A bootstrap method for estimating bias and variance in statistical multispecies modelling frameworks using highly disparate data sets

Lorna Taylor[1], Verena M. Trenkel[2], Vojtech Kupca[3], Gunnar Stefansson[4]

RH-02-2011

Science Institute
University of Iceland
Taeknigardur, Dunhaga 5
107 Reykjavk

May 2011

[1]Science Institute, University of Iceland, Taeknigardur, Dunhaga 5, 107 Reykjavik, Iceland

[2]Ifremer, Nantes, France

[3]HPC2N, MIT-Building, Ume University, SE-901 87 Ume, Sweden

[4]Author for correspondence, Science Institute, University of Iceland, Taeknigardur, Dunhaga 5, 107 Reykjavik, Iceland, gstefans@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 data sets thus avoiding many model–specification issues, which are commonly transferred as assumptions from a likelihood estimation procedure into Hessian–based variance estimation procedures. The technique is demonstrated on a real data set and used to look for estimation bias and the effects of different aggregation levels in population dynamics models. 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.

Key words: bootstrapping; fish population dynamics; nonlinear models; correlated data

1. Introduction

Statistical model 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 and weighting of all data sources. 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.

This paper demonstrates a novel use of bootstrapping to address complex and disparate data issues. The approach is generic, but it is 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 (Björnsson and Sigurdsson, 2003; Taylor et al., 2007). 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) & Stefansson (2003).

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, observations need to be normally distributed. Third, variance assumptions i.e. homoscedasticity and knowledge of the ratios of variances in individual data sets, need to be appropriate. Alternative data distributions can in principle be used as extensively developed in the theory of generalized linear models (McCullagh and Nelder, 1989).

Least-squares estimates are consistent (Wu, 1981) and have established asymptotic properties including normality and the estimated variance-covariance matrix is also a consistent estimator (Jennrich, 1969). 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) 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 as seen 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 can not 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 are not detected with traditional analysis (Stefansson, 2003) but may potantially be detected if histograms of bootstrap parameter estimates also become multimodal (see example in Hannesson et al., 2004).

An alternative approach to Hessian–based parameter variance estimation are 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). A special type of bootstrap, used for time-series models, is the block-bootstrap (see e.g. Lahiri, 1999).

An bootstrap approach similar to the one proposed below (but with much simpler data and simpler models) was developed in Hannesson et al. (2009) where both simulated and real data were used to demonstrate that a bootstrap approach works for the simple case of variation only coming in through tagging data.

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 use several 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 Thordarson, G., Haputhantri, S. and Stefansson, G., 2009) 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) 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.

The following sections describe a methodology to store and handle data in such a manner as to permit bootstrapping for the purpose of variance estimation, using cod in Icelandic waters as a case study.

2. Data handling in multispecies modelling

When modelling population dynamics, the pertinent spatial and temporal scales will depend on the question. Data storage for modelling purposes is therefore best implemented on an aggregation level (data cell) somewhat finer than needed for most modelling purposes. Further, for the chosen aggregation level, between–cell correlations should be negligible. These aggregated data cells will be referred to as **elementary data units**.

The aggregations are made in order to reduce correlations between the elementary data units. For example, to remove within sample correlations (Hrafnkelsson and Stefansson, 2004), only (combinations of) entire length samples are used, rather than lengths of individual fish. Similarly, data are aggregated within fairly large spatial areas and the shortest time–step is 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.

Once all data are available in a data base, where the data have been aggregated to the required spatial and temporal level, any model definition corresponding to an aggregation of these elementary data units also corresponds to simple aggregates of the data. Thus, a simple function transforms the data in a set of elementary data units into input data for modelling. Rerunning the model based on a different aggregation scheme (e.g. different areas, temporal scales or length groupings) becomes a fairly trivial task since the extraction routines from the data base to the model input file formats can be automated.

The aggregation method for elementary data units used to create Gadget input files 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 (2004a).

Figure 1 shows the spatial definition of elementary data units, referred to as subdivision, on the Icelandic coastal shelf developed by Taylor (2003). The spatial structure is based mainly on bathymetry, hydrography and species assemblages with some further disaggregation defined by fishing regulations.

3. Bootstrapping from a data base

A set of elementary data units are aggregated for creating model inputs. These sets of elementary data are then sampled (with replacement) units 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.

