

The original publication is available at <http://www.tandf.co.uk/journals/>

A bootstrap method for estimating bias and variance in statistical fisheries modelling frameworks using highly disparate datasets

BÞ Elvarsson^{a, b, *}, L Taylor^a, VM Trenkel^c, V Kupca^d & G Stefansson^a

^a Science Institute, University of Iceland, Reykjavik, Iceland

^b Marine Research Institute, Reykjavik, Iceland

^c Ifremer [French Research Institute for the Exploitation of the Sea], Nantes, France

^d HPC2N [High Performance Computing Center North], Umeå University, Umeå, Sweden

*: Corresponding author : BÞ Elvarsson, email address : bjarki.elvarsson@gmail.com

Abstract:

Statistical models of marine ecosystems use a variety of data sources to estimate parameters using composite or weighted likelihood functions with associated weighting issues and questions on how to obtain variance estimates. Regardless of the method used to obtain point estimates, a method is required for variance estimation. A bootstrap technique is introduced for the evaluation of uncertainty in such models, taking into account inherent spatial and temporal correlations in the datasets, which are commonly transferred as assumptions from a likelihood estimation procedure into Hessian-based variance estimation procedures. The technique is demonstrated on a real dataset and the effects of the number of bootstrap samples on estimation bias and variance estimates are studied. Although the modelling framework and bootstrap method can be applied to multispecies and multiarea models, for clarity the case study described is of a single-species and single-area model.

Keywords: bootstrapping ; correlated data ; fish population dynamics ; non-linear models

1 Introduction

Statistical models consolidate data from various sources by using them simultaneously to estimate parameters. The importance of using all data in a single model has been emphasised by several authors (Demyanov et al. 2006, Methot 1989) and although the benefits are clear, it is certainly not without problems, including the question of variance estimation, model mis-specification and weighting of all data sources (Francis 2011, Stefansson 2003, Maunder and Punt 2012). In the context of complex population dynamics models of exploited marine species, multiple data sources with widely different properties are routinely used in the estimation process.

Variance estimates of parameters in nonlinear models have commonly been derived from the inverted Hessian matrix at the optimum, when the method of least squares (or maximum likelihood) is employed for parameter estimation. Alternatively the Jacobian matrix of the residuals can be used. Several conditions need to be satisfied for statistical inference, e.g. confidence statements to hold in the finite-sample case. First, the model needs to be correct. Second, variance assumptions i.e. homoscedasticity and knowledge of the ratios of variances in individual data sets, need to be appropriate.

Methods of estimating variances in fish stock assessment models have been discussed and evaluated by many authors including Gavaris et al. (2000), Gavaris and Ianelli (2001), Magnusson et al. (2012) and Patterson et al. (2001). When the distributional properties of the data are not well understood or the models are incorrect, Hessian-based approaches have been seen to fail in several examples in fishery science (Patterson et al. 2001). Although this may seem to contradict the theoretical statements, the assumptions e.g. in Jennrich (1969) include independence of observations, a unique minimum, identically distributed errors and of course the results are only asymptotic. Any of these assumptions may fail. It follows that for problems in fishery science one cannot assume a priori that a Hessian-based method will give reasonable results. For example, disregarding correlation structure when present has been found to potentially lead to incorrect conclusions in single-species assessments, sometimes with serious consequences (Myers and Cadigan 1995). Similarly, multimodal likelihood functions have been seen in real applications (Richards 1991) and typically correspond to incorrect model assumptions that are not detected with traditional analysis (Stefansson 2003) but may potentially be detected if histograms of bootstrap parameter estimates also become multimodal (see example in Hannesson et al. 2009).

Many of the limitations of the Hessian-based approaches have been met by alternative methods. In particular models developed using the Bayesian framework (as discussed in eg. Punt and Hilborn 1997) provide an elegant formulation of uncertainty as posterior distributions of the quantity of interest. In all but trivial cases the posterior distribution must be estimated numerically with methods such as Markov chain – Monte Carlo. With the commoditization of computers in conjunction with the development of frameworks such as BUGS (Spiegelhalter et al. 1996) and ADMB (Fournier et al. 2012), the Bayesian framework has become popular alternative to Hessian-based uncertainty methods. The attraction of the Bayes inference stems, to some degree, from the ability to include prior belief/knowledge into the model as explicit distributions. Various sources (eg. Chen et al. 2000, Millar 2002) suggest, however, that considerable care must be taken when choosing model priors to avoid misspecification and suggest a suite of robust priors applicable in fisheries model setting.

Alternative frequentist approaches to Hessian-based parameter variance estimation include bootstrap methods (Efron 1979, Efron and Tibshirani 1994). The simplest bootstrap method assumes that the data are independent measurements without correlation. However, semi-parametric approaches have also been developed to sample residuals from a model, possibly from a distribution (parametric bootstrap) or with a known correlation structure (Davison and Hinkley 1997).

This paper demonstrates a novel use of bootstrapping to address complex and disparate data issues. The approach is generic, but it has special application to statistical models of (multiple and interacting) marine populations such as those developed within the Gadget framework. Gadget is a statistical age-length structured modelling environment originally proposed by Stefansson and Palsson (1998), combining concepts from several earlier methods (Gavaris 1988, Methot 1989, Tjelmeland and Bogstad 1989, Bogstad et al. 1992), described in Begley (2004) and subsequently

used in multiple fisheries applications (e.g. Björnsson and Sigurdsson 2003, Taylor et al. 2007, Lindstrøm et al. 2009). The protocol used in Gadget to estimate likelihood component weights and optimise model parameters is described in detail in Taylor et al. (2007) and the weighting protocol is based on that described in Stefansson (1998) and Stefansson (2003).

In the following sections the development of an elementary sampling unit used in the bootstrap is described. The methodology is applied to a Gadget model for cod in Icelandic waters (the standard model from Taylor et al. 2007) and contrasted to a more traditional Hessian-based approximation of variance.

2 Development of an elementary sampling unit

Statistical fisheries models may involve the use of a large number of data from a variety of sources. Every sample from each data source can be classified according to sampling location and time. A model such as Gadget operates on certain time-steps and also uses some spatial units. Within any modelled spatio-temporal unit there will normally be several data samples. For any bootstrap method the first question is therefore what the sampling unit should be. A unit of measurement in marine studies tends to be based on a single fish and elementary resampling might bootstrap on individual fish (as in e.g. Gudmundsdóttir et al. 1988). Doing this assumes that all individually measured fish are independent which is invalid for several reasons (Hrafnkelsson and Stefansson 2004, Pennington and Volstad 1994). Resampling entire fish samples (as is done by Singh et al. 2011) can potentially be used to account for this intra-haul correlation. Appropriate analyses of variance can correspondingly be used to evaluate these effects (Helle and Pennington 2004, De Croos and Stefansson 2011) and when combining samples, alternatives to simple sums or means may be needed for aggregation (Babak et al. 2007). However, considering samples as units may not be quite enough, since fish at close geographic locations will also tend to be similar due to a fine-scale spatial structure which can not be easily modelled (e.g. Stefansson and Palsson 1997b).

In addition to the sampling unit problem, one needs to take into account the variety of data sources. Biological samples from commercial catches may be collected on a fine temporal and spatial scale whereas scientific surveys are typically only conducted once or twice a year and different surveys may or may not overlap spatially. Other data sets such as species composition of stomach contents or tagging experiments may be collected at completely different resolutions to age or length data.

Here the proposed sampling unit is based on spatial structure on the Icelandic coastal shelf developed by Taylor (2003), shown in figure 1 where the area within the gridlines are referred to as **subdivisions**. The spatial structure is based mainly on bathymetry, hydrography and species assemblages with some further disaggregation defined by fishing regulations. In this context an **elementary sampling unit** is all data collected inside a subdivision within a time period of interest. In this context subdivisions and elementary sampling units are therefore used interchangeably. In order to reduce correlations between the elementary sampling units aggregations are made. For example, to remove within-sample correlations between length groups (Hrafnkelsson and Stefansson 2004), only (combinations of) entire length samples are used, rather than lengths of individual fish. Similarly, data are aggregated within the fairly large spatial areas and the shortest time-step is at least one month. This should eliminate intra-haul correlations (Pennington and Volstad 1994) and those correlations between age-groups (Myers and Cadigan 1995) which are related to local shoals or small feeding patches.

To generate input files for Gadget a second aggregation method is applied on the elementary sampling units, that is all data from a particular subdivision, which varies somewhat depending on the data source. Some data types, e.g. length distributions, are simply added whereas others, such as mean length-at-age, may go through a computational mechanism involving age-length keys. A description of a fisheries data base, which is able to handle data aggregations in this manner, can be found in Kupca and Sandbeck (2003) and Kupca (2004).

Here the fundamental idea is the aggregation of elementary sampling units in the creation model

Fig. 1
near here

inputs. These sets of elementary sampling units can therefore be sampled (with replacement) before aggregation, with each resample leading to a new model input data set. A typical model run for parameter estimation based on such a resampled data set will result in a resampled parameter estimate. The collection of all such estimates form a bootstrap sample. The procedure could be called a spatio-temporal block bootstrap with unequal block size.

3 A fisheries example

3.1 The setting

The example marine system used in this paper is based on cod in Icelandic waters (fig. 1) with an approach very similar to Taylor et al. (2007). The model consists of two stock components of cod, i.e. mature and immature cod in a single area. Modelling maturity enables the calculation of spawning stock biomass and allows for different weight-length relationships to be used for immature and mature fish.

Two fixed station surveys are used to monitor the stock, in spring and autumn, providing population indices as well as biological samples. Landings information is available from official data bases and raw biological data (length distributions, age compositions) along with survey data in the Marine Research Institutes (MRI) databases (see e.g. ICES 2011, Palsson et al. 1989, Sigurdsson et al. 1997, Taylor et al. 2007, for a description of data and surveys). The technical details of the model are described in appendix A.

