

## ISSUES IN MULTISPECIES MODELS

GUNNAR STEFANSSON

Department of Mathematics

University of Iceland and Marine Research Institute

Skulagata 4, P.O. Box 1390

121 Reykjavik, Iceland

*E-mail address:* `gunnar@hafro.is`

**ABSTRACT.** It is generally acknowledged that models of population dynamics need to include realistic biological and statistical assumptions if they are to be of any use. The “realistic” complexity of these models is defined not only by the underlying biology but also by what the available data can reasonably determine. As complexity increases through inclusion of interactions, new issues arise which need to be resolved before these complex models can be reliably applied. Specific issues in new models for management include proper statistical fitting to the data and incorporation of time-dependent and spatially explicit biological interactions. These current problem issues are presented in this paper and some potential solutions are discussed.

**1. Introduction.** Models of the dynamics of fish populations are usually developed in response to certain questions of interest. The greatest emphasis has been placed on questions related to the management of fisheries and hence the population models have been developed to answer questions relating to effects of different regulations, such as changes in fishing effort, quotas, mesh sizes, protected areas and so on.

As management has been improved or changed in different areas, the questions also change. Thus, once harvest from a target species comes under control, questions arise concerning bycatch species. Similarly, if the target species biomass gets seriously reduced, questions arise concerning stock and recruit relationships.

The last decades of the twentieth century saw the development and implementation of multispecies methods of assessing fish stocks, statistical methods for assessing population abundance and methods of assessing the uncertainty in predicting the development of individual stocks.

Each of these three issues, multispecies interactions, statistical models and uncertainty estimation can have a major impact on the perceived or likely range of population abundance and potential or safe harvest from the resource (Patterson et al. [2000], Patterson et al. [2001], Restrepo et al. [2000], Gavaris et al. [2000] and Stefansson et al. [1997]).

This paper discusses specific modeling issues which arise in large exercises in statistical models of multispecies ecosystems, such as BorMiCon (Stefansson and Palsson [1997a]) and variants thereof, notably those which can be implemented using Gadget, an environment for statistical multispecies modeling (Anon. [2002], Begley [2003]). In accordance with specifications set out by Stefansson and Palsson [1998], this model has been developed as a forward simulation model using statistical estimation through weighted combinations of several log-likelihood criteria. Although that particular framework has been used to generate case-specific multispecies models, the statistical and modeling issues that arise are perfectly general to any statistical multispecies model or, in fact many complex nonlinear statistical models which incorporate several data sources. The issues discussed in this paper have thus become sources of problems when developing, implementing and verifying models based on Gadget.

It will be seen that there are certain issues which need to be addressed in order to ascertain whether the models may potentially be adequate (necessary condition). If these issues are not taken into account, then the models may be quite misleading. As always, models will be improved and new conditions will be found necessitating further improvements.

## **2. Questions relating to management or general biology.**

Traditional fisheries management has been based on one or more regulatory mechanisms. Typically these involve quotas on individual species, effort limitations, closed areas and controls on the types of gear used.

Each of these mechanisms leads to different questions which need to be answered by the models used. Common to all systems is a need to evaluate the long-term yield potential, usually of each individual stock. In addition, system-specific information is required, e.g., a quota system will need advice on absolute catches for each stock, whereas an effort-

control system only needs a suggestion for an overall effort allocation.

The presence of biological and technical interactions implies a second round of questions, regardless of whether the basic system is effort- or quota-based. Thus in the effort system, there is a need to evaluate the effects of a fleet switching from one species to another and in the quota system issues of coordinating quotas in mixed fisheries can become important. Under any management system the effects of prey harvests on predator growth can be important as can the impact on prey abundance of reduced fishing on a predator.

These same and several other multispecies-oriented questions arise when there are closed areas for management purposes. In particular, models incorporating migration are required in order to assess the effectiveness of closed areas. How migration and drift or dispersal parameters are specified and estimated have important consequences in evaluating the impact of closed areas.

In the multispecies context it has been found that the form of prey selection can be quite important, as related to management (Mohn and Bowen [1994]). As a prey species diminishes in abundance, the predator may potentially switch disproportionately to new prey or not switch at all. At low prey abundance such effects can have considerable impact on predictions of the effect of fishing, rebuilding predictions etc.

On a related note, several methods have been used to describe species interactions in general. In particular, different functional forms have been used to describe (local) consumption, with several functional responses available. In rare cases these have been compared formally through adequacy of fits to data (Magnusson and Palsson [1989]) but usually a single model is assumed or an alternative model is tested as a deviant from a baseline (Mohn and Bowen [1994], Danielsson et al. [1997], Stefansson et al. [1997]).

It is therefore seen that management issues give rise to a variety of multispecies-oriented questions which brings forth a need for a whole new breed of models. These new models need to incorporate much more detailed biology than has been considered in assessment models before.

TABLE 1. Some biological issues which need to be brought into future multispecies models.

Model component	Issues
Prey selection/functional responses	Need methods for identifying mechanisms involved
Migration formulation	Parsimonious formulations are needed
Growth update	Need formulations and estimation methods
Tagging data	Methods are needed to incorporate mark-recapture experiments
Life cycle closure	Closing the life cycle of spatially explicit multispecies models has not been done

Many traditional models of fish populations used a single mean length at age or were length-based without considering age. In most applications recruitment is highly variable and this induces a need to track yearclasses through age/time (particularly so for shorter term predictions). On the other hand, the selectivity of the fishing gear as well as food selection is length-dependent and the food supply will affect the growth of the predator. For these reasons modern models need to be able to account for variability in both age and length. Thus there is a need to incorporate growth models where the development of the distribution of length at age is made explicitly dependent on food supply. These models clearly become an order of magnitude more complex than models which are only age-based.

