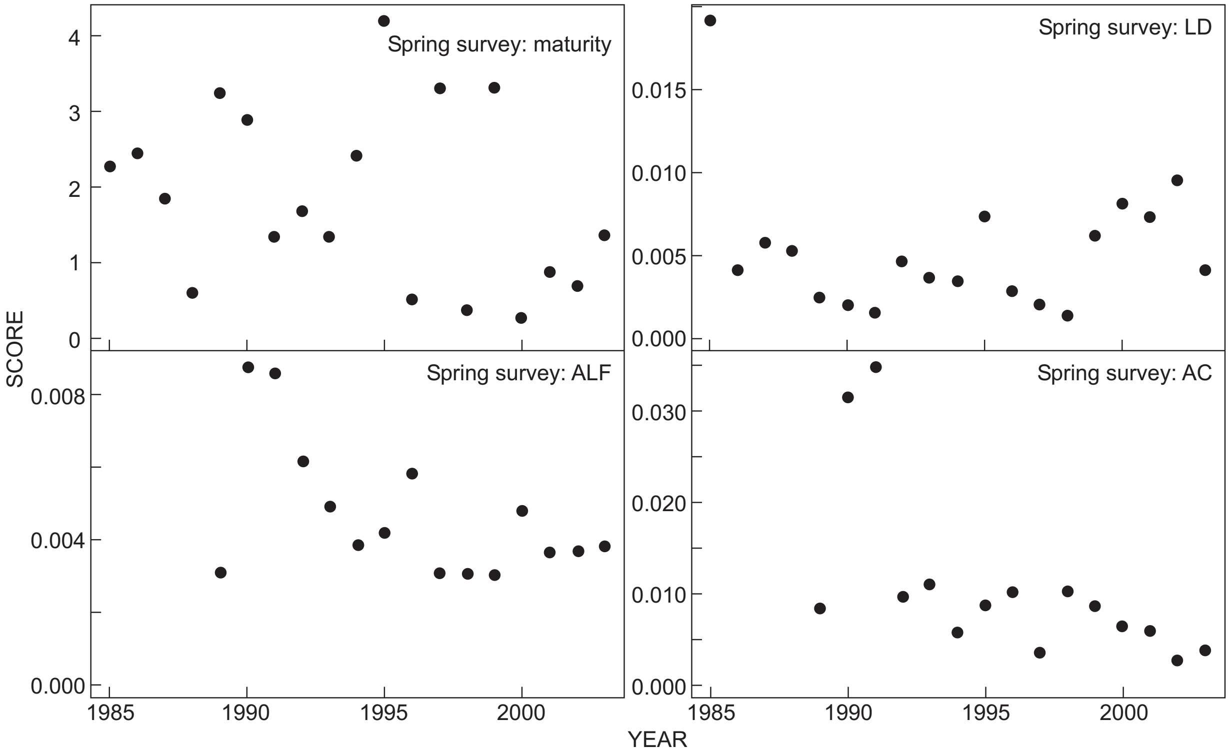

Figure 8: Likelihood component scores by year for the spring survey components: maturity, length distribution (LD), age-length frequency (ALF) and age composition (AC). It should be noted that the age composition data were not used in parameter estimation

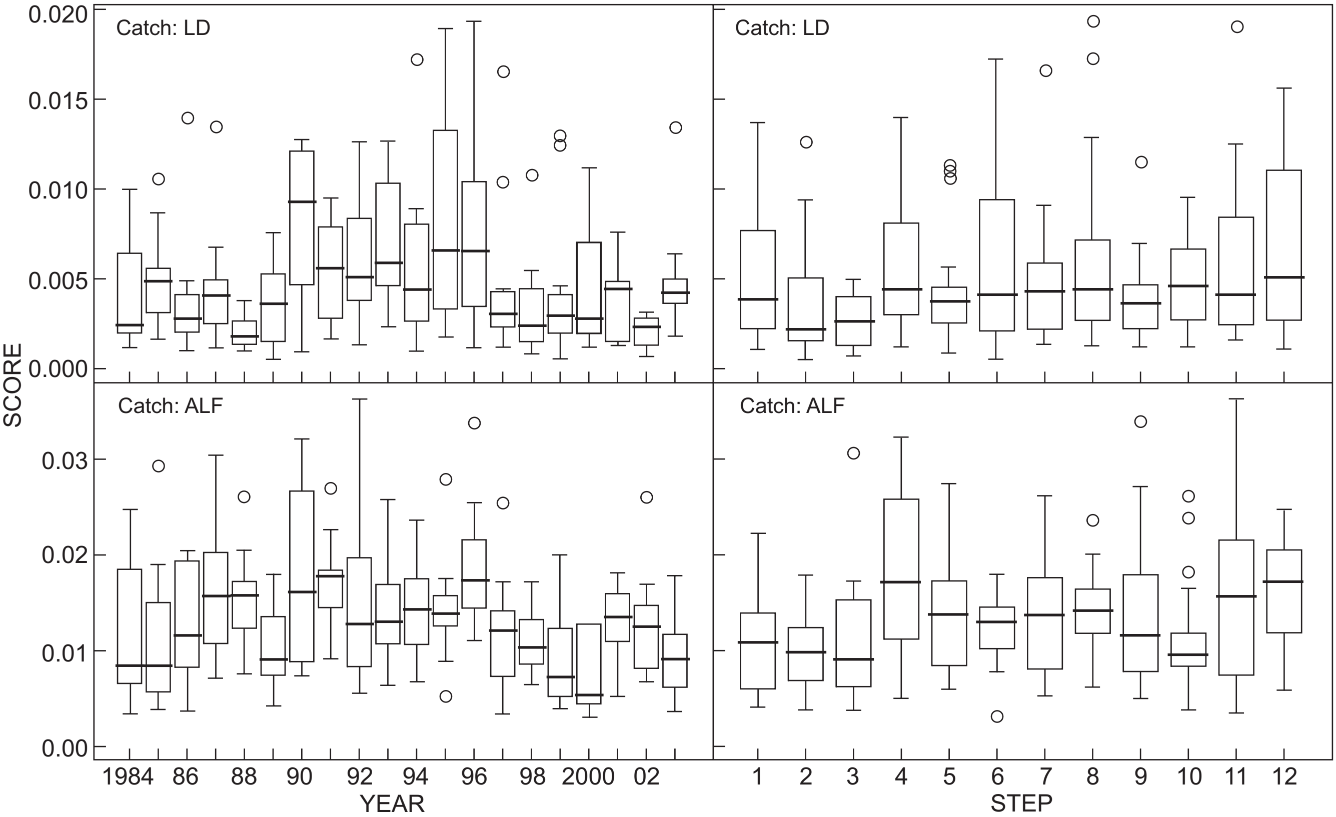

Figure 9: Boxplots of the negative log-likelihood scores by year and step for the commercial catch components: length distribution (LD) and age-length frequency (ALF)

survey length distribution scores are relatively low. In some years there is larval drift of cod from Icelandic waters to Greenland, with migration back to Iceland several years later. There is considered to have been a substantial migration of fish from the 1984 year-class back into Icelandic waters from Greenland in 1990 and 1991 (Shepherd and Pope 1993). The large amount of immigration in 1990 and 1991 could account for the relatively poor fit of the model to the age-length frequency and age-composition data for these years, especially as the cod migrating from Greenlandic waters are smaller at age than those resident in Icelandic waters. The lessening importance of this period of immigration over time might contribute to the decreasing trend in the likelihood component score over the succeeding years. It is possible to include immigration of this type in the model, but it has not been considered in this simple case study.

The model does not fit the catch likelihood data equally well for all years, with higher negative log-likelihood scores for 1990–1996 for the length distribution (Figure 9). There is, however, no pattern in the age-length frequency or any strong season trend.

In addition to these summary scores, the residuals by length can also be calculated, e.g. for the length distribution using ($p_{y,s,l} - \pi_{y,s,l}$) where $p_{y,s,l}$ is the observed length distribution as proportions by year and step and $\pi_{y,s,l}$ the modelled catch length distribution as proportions by year and step.