3.2 The data set and parameters

The model is a parametric and deterministic forward population dynamics simulation model. A single simulation results in a complete population structure, including predictions of all data sets, as described in Begley (2004) and Taylor et al. (2007) and a corresponding evaluation of a (negative log-)likelihood function (sums of squares in the present paper).

With the exception of landings data, data sets are only used in the likelihood components. For simplicity, landings data are used directly in the population models, whereby the populations are simply reduced in numbers to be in accordance with the corresponding landed weight. Note that in the approach proposed here the landings data are not resampled.

An overview of the datasets and model parameters used in this case study is shown in Tables A.1 and A.2 respectively.

3.3 Estimation protocol

The weights on the likelihood components are calculated for each model (i.e. each bootstrap run), according to the protocol described in appendix A.2.4 with arbitrary starting parameters. This is a two stage estimation method, where the error variances, within a data set, are estimated by increasing the weight on that particular component of the total sum of squares, followed by a final minimisation using those inverse variances as weights. For a full description of this procedure refer to the appendix.

The bootstrapping approach consists of the following:

- The base data are stored in a standardized data base:
 - Time aggregation: 3 months
 - Spatial aggregation: subdivision
 - Further disaggregation is based on a range of categories including fishing gear, fishing vessel class, sampling type (e.g. harbour, sea and survey). A full listing of data types used in the case study can be found in table A.1, these data are stored subdivision dis-aggregated to allow for use in a bootstrap.

- To bootstrap the data, the list of subdivisions, depicted in fig. 1, required for the model is sampled (with replacement) and stored. For a multi-area model one would conduct the resampling of subdivisions within each area of the model.
- The list of resampled subdivisions is then used to extract data (with replacement so the same data set may be repeated several times in a given bootstrap sample).
- For a single bootstrap Gadget model, the same list of resampled subdivisions is used to extract each likelihood dataset i.e. length distributions, survey indices and age-length frequencies are extracted from the same spatial definition.
- A Gadget model is fitted to the extracted bootstrap dataset using the estimation procedure described above.
- The resampling process is repeated until the desired number of bootstrap samples are extracted.

When resampling, data are forced to remain in the correct year and time-step so resampling is based on sampling spatially the elementary data units within a given modelled unit of time and space. Thus, within a modelled spatial unit the bootstrap is a resampling of subdivisions. This implicitly assumes data contained within each area of the model to be independent and identically distributed. Independence is justified by the definition of subdivisions. Furthermore treating them as they were from the same distribution, i.e. bootstrap replicates, appears to have little negative effect when compared to more traditional methods (Taylor 2002).

The entire estimation procedure is repeated for each bootstrap sample. In particular, since the estimation procedure includes an iterative reweighting scheme, this reweighting is repeated for every bootstrap sample. The point of this is that the bootstrap procedure is no longer conditional on the weights. The procedure as a whole is quite computationally intensive but can easily be run in parallel, e.g. on a computer cluster.

In stark contrast to this, Hessian-based approaches usually only compute the Hessian at the final solution. Thus, they completely omit the effect of reweighting likelihood components when estimating uncertainty. Such methods are thus conditional on the weights obtained in a pre-estimation stage.

3.4 Application of the bootstrap procedure and its variants

The bootstrap procedure presented here is, as noted earlier, quite computationally demanding as the number of bootstrap samples increases. In this exercise 1000 bootstrap samples were chosen as the **baseline** simulation. This number of iterations was chosen as a practical upper limit, as a single optimisation run for a Gadget model takes a substantial amount of time. In addition to the baseline simulation two sensitivity tests are considered in the present case study. Here it is of considerable interest to study possible reduction in the number of bootstrap samples and other means to reduce the amount of calculations. An interesting comparison to the baseline simulation would be to reduce the number of bootstrap samples to 100 samples. A more thorough analysis of the effects of sample size is described in subsection 3.6.

Another interesting sensitivity test would be a bootstrap procedure conditional on weights obtained at the pre-estimation stage, i.e. use the same (fixed) likelihood weights throughout the simulation. The reason for this comparison is twofold, it is computational as the amount of calculations required would be drastically reduced and also in relation to Hessian-based approaches. One should note however that with this bootstrap the estimation is not the same function of the data as the procedure where the weighting takes place for each dataset. This may lead to inappropriate weights for a given dataset which in turn can, as mentioned earlier, lead to inaccurate parameter estimates.

3.5 Hessian-based inference

For illustrative purposes the inferences arising from the bootstrap procedure presented here is compared to a Hessian-based confidence interval (described by Tinker et al. 2006, and references therein). In particular central differences were used to calculate the needed second derivatives used to obtain an estimate of the variance-covariance matrix and a multivariate delta method (Oehlert 1992) to obtain the confidence interval for derived biomass.

The effects of samples size on the inferences obtained from the inverted Hessian matrix were studied by an artificial increase in measurements. The timestep length was varied between 1, 2 and the baseline 3 months, with input files being adjusted accordingly. The resulting CVs for the recruitment parameters were estimated and the effects of the different step lengths contrasted. Similar analysis was conducted for the proposed bootstrap procedure but is, for the sake of clarity, only discussed in connection to the Hessian based approach.

3.6 Number of bootstrap samples

With regards to the bootstrap procedure itself this study also examines the effect of the number of bootstrap samples on the variance and bias estimates using a retrospective bootstrap. For a sample number n , ranging from 25 to 1000 bootstrap samples, n vectors of parameter estimates from the baseline bootstrap were sampled with replacement 100 times. From those 100 samples the coefficient of variation (CV) was calculated for the mean and standard deviation of each parameter. Uncertainty in bias estimation is harder to quantify in a similar way as parameter bias is often estimated close to zero.

3.7 Model output

Given the optimised parameter estimates it is possible to output a wide range of descriptors of the model ecosystem as Gadget operates on and stores the number in each age-length cell for each time-step of the model. For this study, the estimated parameters along with a derived biomass trajectory (age 4+) are considered. Comparisons of uncertainty estimates will be, as noted earlier, made using the three bootstrap variants, i.e. both 1000 and 100 bootstrap simulations with the iterative reweighting procedure applied to all bootstrap samples and 1000 bootstrap simulations using fixed weights, and the Hessian-based approach. A schematic overview of all calculations performed here is shown in figure 2.

Fig. 2
near here

4 Results

The simplest model outputs are the point estimates of model parameters. Fig. 3 gives histograms of bootstrap estimates of several parameters. It compares the distributions of those parameter estimates from 1000 bootstrap samples, either using reweighting for each dataset or fixed weights, to those using only 100 samples with reweighting. For each parameter, the point estimate from the full data set, the median of the bootstrap estimates and 95% confidence intervals from a hessian-based approximation are indicated. The differences between the point estimate and the bootstrap mean can be seen to be relatively minor, i.e. there is no obvious sign of an estimation bias, in all cases except for the length update (see β in eq. 3). It should be noted that the maturation parameters are correlated, affecting the relationship between the point estimate and bootstrap mean for the maturation. The different bootstrap methods exhibit similar distribution of parameter estimates with the exception of the length update where the bootstrap mean based on the original weights falls closer to the point estimate thus failing to detect bias in the length update.

Fig. 3
near here

Boxplots can be used to illustrate bootstrapped trajectories of various abundance or biomass measures. The estimate of the 4+ biomass is shown in fig. 4. It is seen that the main variation appears, in absolute terms, in the initial and final years, while only the final year shows considerable amount of variation in terms of CV. The initial and final years are of course considerably different

Fig. 4
near here

from the intermediate ones, but in different ways. The number of fish in the initial year are part of the estimation procedure and therefore of a different nature when compared to subsequent years. Further, the survey starts in 1985 (with the model starting in 1984), which makes the initial conditions somewhat poorly determined. The final years are on the other hand poorly determined since there is relatively little information in the objective function for the younger year classes as they have only been surveyed for a few years.

The same effects are seen for estimated recruitment at age 1 (fig. 5) where there is less variation in the earlier and intermediate years than the later years. As for the other parameters the Hessian-based confidence estimates are considerably smaller than those obtained using bootstrap methods. The CV of the Hessian-based approach roughly followed the same pattern as for those arising from the various bootstrap approaches but were generally around 12% of the corresponding bootstrap CV.

In figure 6 CVs for the mean and standard deviation of the model parameters are shown as function of the number of bootstrap samples, n , where the separate panels show different groups of parameters. The CV-estimates appear to fall close to $\frac{1}{\sqrt{n}}$, as shown in the figure, and that most of them are less than 15% for 100 bootstrap samples. The initial conditions, that is the numbers at age in 1984, had a somewhat higher CV for the mean and standard deviation as the other parameter groups. The initial numbers at age 8 and 9 in 1984 in particular, showed a considerably higher CV for all sample sizes. Those two age groups were, as noted earlier, poorly determined, and had a very low estimate compared to other initial numbers, as the corresponding year classes were only present in the data for the first few years of the model.

Hardly any biases were observed in this analysis. Notable exceptions were the length update parameter, shown in figure 3, and the first two years the 4+ biomass appeared to have a measurable bias. This was only detected in the bootstrap simulations where the iterative reweighting scheme was applied to all bootstrap samples. The fixed weight run and the Hessian-based approach failed to detect these differences.

The effects of the number of timesteps within a year can be seen in figure 7. There the CV of recruitment is illustrated as a function the number of (intra-year) timesteps in the model. The number of timesteps appears to be inversely proportional to the CV size. These effects were not, when varying the timestep, observed when conducting similar analysis using the bootstrap.