Details of an implementation of such bootstrap extraction utilities from a data base are described in Kupca (2004b). It should be noted that such a procedure can not be implemented as a standard data base access procedure, as repeated extractions of the same elements of a table are required for bootstrapping which are disallowed by standard SQL statements.

With this approach the "sampling unit" in the context of resampling becomes the entire set of all measurements within a spatio–temporal "elementary data unit" in the data base. This will be a "ragged array" of widely different data types, and commonly not all data types will appear in a given bootstrap sample. In the context of bootstrap theory one can expand the "ragged array" with missing values to the largest possible size and thus consider the resampling to be of k-dimensional vectors. Viewed in this manner, the initial "sample" to be bootstrapped within a time step consists of independent k-vectors, $\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$ where n is the number of subdivisions. If F_n denotes the (multivariate) empirical distribution function (e.d.f.) of these data, then the estimator is a function of the e.d.f. viz $\hat{\theta} = g(F_n)$ and the bootstrap estimates are the same function of each bootstrapped sample, $\hat{\theta}^* = g(F_n^*)$ as required.

4. A fisheries example

4.1. 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 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. The ultimate aim of a simple model is often be to be included in a more complex model and by modelling maturity the uses of the model are increased.

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 MRI data bases (see e.g. MRI, 2004; Palsson et al., 1989; Sigurdsson et al., 1997; Taylor et al., 2007, for a description of data and surveys).

Routine assessments of this stock tend to use a time and age scale of a year and aggregate data into a single area (MRI, 2004). These assessments do not explicitly model body growth but directly include measurements of e.g. mean weight at age. The present model includes a body growth model which is implemented by "moving" fish from a source length group into several other length groups so as to obtain the appropriate average growth. In this setting a scale needs to be chosen for modelling fish length. If this scale is too coarse in relation to the time scale then on each time–step some fish will not grow whereas others will inevitably grow too much, resulting in an incorrect distribution of length at age for older age groups.

As pointed out by Vandermeer (1978), there is a balance to be found between estimation errors due to too small size classes and distribution error caused by too large size classes. 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) and this can be done with the approach proposed here.

4.2. The data set

The model is a parametric and deterministic forward 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.

The likelihood data used are:

- Length distributions from the March survey (1985 — 2003), October survey (1995 — 2003) and sampling from the commercial fishery (1984 — 2003).

- Age–length frequencies from the March survey (1989 — 2003), October survey (1995 — 2003) and sampling from the commercial fishery (1984 — 2003).

- Survey indices from the March survey (1985 — 2003) and October survey (1995 — 2003). These are calculated from the length distributions and are disaggregated ("sliced") into three groups which correspond roughly to age 1, age 2 and age3+ (as in Taylor et al., 2007).

- The ratio of immature:mature by length group from the March survey (1985 — 2003).

The mean length at age of the recruits and the population in the first timestep of the initial year can either be defined using *ad hoc* methods e.g. from observed data or calculated using the growth model.

4.3. Population models, likelihoods and parameters

When population abundance is estimated, one particularly important facet of likelihood functions is how survey abundance numbers (U) are compared to the modelled population numbers (N). The simplest approach is to assume a multiplicative relationship viz $E[U] = qN$ where the unknown q is commonly termed the catchability coefficient. In many cases it is seen that this is inadequate due to non–Gaussian behaviour as well as apparent nonlinearity in the relationship (e.g. Stefansson (1992)). On the other hand, estimation of parameters in $E[U] = qN^\beta$ or (more appropriately) $ln(U) = \alpha + \beta ln(N) + error$ can result in poorly determined power/slope coefficients, β (Taylor et al., 2007).

Confidence intervals for the slope based on regression of model outputs commonly imply them to be (significantly) different from 1.0 but typically without a fixed pattern (see Taylor et al., 2007, and references therein) and it follows that the choice of which slopes to estimate and which to fix at a value of 1.0 is more an art than a science. A common approach, however, is to fix the slope to 1.0 at older ages and to estimate it for the youngest ages. This takes into account possible nonlinearities due to behavioural differences as well as possible density dependent natural mortality, both of which are (more) likely for the youngest ages (e.g. Myers and Cadigan, 1993).