This implementation of Gadget may be insufficiently flexible to fully describe the length structure of the population, as observed by the surveys. Figure 10 indicates that the model tends to underpredict the number of fish between 20cm and 30cm and overpredict the number of fish over 80cm. This pattern is also influenced by the likelihood function not taking the number of data in each length-class into account. It is, however, to be expected that the length distributions from the modelled population will be smoother than those observed because samples are affected by intra-haul correlation (Pennington and Volstad 1994), with fish of a similar length being caught together (Hrafnkelsson and Stefansson 2004). A model with disaggregated fleets, cannibalism, or with growth affected by consumption might improve the fit to the length distribution data, as might variable growth rates for different years or time periods.

Model comparison

Models can be compared using the fit of the modelled population to the observed data, e.g. through the total negative log-likelihood scores for each component as shown in Figure 11. A direct comparison is, however, only statistically valid when the parameters estimated are the same for all models. Figure 11 shows that Alternatives 6 and 7 give a considerable improvement in terms of the fit to the survey indices for Ages 1 and 2. This is not surprising because Alternatives 6 and 7 estimate the slopes of the regression lines through these two datasets and thus have more flexibility to accommodate these specific data. Interestingly, the improvement in fit to these data has a negative effect on the fit to several other datasets, in particular the age compositions (AC_s, AC_c and AC_a, none of which are used in the fitting procedure).

Although diagnostic plots are useful and important, it is also important to consider the difference in the population dynamics as predicted by these models, e.g. as given by the abundance or biomass (Figure 12). These plots give one example of how these models differ from common Virtual Population Analysis (VPA)-based models (Gulland 1965) in that the backwards convergence (Pope 1972,

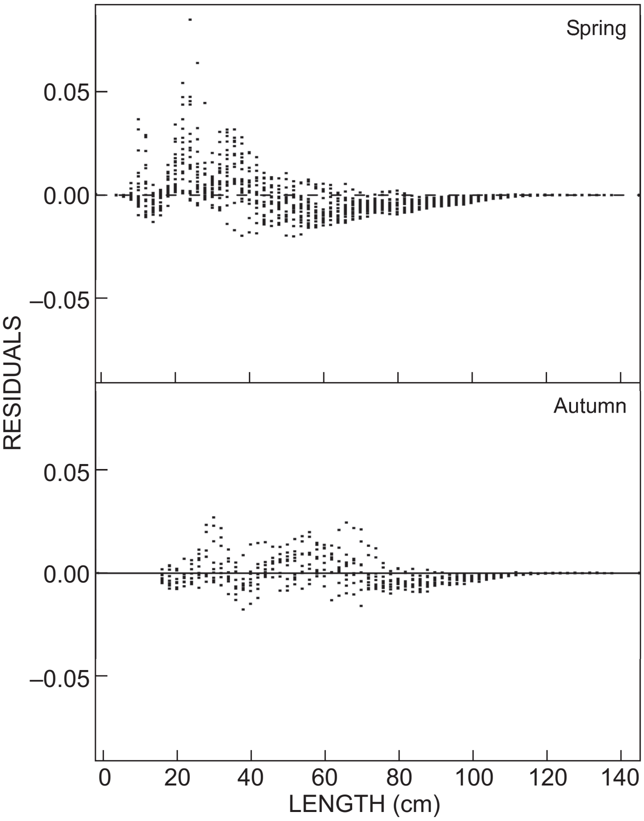

Figure 10: Length distribution residuals, by length-group, from the spring and autumn surveys

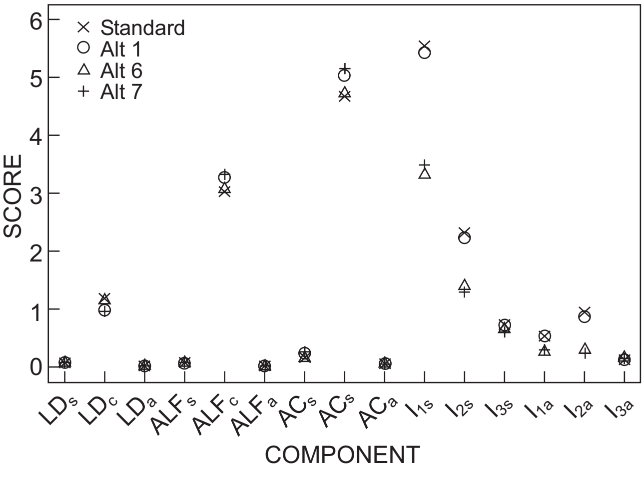

Figure 11: The negative log-likelihood scores (sums of squares) for four models: the standard model, and Alternatives (Alt) 1, 6 and 7. See Table 1 for component names

Ulltang 1977) is not seen here. Rather, differences in model behaviour appear as shifts throughout the entire time period. Thus, the standard model consistently gives the highest biomass level whereas Alternative 7 is consistently lower (Figure 12).

Results from the iterative reweighting scheme and direct use of inverse sums of squares

If the negative log-likelihood scores from the inverse SS and standard runs are compared (Figure 13), the inverse SS model is a better fit to the survey index data but the standard run is a better fit to most of the sampling data, with the greatest deviations for the largest components. This reflects the difference in weighting the components using the inverse SS alone or by taking the size of the dataset into account as use of the inverse SS results in smaller datasets having a disproportionately large influence. The relative difference in the weights of the inverse SS and standard models can be seen in Table 3.

Weighting the age-length frequency

Two comparisons are made of the degrees of freedom used to weight the age-length frequencies: the standard model with Alternative 1, in which the survey power is fixed to 1 for all length-groups, and Alternatives 6 and 7 in which the power is only fixed for Length-group 3. In both comparisons the fit of the model to the indices (Figure 11) is similar, but unsurprisingly the models with more weight on the age-length frequencies result in a better fit to the commercial catch components containing age data. The trade off is that the fit of the model to the length distributions is slightly worse.

Total biomass estimated by Alternative 1 is 5.6% less than for the standard model and the mean spawning stock biomass is 11% less (Figure 12), with the mean percentage difference in abundance of mature fish being 6.2%. The difference between the biomass estimated by Alternatives 6 and 7 displays a similar pattern to that between the standard model and Alternative 1.

Estimation of the power in the survey index

The full iterative reweighting procedure was applied to five models with the power of the index of survey Length-groups 1 and 2 estimated. Three starting parameters types were used: arbitrary values (Alternatives 2 and 3), values estimated from the inverse SS run (Alternatives 6 and 7) for the equivalent model with the power term fixed and the final values estimated from the standard model (Alternative 8). The survey index minimum sums of squares from the iterative reweighting scheme for the standard model and Alternative 6 show that the greater flexibility of the model reduces the inconsistencies between the survey length-groups (Table 4).

Starting the procedure from the arbitrary parameter set (Alternatives 2 and 3) fails to find a reasonable solution.