When models are made age- and length-based a growth update mechanism needs to be included. Several approaches exist for this (Bogstad et al. [1992], Stefansson and Palsson [1997a]), but the lack of decent data on individual growth has hampered the validation of such models. It would seem, however, that overly simple models of growth updates can fail to adequately explain the length distribution at age

in samples and therefore parametric models with more flexibility are needed (Stefansson [2001]).

Dynamic multispecies models which have been used most extensively for assessment purposes have not included estimation of a recruitment function or attempted to close the life cycle, e.g., MSVPA or other models following the work of Helgason and Gislason, [1979]. Closure of the life cycle is important, however, when attempting to estimate medium- or long-term yield potential. This has been investigated in the context of limit and precautionary reference points under stability assumptions, and the multispecies effects have been found not only to be quite important but also nontrivial (Collie and Gislason [2001]). When the models are made spatially explicit and dynamic, it is no longer obvious how to proceed with closing the life cycle, but some proposals can be made. For common fish species which survive for several years, spawning annually after a certain age (and/or size), it is hardly feasible or desirable to use models which are based on very fine (weekly) time-steps and methods for statistical inference from individual-based models are almost non-existent. It follows that new approaches need to be developed for closing the life cycle when viewing the fish species in a spatially explicit multispecies context. It may be possible to proceed by defining the number of spawned individuals (eggs and larvae) as a direct (linear) function of the spawning biomass in each area, adding (immediate and high) density-dependent natural mortality along with larval drift, modeled in the same manner as the migration of adults. In order to get values anywhere near realism in each area, it is likely that in the first round some ad-hocery will be needed, such as estimating some simple natural mortality function which brings the number of young fish to the sorts of levels observed in the first surveys after settlement. It is clear that estimating these artificial just-before-the-survey mortality rates for each yearclass and each area can be used to fit exactly to the first survey indices. This is of course inadequate due to the large number of parameters but, at the other extreme, it should be possible to estimate a single migration mortality in this time period, which should result in the correct average yearclass size, though incorrectly distributed in time and space. To get started with closing the life cycle it should be possible to go ahead with such models, intermediate in parsimony, until a more appropriate formulation is found.

On rare occasions migration modules have been incorporated in multispecies models. Typically, adding migration matrices allows too much flexibility and the number of parameters becomes too large. Solving partial differential equations for migration directly may be prohibitive, so methods are needed for a realistic, yet parsimonious, representation of migration. Subsequent estimation of parameters describing migration rates need to be linked to tagging data where tagged fish in the model should be affected by the same (or similar) processes as the remainder of the population. This leads to a considerable exercise in modeling and computing resources, but it is not clear initially how well such parameters can be estimated. New methods of assessing estimability and uncertainty estimation are needed. Methods involving bootstrapping entire tagging experiments may possibly be of use in this context.

These biological issues are brought into models through the inclusion of model complexity and more data sources. This leads to new classes of problems relating to statistical and numerical estimation. As will be seen below, new methods are needed to enhance the ability to detect inconsistencies between models and data, the ability to detect the appropriate functional relationships to be included in different model components, measures to indicate deviations from data, describe estimability of parameters and so on. These methods to address the above biological issues are quite general and statistical in nature and need to be developed as such.

### 3. Statistical issues.

**3.1 Basic model assumptions.** There are many cases when basic model assumptions used in single species assessments are clearly violated, yet there is no easy way to fix the problem. For example, it is common in several assessment models that a nonlinear (power) relationship between an abundance index and modeled population size is obtained. In principle the exponent in this nonlinear relationship should be fixed at 1.0 for surveys which are designed to provide abundance indices. Obtaining an exponent which is significantly different from 1 is not a problem per se, but when the estimated exponents are large for some age groups and small for others without regularity, apparent, e.g., in Stefansson [1992], as noted by Butterworth [pers.

comm.]) there would appear to be some effect present which is not captured by the model but is being picked up by such parameters.

Such unquantifiable effects are somewhat worrisome since there is no clear reason why they should appear. In some cases it appears that changing a parameter from being a constant to being a time series effect gives a considerably improved model fit. It is clear that many concepts included in fisheries models are not constant in reality. Some of these correspond to upheavals due to regime shifts, one example of which is influx of cold water into an otherwise delicately balanced area. Such events can have detrimental effects on processes such as recruitment, and this has sometimes been handled by adding a single factor to the models, e.g., Jakobsson et al. [1993]. A different and much more common effect is that of changes which should be allowed to vary smoothly with time. This is known to be true of such simple things as the annual selection pattern (Gudmundsson [1994]) which clearly violates the commonly used separability assumption on fishing mortality, e.g., Deriso et al. [1985]. Similarly, migration rates will also change with time (as modeled in MULTSPEC, cf. Bogstad et al. [1997]), though probably not completely at random or independently from year to year but with some time series structure.

It is quite possible that the real problem which manifests itself in “strange” parameter estimates is really due to the lack of flexibility and lack of time series parametrizations. Such time series effects have been estimated using either Kalman filters (Gudmundsson [1994]) or as parameters in nonlinear models (Ianelli and Fournier [1998], Gavaris and Ianelli [2001]). The “best” mechanisms for bringing time series effects into the fisheries models need to be investigated as needs the likelihood of obtaining sensible estimates in a variety of different scenarios. The truly worrisome situation is when a simple assessment model is used with no diagnostics to indicate that there is a problem of this form.