5 Discussion

This paper has presented a novel bootstrap method suitable for models of population dynamics. Several modifications and alternatives to the original bootstrap methodology (Efron 1979, Efron and Tibshirani 1994) have been presented. For example, to account for correlations in simple non-replacement sampling schemes (as used for most questionnaires or “sample surveys”), without-replacement bootstraps and with-replacement bootstraps have been suggested along with somewhat more general resampling procedures for complex survey data (McCarthy and Snowden 1985, Gross 1980, Rao and Wu 1988, Sitter 1992). Theoretical assumptions and derivations behind these approaches do not easily extend to the present situation with disparate data sets, composite likelihoods in the estimation phase and last but not least the highly nonlinear population dynamics models used as a basis for obtaining predicted values and error sums of squares or likelihood functions. The “trick” in the current proposal is not a theoretical development but the methodology of having the bootstrap sampling unit y_i as a collection of all relevant datasets sufficiently aggregated such that they can be assumed to be independent.

Some of the modifications of the original bootstrap have been developed for marine surveys (Smith 1997) but this has been intended to reflect e.g. the sampling design used for the surveys and simple estimation of quantities such as a stratified mean. In the present setting the data need to go through an aggregation procedure to be used in a nonlinear population dynamics model and it is the output of this model which is of interest, not variances in the input. Thus there is a need for the bootstrap to mimic this aggregation procedure for the full data from raw data or finer-scale aggregates. This is the case with any population dynamics or assessment model, used in fisheries

Fig. 5
near here

Fig. 6
near here

Fig. 7
near here

or other areas of resource harvesting particularly in a multispecies and multi-area context.

The methodology proposed here is certainly computationally intensive. However this is also the case for many other methods. For example, the MCMC evaluation of a Bayesian posterior involves a simulation of a correlated time-series whose stationary distribution is the posterior. This process is not trivially parallelizable over an arbitrary grid of computers (some of the difficulties are described in Wilkinson 2006). In comparison the bootstrap approach described here is fairly trivially distributed onto a computer cluster.

To make the bootstrap proposed here more feasible one could reduce the number of resampled datasets. Using 100 bootstrap replicates instead of 1000 yields satisfactory results in terms of variance estimation, allowing a drastic reduction in the computing time needed. Conditioning on the weights from the original sample could further reduce the time needed but, judging by the results presented here, possible estimation biases may be harder to detect.

When compared to the bootstrap the Hessian-based approximation appears to underestimate the uncertainty by a factor of 8. This may seem contrary to previous results. Magnusson et al. (2012), using a simple catch-at-age simulation model, concluded that MCMC-method and the Hessian-based approach performed similarly. And recently in Stewart et al. (2012) an MCMC and a Hessian-based approach performed similarly for real applications. The notable difference between the model described here and the aforementioned approaches is the objective function used here and total number of data points (defined in appendix A.2.4) used in the estimation process. The objective function consists of simple sums of squares that ignores potential correlations and tends to exaggerate the confidence level in the Hessian-approach as the number of datapoints increases. This is illustrated in figure 7 where it appears that the main factor in determining the size of the CV is the number of datapoints in the input files. Scale changes, such as aggregating data to larger lengthgroups or increasing the size of the plus group by lowering the modelled maximum age, would in this case increase the size of the CV by simply reducing the number of datapoints. In contrast, to the approach used here, a multinomial model, where the degrees of freedom are estimated, is often employed on catch at age (eg. Trenkel et al. 2012) but length distributions, in the case of Icelandic cod, have serious distributional problems (Hrafnkelsson and Stefansson 2004). Future work on the model could potentially evaluate different distributional assumptions similar to those suggested above using the proposed bootstrap approach.

In this particular case study there were no discernible biases detected. Thus the consequences of the Hessian-based approach appear to be mostly restricted to narrower confidence intervals. However it is reasonable to assume that inconsistencies arising from conflicting data sources (eg. in Stefansson 2003, Schnute and Hilborn 1993) would not be detected without analysing the effects of their relative weights. On the other hand, incorrect variance estimates may directly affect how annual catches are set. This occurs for example if a harvest control rule were to be based on a probabilistic measure such as a that of a biomass not falling below a threshold or a TAC not deviating too much from a target.

It is of considerable interest to compare the proposed bootstrap method to MCMC methods used in the Bayesian framework. This is however outside of the scope of this study as it would require a considerable effort to adapt the Gadget framework to the Bayesian one. Future work could potentially focus on the evaluation of the two methodologies both on simulated datasets and for real applications similar as was done in Hannesson et al. (2009).

It is reassuring that the modelled years in which the greatest uncertainty is found are those where it is expected i.e. the initial year and then increasing towards the end of the modelled time period. The first year is the most data poor with no survey data or age-length compositions and towards the end of the time period there are fewer cohorts with data available for most ages.

The method described here is designed to alleviate several known problems with other methods of uncertainty estimation. Several issues remain, however. For example if a model is too “stiff” through fixing parameters or other assumptions, then this may not be detected here except in special cases. These considerations could be explored by different models, e.g. split the commercial fleet component by gears, which can be implemented within the Gadget framework. On a related note there is also a balance to be found between estimation errors due to too small size classes and distribution error caused by too large size classes (Vandermeer 1978). It is therefore of interest to

investigate the effects of the choice of scale such as size class width but also time step (Drouineau et al. 2009). The relative merits of these models can then be evaluated using an approach similar to the one proposed here. Similarly different modelling approaches, such as the different data weighting discussed in Francis (2011) or Hu and Zidek (2002), can be also be compared using the bootstrap technique presented here. Ultimately, each reweighting scheme is a different method for obtaining a point estimate and the bootstrap is a perfectly general method to obtain variance estimates.

When designing an aggregated data base to be used for modelling several issues need to be taken into account. The most important statistical condition on the choice of the “data units” is that correlations between them should be minimal. On the other hand there also needs to be a fair number of them within each model area if the bootstrap mechanism is to provide some variation in results. For a given measurement type one can in many cases investigate spatial correlation or variograms to determine the distances at which those become negligible (Petitgas 2001). This can not easily be done for many data types, however (age-length tables, tagging experiments etc). In fact, the original reasoning for the areas used in this paper was ecological (Stefansson and Palsson 1997a, Taylor 2003) rather than based on spatial correlation, and it is likely that in most real situations data will be aggregated according either to such criteria or pragmatically into “statistical rectangles” of some form.

Simple bootstrap resampling usually assumes that the elementary data units, $\{\mathbf{y}_1, \dots, \mathbf{y}_n\}$, behave like independently and identically distributed samples. Data in fisheries tends to be collected in a somewhat stratified manner, ranging from formal stratification to attempts to “spread out” sampling, across gears, time and space. In the present setup this is simply ignored. This can be justified when the data are aggregated in a simple manner (through sums or averages) anyway since the bootstrap method then mimics the computation accordingly and/or when there are a large number of data units which can be viewed as representing a population of such units. In cases when one or a few of the subdivisions represent e.g. a spawning area and the intended analysis is stratified accordingly, this approach can clearly not be used since then the bootstrap resampling does not reflect the computational method in use. When such issues arise, whether with respect to fishing gear, space or other units, an appropriate approach is to include these elements in the model. For example the likelihood function can incorporate the various fishing gears, modelling each selectivity separately. The resampling then takes place separately for each gear.

Acknowledgements

Much of the work described here was undertaken while authors BPE, LT, VK and GS were employed at the MRI (Marine Research Institute, Reykjavik) and uses data from the MRI databases. The Gadget code has been in development for more than a decade by many programmers at the MRI and IMR (Institute of Marine Research, Bergen). The work was supported in part by EU grants QLK5-CT1999-01609 and FP6 TP8.1 502482, as well as a grant from The Icelandic Centre for Research (Rannis).

The authors would like to thank Dr S. Gavaris for useful discussions, which have considerably improved the paper.

References

- Babak, O., Hrafnkelsson, B., and Palsson, O. 2007. Estimation of the Length Distribution of Marine Populations in the Gaussian-multinomial Setting using the Method of Moments. *Journal of Applied Statistics*, 34(7-8):985–996.
- Begley, J. 2004. Gadget User Manual. Technical Report 120, Marine Research Institute.
- Björnsson, H. and Sigurdsson, T. 2003. Assessment of golden redfish (*Sebastes marinus* L.) in Icelandic waters. *Scientia Marina*, 67(Suppl 1):301–314.
- Bogstad, B., Tjelmeland, S., Tjelda, T., and Ulltang, O. 1992. Description of a multispecies model for the Barents Sea (multspec) and a study of its sensitivity to assumptions on food preferences and stock sizes of minke whales and harp seals. Technical report, SC/44/O 9.
- Chen, Y., Breen, P., and Andrew, N. 2000. Impacts of outliers and mis-specification of priors on bayesian fisheries-stock assessment. *Canadian Journal of Fisheries and Aquatic Sciences*, 57(11):2293–2305.
- Davison, A. and Hinkley, D. 1997. *Bootstrap Methods and Their Application*. Cambridge University Press.
- De Croos, M. and Stefansson, G. 2011. A case study of sampling strategies for estimating the length composition of commercial catches: The sri lankan shrimp trawl fishery. *Crustaceana*, 84, 12(13):1581–1591.
- Demyanov, V., Wood, S. N., and Kedwards, T. J. 2006. Improving ecological impact assessment by statistical data synthesis using process-based models. *Journal of the Royal Statistical Society Series C*, 55(1):41–62.
- Drouineau, H., Mahévas, S., Bertignac, M., and A., F. 2009. Assessing the impact of discretisation assumptions in a length-structured population growth model. *Fisheries Research*, 91(2-3):160–167.
- Efron, B. 1979. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1):1–26.
- Efron, B. and Tibshirani, R. J. 1994. *An Introduction to the Bootstrap*. Chapman & Hall/CRC.
- Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., Nielsen, A., and Sibert, J. 2012. Ad model builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, 27(2):233–249.
- Francis, R. 2011. Data weighting in statistical fisheries stock assessment models. *Canadian Journal of Fisheries and Aquatic Sciences*, 68(6):1124–1138.
- Gavaris, S. 1988. An adaptive framework for the estimation of population size. *CAFSAC Res. Doc*, 88(29):12.
- Gavaris, S. and Ianelli, J. N. 2001. Statistical issues in fisheries stock assessment. *Scand. J. Statistics: Theory and Appl.*, 29, pages 245–272.
- Gavaris, S., Patterson, K. R., Darby, C. D., Lewy, P., Mesnil, B., Punt, A. E., Cook, R. M., Kell, L. T., O’Brien, C. M., Restrepo, V. R., Skagen, D. W., and Stefansson, G. 2000. Comparison of uncertainty estimates in the short term using real data. ICES (Int. Counc. Explor. Sea) Document C.M. 2000/V:03.
- Gross, S. 1980. Median estimation in sample surveys. *Proceedings of the Section on Survey Research Methods*, pages 181–184.