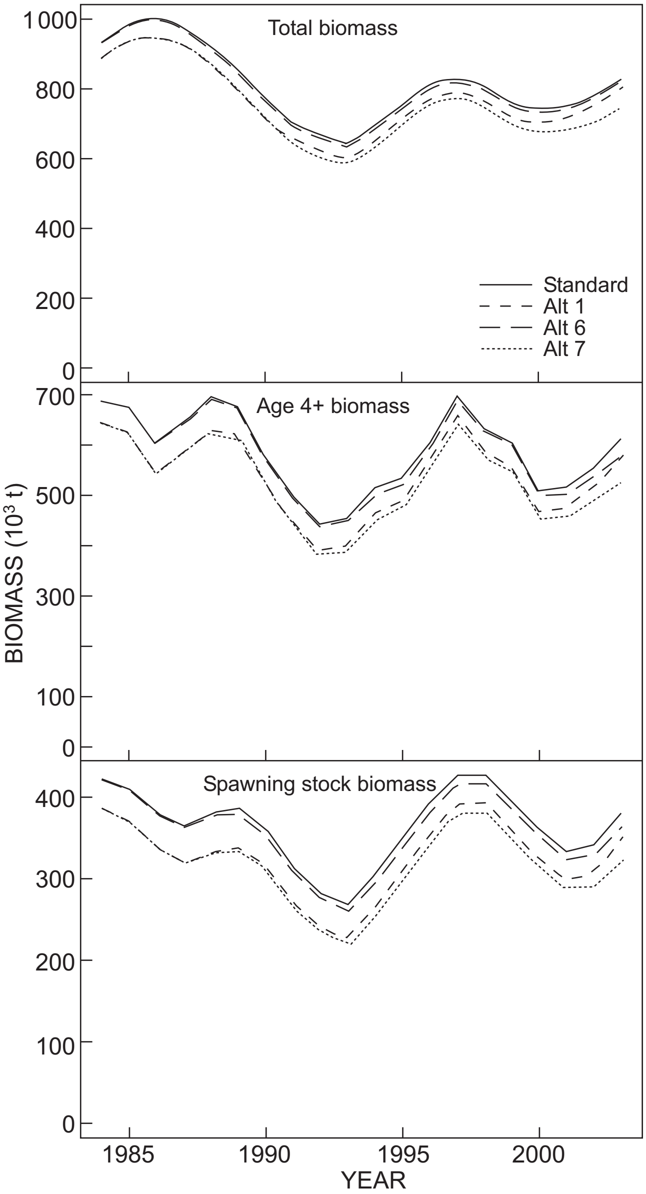

Figure 12: End of year biomass trajectories for four models: standard, Alternatives 1, 6 and 7 for: total (Age 1+) biomass, Age 4+ biomass (considered the fishable biomass) and spawning stock biomass

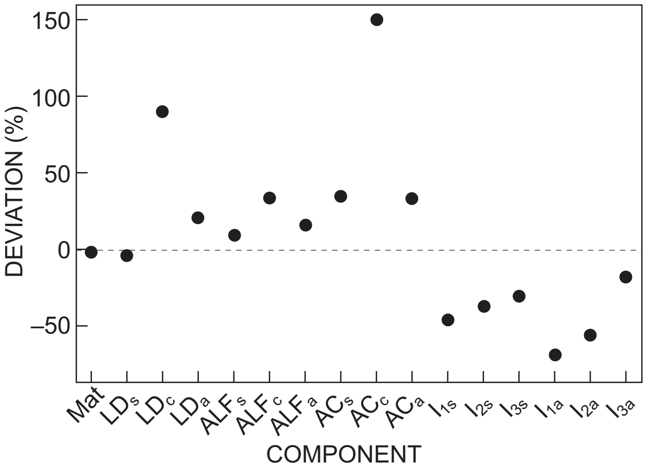

Figure 13: The difference in negative log-likelihood scores by component between the standard and inverse SS models, expressed as a percentage of the standard model [(inverse – standard)/standard] x 100. See Table 1 for component names

From Alternative 2, the fit of the model to the spring survey indices is very poor (Figures 14 and 15). The relationship between Length-group 3 and the model is clearly negative despite the slope being fixed to 1, and for Length-group 2 the slope is approximately 0 with positive indices. The fit of the model to the autumn survey is considerably better (Figure 14). Given arbitrary starting values the less constrained models have more difficulty finding a solution for those years with only one survey than the equivalent models with the slopes of the indices fixed.

Conversely, the optimised models with non-arbitrary starting parameters optimise to solutions that fit the survey

Table 4: Minimum sums of squares from iterative optimisations for the standard run and Alternative 6, with each column representing a run with that component heavily weighted. The values in bold are the sums of squares of the heaviliy weighted component in each optimisation. Abbreviations as in Table 1

Minimum	Standard fixed power			Alternative 6 estimated power (1 and 2)		
sums of squares	I_1	I_2	I_3	I_1	I_2	I_3
I1s	**1.09**	32.15	91.44	**1.31**	11.12	15.23
I2s	2.08	**0.03**	40.95	4.74	**0.06**	7.26
I3s	0.75	1.49	**0.35**	2.38	1.08	**0.34**
I1a	**0.08**	9.57	19.07	**0.02**	2.11	3.33
I2a	0.88	**0.03**	6.36	2.38	**0.01**	2.38
I3a	0.06	0.16	**0.05**	0.49	0.17	**0.02**

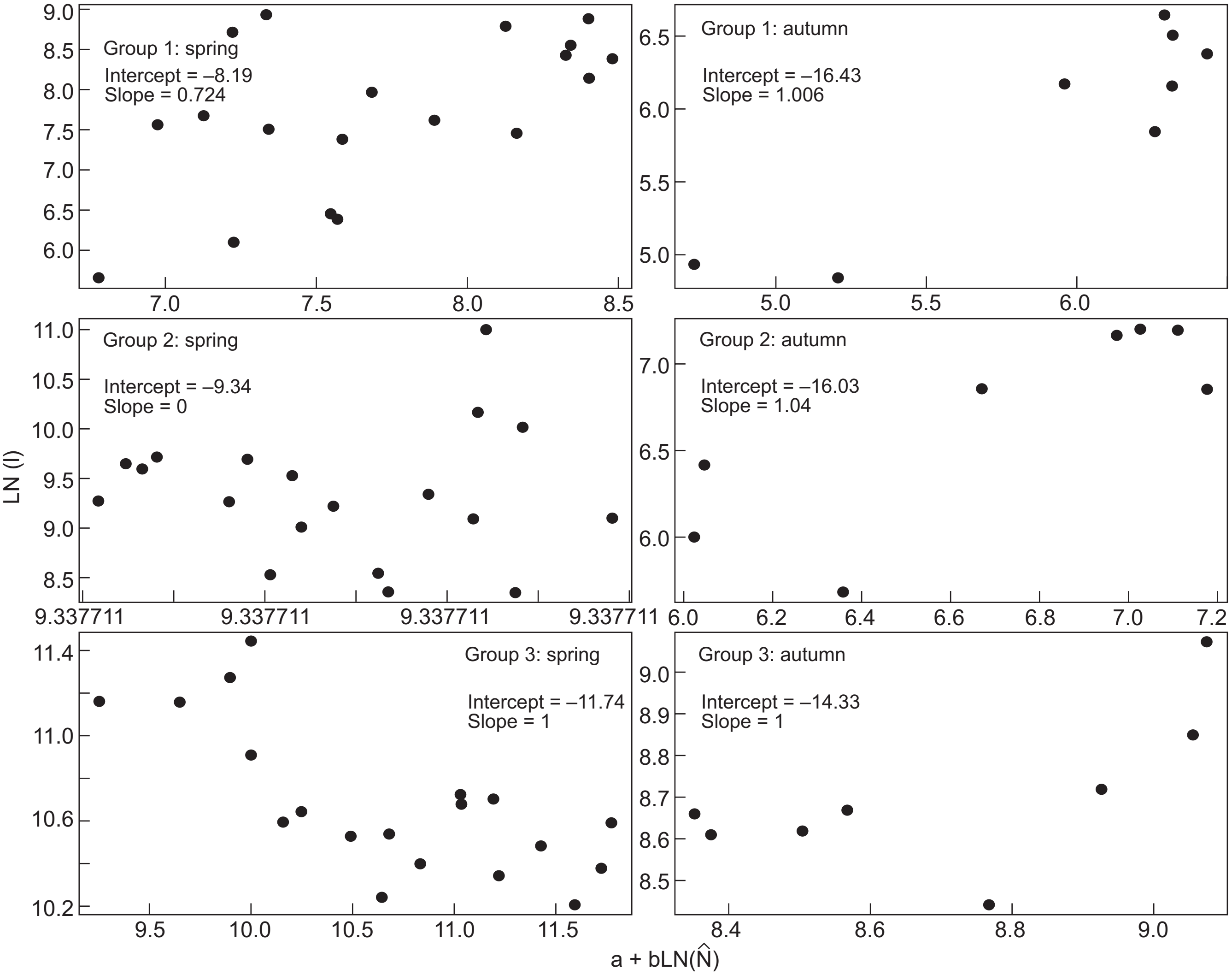

Figure 14: Alternative 2: plots of the log-linear regression as estimated in Gadget, with one plot for each negative log-likelihood component and the corresponding slope (estimated for Length-groups 1 and 2). Ln(I) is the index and a + b[ln($\hat{N}$)] the fitted regression to the modelled population

indices better (with the exception of Group 3 of the autumn survey) than the equivalent fixed-slope models, with little impact on the fit to the other components (Figures 11, 16). The model started from the end point of the standard run (Alternative 8) shows the least difference. Slopes estimated from the different runs show a similar pattern (Table 5), demonstrating consistency in the optimisation despite the differences in the models.