Section 3.4 below addresses the issue of detecting some of these problems when different data sets allow the identification of model inconsistencies. The generic model identification issue (defining the mean function) is somewhat different. In simple cases such as linear regression with repeated measurements it is possible to use lack-of-fit tests (Neter et al. [1996]) to identify whether the assumption of linearity is inappropriate. In constrained linear or nonlinear optimization it is

possible to use the concept of a dual problem, giving “shadow values” or derivatives of the optimum value with respect to each constraining value. These indicate the effects on the objective function of alleviating the constraints (e.g., Bertsekas [1999]).

If the nonlinear model is statistical it may be possible to use factors to investigate whether the functional form is incorrect. Although in complex models it is sometimes possible to change a parameter to a time series effect, there is a more general need to identify where such model changes will work to provide a better fit. The multispecies models become quite complex with modules to describe growth, migration, consumption, fishing, etc. The equivalent of “constraints” is the linking mechanism between these modules. Basically, what is needed is a generic method to indicate how much the fit will improve if some of the links between these complex modules are relaxed or removed.

The methods described in Richards [1991] and the weighting methods described below provide one approach to investigating such problems but more generic approaches are needed as the incorrect links may not correspond directly to data sets but only to pairs of biological modules.

**3.2 Weighting data sources.** The typical statistical multispecies model is based on several data sources. Although this phenomenon is common in fishery science, the number of data sources tends to go up more than linearly as the number of species increases. Thus, there are data sets on length measurements, age readings, tagging experiments, stomach contents, landings, survey abundance and so on, all of which play a role in determining parameters of these models. The issue at hand is to combine all these data sources through a single model. Rather than attempt to investigate this in a life-size model, the issue can be examined using a much simpler model.

Considering first the simple independent-Gaussian assumption on the probability distribution (possibly after log-transformation), it is more common than not that data come from several different sources in which case the total likelihood function becomes a product of the likelihoods of the different components (assuming independent sampling from the various data sets). The objective function to be minimized for estimation purposes is the negative log likelihood which becomes the



TABLE 2. Some generic modeling issues which need to be resolved in future multispecies models.

Statistical aspect	Issues
Weighting of data sources	New methods are needed (and proposed here), to estimate the correct weighting of data sources.
Evaluating model inconsistencies	Methods are needed (and suggested here) to find where model inconsistencies exist.
Goodness-of-fit methods	Some such methods are available but many more will be needed as the models improve. The new methods need to work on fairly arbitrary likelihood functions, not just independent Gaussian or multinomials.
Identifying relaxing methods	This is an open question: How does one find which part of a complex nonlinear model is too restrictive?
Estimability of parameters	Some methods are available but need to be put into routine use.
Time series components	Need to investigate estimability and methods of estimation
New likelihood functions	Entire classes of likelihood functions need to be developed for use with categorical data in fisheries
Degrees of freedom	At present there is no reliable method available to ascertain the degrees of freedom corresponding to a data source.

sum of the log components (or weighted sum of squares if the variances are assumed known).

When this Gaussian assumption is used for the different data sources, it is in principle possible to estimate all variances involved using

maximum likelihood. In practice, however, this can be quite tricky. This is particularly the case when the model cannot explain all the data sources simultaneously.

The actual problem in this case consists of two parts. Firstly, the weights, usually inverse variances, to be given to each data source are unknown and secondly the model is incorrect. A large model is like a black box, where parameters are put in and out come values of likelihood components or sums of squares. The effect of an incorrect black box in this case is that a simultaneous good fit to all data sources cannot be obtained. This has been a known problem in fishery science for some time (Richards [1991], Schnute and Hilborn [1993]).

To mimic this issue, take the simple true state of nature (the data generator) to consist of independent measurements around two straight lines

$$Y_{ij} \sim n(\alpha_i + \beta_i x_{ij}, \sigma_i^2), \quad j = 1, \dots, n_i, \quad i = 1, 2.$$

At the outset it might seem that one can simply test whether the model is correct, by testing whether  $\beta_1 = \beta_2$  using an ordinary F-test (or t-test). However, the whole point is that the estimation procedure does not know which part of the model is incorrect. The only consistent approach would be to test whether  $\theta^{(1)} = \theta^{(2)}$  where the superscript indicates a reference to each data set,  $i = 1, 2$ . Estimates of these vectors of all model parameters, i.e.,  $\theta' = (\alpha_1, \alpha_2, \beta)$ , are first obtained by fitting to each data set separately and subsequently using the combined data set.

The situation here is worse, though, since the assumption of homogeneity of variances does not hold and hence none of the usual F-tests will work without modification. This is seen by considering the simplest F-test which assumes constant variances (even assuming knowledge of the potential model error and thus testing  $\beta_1 = \beta_2$ ). This test fails miserably in the case when one of the data sets is considerably larger than the other and the variance of the larger data set is the smaller one. In this case the traditional F-test fails to control its error rate. This is obvious when considering that the F-test is based on a common estimate of the variance and the large data set will dominate the variance estimate. The resulting low variance estimate will indicate that the smaller sample deviates too much from the common slope. In addition to this problem, an attempt to estimate variances in an obvious

manner will lead to problems in the general case since the usual F-test for  $\theta^{(1)} = \theta^{(2)}$  is based on the same sums of squares as are used in the variance estimation. These statements are easily quantified using simulations based on the simple model.

In reality, the black box model assumed will estimate the variances. Worse still, the model is typically incorrect by assuming that  $\beta_1 = \beta_2$ . In this scenario, maximum likelihood estimation of all parameters including the variances can be used, but the values of the log-likelihood surfaces are plotted in Figure 1 as individual curves for each data set as well as the sum for simulated data with  $n_i = 50$  data points in each set. This trivial example suffices to illustrate the problem of apparently incompatible data sets.