The parameters estimated are:

- one parameter for the growth function

- one parameter for the length update (i.e. growth transition matrix)

- one parameter for each fleet selection pattern (commercial catch, March survey and October survey)

- two parameters defining the maturation ogive (which are correlated)

- the number of recruits (abundance of age 1) for each year (1984 — 2003)

- the abundance at ages 2 — 11 at the start of the model in 1984

4.4. Estimation protocol

The weights on the likelihood components are calculated for each model (i.e. each bootstrap run), according to the protocol described in Taylor et al. (2007) with arbitrary starting parameters.

The bootstrapping approach consists of the following.

- The base data are stored in a standardised data base:
    - Time aggregation: 1 month
    - 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).

- To bootstrap the data, the list of subdivisions required for the model is sampled (with replacement) and stored. For a multi–area model resampling of subdivisions is conducted 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.

- The full dataset is extracted and 100 bootstrap datasets.

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.

Notably, 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. Also, it is by now well–established that incorrect model assumptions can seriously affect parameter estimates through incorrect weights (Stefansson, 2003) and simple Hessian–based methods will not capture this as a part of the uncertainty estimation procedure.

4.5. 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 derived biomass trajectories (end of year total biomass, age 4+ biomass and spawning stock biomass) are considered.

5. Results

A wide range of model output can be used to investigate parameter estimation of individual models and to compare models.

The simplest (and least interesting) model outputs are the point estimates of model parameters. Fig. 2 gives histograms of bootstrap estimates of several parameters. For each parameter, the point estimate from the full data set and the mean of the bootstrap estimates are also 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. It should be noted that the maturation parameters are correlated, affecting the relationship between the point estimate and bootstrap mean for L50.

Boxplots can be used (fig. 3) to illustrate bootstrapped trajectories of various abundance or biomass measures. It is seen that the main variation appears in the initial and final years. The initial and final years are of course considerably different 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. 4) where there is less variation in the intermediate years than the earliest or later years.

When comparing the slope and log–intercept of the observation model across surveys (fig. 5), the difference in basic catchability levels is not of interest as in both cases these are indices, subject to arbitrary scaling. More interesting, however, is a comparison of the variation in the estimates. There is considerably more variation in the parameter estimates obtained for the autumn survey, which is in accordance with the reduced number of tows in autumn (Palsson et al., 1989; Sigurdsson et al., 1997).
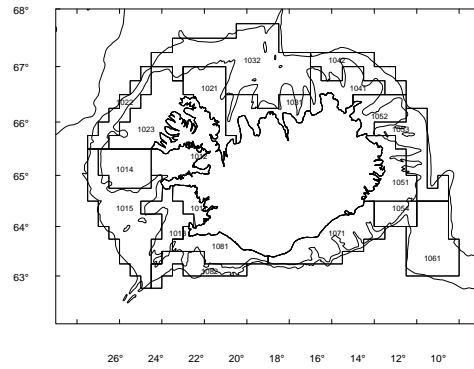
Figure 1: The spatial structure of data storage on the Icelandic coastal shelf along with 200m and 500m depth contours. These areas are referred to as "subdivisions". A given year, month and subdivision is referred to as a "elementary data unit".
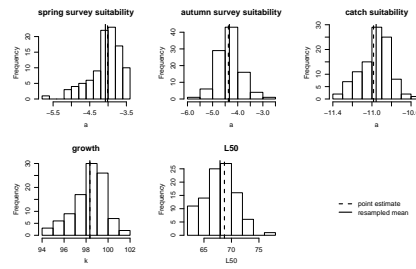