Estimating the survey index power reduces the variation in annual recruitment for the final years of the model. Differences in all other parameters are negligible. In turn, there is little impact on the population biomass (Figure 12). The difference in biomass between the models with the power fixed or estimated is predominantly towards the end of the time-series (Figure 17), which is a period of the model for which there is less information and fewer constraints on parameter estimation. The least difference is for the model starting at the end point of the standard model and the greatest difference for the models with lower weights on the age-length frequency data.

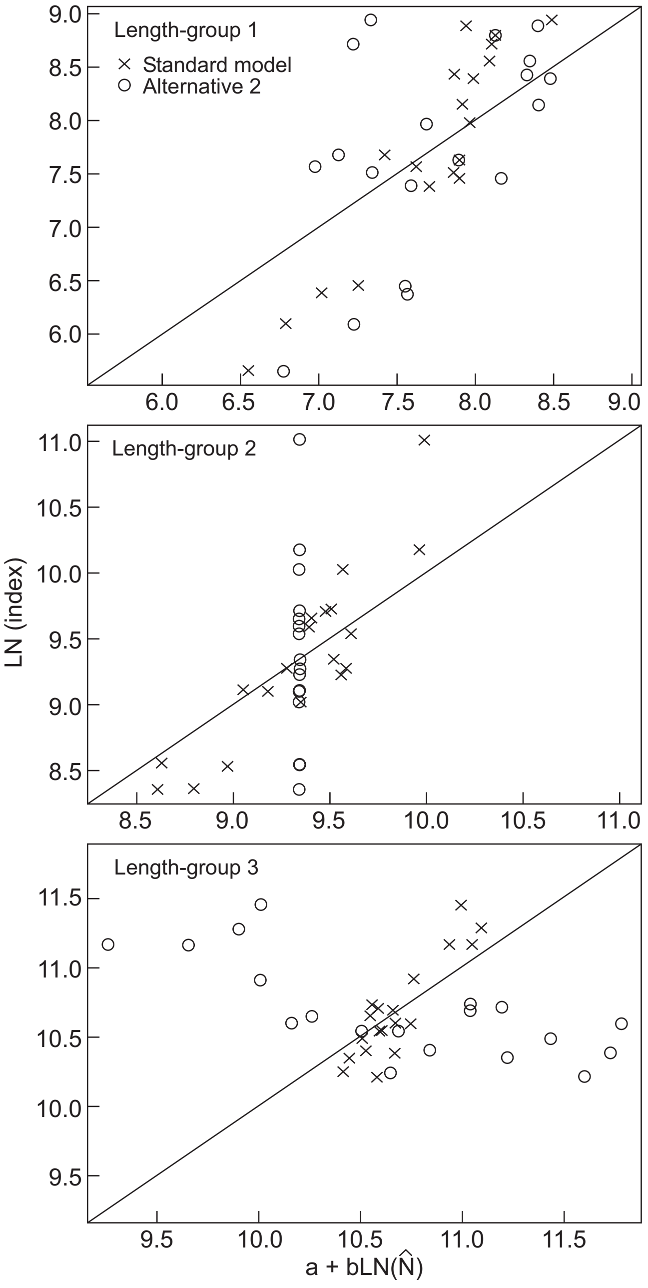

Figure 15: Comparison of the relationship between the modelled population and observed spring survey indices for the standard model and Alternative 2

Importance of starting parameters

The choice of starting parameters for the full estimation procedure can affect the model optimisation, as demonstrated by Alternative 2. With the power fixed, however, the importance of the initial parameters is considerably less. In addition, when the initial parameters are of the correct order of magnitude, the resulting optimised models are equivalent, even with the power estimated. The mean difference between Alternatives 6 and 8 is only 0.41% and 0.44% for total and spawning stock biomass respectively.

Use of age data

Two models were run without use of age-length frequency data in the objective function: one equivalent to the other models with the growth rate estimated (Alternative 4) and the other with the growth rate fixed (Alternative 5).

Without age data, estimating the spring survey selection pattern is a problem, but both models without age data estimate the same selection pattern. Catch selection, which is

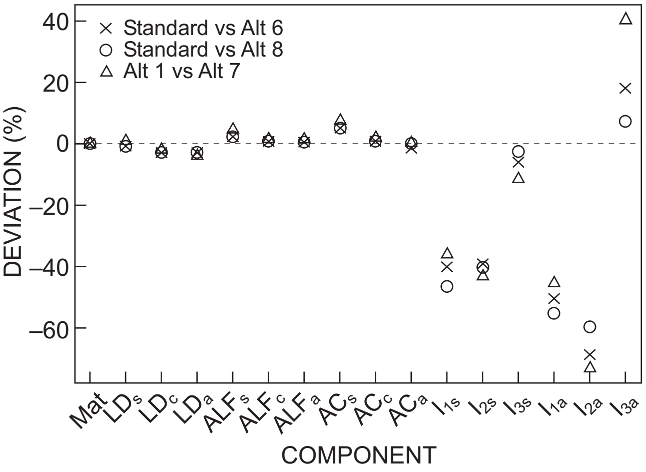

Figure 16: Comparison of the total negative log-likelihood scores by component for models with the power fixed vs models with the power estimated. See Table 1 for component names

Table 5: Estimated slopes when predicting log-scale survey indices

	Spring survey		Autumn survey	
Run	Group 1	Group 2	Group 1	Group 2
Alt 6	1.98	1.70	1.77	2.10
Alt 7	2.03	1.75	1.84	2.21
Alt 8	1.99	1.69	1.73	1.97

more important for model dynamics (because it determines the length distribution of removals by the fishing fleet) is similar to that estimated by the other models. Because the standard estimation procedure did not optimise for Alternative 4 (K estimated), a further run was done with an additional numerical optimising routine (BFGS, a quasi-Newton gradient search method) subsequent to the standard optimisation method. This run converged to a solution similar to that for Alternative 5 and with the growth rate K within 1.2% of the value estimated by the standard model.

Without age data in the objective function, the main deviation from the standard model is in estimation of the initial population abundance at age — parameters for which the objective function contains little information. The resulting biomass trends are lower (mean of 17% and 32% lower as a percentage of the standard model for total biomass and spawning stock biomass respectively), but follow the same pattern as the other models. Comparing abundance rather than biomass, the respective mean differences are 0.8% and 20.4%.

Comparison of the negative log-likelihood component scores for the standard model and Alternative 4 (which are equivalent to those for Alternative 5) indicates that, without being constrained by the age-length frequencies, the model fit to most of the survey indices and the length distributions is improved (Figure 18). Unsurprisingly, the model is, however, a worse fit to most of the age data and Length-group 3 of the spring survey.

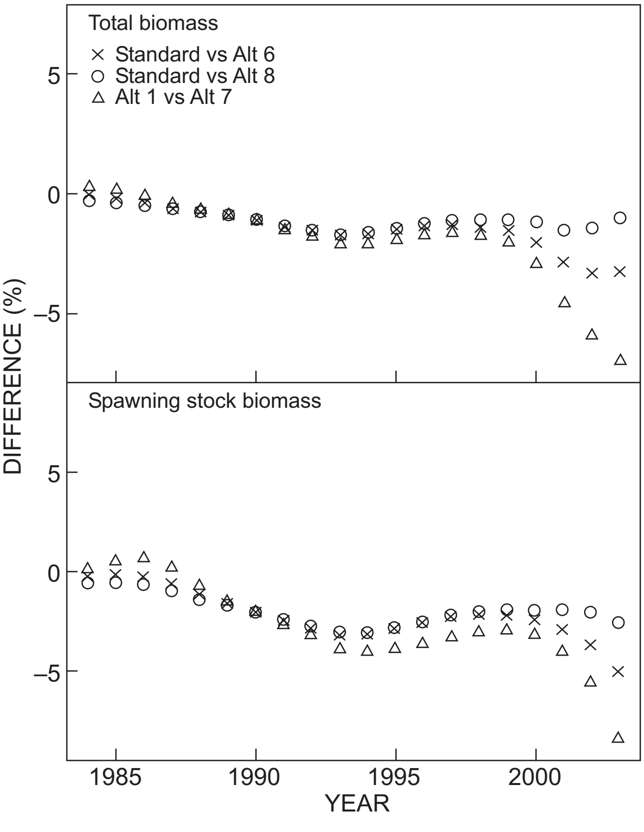

Figure 17: Comparison of the total and spawning stock biomass of models with the survey index power fixed or estimated. Plotted as the difference in biomass as a percentage of the biomass predicted by the model with the power fixed

Estimation of recruitment parameters

For a population with the growth, maturation and fleet selection parameters considered known, and with annual recruitment and a scaling factor for the initial population being the only estimated parameters, it might be considered that fewer data types are required in the objective function. In particular, models of this parameter structure might be used to increase the time span of an existing model if there is no reason to believe that the selection pattern of the fleet or the growth rate of the population has changed. Three additional models are used to consider this, each using the parameter estimates from the standard model as initial values. The objective functions contain either only survey indices, survey indices with three length distributions or survey indices with three age-length frequencies.