Of course when the slopes are similar there is no problem, but as they move apart, the situation becomes as often seen in real data where the model fails to fit both data sets. Figure 1 shows that when the true variances are similar, the maximum likelihood estimate can flip between being close to each of the two true slopes and correspondingly the variance estimators flip from one or the other being correct. This is similar to the results found by Richards [1991]. It is also seen in this case that it is not particularly feasible to obtain (local) diagnostics in terms of, e.g., the slope of each objective function since the slope can be close to zero at the minimum of the other (this is due to the logarithmic nature of the likelihood as a function of the variance).

In terms of variance estimation, there is light at the end of the tunnel, however, since it is technically feasible to estimate the variance in each data set under a Gaussian assumption. This can be done by first reducing any single variance towards zero (i.e. diminishing the influence of any other data sets) and using the corresponding resulting sum of squares in estimating that variance. Using this technique iteratively across data sets will result in unbiased estimates of the appropriate weighting factors, even if the model cannot fit all data sets simultaneously, as long as each individual data set can be explained. A problem with degrees of freedom will remain but at least in cases when the number of data sources and samples is large it will be minor.

For example, a multispecies model may not be able to reconcile growth data on a predator with abundance data on a prey. In this case, fitting to the predator growth data alone may force modeled prey

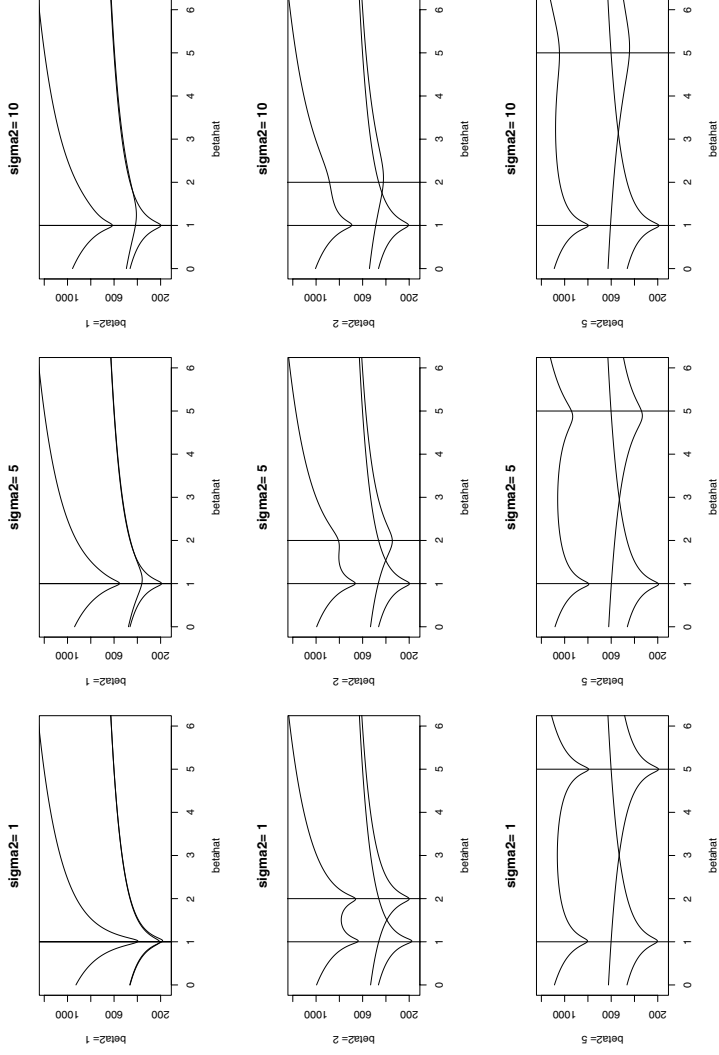


FIGURE 1. Minimum negative log-likelihood value from simulated data using assumed model. Each panel gives three curves, the total negative log-likelihood (top) and the two component curves, plotted against  $\hat{\beta} \in [0, 6]$ , after having minimized over all other parameters ( $\alpha_1, \alpha_2, \sigma_1, \sigma_2$ ) in the assumed model. True values of  $\beta_1 = 1$  and  $\sigma_1 = 1$  are fixed but panels indicate varying true values of  $\sigma_2 = 1, 5, 10$  (across) and  $\beta_2 = 1, 2, 5$  (down).

abundance to take on unrealistic values, but estimates of the variance of growth measurements will be obtained.

To stabilize estimation in real applications, it is likely that the various alternative data sources cannot be completely omitted but included with very low relative weights compared to the one where the variance is being estimated.

Since it is thus possible to obtain estimates of the variability of each data set, and thus the weightings, it remains to develop methods to elucidate the “incorrectness” of the black box, to investigate alternative likelihood functions and investigate the effects of non-independent sources of measurements.

**3.3 Likelihood functions.** Likelihood functions in fishery science are commonly based on traditional probability distributions, such as the Gaussian, multinomial, negative binomial or gamma distributions.