Figure 2: Histograms of some optimised parameters from 100 bootstrap model runs with the point estimate and bootstrap mean indicated.
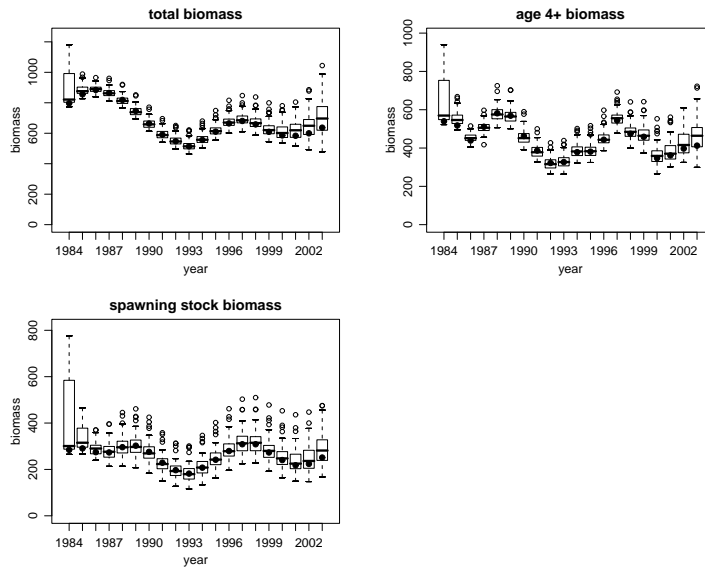
Figure 3: Boxplots of the end of year biomass estimated by the 100 bootstrap models with the point estimate indicated by the central dot of each box. The box indicates the interquartile range and the whiskers to the data point which is no more than 1.5 times the interquartile range from the box. Any further outlying data points are indicated as points. 5
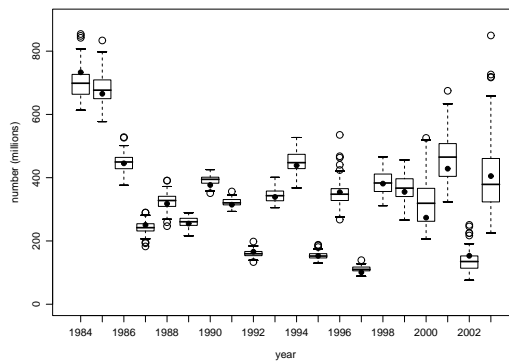

Figure 4: Boxplots of the number of recruits (age 1) in each year estimated by the 100 bootstrap models with the point estimate indicated by the central dot of each box.
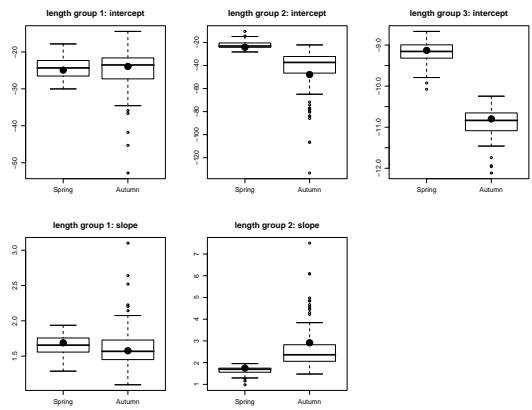
Figure 5: Boxplots of the survey intercept (ln(catchability)) and slope (power, β) estimated by the 100 bootstrap models for Structure 2 with the point estimate indicated by the central dot of each box.

6. Discussion

Data bases described in this paper enable resampling–based extraction methods. Thus the data bases generate bootstrapped data sets providing uncertainty estimates from highly complex models of marine ecosystems. Since traditional Hessian–based methods have obvious and serious issues due to the assumptions required, the bootstrap approach clearly has considerable advantages (e.g Demyanov et al. 2006).

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 be a "ragged array" (or data "list") defined in a sufficiently aggregated manner that these objects can be assumed to be independent.

Some of these 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 or other areas of resource harvesting particularly in a multispecies and multi–area

Scale changes are a fairly trivial change when data and models are set up as described here, but such changes tend to constitute a major upheaval in traditional modelling environments, where data from different sources are manually adapted to input formats for the models.

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 methodology proposed here is certainly quite computationally intensive. However this is also the case for many other methods. For example, the Hessian–based approach which has been extensively used in fishery science and general ecology needs to be quite carefully implemented since the Hessian matrix needs to be estimated quite accurately (see e.g. Tinker et al., 2006, who use expensive central differencing to estimate the Hessian matrix). For these particular case studies, bootstrapping verified the suitability of less computationally intensive models.

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.

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 need 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, $\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$, behave like i.i.d. 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 mimicks 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.

References

Babak, O., Hrafnkelsson, B., 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. Tech. Rep. 120, Marine Research Institute.

Björnsson, H., 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., 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. Tech. rep., SC/44/O 9.

Davison, A., Hinkley, D., 1997. Bootstrap Methods and Their Application. Cambridge University Press.

Demyanov, V., Wood, S. N., 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., Mahvas, S., Bertignac, M., 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., Tibshirani, R. J., 1994. An Introduction to the Bootstrap. Chapman & Hall/CRC.

Gavaris, S., 1988. An adaptive framework for the estimation of population size. CAFSAC Res. Doc 88 (29), 12.