The negative log-likelihood scores that differ most from those of the standard model are from the model with only survey indices in the objective function (Figure 19). As would be expected, inclusion of the age-length frequencies in the objective function produces a better fit to all the age components than for the model with survey indices and length distributions. In addition, the combination of length distributions with the survey indices results in a considerably better fit to the age components than for the indices alone. When the only data in the objective function are the length-disaggregated indices, the estimated recruitment parameters display much greater deviation from the standard model parameter estimates and the total biomass trajectory is markedly different (Figure 20).

The combination of age-length frequencies with the indices generates higher population biomass levels than that from the model combining length distributions with the indices. As was shown in the comparison of age-length frequency weights, use of the age-length frequencies (or

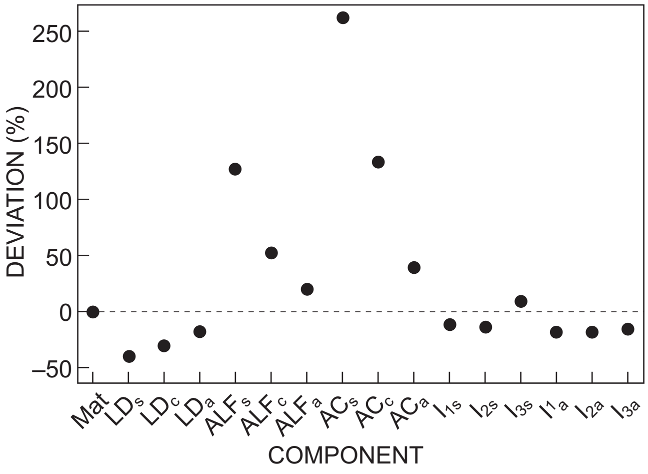

Figure 18: The difference in negative log-likelihood scores by component between the standard model and Alternative 4 (no age data in objective function), expressed as a percentage of the standard model. See Table 1 for component names

higher weights on the age-length frequencies) indicates higher population biomass than the length data.

Clearly, the information on population structure in the length and age-length frequencies is an important addition to the survey indices when estimating annual recruitment.

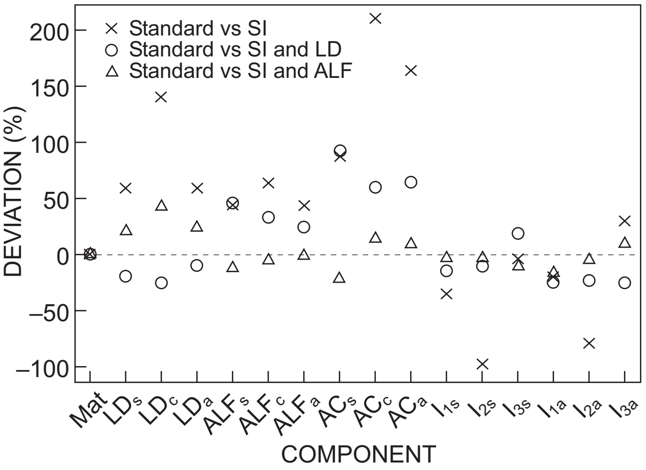

Figure 19: The difference in negative log-likelihood scores by component (as a percentage of the standard model) between the standard and inverse SS models, where the only parameters estimated are recruitment and the initial population multiplier. See Table 1 for component names; SI — survey indices; LD — length distribution; ALF — age-length frequency

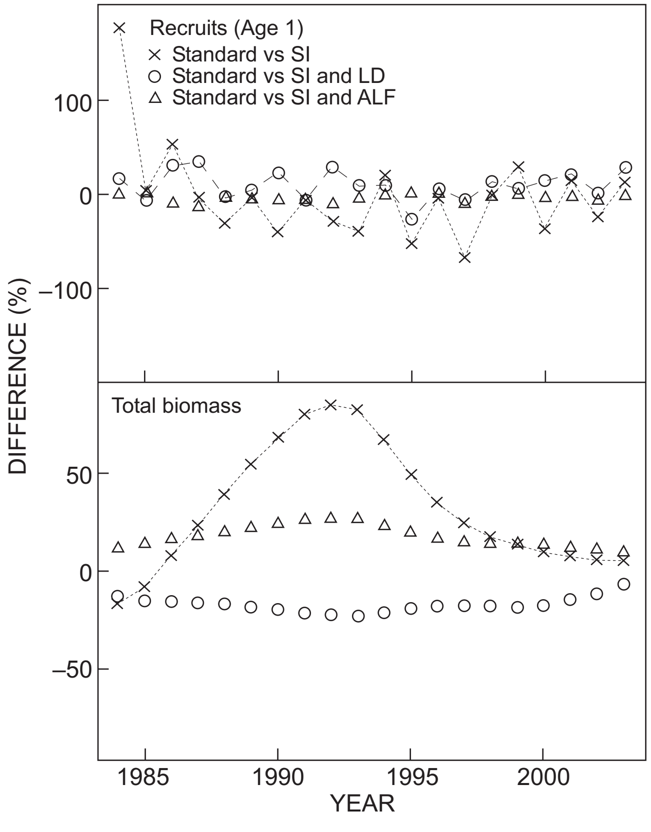

Figure 20: The difference in negative log-likelihood scores by component (as a percentage of the standard model) for the standard and subsequent models where the only parameters estimated are recruitment and the initial population multiplier: SI — survey indices; LD — length distribution; ALF — age-length frequency

Discussion

This paper has given examples of how one can parameterise a statistical fisheries model and estimate those parameters using a formal statistical method. The case study presented is a fairly simple one — a single-area model of a stock with two components. This work should be considered a prelude to much more complex models using Gadget. A multispecies model within Gadget is built up from single-species models, each of which should be fitted and evaluated using the methods incorporated in this paper before attempting to link them together.

A statistical model such as Gadget uses observed data directly and hence any discrepancies between the model and measurements of the population can be identified. Identifying such discrepancies provides an opportunity to improve the model and to assess the importance of increasing the model complexity. As an example of this is, one can consider the slope parameters in Equation 2 describing the log-log relationship between stock and index. A simple model using only catchability is indisputably simpler, but it may not be adequate, as noted for example by Stefansson (1992), as these slopes may be poorly estimated (DS Butterworth, University of Cape Town, pers. comm. cited in Stefansson 1992). The present results suggest that it is possible to estimate the slope parameters from the beginning of the estimation, but reasonable initial values of the parameters are required. The model with slopes thus estimated fits the survey indices better for Length-groups 1 and 2 than with the slopes fixed.

This formal approach of comparing models to individual datasets also allows for comparisons not possible when highly processed or aggregated data are used. In particular, using the weighting schemes employed, it is possible to identify inconsistencies between the model and datasets in a formal manner. This is done by evaluating whether increased weight on one dataset results in a considerably poorer fit to another set. Similarly, one can evaluate simple questions such as whether individual datasets are pulling results such as recruitment estimates or biomass trends in different directions.

Placing heavier weights on the age-length frequencies results in the model being a closer fit to the catch age-based likelihood components and generates a slightly higher fishable biomass. Placing less weight on age data results in a lower projected biomass. With no age data in objective function, the projected biomass is even lower. In general, models that fit the catch age-length distribution better tend to contain relatively more fish in older age groups.