The negative binomial, lognormal and gamma distributions are sometimes used for abundance data. It has frequently been noted that these distributions may not be appropriate for these data, even when restricting to tows with positive numbers of fish, e.g., Steinarsson and Stefansson [1986], Stefansson [1996] and Brynjarsdottir [2002]. Similarly, it has been found that even the simple Bernoulli assumption is not an appropriate model for the extremely simple presence/absence of fish in each tow (Stefansson and Palsson [1997b]). Complex models have also been used, combining several distributions (Lo et al. [1992]) but these have also been found wanting (Stefansson [1996]). In fish stock assessment models it is not uncommon to assume the catches in numbers at age or length to be lognormal, but in other settings a multinomial distribution is preferred. Now, it is indeed clear that even if the entire catch was sampled for age, the distribution would not be lognormal, but the multinomial assumption seems quite popular for counts, particularly for length distributions, e.g., (MacDonald and Pitcher [1979]). Unfortunately this assumption is usually not valid for real data sets.

As an example, consider the simplest case of length distributions from a groundfish survey. To standardize the measurements and simplify binomial variances, it has been proposed to randomly subsample a fixed number ( $n$ ) of length-measured fish (here 25) from each station

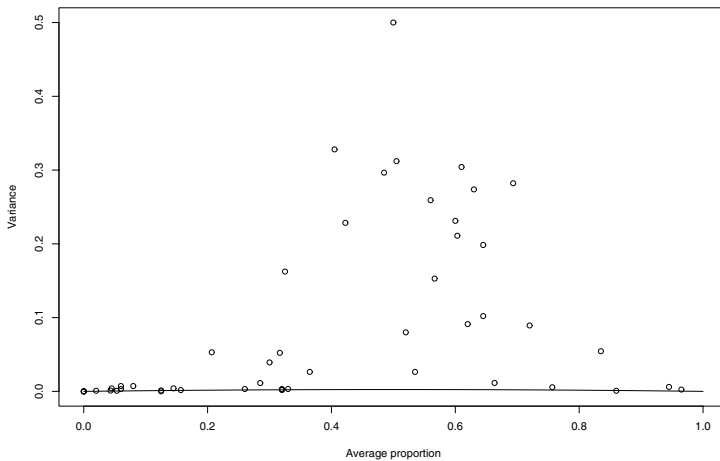


FIGURE 2. Variance-mean relationship for proportion ( $p$ ) of cod per tow below 45 cm, theoretical (solid curve,  $Var[\hat{p}] = pq/n$ ) and observations (circles). Each point represents the mean and variance of the number of  $n = 100$  randomly selected fish below 45 cm caught at different tow locations within a small area.

(Hrafnkelsson and Stefansson [2004]). If these measurements were from a multinomial distribution, then the variance of the number of fish,  $X$ , per length cell would be  $V[X] = npq$  where  $p$  is the true proportion of fish in that length cell and  $q = 1 - p$ . On the other hand the theoretical mean is  $E[X] = np$ . Within a small area (a statistical sub-rectangle of 30 feet by 15 feet) the data from different tow locations can be used to compute the observed means and variances. These data and theoretical curve are given in Figure 2.

It is clear from this analysis that the multinomial assumption is completely incorrect in terms of the variance function. This can in principle be corrected using an overdispersed multinomial or adjusting the sample size, but this will not be enough: Simple further analysis along the same lines indicates that the correlations between length cells is similarly incorrect. Since this is in fact fairly obvious when the biology is considered (the correlation between the abundance of fish in adjacent size groups is bound to be higher than between large and small size groups), the general use of a multinomial is questionable to say the least.

At present there is no generally adopted method to amend the multinomial to incorporate this general correlation structure. Using a multivariate Gaussian density with autoregressive correlation structure might suffice but this needs to be formally evaluated. The suggestions of Schnute and Richards [1995] of using what could be termed a multivariate logit-normal (Schnute [pers. comm.]), is one approach which should certainly be evaluated in this context.

The issue of different likelihood functions is without doubt important. In conjunction with model misspecification, this can lead to completely incorrect assessments. However, at least this is one issue which can in principle be solved through fairly standard methods, i.e., developing alternative likelihood functions which fit the data when using a flexible model.

**3.4 Apparent incompatibilities.** As noted above, in many real applications the pictures indicated by the different data sources appear to be inconsistent. Thus time trends of catch-per-unit-effort (CPUE) from different fishing gear can indicate different trends in abundance. The true reasons for this can be simply that there is a time trend in catchability of the commercial fleet, which is not incorporated in the model. In terms of fitting (incorrect) models, what is seen is an incompatibility between data sources. Methods are required to detect such apparent incompatibilities.

Sometimes time series components can be added to the models to alleviate problems such as these. Thus, simple random walk effects in catchabilities have the potential to describe persistent changes in catchability. Naturally, generic time series effects are no substitute for an appropriate physical model of the process, but in the absence of such information the time series approach has considerable potential.

The generality of these apparent inconsistencies between data sources is unknown but several examples have been documented where these effects are important enough to give quite different results. These effects are not restricted to complex models as they have been found in simple single-species models, e.g., different abundance trends or different results solely due to different weighting of individual age groups in the same survey (Stefansson [1998]).

In some cases fitting to more than one data set can affect important

quantities such as population trends. Some approaches to this issue have been proposed, such as fitting to each data series separately, e.g., Polacheck et al. [1993]. A more general approach is to investigate a series of weightings to each data set (Stefansson [1998]), as detailed in Section 3.2. This can potentially give estimates of appropriate weightings for each data set.

Having obtained the weights, a selected parameter can be estimated using only one data source and then all the remaining ones. If there is enough information in the “remaining” data sources, then one can treat that estimate as fixed and test whether the estimate using the single data source deviates significantly from the fixed value using a standard t-test based on the Hessian matrix or a likelihood ratio test. Each approach depends on assumptions, but this would be a way forward to investigate the inconsistencies quantitatively and objectively. Alternative methods based on testing the entire parameter vectors are under development.