- Gudmundsdóttir, Á., Steinarsson, B. Æ., and Stefánsson, G. 1988. A simulation procedure to evaluate the efficiency of some otolith and length sampling schemes. *ICES (Int. Counc. Explor. Sea) C.M. 1988/D:14*, 500:14.
- Gudmundsson, G. 2013. Fish stock assessment by time series analysis. Presented at ICES NWWG.
- Hannesson, S., Jakobsdottir, A., Begley, J., Taylor, L., and Stefansson, G. 2009. On the use of tagging data in statistical multispecies multi-area models of marine populations. *ICES Journal of Marine Science*, 65(9):1762–1772.
- Helle, K. and Pennington, M. 2004. Survey design considerations for estimating the length composition of the commercial catch of some deep-water species in the northeast Atlantic. *Fisheries Research*, 70(1):55–60.
- Hrafnkelsson, B. and Stefansson, G. 2004. A model for categorical length data from groundfish surveys. *Canadian Journal of Fisheries and Aquatic Sciences*, 61(7):1135–1142.
- Hu, F. and Zidek, J. V. 2002. The weighted likelihood. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 30(3):pp. 347–371.
- ICES, C. 2011. Report of the north western working group. Technical report, International Council for the Exploration of the Sea.
- Jennrich, R. I. 1969. Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, 40(2):633–643.
- Kupca, V. 2004. A standardized database for fisheries data. ICES (Int. Counc. Explor. Sea) Document C.M. 2004/FF:15.
- Kupca, V. and Sandbeck, P. 2003. dst^2 Datawarehouse structure and data import. In *dst²: Development of structurally detailed statistically testable models of marine populations*, volume 98 of *Technical Report*, pages 37–46. Marine Research Institute, Reykjavik.
- Lindstrøm, U., Smout, S., Howell, D., and Bogstad, B. 2009. Modelling multi-species interactions in the barents sea ecosystem with special emphasis on minke whales and their interactions with cod, herring and capelin. *Deep Sea Research Part II: Topical Studies in Oceanography*, 56(21):2068–2079.
- Magnusson, A., Punt, A. E., and Hilborn, R. 2012. Measuring uncertainty in fisheries stock assessment: the delta method, bootstrap, and mcmc. *Fish and Fisheries*.
- Maunder, M. N. and Punt, A. E. 2012. A review of integrated analysis in fisheries stock assessment. *Fisheries Research*.
- McCarthy, P. J. and Snowden, C. B. 1985. The bootstrap and finite population sampling. *Vital Health Stat 2*, 95:1–23.
- Methot, R. D. 1989. Synthetic estimates of historical abundance and mortality for northern anchovy. *Amer. Fish. Soc. Symp.*, 6:66–82.
- Millar, R. B. 2002. Reference priors for bayesian fisheries models. *Canadian Journal of Fisheries and Aquatic Sciences*, 59(9):1492–1502.
- Myers, R. A. and Cadigan, N. G. 1995. Statistical analysis of catch-at-age data with correlated errors. *Canadian journal of fisheries and aquatic sciences(Print)*, 52(6):1265–1273.
- Oehlert, G. 1992. A note on the delta method. *American Statistician*, pages 27–29.
- Pálsson, O. K., Schopka, S. A., Stefansson, G., and Steinarsson, B. A. 1989. Icelandic groundfish survey data used to improve precision in stock assessments. *J. Northw. Atl. Fish. Sci.*, 9:53–72.

- Patterson, K., Cook, R., Darby, C., Gavaris, S., Kell, L., Lewy, P., Mesnil, B., Punt, A., Restrepo, V., Skagen, D. W., and Stefansson, G. 2001. Estimating uncertainty in fish stock assessment and forecasting. *Fish and Fisheries*, 2(2):125–157.
- Pennington, M. and Volstad, J. H. 1994. Assessing the effect of intra-haul correlation and variable density on estimates of population characteristics from marine surveys. *Biometrics*, 50:1–8.
- Petitgas, P. 2001. Geostatistics in fisheries survey design and stock assessment: models, variances and applications. *Fish and Fisheries*, 2(3):231–249.
- Punt, A. E. and Hilborn, R. 1997. Fisheries stock assessment and decision analysis: the bayesian approach. *Reviews in Fish Biology and Fisheries*, 7(1):35–63.
- Rao, J. N. K. and Wu, C. F. J. 1988. Resampling inference with complex survey data. *Journal of the American Statistical Association*, 83(401):231–241.
- Richards, L. 1991. Use of contradictory data sources in stock assessments. *Fisheries research*, 11(3-4):225–238.
- Schnute, J. T. and Hilborn, R. 1993. Analysis of contradictory data sources in fish stock assessment. *Canadian Journal of Fisheries and Aquatic Sciences*, 50(9):1916–1923.
- Sigurdsson, T., Hjorleifsson, E., Bjornsson, H., and Palsson, O. K. 1997. Fall groundfish survey in icelandic waters (stofnmaeling botnfiska a Islandsmidum haustid 1996). Technical Report 61, Marine Research Institute. (in Icelandic).
- Singh, W., Thordarson, G., Haputhantri, S., and Stefansson, G. 2011. Identification of year classes in length distributions.
- Sitter, R. R. 1992. A resampling procedure for complex survey data. *Journal of the American Statistical Association*, 87(419):755–65.
- Smith, S. J. 1997. Bootstrap confidence limits for groundfish trawl survey estimates of mean abundance. *Canadian Journal of Fisheries and Aquatic Sciences*, 54(3):616–630.
- Spiegelhalter, D., Thomas, A., Best, N., and Gilks, W. 1996. Bugs 0.5: Bayesian inference using gibbs sampling manual (version ii). *MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK*.
- Stefansson, G. 1998. Comparing different information sources in a multispecies context. In Funk, F., II, T. Q., Heifetz, J., Ianelli, J., Powers, J., Schweigert, J., Sullivan, P., and Zhang, C., editors, *Fishery Stock Assessment Models: Proceedings of the international symposium; Anchorage 1997*, 15th Lowell Wakefield Fisheries Symposium, pages 741–758.
- Stefansson, G. 2003. Issues in multispecies models. *Natural Resource Modeling*, 16(4):415–438.
- Stefansson, G. 2005. Growth models in population simulations. In *Final Report: dst^2 : Development of structurally detailed statistically testable models of marine populations*, volume 118 of *Technical Report*, pages 73–83. Marine Research Institute, Reykjavik.
- Stefansson, G. and Palsson, O. K. 1997a. BORMICON A Boreal Migration and Consumption model. Technical Report 58, Marine Research Institute.
- Stefansson, G. and Palsson, O. K. 1997b. 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. and Palsson, O. K. 1998. A framework for multispecies modelling of boreal systems. *Reviews in Fish Biology and Fisheries*, 8:101–104.

- Stewart, I. J., Hicks, A., Taylor, I., Thorson, J. T., Wetzel, C., and Kupschus, S. 2012. A comparison of stock assessment uncertainty estimates using maximum likelihood and bayesian methods implemented with the same model framework. *Fisheries Research*.
- Taylor, L. 2002. Stock dynamics and long term assessment of haddock (*Melanogrammus aeglefinus*) in icelandic waters. Presented at ICES North-Western working group (available as <http://www.hi.is/gunnar/papers/ltdyn.pdf>).
- Taylor, L. 2003. Definition of areas in Icelandic waters. In *dst²: Development of structurally detailed statistically testable models of marine populations*, volume 98 of *Technical Report*, pages 222–230. Marine Research Institute, Reykjavik.
- Taylor, L., Begley, J., Kupca, V., and Stefansson, G. 2007. A simple implementation of the statistical modelling framework Gadget for cod in Icelandic waters. *African Journal of Marine Science*, 29(2):223–245.
- Tinker, M. T., Doak, D. F., Estes, J. A., Hatfield, B. B., Staedler, M. M., and Bodkin, J. L. 2006. Incorporating diverse data and realistic complexity into demographic estimation procedures for sea otters. *Ecol Appl*, 16(6):2293–312.
- Tjelmeland, S. and Bogstad, B. 1989. *MULTSPEC: The manual*. Institute of Marine Research, Bergen.
- Trenkel, V. M., Bravington, M. V., and Lorance, P. 2012. A random effects population dynamics model based on proportions-at-age and removal data for estimating total mortality. *Canadian Journal of Fisheries and Aquatic Sciences*, 69(11):1881–1893.
- Vandermeer, J. 1978. Choosing category size in a stage projection matrix. *Oecologia*, 32(1):79–84.
- Wilkinson, D. J. 2006. Parallel bayesian computation. *Statistics textbooks and monographs*, 184:477.