Even with the selection pattern and growth fixed, inclusion of length and/or age-length composition data in the objective function improves the model parameterisation because these data contain additional information on the population structure not available from the indices.

Whereas it may seem reasonable, when only estimating recruitment parameters, to restrict the data in the objective function to survey indices, it is clear that for this population (with this model structure), these data alone are insufficient.

The examples come from a data-rich situation, but with judicious use of available data the method is seen to be promising for less data-rich environments. In particular, as demonstrated in alternatives, the lack of age data does not adversely affect estimates of recruitment or stock biomass trends which are the most important outputs discussed with respect to fishery management.

There are indications that these implementations are too 'stiff' with respect to growth. These issues may possibly be resolved by disaggregating the commercial fleet and/or allowing for variations in growth rate (both of which are currently possible within the Gadget framework).

Gadget is a framework that is flexible in terms of the structural models that can be developed and the range of data types which can be included. In fact, although there are several built-in functions to describe biological processes such as growth and fleet selection patterns, new functions can easily be added. Similarly, only certain likelihood functions are used in the present paper, but other options are available within Gadget and more could be added. Different datasets may be available for different stocks, and different growth, suitability or maturation functions may be used for different species, but any species can be in any role. This symmetric approach implies that each species can take on the role of predator or prey, depending on which other species are in the model, without changing the computer program itself. This is fundamentally different from several more traditional approaches where a single model is implemented as a computer program that then becomes tightly coupled to the exact system under consideration. In all the works presented in Stefansson *et al*. (1998), Tjelmeland and Bogstad (1998) and Mori and Butterworth (2004), a fundamentally different (and applicable) model is chosen for each species. Although this may adequately answer each of the questions being asked, it does not provide a general framework for modelling and testing assumptions that Gadget provides.

As an extension to this work, there is a need to demonstrate the method for the more complex multi-area case. Extensions to more areas open up many possibilities, even for the single-species case. Notably, a multi-area model can explicitly take into account variation in growth rates and area-dependent fleet behaviour, both of which can be difficult to encapsulate in a single-area model.

Additional future work involves formal methods for the estimation of uncertainty. It has been shown here that there are inherent problems in determining an adequate composite likelihood function or in defining correct weights to sums of squares. It is therefore not a trivial issue to define a plausible measure of uncertainty using Hessian-based methods. For this reason, a bootstrap approach seems most promising for estimating uncertainty and this is in accordance with preliminary simulations with Gadget (Hannesson *et al*. 2004) as well as other population dynamics/assessment models (e.g. Gavaris *et al*. 2000).

The objective function (negative log-likelihoods) for each dataset consisted of a simple sums of squares. The reasons for this choice are given in the case study, but they include serious problems with using common distributional assumptions in likelihood functions, as noted by e.g. Gavaris *et al*. (2000), Patterson *et al*. (2000), MacCall (2003) and Hrafnkelsson and Stefansson (2004). Future work should evaluate the benefits of using a multivariate Gaussian assumption, and preliminary tests indicate that this may be a considerable improvement on other assumptions (Elvarsson 2005). Such work needs to evaluate whether alternative distributional assumptions really provide better estimates, and this is probably best done using the bootstrap methods outlined above.

Acknowledgements — This work was supported in part by EU grant QLK5-CT199-01609.

References

Baranov T (1918) One the question of biological basis of fisheries. *Nauch. issledov. iktiol. inst. Izv.* **1**: 81–128 ((in Russian)

Begley J (2005) Gadget User Manual. *Technical Report* **120**, Marine Research Institute, Reykjavik

Beverton RJH, Holt SJ (1957) *On the Dynamics of Exploited Fish Populations.* Fisheries Investigations, Ser. 2, Vol. 19, UK Ministry of Agriculture and Fisheries, London

Björnsson H, Sigurdsson T (2003) Assessment of golden redfish (*Sebastes marinus* L.) in Icelandic waters. *Scientia Marina* **67**(Suppl. 1): 301–314

Elvarsson B (2005) An application of the multivariate normal distribution in a statistical fisheries model. *Final Report DST2: Development of Structurally Detailed Statistically Testable Models of Marine Populations*, *Technical Report* **118**, Marine Research Institute, Reykjavik, 127–131

Gavaris S (1988) An adaptive framework for the estimation of population size. *CAFSAC* (Canadian Atlantic Fisheries Scientific Advisory Committee) *Research Document* **88**(29): 12pp

Gavaris S, Patterson KR, Darby CD, Lewy P, Mesnil B, Punt AE, Cook RM, Kell LT, O'Brien CM, Restrepo VR, Skagen DW, Stefansson G (2000) Comparison of uncertainty estimates in the short term using real data. **ICES CM 2000/V:03**

Gudmundsson G (1994) Time series analysis of catch-at-age observations. *Applied Statistics* **43**(1): 117–126

Guldbrandsen Frøysa K, Bogstad B, Skagen D (2002) Fleksibest-an age-length structured fish stock assessment model. *Fisheries Research* **55**(1–3): 87–101

Gulland J (1965) Estimation of mortality rates. Annex to Arctic Fisheries Working Group Report. ICES CM D:3 (mimeo).

Hannesson S, Jakobsdottir A, Begley J, Taylor L, Stefansson G (2004) On the use of tagging data in statistical multispecies multi-area models of marine populations. **ICES CM 2004/FF:30**

Helgason T, Gislason H (1979) VPA-analysis with species interaction due to predation. **ICES CM 1979/G:52**

Hrafnkelsson B, Stefansson G (2002) Analysis of categorical length data from groundfish surveys. *Development of Structurally Detailed Statistically Testable Models of Marine Populations*, *Technical Report* **87**, Marine Research Institute, Reykjavik, 111–133

Hrafnkelsson B, Stefansson G (2004) A model for categorical length data from groundfish surveys. *Canadian Journal of Fisheries and Aquatic Sciences* **61**(7): 1135–1142

King M (ed) (1995) *Fisheries Biology, Assessment and Management*. Fishing News Books, Oxford

Kupca V (2005) Standardised fisheries database. In *Final Report: DST2: Development of Structurally Detailed Statistically Testable Models of Marine Populations*, *Technical Report* **118**, Marine Research Institute, Reykjavik, 1–57

Kupca V, Sandbeck P (2003) Data warehouse structure and data import. In *Final Report: DST2: Development of Structurally Detailed Statistically Testable Models of Marine Populations*, *Technical Report* **98,** Marine Research Institute, Reykjavik, 37–46

MacCall A (2003) Status of bocaccio off California in 2003. Technical Report, Pacific Fishery Management Council

MacDonald P, Pitcher T (1979) Age-groups from size-frequency data: a versatile and efficient method of analyzing distribution mixtures. *Fisheries Research Board of Canada* **36**: 987–1001

Methot RD (1989) Synthetic estimates of historical abundance and mortality for northern anchovy. *American Fishery Society Symposium* **6**: 66–82

Mori M, Butterworth DS (2004) Consideration of multispecies interactions in the Antarctic: a preliminary model of the minke whale–blue whale–krill interaction. In: Shannon LJ, Cochrane KL, Pillar SC (eds) *Ecosystem Approaches to Fisheries in the Southern Benguela*. *African Journal of Marine Science* **26**: 245–259

Palsson OK, Schopka SA, Stefansson G, Steinarsson BA (1989) Icelandic groundfish survey data used to improve precision in stock assessments. *Journal of Northwest Atlantic Fisheries Science* **9**: 53–72

Patterson K, Cook R, Darby C, Gavaris S, Kell L, Lewy P, Mesnil B, Punt A, Restrepo V, Skagen D, Stefansson G (2001) Estimating uncertainty in fish stock assessment and forecasting. *Fish and Fisheries* **2**(2): 125–157

Patterson K, Cook R, Darby C, Gavaris S, Mesnil B, Punt A, Restrepo V, Skagen D, Stefansson G, Smith M (2000) Validating three methods for making probability statements in fisheries forecasts. **ICES CM 2000/V:06**