It follows that the question of apparently incompatible data sets can to some extent be detached from the issue of weights given to each data set. However, the issue of selecting likelihood functions remains problematic and intertwined with apparent incompatibility: Incompatibility is detected if a likelihood component increases “too much,” i.e., significantly, when a parameter is fixed at the value estimated by the remainder of the data. If the likelihood function is inappropriate then it is quite possible that the value of “too much” is inappropriate, i.e., rejection occurs too frequently. The over-dispersion issue in the multinomial above is precisely such a case.

In fisheries it is quite common that most standard likelihood functions are rejected as inappropriate. As more complex models are considered these issues will become steadily more important and means to resolve them need to be developed.

As indicated in this section, to obtain an orderly development of these complex models a new methodology is needed to verify model adequacy. In simple linear regression with repeated measurements this is done through lack-of-fit tests (Neter et al. [1996]). In the present setting these verifications need to be formalized by comparing point estimates through different weighting schemes. Such statistical comparisons are outlined in Stefansson [in prep.].



**3.5 Goodness-of-fit tests.** As mentioned above, several likelihood functions are rejected when statistical measures of goodness-of-fit are applied. Reasons for this may vary, but they certainly include over-dispersion due to patchiness, correlation due to the animals' behavior, a model's basic inability to fit the data (i.e., error in the mean function) and so on.

Assuming that the issues above of selecting appropriate likelihood functions are alleviated, the functional relationships are properly modeled, appropriate weights are estimated for each data set and so on, then the goodness-of-fit tests should have a clear interpretation.

The multispecies model BorMiCon can use a large data set, as seen in the first implementation described in Stefansson and Palsson [1997a]. Initial goodness-of-fit tests were carried out for the outputs from this 3-species, multi-area model. These included tests for normality (for assumed Gaussian residuals), tests for the multinomial assumptions for length distributions and so on. For some of the classes of tests, most tests rejected the appropriateness of the distributional assumptions. In no cases were the proportion rejected less than (or even close to) 5% when using a level of 5%. It follows that there is considerable work that needs to be undertaken to make these ecosystem models consistent with observations.

Work on goodness-of-fit tests will thus become crucial as the likelihood functions, model identification schemes, and weighting methods are improved. At present, the apparently incompatible data sets, inadequate likelihood functions and not-fully-developed weighting mechanisms completely override any goodness-of-fit test development.

The goodness-of-fit tests to be used in the fishery science of the future will have to be able to handle generic likelihood functions. Ideally a goodness-of-fit test should be able to deal with deviations from different likelihood components in a generic manner, rather than the distribution-specific manner used in today's tests. At present this is an open problem.

**3.6 Degrees of freedom.** It is seen from the previous sections that the multinomial distribution frequently gives a much poorer fit than expected, partly due to over-dispersion. One way to interpret this is that the true sample size is really lower than the one usually stated, at least

partly due to the intra-haul correlation as discussed by Pennington and Volstad [1994]. By the same token, the degrees of freedom associated with sums of squares may frequently be considerably different from those expected from sample size considerations alone.

In the complex models under consideration here, further problems arise when the degrees of freedom associated with a particular model are considered. When the size of the data set is very large, this is not a particular problem, but when the data set under consideration is small compared to the total number of parameters, the degrees of freedom will be hard to ascertain.

**3.7 Data preparation.** An important part of fisheries data consists of length distributions (LDs) and age-length keys (ALKs). In some cases fish are randomly sampled for age and length, resulting in age-length (frequency) tables (ALTs), in which case there may not be separate length distributions.

LDs and ALKs (or ALTs) are used to derive information such as the mean length at age and proportion at age.

A multinomial, or extension thereof, can be used to model the length distribution and also the age distribution within a length group in the ALK or ALT. However, the mean length at age can also be modeled using a Gaussian distribution and the distribution of proportions at age can similarly be approximated using a (possibly weighted or multivariate) Gaussian distribution.

It is of course clear that not all of these can be used simultaneously since this results in duplicate use of data, with unknown weighting effects. However, in principle the mean lengths and proportions can be used in place of the LDs and ALKs, for example. Since there is no accepted notion of sufficient statistics in these data sets, it is not known a priori, which approach is best. It is, however, clear that the different summary statistics will correspond to an emphasis on different model attributes.

Initial use of BorMiCon and Gadget seem to imply that it may be quite important to select very carefully the data summaries used and make these correspond to the model under consideration. Thus it appears that the use of length-based indices of abundance for each length group separately may in some cases be more stable than the use

of length distribution in multinomial form along with a total index of abundance (Bjornsson [pers. comm.]). This is an area which requires further research.

**4. Estimability and uncertainty.** In multiple linear regression the concept of an “estimable function” or parameter is a well-defined mathematical concept (Scheffe [1959]) as it basically denotes a parameter which is uniquely defined in the model. Similarly, but not equivalently, the structure of the  $\mathbf{X}$ -matrix defines what parameters are well estimated etc. In the typical nonlinear highly complex spatially disaggregated multispecies statistical fisheries models, this is not so simple. Examples of hard-to-estimate scenarios abound and some of these have been studied in some detail (see e.g., Schnute [1991]).

It is clear within several models that some subsets of parameters are more difficult to estimate or disentangle than others. Notably, the selection pattern in a survey is closely linked to the growth curve, particularly if only mean length at age is available.