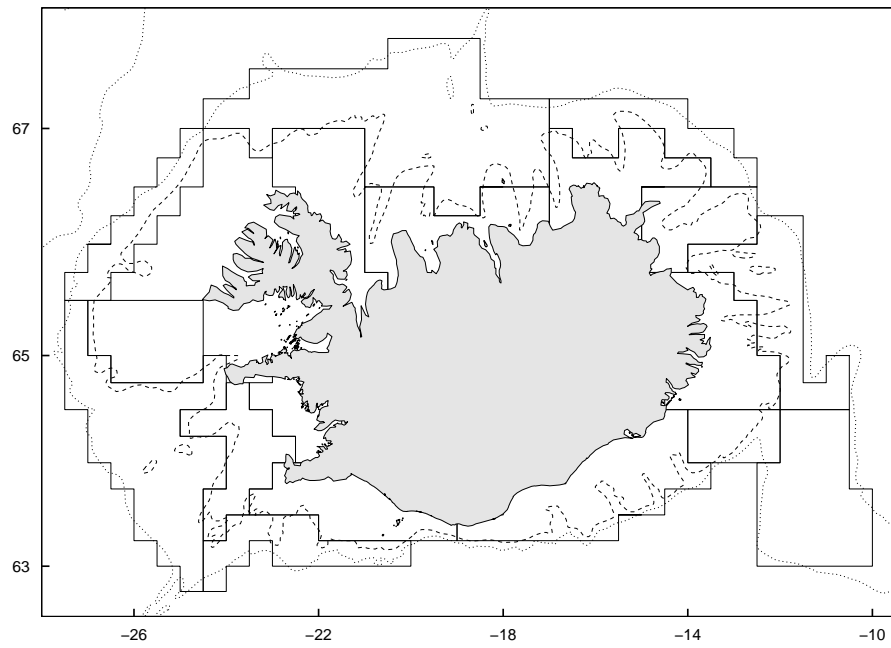


Figure 1: The spatial structure of data storage on the Icelandic coastal shelf along with 200m (broken line) and 500m depth (pointed line) contours. These areas are referred to as “subdivisions”. A given timeperiod, timestep size and subdivision is referred to as a “elementary sampling unit”.

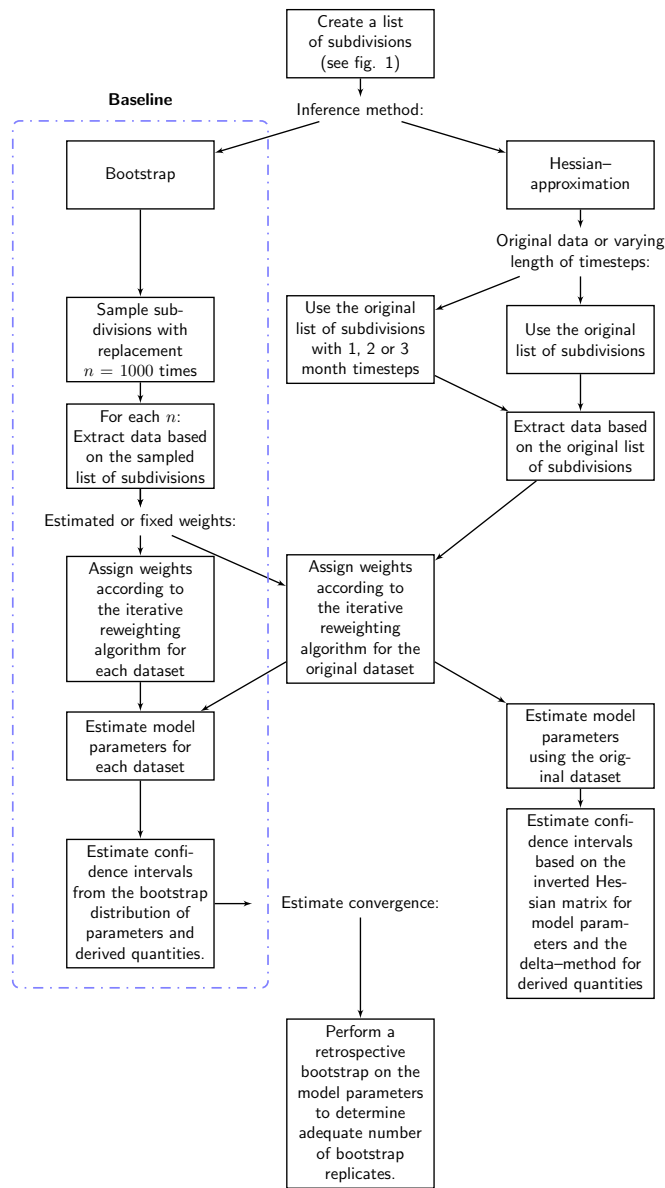


Figure 2: A flowchart of the calculations performed. Boxes indicate action and unbounded text possible uncertainty estimation variants or decisions.

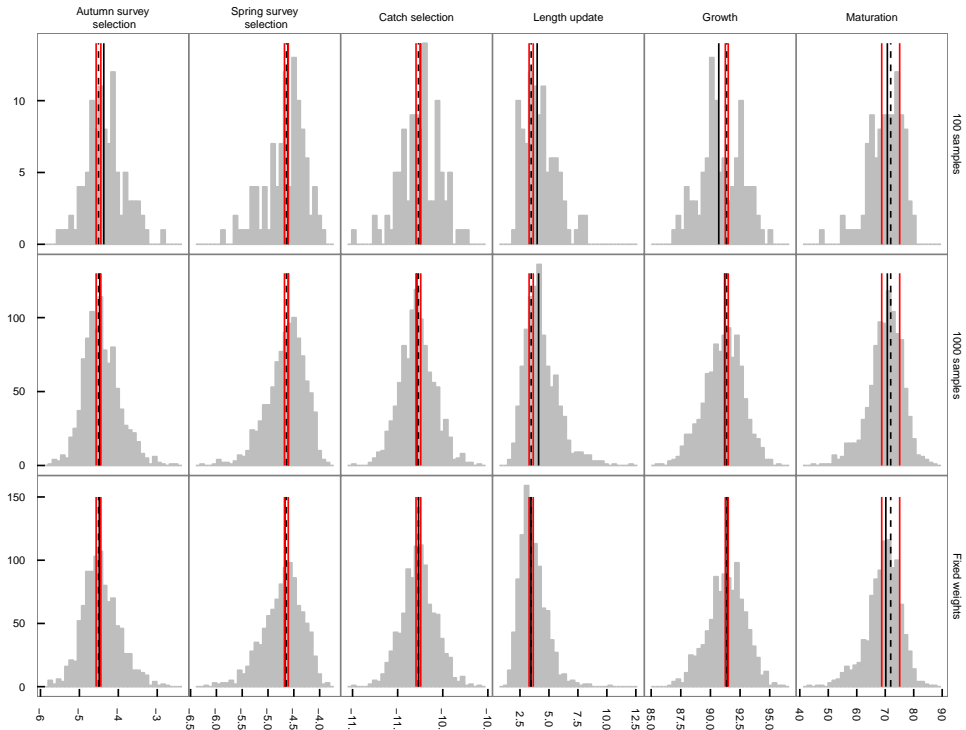


Figure 3: Histograms of the estimated fleet selection parameter a_f for the three fleets (October survey, March survey, Commercial catch), β the parameter defining the length update matrix, k the growth rate and the maturity l_{50} . The parameter estimates were obtained from 1000 bootstrap samples, compared to a smaller number of bootstrap samples, 100, where for the two number of samples iterative weighting applied to all bootstrap samples. This is then all compared to 1000 bootstrap samples where in the parameter estimation the weighted likelihood function is conditioned on the original weights. The point estimate (black broken line) and bootstrap mean (black solid line) along with 95% confidence bound obtained from a Hessian-based approximation to the variance covariance matrix (red solid lines) are indicated.