Gavaris, S., Ianelli, J. N., 2001. Statistical issues in fisheries stock assessment. Scand. J. Statistics: Theory and Appl., 29, 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., 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, 181–184.

Gudmundsdóttir, Á., Steinarsson, B. Æ., 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.

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 (Int. Counc. Explor. Sea) Document C.M. 2004/FF:30.

Hannesson, S., Jakobsdottir, A., Begley, J., Taylor, L., 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., 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., Stefansson, G., 2004. A model for categorical length data from groundfish surveys. Canadian Journal of Fisheries and Aquatic Sciences 61 (7), 1135–1142.

Jennrich, R. I., 1969. Asymptotic properties of non-linear least squares estimators. The Annals of Mathematical Statistics 40 (2), 633–643.
URL http://www.jstor.org/stable/2239482

Kupca, V., 2004a. A standardized database for fisheries data. ICES (Int. Counc. Explor. Sea) Document C.M. 2004/FF:15.

Kupca, V., 2004b. Bootstrapping of areas. In: Final Report: dst^2: Development of structurally detailed statistically testable models of marine populations. Vol. 118 of Technical Report. Marine Research Institute, Reykjavik, pp. xx–xx.

Kupca, V., Sandbeck, P., 2003. dst^2 Datawarehouse structure and data import. In: dst^2: Development of structurally detailed statistically testable models of marine populations. Vol. 98 of Technical Report. Marine Research Institute, Reykjavik, pp. 37–46.

Lahiri, S., 1999. Theoretical comparisons of block bootstrap methods. ANNALS OF STATISTICS 27, 386–404.

McCarthy, P. J., Snowden, C. B., 1985. The bootstrap and finite population sampling. Vital Health Stat 2 95, 1–23.

McCullagh, P., Nelder, J. A., 1989. Generalized Linear Models, 2nd Edition. Chapman & Hall/CRC, London, 511 p.

Methot, R. D., 1989. Synthetic estimates of historical abundance and mortality for northern anchovy. Amer. Fish. Soc. Symp. 6, 66–82.

MRI, 2004. State of marine stocks in Icelandic waters 2003/2004. Prospects for the quota year 2004/2005. Tech. Rep. 102, Marine Research Institute.

Myers, R. A., Cadigan, N. G., 1993. Density-dependent juvenile mortality in marine demersal fish. Canadian journal of fisheries and aquatic sciences(Print) 50 (8), 1576–1590.

Myers, R. A., 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.

Palsson, O. K., Schopka, S. A., Stefansson, G., 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., Stefansson, G., 2001. Estimating uncertainty in fish stock assessment and forecasting. Fish and Fisheries 2 (2), 125–157.

Pennington, M., 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.

Rao, J. N. K., 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.

Sigurdsson, T., Hjorleifsson, E., Bjornsson, H., Palsson, O. K., 1997. Fall groundfish survey in icelandic waters (stofnmaeling botnfiska a Islandsmidum haustid 1996). Tech. Rep. 61, Marine Research Institute, (in Icelandic).

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.

Stefansson, G., 1992. Notes on the stock-dynamics and assessments of the Icelandic cod. ICES (Int. Counc. Explor. Sea) Document C.M. 1992/G:71.

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., 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. K., 1997a. BORMICON A Boreal Migration and Consumption model. Tech. Rep. 58, Marine Research Institute.

Stefansson, G., 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., Palsson, O. K., 1998. A framework for multispecies modelling of boreal systems. Reviews in Fish Biology and Fisheries 8, 101–104.

Taylor, L., 2003. Definition of areas in Icelandic waters. In: dst^2: Development of structurally detailed statistically testable models of marine populations. Vol. 98 of Technical Report. Marine Research Institute, Reykjavik, pp. 222–230.

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 (2), 223–245.

Thordarson, G., Haputhantri, S. and Stefansson, G., 2009. A data sampling strategy for length based stock assessments - Sample size vs. number of samples.

Tinker, M. T., Doak, D. F., Estes, J. A., Hatfield, B. B., Staedler, M. M., 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., Bogstad, B., 1989. MULTSPEC: The manual. Institute of Marine Research, Bergen.

Vandermeer, J., 1978. Choosing category size in a stage projection matrix. Oecologia 32 (1), 79–84.

Wu, C.-F., 1981. Asymptotic theory of nonlinear least squares estimation. The Annals of Statistics 9 (3), 501–513.
URL http://www.jstor.org/stable/2240815