However, the notion that the large and complex models are automatically inestimable due to their size is not true in general. This is a result of the simple fact that the individual components of the model may each be estimable from individual data sets. Thus, parameters of a von Bertalanffy curve are estimable from mean length at age alone, a selection pattern at age can be estimated from catch composition at age and annual fishing mortality (smoothed) can be estimated when total catches are included. This can be done for each individual species if the data is available. Even with no interactions this implies that up to hundreds of parameters are easily estimable.

For some parameters, it is not clear whether data exist to estimate them properly. Migration parameters are an example where this may be difficult, though of course some bounds are placed on these parameters through regional catch history alone and some minimal information will be obtained through tagging experiments. Although it is not a priori clear at all just how accurately migration parameters can be estimated, it is clear that they are linked to very many other parameters of population models. For example different areas are intrinsically associated with different growth rates through variable food supply and temperature. In this case modeled or real changes in migration rates

will also manifest themselves in corresponding changes in regional body size and weight. It follows that the migration rates form a crucial part of the multispecies models and resolving estimation issues for these parameters is of major importance.

Similarly, it is not clear how the various parameters of suitability functions can be estimated, or of selection patterns at length, when growth is estimated simultaneously.

In order to clear up such questions of estimability in complex models, simulated data sets will most likely be required. In principle these can be used to determine what quantities stand a remote chance of being estimated.

Estimating uncertainty is related to the question of estimability. It is not clear how uncertainty should be estimated in complex models with many data sources. However, bootstrapping all the various data sources may eventually provide the uncertainty estimates required. Even the most complex multispecies models are based on areas which can be taken to consist of sub-units, which are aggregated before they are entered into the model. The entire data set from a given time point in a sub-unit can possibly be taken as the data vector to be sampled for bootstrapping purposes.

If reliable estimates of uncertainty can be obtained, then this will give not only the variances but also the correlations between parameter estimates, thus providing first indications of which parameters are not estimable. This may not suffice, however, since complex confounding may arise due to the nonlinearities involved.

**5. Discussion.** To ensure some level of objectivity, the way forward with complex fisheries models needs to be statistically based. Statistical methods need to be developed to ensure orderly buildups of these nonlinear models through increasing complexity as well as selection of appropriate functional relationships and statistical tests are needed to evaluate whether models explain all available data sets adequately.

This paper has raised several issues which arise when complex nonlinear models are built. It has also been indicated how some of these problems can be solved, but some are still unsolved. Finding appropriate solutions to these issues is a requirement for knowing with some level of certainty that adequate model fits have been obtained.

## REFERENCES

- Anon. [2002], *Development of Structurally Detailed Statistically Testable Models of Marine Populations* (dst2), QLK5-CT1999-01609, Progress Report for 1 January 2001 to 31 December 2001, MRI Tech. Report 87.
- D.P. Bertsekas [1999], *Nonlinear Programming*, Athena Scientific, 708 pp.
- J. Begley [2003], *Gadget User Manual*, available at: <http://www.hafro.is/gadget>.
- R.J.H. Beverton and S.J. Holt [1957], *On the Dynamics of Exploited Fish Populations*, Fish and Fisheries Series **11**, Chapman and Hall, London.
- B. Bogstad, K. Hiis Hauge and O. Ulltang [1997], MULTSPEC—A Multi-Species Model for Fish and Marine Mammals in the Barents Sea, J. Northwest Atl. Fish. Sci. **22**, 317–341.
- B. Bogstad, S. Tjelmeland, T. Tjelda and O. Ulltang [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, SC/44/O 9.
- J. Brynjarsdottir [2002], *Statistical Analysis of Cod Catch Data from Icelandic Groundfish Surveys*, MRI Tech. Report **91**, Reykjavik, 81 pp.
- J.S. Collie and H. Gislason [2001], *Biological Reference Points for Fish Stocks in a Multispecies Context*, Can. J. Fish. Aquat. Sci. **58**, 2167–2176.
- A. Danielsson, G. Stefansson, F.M. Baldursson and K. Thorarinsson [1997], *Utilization of the Icelandic Cod Stock in a Multispecies Context*, Marine Resource Economics **12**, 329.
- R.B. Deriso, T.J. Quinn and P.R. Neal [1985], *Catch-Age Analysis with Auxiliary Information*, Can. J. Fish Aquat. Sci. **42**, 815–824.
- S. Gavaris and J.N. Ianelli [2001], *Statistical Issues in Fisheries Stock Assessment*, Scand. J. Statistics: Theory and Appl., **29**, 245–272.
- S. Gavaris, K.R. Patterson and C.D. Darby, P. Lewy, B. Mesnil, A.E. Punt, R.M. Cook, L.T. Kell, C.M. O'Brien, V.R. Restrepo, D.W. Skagen, and G. Stefansson [2000], *Comparison of Uncertainty Estimates in the Short Term Using Real Data*, ICES C.M. 2000/V:03.
- G. Gudmundsson [1994], *Time-Series Analysis of Catch-at-Age Observations*, Applied Statist. J. Royal Statist. Soc. Ser. C **43**, 117–126.
- T. Helgason and H. Gislason [1979], *VPA-Analysis with Species Interaction Due to Predation*, ICES C.M. 1979/G:52.
- Hrafnkelsson and Stefansson [2004], *A Model for Categorical Length Data from Groundfish Surveys*, Can. J. Fish. Aquat. Sci., accepted.
- J.N. Ianelli and D.A. Fournier [1998], *Alternative Age-Structured Analyses of the NRC Simulated Stock Assessment Data*, in *Analyses of Simulated Data Sets in Support of the NRC Study on Stock Assessment Methods* (V.R. Restrepo, ed.), NOAA Tech. Memo. NMFS-F/SPO-30, 96 pp.
- J. Jakobsson, A. Gudmundsdottir and G. Stefansson [1993], *Stock Related Changes in Biological Parameters of the Icelandic Summer Spawning Herring*, Fish. Oceanogr. **2**, 260–277.