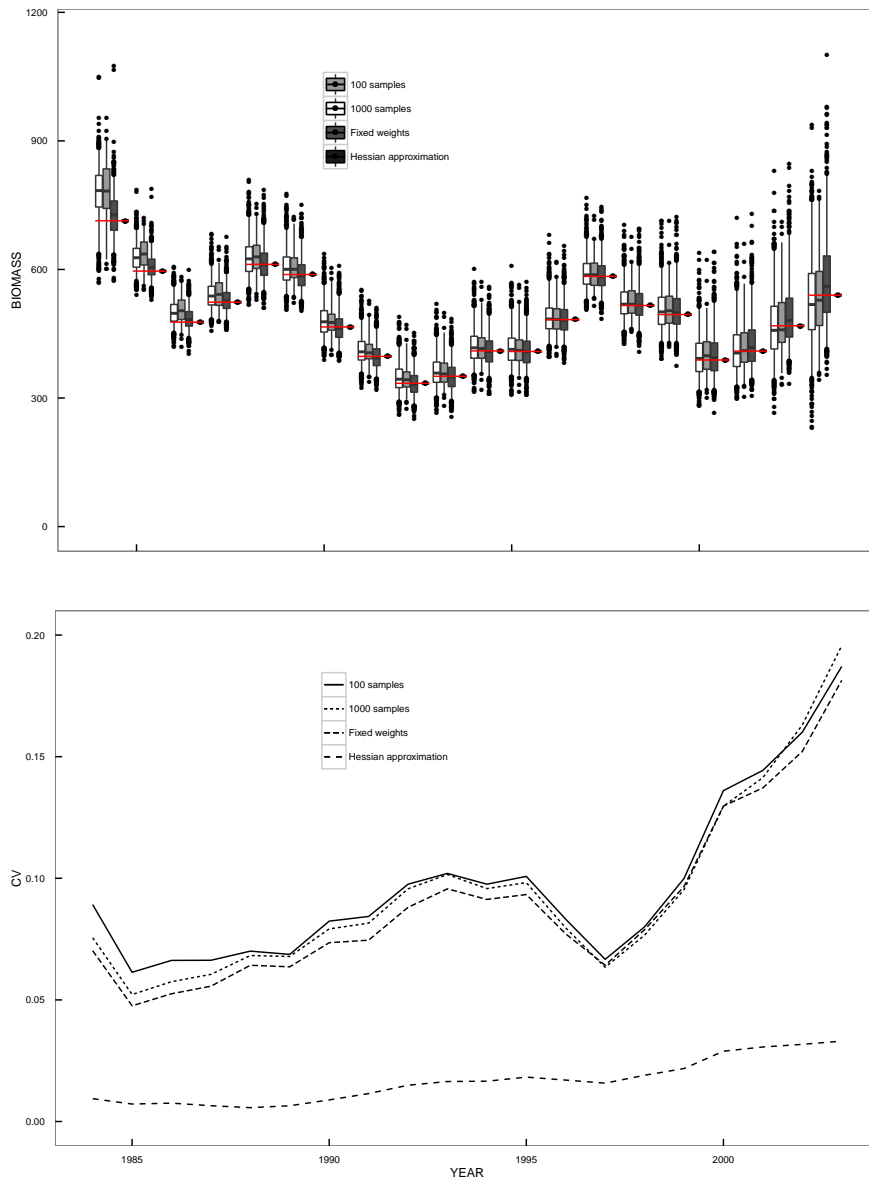


Figure 4: Boxplot (top panel) of the end of year biomass for cod of age 4 and older estimated on 1000 bootstrap samples, both using iterative weighting for each sample and using the fixed weights for all samples, compared to 100 bootstrap samples. The fixed weights were obtained using iterative weighting for the original dataset. The point estimate is indicated by the central red line through the boxes. The box indicates the interquartile range and the whiskers 95% confidence intervals. Any further outlying data points are indicated as points. Bottom panel shows the estimated CV for the age 4⁺ biomass using the same methods as above.

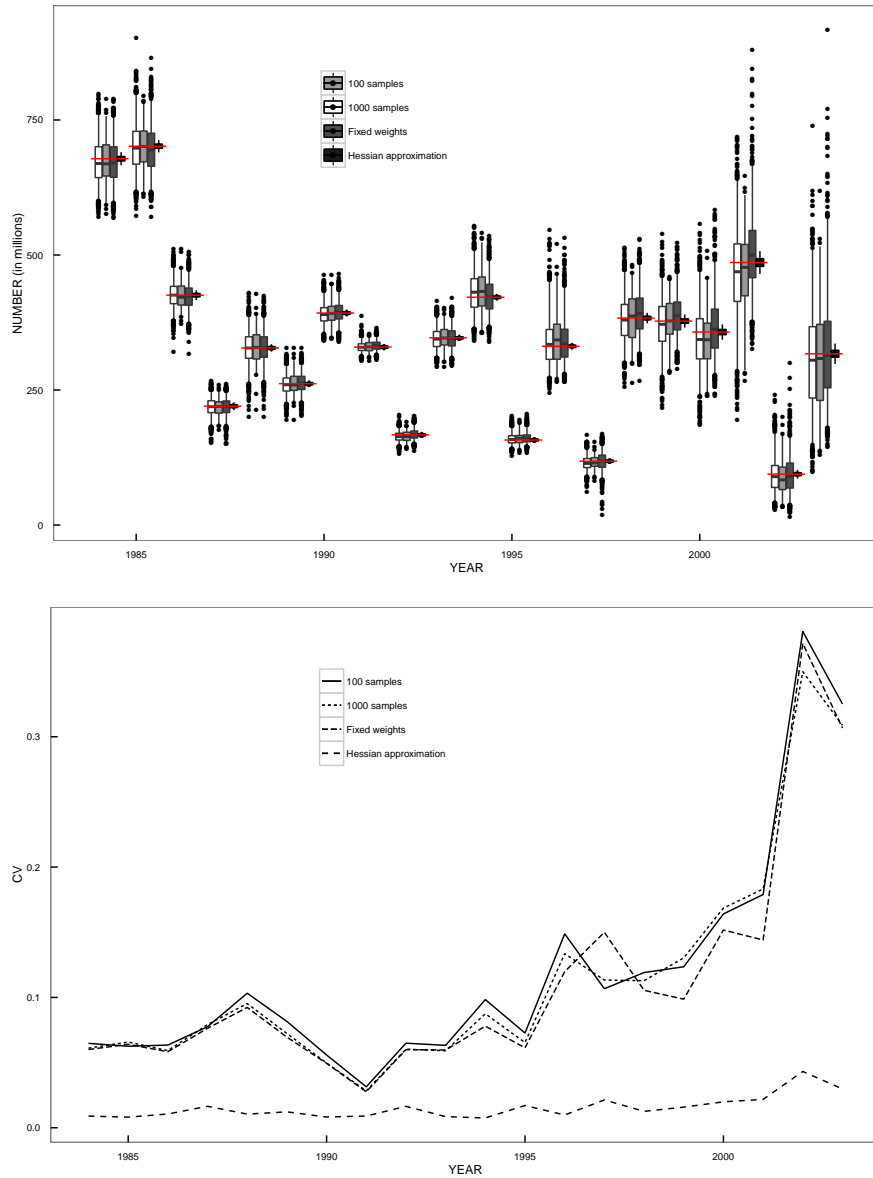


Figure 5: Boxplot (top panel) of the number of recruits (age 1) in each year estimated by 1000 and 100 bootstrap models compared to 1000 bootstraps with fixed weights and a Hessian-based approximation to the 95% confidence interval. The point estimate indicated by a central red line through the boxes. Bottom panel shows the estimated CV for the recruitment using the same methods as above.

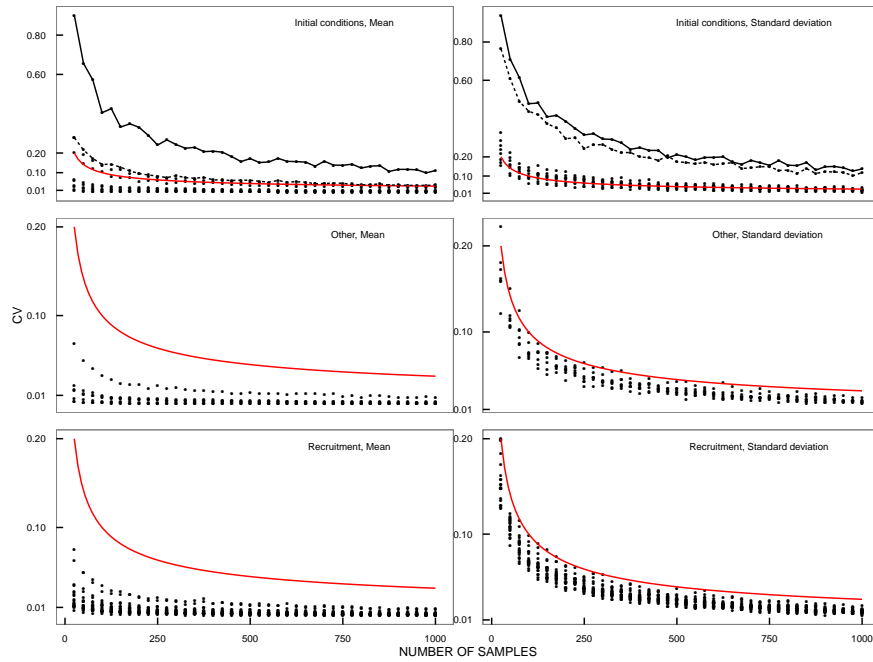


Figure 6: Results of a retrospective bootstrap sampling on the parameter estimates from the 1000 bootstrap samples, with iterative weighting applied to all samples. This retrospective bootstrap studies the variation of the mean and standard deviation of each parameter estimate by calculating the coefficient of variation (CV) as function of the number of bootstrap samples, n , of both the mean and standard deviation (SD). A point on the graph shows the CV of the mean (panels on the left hand side) or SD (panels on the right hand side) for a particular parameter and number of samples, n . The different panels contain the CVs of the initial number at age (ν_a in eq. 6), “Other” variables i.e. the variables which are shown in figure 3, and yearly recruitment shown in figure 5 (R_y in eq. 5). CV of the initial number at ages 8 and 9 are illustrated with solid and broken lines respectively. For comparison $1/\sqrt{n}$ is shown (red solid line) on all panels.