Pennington M, Volstad JH (1994) Assessing the effect of intra-haul correlation and variable density on estimates of population characteristics from marine surveys. *Biometrics* **50**: 1–8

Pope JG (1972) An investigation of the accuracy of virtual population analysis using cohort analysis. *ICNAF Research Bulletin* **9**(10): 65–74

Ramberg JS, Dudewicz EJ, Tadikamalla PR, Mykytka EF (1979) A probability distribution and its uses in fitting. *Technometrics* **21**(2): 201–214

Shepherd J (1999) Extended survivors analysis: an improved method for the analysis of catch-at-age data and abundance indices. *ICES Journal of Marine Science* **56**(5): 584–591

Shepherd J, Pope J (1993) Alternative methods for the estimation of immigration to the Icelandic cod stock. *Fisheries Oceanography* **2**(3): 254–259

Sigurdsson T, Hjorleifsson E, Bjornsson H, Palsson OK (1997) Stofnmaeling botnfiska a Islandsmidum haustid 1996. *Technical Report* **61**, Marine Research Institute, Reykjavik

Stefansson G (1992) Notes on the stock-dynamics and assessments of the Icelandic cod. **ICES CM 1992/G:71**

Stefansson G (1998) Comparing different information sources in a multispecies context. In: Funk F, II TQ, Heifetz J, Ianelli J, Powers J, Schweigert J, Sullivan P, Zhang C (eds) *Fishery Stock Assessment Models: Proceedings of the International Symposium, Anchorage 1997*, 15th Lowell Wakefield Fisheries Symposium, pp 741–758

Stefansson G (2003) Issues in multispecies models. *Natural Resource Modeling* **16**(4): 415–438

Stefansson G, Palsson O (1997a) Statistical evaluation and modelling of the stomach contents of Icelandic cod (*Gadus morhua*). *Canadian Journal of Fisheries and Aquatic Sciences* **54**(1): 169–181

Stefansson G, Palsson OK (1997b) BORMICON a Boreal Migration and Consumption model. *Technical Report* **58**, Marine Research Institute, Reykjavik

Stefansson G, Palsson OK (1998) A framework for multispecies modelling of boreal systems. *Reviews in Fish Biology and Fisheries* **8**: 101–104

Stefansson G, Sigurjonsson J, Vikingsson G (1998) On dynamic interactions between some fish resources and cetaceans off Iceland based on a simulation model. *Journal of Northwest Atlantic Fisheries Science* **22**: 357–370

Taylor L (2005) Definition of areas in Icelandic waters. In: *Final Report: DST2 Development of Structurally Detailed Statistically Testable Models of Marine Populations*. *Technical Report* **119**. Marine Research Institute, Reykjavik, pp 330–339

Tjelmeland S, Bogstad B (1989) *MULTSPEC: The Manual*. Institute of Marine Research, Bergen

Tjelmeland S, Bogstad B (1998) MULTSPEC — a review of a multispecies modelling project for the Barents Sea. *Fisheries Research* **37**(1): 127–142

Ulltang O (1977) Sources of errors in and limitations of Virtual Population Analysis (cohort analysis). *ICES Journal of Marine Science* **37**(3): 249–260

Wang X, Zidek J (2005) Selecting likelihood weights by cross-validation. *Annals of Statistics* **33**(2): 463–500

Manuscript received December 2006; accepted May 2007

Appendix: Growth, predation and fishing formulation

Basic Formulation of Growth

A simple growth model

In a multispecies context, growth in Gadget may depend on consumption, but in simpler models it may be assumed to follow a von Bertalanffy curve:

$$L_t = L_\infty \left(1 - e^{-Kt}\right), \quad t \in R_+ \qquad \text{(App.1)}$$

Suppose fish of a given age a are of length l. Growth according to Equation App.1 is then given by

$$\Delta L = L_\infty \left(1 - \frac{L}{L_\infty}\right)\left(1 - e^{-K\Delta t}\right) \qquad \text{(App.2)}$$

Growth in weight and growth in length can be assumed linked and a length-weight relationship of the form

$$W = cL^b \qquad \text{(App.3)}$$

used, where typically the power b is close to 3 and if b = 3, the constant c is the condition factor.

It should be noted that when l is greater than L_∞, growth becomes negative according to the growth equation. In this case it is assumed that the fish do not grow further. In practice, L_∞ is constrained to be greater than the maximum length of the population.

Also, the average growth of a group of fish needs to be translated into a new distribution of these fish in the following time step. This will be handled in subsequent sections.

On updating length distributions

Fish population dynamics are modelled in MULTSPEC and Gadget through forward simulations of fish populations, allowing fish to migrate between areas, grow, mature, spawn and die. The basic unit in these models is the number of fish in a certain model unit or 'cell'. The fish in a 'cell' are in the same age- and size-group, in the same region and time step. When this basic model formulation is used, the numbers in a 'cell' need to be updated during a given time step, so as to reflect all processes being modelled. In addition to growth, these processes include migration, spawning, natural and fishing mortality.

Starting with a specified average length increment (ΔL), fish need to be redistributed from their initial length-class into upper length-classes in a reasonable manner. Simple techniques may use only a few upper length intervals and a simple *ad hoc* update scheme. The update scheme should eventually be evaluated in terms of its ability to provide adequate eventual length distributions. This sets some immediate bounds on the dispersion at each time step, because an overly high or low variance in the length update will quickly result in inadequate final length distributions at age for the oldest ages.

The length update scheme can most easily be implemented through a predetermined (from a distribution, e.g. multinomial) discrete set of ΔL-values to reallocate fish in an initial length-group, when the desired average growth is to be ΔL. This approach, used in earlier Bormicon, MULTSPEC and Fleksibest implementations (Stefansson and Palsson 1997b), is undesirable for many reasons. First, the setup is completely rigid because there is no built-in parameter to describe possible deviations of growth from the specified distribution and therefore data on growth may adversely affect parameters in other parts of a complex model, because of the incorrect specification of the rigid relationship. Second, a simple discrete (rounded) lookup provides a non-differentiable likelihood function that will result in estimation problems later on (e.g. Guldbrandsen Frøysa *et al*. (2002)).

What is needed is a way to specify a flexible parametric distribution with enough parameters to allow minimal flexibility to track length distributions of an age-group, yet with enough parsimony in parameters to make the parameters involved estimable.

Consider fish of a specified length, L, which are destined to grow on average ΔL, according to the growth model. A model for the update will assign probabilities p_j of a fish in the original cell growing by j length-groups. Denote the absolute length increment corresponding to j length-groups by δl_j. These probabilities must then satisfy the obvious restrictions

$$\sum_j p_j = 1$$

$$\sum_j p_j \delta l_j = \Delta L$$

Growth in weight

Because the approach described in this section is length-based, the weight needs to be updated in synchrony with the length increase. Consider, therefore, fish of length L (with an average weight of W) that will increase in length to $L+\delta l_j$. Although it is possible to use Taylor approximations to the length-weight relationship, as implemented in Bormicon, a more internally and mathematically consistent approach is to maintain length-weight relationships exactly. This can be done by defining the upcoming weight increment of this group of fish as

$$\delta w_j = c\left(L + \delta l_j\right)^b - cL^b$$

The mean weight in the receiving group is subsequently updated using prior and incoming numbers along with the current mean weight in the group.

The result of this approach is that, if fish start out so that the length-weight relationship holds exactly for all length cells, then this relationship is maintained throughout the simulation.

A specific model for the length update

Although a first step might be to attempt to estimate individual probabilities p_j, this would result in too many parameters. Another approach would be to estimate variance, skewness and kurtosis and go from these to transition probabilities, but there is no trivial transformation between the two.

A flexible probability distribution such as the 4-parameter inverse lambda distribution could probably be used (Ramberg *et al*. 1979), but parameter estimation tends to be difficult. Similarly, a binomial distribution (or even a (truncated) Poisson) can be used, but both are completely rigid, because the value of ΔL completely specifies the single free parameter in each of these distributions (assuming the number of permissible length-group increments to be fixed).

The beta-binomial distribution can be used as a simple alternative. This approach can be formulated so as to provide a single estimable parameter in addition to the mean, which is specified by ΔL.