N.C.H. Lo, L.D. Jacobson and J.L. Squire [1992], *Indexes of Relative Abundance from Fish Spotter Data Based on Delta-Lognormal Models*, Can. J. Fish. Aquat. Sci. **49**, 2515–2526.

P.D.M. MacDonald and T.J. Pitcher [1979], *Age-Groups from Size-Frequency Data: A Versatile and Efficient Method of Analyzing Distribution Mixtures*, J. Fish. Res. Board Can. **36**, 987–1001.

K.G. Magnusson and O.K. Pálsson [1989], *Trophic Ecological Relationships of Icelandic Cod*, Rapp. P.-v. Reun. Cons. int. Explor. Mer, **188**, 206–224.

R. Mohn and W.D. Bowen [1994], *A Model of Grey Seal Predation on 4VsW Cod in Its Effects on the Dynamics and Potential Yield of Cod*, DFO Atlantic Fisheries Research Document **94/64**.

J. Neter, W.H. Kutner, C.J. Nachtsheim and W. Wasserman [1996], *Applied Linear Statistical Models*, McGraw-Hill, Boston, 1408 pp.

K.R. Patterson, R.M. Cook and C.D. Darby, S. Gavaris, V.R. Restrepo, A.E. Punt, B. Mesnil, D.W. Skagen, G. Stefansson and M. Smith [2000], *Validating Three Methods for Making Probability Statements in Fisheries Forecasts*, ICES C.M. 2000/V:06.

K. Patterson, R. Cook, C. Darby, S. Gavaris, L. Kell, P. Lewy, B. Mesnil, A. Punt, V. Restrepo, D.W. Skagen, and G. Stefansson [2001], *Estimating Uncertainty in Fish Stock Assessment and Forecasting*, Fish and Fisheries **2**, 125–157.

M. Pennington and J.H. Volstad [1994], *Assessing the Effect of Intra-haul Correlation and Variable Density on Estimates of Population Characteristics from Marine Surveys*, Biometrics **50**, 725–732.

T. Polacheck, R. Hilborn and A.E. Punt [1993], *Fitting Surplus Production Models: Comparing Methods and Measuring Uncertainty*, Can. J. Fish. Aquat. Sci. **50**, 2597–2607.

V.R. Restrepo, K.R. Patterson and C.D. Darby, S. Gavaris, L.T. Kell, P. Lewy, B. Mesnil, A.E. Punt, R.M. Cook, C.M. O'Brien, D.W. Skagen, and G. Stefansson [2000], *Do Different Methods Provide Accurate Probability Statements in the Short Term?*, ICES C.M. 2000/V:08.

L.J. Richards [1991], *Use of Contradictory Data Sources in Stock Assessments*, Fish. Research **11**, 225–238.

H. Scheffe [1959], *The Analysis of Variance*, John Wiley & Sons, New York., 477 pp.

J.T. Schnute [1991], *The Importance of Noise in Fish Population Models*, Fish. Research **11**, 197–223.

J.T. Schnute and R. Hilborn [1993], *Analysis of Contradictory Data Sources in Fish Stock Assessment*, Can. J. Fish. Aquat. Sci. **50**, 1916–1923.

J.T. Schnute and L.J. Richards [1995], *The Influence of Error on Population Estimates from Catch-Age Models*, Can. J. Fish. Aquat. Sci. **52**, 2063–2077.

G. Stefansson [1992], *Notes on the Stock-Dynamics and Assessments of the Icelandic Cod*, ICES C.M. 1992/G:71.

G. Stefansson [1996], *Analysis of Groundfish Survey Abundance Data: Combining the GLM and Delta Approaches*, ICES J. Marine Science **53**, 577–588.

G. Stefansson [1998], *Comparing Different Information Sources in a Multispecies Context*, in *Fishery Stock Assessment Models: Proc. of Internat. Symposium* (F. Funk, et al., eds.), Alaska Sea Grant College Program, AK-SG-98-01, 741–758.

G. Stefansson [2001], *An Implementation of Growth*, in *Development of Structurally Detailed Statistically Testable Models of Marine Populations* (dst2), Anon. [2002], QLK5-CT1999-01609, Progress Report for 1 January 2000 to 31 December 2000, MRI Tech. Report 78.

G. Stefansson [in prep.], *Consistent Combination of Likelihood Components*.

G. Stefansson and O.K. Palsson, eds. [1997a], *Bormicon: A Boreal Migration and Consumption Model*, Marine Research Institute Tech. Report, 253 pp.

G. Stefansson and O.K. Palsson [1997b], *Statistical Evaluation and Modelling of the Stomach Contents of Icelandic Cod* (*Gadus morhua*), Can. J. Fish. Aquat. Sci. **54**, 169–181.

G. Stefansson and O.K. Palsson [1998], *A Framework for Multispecies Modelling of Boreal Systems*, Reviews in Fish Biology and Fisheries, **8**, 101–104.

G. Stefansson, J. Sigurjonsson and G.A. Vikingsson [1997], *On Dynamic Interactions Between Some Fish Resources and Cetaceans off Iceland Based on a Simulation Model*, J. Northwest Atl. Fish. Sci. **22**, 357–370.

B.Ae. Steinarsson and G. Stefansson [1986], *Comparison of Random and Fixed Trawl Stations in Icelandic Groundfish Surveys and Some Computational Considerations*, ICES C.M. 1986/D:13.