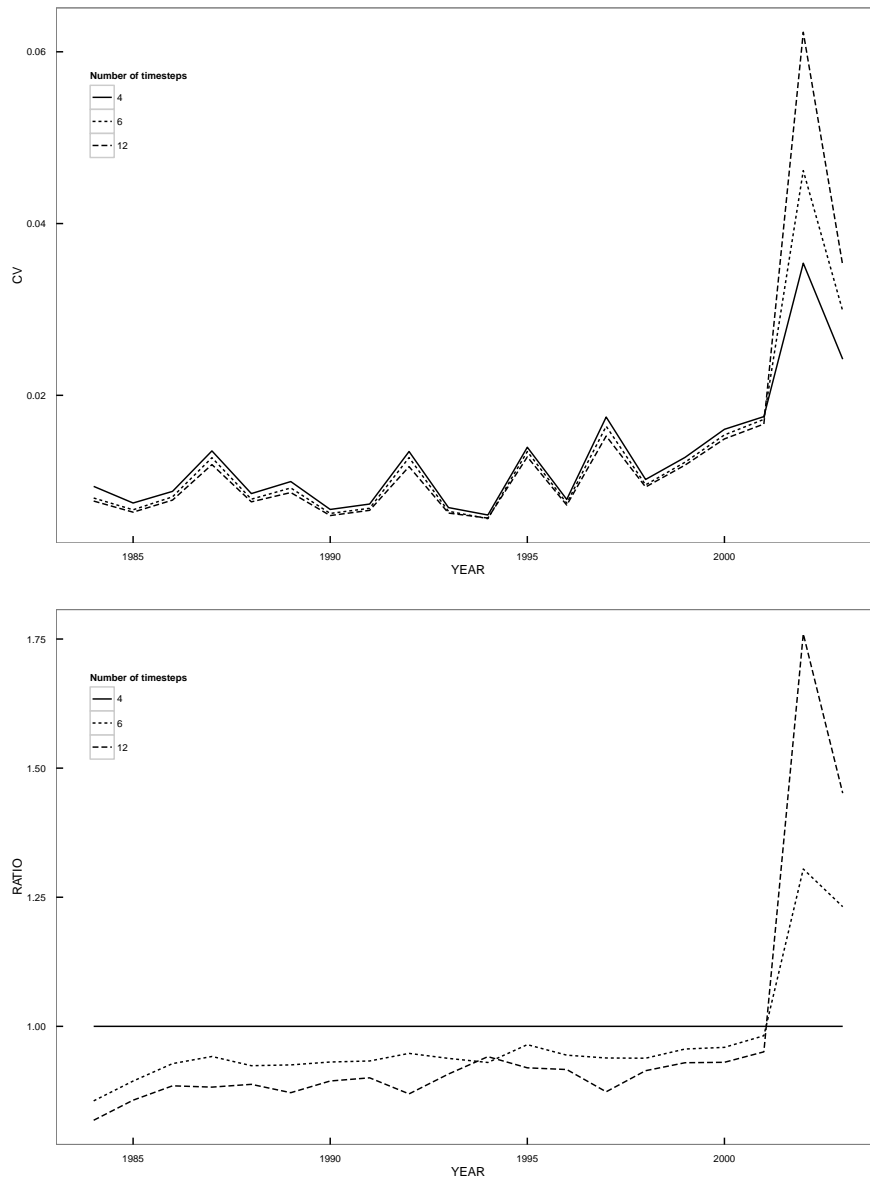


Figure 7: The CV of recruitment arising from the inverted Hessian (upper panel) by year as the number of intra-year timesteps are increased. The bottom panel shows the ratio of the CV of the model with 3 month timesteps to the models with 2 months and 1 month steps.

A Model description

A.1 Simulation model

The following describes the technical details of the Gadget model used in this case study. The model was previously described in Taylor et al. (2007) along with a statistical assessment of its fit to data. In the model the simulated quantity is the number of individuals, N_{alsyt} , at age $a = 1 \dots 12$, in a lengthgroup l , representing lengths ranging between 4 and 140 cm in 2 cm lengthgroups, stock component s where $s = 0, 1$ denotes the mature and immature stock component respectively, at year y which is divided into quarters $t = 1 \dots 4$. The length of the timestep is denoted Δt . The population is governed by the following equations:

$$\begin{aligned}
 N_{alsy,t+1} &= \sum_{l'} G_l^{l'} [(N_{al'syt} - C_{fal'st})e^{-M_a\Delta t} + I_{al'syt}] && \text{if } t < 4 \\
 N_{a+1,ls,y+1,1} &= \sum_{l'} G_l^{l'} [(N_{al'sy,4} - C_{fal's,4})e^{-M_a\Delta t} + I_{al'sy,4}] && \text{if } t = 4 \text{ and } a < 12 \\
 N_{a,ls,y+1,1} &= \sum_{l'} G_l^{l'} (N_{al'sy,4} - C_{fal'sy,4} + N_{a-1,l'sy,4} - C_{f,a-1,l'sy,4})e^{-M_a\Delta t} && \text{if } t = 4 \text{ and } a = 12
 \end{aligned} \tag{1}$$

where $G_l^{l'}$ is the proportion in lengthgroup l that has grown $l' - l$ lengthgroups in Δt , C_{falsyt} denotes the catches by fleet $f \in \{S, A, C\}$, S, A and C denote the spring and autumn surveys and commercial fleets respectively¹, M_a the natural mortality at age a and I_{alsyt} denotes the movement from the immature to the mature stock components. A short note on notation, here l is used interchangeably as either the lengthgroup or the midpoint of the length interval for that particular lengthgroup, depending on the context.

A.1.1 Growth

Growth in length is modeled as a two-stage process, an average length update in Δt and a growth dispersion around the mean update (as described in Stefansson 2005). The average length update per time step is set according to a simplified form of the Von Bertalanffy equation:

$$\Delta l = (l_\infty - l)(1 - e^{-k\Delta t}) \tag{2}$$

where l_∞ is the terminal length and k is the annual growth rate. In the second step the growth is dispersed according to a beta-binomial distribution parameterised by the following equation:

$$G_l^{l'} = \frac{\Gamma(n+1)}{\Gamma((l' - l) + 1)} \frac{\Gamma((l' - l) + \alpha)\Gamma(n - (l' - l) + \beta)}{\Gamma(n - (l' - l) + 1)\Gamma(n + \alpha + \beta)} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \tag{3}$$

where α is subject to

$$\alpha = \frac{\beta\Delta l}{n - \Delta l} \tag{4}$$

where n denotes the maximum length group growth and $(l' - l)$ the number of lengthgroups grown.

A.1.2 Recruitment and initial abundance

A fairly simple model is used for recruitment in this exercise. Recruitment enters to the population according to:

$$N_{110yt'} = R_y p_l \tag{5}$$

where t' denotes the recruitment time-step and R_y is the yearly recruitment. p_l is the proportion in lengthgroup l that is recruited which is determined by a normal density with mean according to the growth model and variance σ_y^2 .

¹The survey fleets catches are given a nominal catch to allow for survey age and length distribution predictions.

Equally simple is the formulation of initial abundance in numbers for each age group in length-group l is:

$$N_{als11} = \nu_a q_l \quad (6)$$

where ν_a is the initial number at age a in the initial year and q_l the proportion at lengthgroup l which is determined by a normal density with a mean according to the growth model in equation 2 and variance σ_a^2 .

A.1.3 Maturation

Two stage maturity is modeled and represented by the two stock components. The movement between the two components is formulated as

$$I_{alsyt} = \begin{cases} \sum_{l'} N_{al'0y,t-1} \times m_{l'}^l & \text{if } s = 1 \text{ and } t > 1 \\ \sum_{l'} N_{al'0y-1,4} \times m_{l'}^l & \text{if } s = 1 \text{ and } t = 1 \\ -\sum_{l'} N_{al'0y,t-1} \times m_{l'}^l & \text{if } s = 0 \text{ and } t > 1 \\ -\sum_{l'} N_{al'0y-1,4} \times m_{l'}^l & \text{if } s = 0 \text{ and } t = 1 \end{cases} \quad (7)$$

where $s = 0$, as noted above, denotes the immature stock component. and $m_{l'}^l$ is the proportion of immatures that mature between the lengths l and l' defined as:

$$m_{l'}^l = \frac{-\alpha G_{l'}^l (l - l') e^{-\lambda(l' - l_{50})}}{1 + e^{-\lambda(l' - l_{50})}} \quad (8)$$

A.1.4 Fleet operations

Catches are simulated based on reported total landings and a length based suitability function for each of the three fleets, commercial fleet and the autumn and spring survey. Total landings are assumed to be known and the total biomass is simply offset by the landed catch. The catches for lengthgroup l , fleet f at year y and time-step t are calculated as

$$C_{flsyt} = E_{ft} \frac{S_f(l) N_{lsyt} W_{ls}}{\sum_{s'} \sum_{l'} S_f(l') N_{l's'yt} W_{l's'}} \quad (9)$$

where E_{ft} is the landed biomass at time t and $S_f(l)$ is the suitability of lengthgroup l by fleet f defined as:

$$S_f(l) = \frac{1}{1 + e^{(-a_f - b_f l)}} \quad (10)$$

The weight, W_{sl} , at lengthgroup l is calculated according to the following stock component specific length – weight relationship:

$$W_{sl} = \mu_s e^{\omega_s l} \quad (11)$$

A.2 Observation model

In Gadget data are assimilated using a weighted log-likelihood function. Here four types of data enter the likelihood, length-based survey indices, length distributions from survey and commercial fleets, age – length distribution from the survey and commercial fleets and maturity at length for 3 year olds.

A.2.1 Survey indices

The survey indices are defined as the total number of fish caught in a survey within a certain length interval. The intervals used here are 16 – 25 cm, 26 – 38 cm and larger than 38 cm. These intervals are chosen such that they roughly represent age 1, age 2 and age 3⁺.

For each length range g the survey index is compared to the modelled abundance at year y and time-step t using:

$$l_{gf}^{SI} = \sum_y \sum_t (\log I_{gfy} - (\log q_f + \log \widehat{N}_{gyt}))^2 \quad (12)$$

where

$$\widehat{N}_{gyt} = \sum_{l \in g} \sum_a \sum_s N_{alsyt}$$

The above formulation assumes that the length-based indices are independent and uniform selectivity by survey gear. This is seen as a fair assumption as little correlation is observed in the data for the first two lengthgroups. With regards to survey selectivity this has been estimated to be fairly constant during the model time period (Gudmundsson 2013). For further implementation details see Taylor et al. (2007).

A.2.2 Maturity at length

Length at maturity comparison uses the number of mature cod that are age 3 observed in the Icelandic ground fish survey. The observed proportions are compared to the modelled proportion using sum of squares:

$$l^M = \sum_y \sum_t \sum_l (\pi_{lyt} - \hat{\pi}_{lyt})^2 \quad (13)$$

where

$$\pi_{lyt} = \frac{\sum_a O_{al1yt}}{\sum_a \sum_{l'} \sum_s O_{alsyt}}$$

and

$$\hat{\pi}_{lyt} = \frac{\sum_a N_{al1yt}}{\sum_a \sum_{l'} \sum_s N_{alsyt}}$$

i.e. the observed and modelled proportions mature respectively in length group l , year y and timestep t .