First consider the binomial distribution which is defined for integers, x = 0,…,n by

$$\binom{n}{x} p^x (1-p)^{n-x} = \frac{\Gamma(n+1)}{\Gamma(x+1)\Gamma(n-x+1)} p^x (1-p)^{n-x}$$

Using this probability distribution for the issue at hand, for a given n, the other parameter p of this distribution is fully defined because $\mu = np$ and the mean growth is given as the specified ΔL, which fixes $p = \Delta L/n$. Although this distribution can certainly be used, it is clear that no flexibility is allowed at all, and in fact it would be quite unlikely for such a rigid distribution to satisfy the specified requirements of attaining the correct final distribution of length-at-age.

A common approach to more flexibility is to allow the parameter p itself to come from another distribution, often the beta distribution. The beta distribution is defined for arbitrary values of $\alpha > 0$ and $\beta > 0$ by

$$f(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}, \; 0 \le p \le 1 \qquad \text{(App.4)}$$

and it is known that the mean of this distribution is given by

$$E(p) = \frac{\alpha}{\alpha+\beta} \qquad \text{(App.5)}$$

Thus, rather than using a fully specified binomial distribution, more flexibility is obtained by using this combined beta-binomial distribution. This approach results in the following marginal distribution of the length increments:

$$P[X=x] = \int_{p=0}^{1} P[X=x \mid p] f(p) dp$$
$$= \frac{\Gamma(n+1)}{\Gamma(n-x+1)\Gamma(x+1)} \frac{\Gamma(\alpha+\beta)}{\Gamma(n+\alpha+\beta)} \frac{\Gamma(n-x+\beta)}{\Gamma(\beta)} \frac{\Gamma(x+\alpha)}{\Gamma(\alpha)}$$

It should be noted that in the case of growth by length-groups within a short time interval, only low values of n and X are needed and the above products become correspondingly simple to generate. Probabilities can, therefore, be readily generated from this beta-binomial distribution given specified values of α, β and n. The latter (n) will usually be assumed known, outside an estimation procedure.

It is also reasonably easy to see that the mean of the beta-binomial distribution is given by

$$\mu = nE[p] = \frac{n\alpha}{\alpha+\beta}$$

If β is taken as a parameter to be estimated, the requirement $\Delta L = \mu$ therefore implies

$$\alpha = \frac{\beta \Delta L}{n - \Delta L}$$

This approach is implemented in Gadget by defining a growth function with a single estimated parameter β.

Alternative growth formulations

An alternative set of growth functions is obtained by basing length growth on weight growth and computing weight growth from consumption. In order to verify different methods of implementing growth, it is useful to be able to link these approaches together, which can be done in Gadget (Begley 2005).

The general case will involve consumption and therefore growth in length and in weight will not always conform strictly to a length-weight relationship. Notably, if fish do not get enough food, they will typically lose weight but not get shorter and this effect can be accounted for within Gadget (but is not implemented in the case study presented).

Predation, Including Fishing and Consumption

Although this paper does not include predation by fish, it does include fishing operations which are modelled in the same manner.

In order to distinguish between predator and prey species in an interaction, the following notations are used:

s = a general species
p = a prey species
P = a predator species
l = a prey length-group
L = a predator length-group

Accordingly, N_{lp} denotes the number of individuals of a prey species p of length l and N_{LP} the numbers in length-group L of a predator P. The notation is solely for convenience because there is no reason mathematically to exclude one predator's prey from being another species' predator. When fishing is implemented as predation, P is a fleet, in which case L has no meaning.

The consumption of a prey by a predator is defined through the below-mentioned equations. Consider a fixed area and time step so the notation can be simplified to indicate only the length of predator and prey.

The suitability function $S_{P,p}$ describes the suitability of length-group l of prey p as food for length-group L of predator P. This is usually viewed as a function of two variables, the length of the predator and prey. A common suitability function is

$$S_{P,p}\left(l,L\right)=\frac{\phi_4}{1+e^{-\phi_1-\phi_2 l-\phi_3 L}} \qquad \text{(App.6)}$$

where ϕ_1–ϕ_4 are parameters to be determined (or fixed at specified values).

For a fleet, the 'suitability function' in Equation App.6 is simply proportional to a 'selection pattern' (Beverton and Holt 1957, King 1995) as usually defined in fishery science (see e.g. Section 5.2 in Stefansson and Palsson 1997b).

As a function of l, over the domain of prey length-groups, the form of the suitability function can in principle be either strictly increasing or dome-shaped, although Equation App.6 is always increasing. In general, the function needs to reflect the relative desirability of one prey size-group to another for a given predator size-group.

A suitability function for predation of a predator on a prey could be given by

$$S_{P,p}\left(l,L\right)=\begin{cases}\eta_0+\eta_2 e^{-\frac{\left(\ln\frac{L}{l}-\eta_1\right)^2}{\eta_4}} & \text{if } \ln\frac{L}{l}\le\eta_1\\ \eta_0+\eta_2 e^{-\frac{\left(\ln\frac{L}{l}-\eta_1\right)^2}{\eta_3}} & \text{if } \ln\frac{L}{l}>\eta_1\end{cases} \qquad \text{(App.7)}$$

with five parameters ($\eta_0-\eta_4$) which allows for an asymmetric relationship about the optimal predator:prey size ratio. This is implemented in Gadget but not used in the case study presented.

The auxiliary function ϕ_P is defined by

$$\phi_P\left(l,L,p\right):=\left(S_{P,p}\left(l,L\right)E_p N_p\left(l\right)W_p\left(l\right)\right)^{d_{Pp}} \qquad \text{(App.8)}$$

where E_p is the energy content (in kJ kg^{-1}) of the prey and d_{Pp} is the preference of the predator for the prey (most commonly $d_{Pp}=1$). This simply weights the prey biomass according to its suitability and is used to define the *feeding level*

$$\psi_P\left(L\right):=\frac{\sum_{l,p}\phi_P\left(l,L,p\right)}{\sum_{l',p'}\phi_P\left(l',L,p'\right)+A} \qquad \text{(App.9)}$$

where A is the biomass of prey required to allow the predator to consume half the maximum consumption.

The *target consumption* can be considered as the amount that a predator would consume of a prey, assuming that enough food is available. The total target consumption across all prey types is given by

$$T_P\left(L\right)=N_P\left(L\right)H_P\left(L,T\right)\psi_P\left(L\right) \qquad \text{(App.10)}$$

here T denotes area-specific temperature and the function H is termed the *maximum consumption*, which is given by

$$H_P\left(L,T\right)=m_0 e^{\left(m_1 T-m_2 T^3\right)}L^{m_3} \qquad \text{(App.11)}$$

for constants m_0–m_3. Given the total across all prey types being $T_P(L)$, then the amount that length-group L of predator P will consume of length-group l of prey p is given by the equation

$$T_{P,p,L,l}=T_P\left(L\right)\frac{\phi_P\left(l,L,p\right)}{\sum_{l',p'}\phi_P\left(l',L,p'\right)} \qquad \text{(App.12)}$$

Although generic, the consumption formulae are only used in the present paper to describe fishing, in which case a simple model replaces $T_P(L)$ in Equation App.12 by the observed landings.

For each prey, an upper limit is set on the total amount consumed by all predators, which restricts consumption to a maximum of 95% of the available biomass. This is simply implemented by scaling target consumption by all predators. A corresponding penalty is subsequently added to the negative log-likelihood function when this occurs (see Equation 4).

Commercial catches can be implemented using the predation procedure, by setting the *target catch* directly equal to the observed catch and using the same scaling mechanism as for (other) predators when prey biomass is too low to sustain total predation (including catches).

Alternatively, commercial or survey catches can be set to be linearly related to the biomass of the preys. In this case, the amount caught of length-group l of the prey p on the area a and time step t by fleet P is

$$T_{P,p,L,l}=cE_{t,a}\phi_P\left(l,L,p\right) \qquad \text{(App.13)}$$

$$=cE_{t,a}S_{p,P}\left(l,L\right)N_p\left(l\right)W_p\left(l\right) \qquad \text{(App.14)}$$

where the effort measure $E_{t,a}$ is a parameter, c is a constant and L is a number appropriate to scale the suitability function into a selection pattern for the fleet.