A.2.3 Fleet data

Length distributions are compared using 2 cm lengthgroups for both commercial and survey fleets using

$$l_f^{LD} = \sum_y \sum_t \sum_l (\pi_{lyt} - \hat{\pi}_{lyt})^2 \quad (14)$$

where f denotes the fleet where data was sampled from and

$$\pi_{lyt} = \frac{\sum_a \sum_s O_{alsyt}}{\sum_a \sum_{l'} \sum_s O_{alsyt}}$$

and

$$\hat{\pi}_{lyt} = \frac{\sum_a \sum_s N_{alsyt}}{\sum_a \sum_{l'} \sum_s N_{alsyt}}$$

i.e the observed and modelled proportions in lengthgroup l respectively at year y and timestep t . Similarly age – length data are compared using 4 cm length groups:

$$l_f^{AL} = \sum_y \sum_t \sum_a \sum_l \sum_s (\pi_{falsyt} - \hat{\pi}_{falsyt})^2 \quad (15)$$

where

$$\pi_{alyt} = \frac{\sum_s O_{alsyt}}{\sum_a \sum_{l'} \sum_s O_{alsyt}}$$

and

$$\hat{\pi}_{alyt} = \frac{\sum_s N_{alsyt}}{\sum_a \sum_{l'} \sum_s N_{alsyt}}$$

A.2.4 Iterative re-weighting

The total objective function used the modelling process combines equations 12 to 15 using the following formula:

$$l^T = \sum_g \sum_{f \in \{S,A\}} w_{gf}^{SI} l_{gf}^{SI} + \sum_{f \in \{S,A,C\}} (w_f^{LD} l_f^{LD} + w_f^{AL} l_f^{AL}) + w^M l^M \quad (16)$$

where $f = S, A$ or C denotes the spring survey, autumn survey and commercial fleets respectively and w 's are the weights assigned to each likelihood components.

The weights, w_i , are necessary for several reasons. First of all it is used to prevent some components from dominating the likelihood function. Another would be to reduce the effect of low quality data. It can be used as an *a priori* estimates of the variance in each subset of the data.

Assigning likelihood weights is not a trivial matter, has in the past been the most time consuming part of a Gadget model. Commonly this has been done using some form of 'expert judgement'. General heuristics have recently been developed to estimate these weights objectively. Here the iterative re-weighting heuristic introduced by Stefansson (2003), and subsequently implemented in Taylor et al. (2007), is used.

The general idea behind the iterative re-weighting is to assign the inverse variance of the fitted residuals as component weights. The variances, and hence the final weights, are calculated according to the following algorithm:

1. Calculate the initial sums of squares (SS) given the initial parameterization for all likelihood components. Assign the inverse SS as the initial weight for all likelihood components.
2. For each likelihood component, do an optimization run with the initial SS for that component set to 10000. Then estimate the residual variance using the resulting SS of that component divided by the degrees of freedom (df^*), i.e. $\hat{\sigma}^2 = \frac{SS}{df^*}$.
3. After the optimization set the final weight for all components as the inverse of the estimated variance from the step above (weight = $1/\hat{\sigma}^2$).

The number of non-zero data-points (df^*) is used as a proxy for the degrees of freedom. While this may be a satisfactory proxy for larger datasets it could be a gross overestimate of the degrees of freedom for smaller datasets. In particular, if the survey indices are weighed on their own while the yearly recruitment is estimated they could be over-fitted. In general problems such as these can be solved with component grouping, that is in step 2 the likelihood components that should behave similarly, such as survey indices, should be heavily weighted and optimized together. This approach is used here for the spring and autumn survey indices.

A.3 Fit to data

A detailed discussion of the model fit to the datasets listed in Table A.1 is beyond the scope of this discussion. This analysis can be found in Taylor et al. (2007). However for illustrative purposes a comparison between survey indices is shown in Figure A.1. The model appears to follow the general features of the survey indices, in particular the age 3+ groups. The indices representing age one are in general seen to be more variable than for the latter groups.

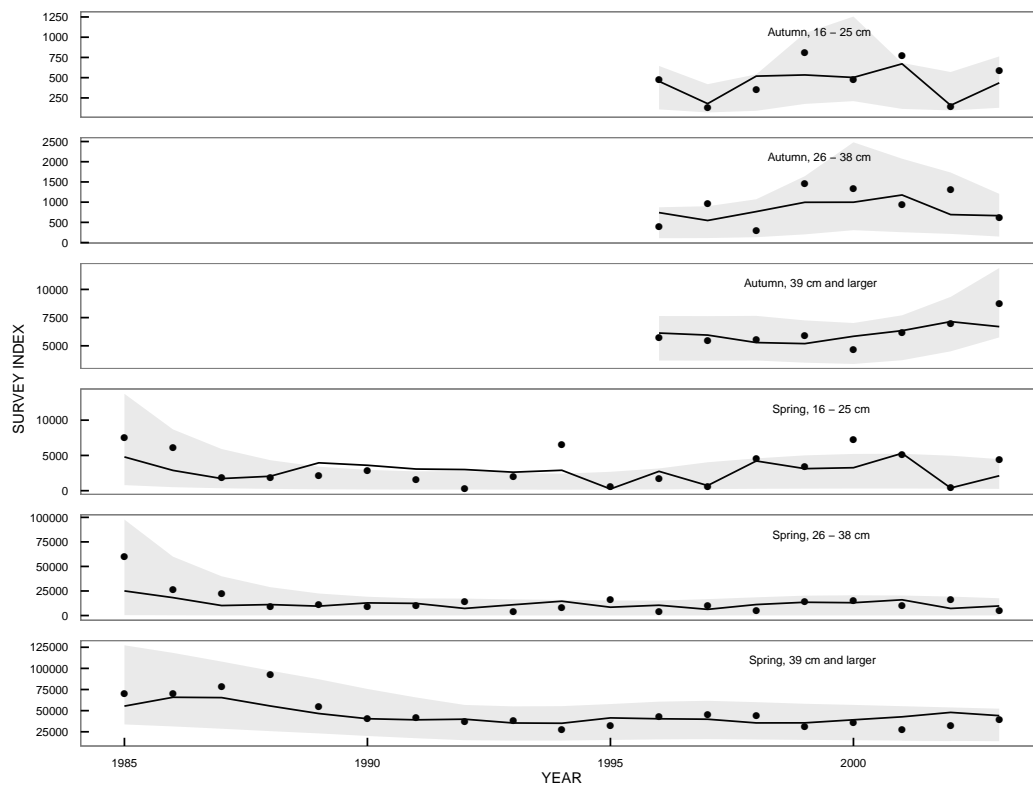


Figure A.1: Survey indices (points) by year compared to the model fit (lines) split by season and length groups. The length groups roughly represent the one, two and older agegroups. 95% confidence intervals from a loess smoother are shown for comparison.

Origin	Timespan	Length group size	Num. data-points	Likelihood function
	Length distributions:			
March Survey	1 st quarter, 1985 – 2003	2 cm	1292	See eq. 14
October Survey	4 th quarter, 1995 – 2003	2 cm	558	See eq. 14
Commercial catches	All quarters, 1984 – 2003	2 cm	5202	See eq. 14
	Age – length frequencies			
March Survey	1 st quarter, 1989 – 2003	4 cm	6120	See eq. 15
October Survey	4 th quarter, 1995 – 2003	4 cm	3348	See eq. 15
Commercial catches	All quarters, 1984 – 2003	4 cm	31042	See eq. 15
	Survey indices			
March Survey	1 st quarter, 1985 – 2003	–	57	See eq. 12
October Survey	4 th quarter, 1995 – 2003	–	24	See eq. 12
	Ratio of immature:mature at age 3 by length group			
March Survey	1 st quarter, 1985 – 2003	2 cm	1672	See eq. 13

Table A.1: Overview of the likelihood data used in the model. Survey indices are calculated from the length distributions and are disaggregated (“sliced”) into three groups which correspond roughly to age 1, age 2 and age 3+ (as in Taylor et al. 2007). Number of datapoints refer to aggregated data used as inputs in the Gadget model and represent the original dataset. All data can be obtained from the Marine Research Institute, Iceland.

Description	Notation	Comments	Formula
Natural mortality	M_a	Fixed at 0.2 for ages 3 to 9 while 0.5, 0.35 for the first two ages and 0.3, 0.5 and 0.7 for ages 10 – 12	See eq. 1
Growth function	k	l_∞ is fixed at 120 cm	See eq. 2
Growth implementation	β	n is fixed at 15 lengthgroups	See eq. 3
Fleet selection	a_f	b_f fixed at 0.09 for the survey fleets while 0.19 for the commercial fleet	See eq. 10
Maturity ogive	λ, l_{50}	l_{50} in cm	See eq. 7
Number of recruits by year	R_y	$y \in [1984, 2003]$. σ_y^2 , i.e. variance in recruitment length, based on length distributions obtained in the spring survey	See eq. 5
Initial abundance at ages 2 – 11 in 1984	η_a	$a \in [2, 11]$. σ_a^2 , i.e. variance in initial length at age a , based on length distributions obtained in the spring survey	See eq. 6
Survey catchability	q_f	Intercept term in a log–linear relationship with abundance. The slope term is assumed to be 1 for all indices	See eq. 12
Length–weight relationship	μ_s, ω_s	Different values by stock component, estimated outside of the model	See eq. 11

Table A.2: An overview of the estimated parameters in the model. For those parameters with fixed values a description of how these values were derived can be found in Taylor et al. (2007) and references therein.

References

Gudmundsson G. 2013. Fish stock assessment by time series analysis. Presented at ICES North West Working Group, 25 April – 2 May 2013, Copenhagen.

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

Stefansson G. 2005. Growth models in population simulations. In: Final Report. *dst² : Development of structurally detailed statistically testable models of marine populations*. Technical Report 118. Marine Research Institute, Reykjavik. pp 73 – 83.

Taylor L, Begley J, Kupca V, Stefansson G. 2007. A simple implementation of the statistical modelling framework Gadget for cod in Icelandic waters. *African Journal of Marine Science* 29: 223 – 245.