Are stock assessment methods too complicated?

A J R Cotter[1], L Burt[2], C G M Paxton[2], C Fernandez[3], S T Buckland[2] & J-X Pan[4]

[1]Centre for Environment, Fisheries and Aquaculture Science, Lowestoft NR33 0HT; [2]School of Mathematics and Statistics, University of St Andrews, St Andrews KY16 9SS; [3]Department of Mathematics and Statistics, Fylde College, Lancaster University, Lancaster LA1 4YF; [4]Department of Mathematics, The University of Keele, Staffordshire ST5 5BG, UK

Abstract

This critical review argues that several methods for the estimation and prediction of numbers-at-age, fishing mortality coefficients *F*, and recruitment for a stock of fish are too hard to explain to customers (the fishing industry, managers, etc.) and do not pay enough attention to weaknesses in the supporting data, assumptions and theory. The review is linked to North Sea demersal stocks. First, weaknesses in the various types of data used in North Sea assessments are summarized, i.e. total landings, discards, commercial and research vessel abundance indices, age-length keys and natural mortality (*M*). A list of features that an ideal assessment should have is put forward as a basis for comparing different methods. The importance of independence and weighting when combining different types of data in an assessment is stressed. Assessment methods considered are Virtual Population Analysis, *ad hoc* tuning, extended survivors analysis (XSA), year-class curves, catch-at-age modelling, and state-space models fitted by Kalman filter or Bayesian methods. Year-class curves (not to be confused with 'catch-curves') are the favoured method because of their applicability to data sets separately, their visual appeal, simple statistical basis, minimal assumptions, the availability of confidence limits, and the ease with which estimates can be combined from different data sets after separate analyses. They do not estimate absolute stock numbers or *F* but neither do other methods unless *M* is accurately known, as is seldom true.

Keywords catch-at-age modelling, fisheries data, natural mortality, stock assessment, virtual population analysis, year-class curves

Correspondence:
John Cotter, CEFAS Laboratory, Lowestoft NR33 0HT, UK
Tel.: + 44 1502 524323
Fax: + 44 1502 513865
E-mail: a.j.cotter@cefas.co.uk

Received 17 June 2003
Accepted 7 June 2004

Introduction

Commercial fisheries for marine species are often managed with the help of advice from stock assessment working groups (WGs) that sit annually to estimate numbers-at-age, coefficients of mortality, and annual recruitments for each relevant stock. Although an extensive literature now exists on mathematical methods for this work (e.g. Pope 1979; Megrey 1989; Fréon and Misund 1999; Quinn and Deriso 1999), many WGs persist with one basic method for years and find it difficult to make an objective case to change other than by comparing the results of different methods applied to the available data. Researchers supporting WGs have made use of simulated fish stocks to compare assessment methods because the parameters estimated by each method may then be compared with the known values used for the simulation (e.g. Kimura 1990; National Research Council 1998). Both of these empirical approaches are scientifically constrained. The test data used by WGs may have features that favour one method over another, while simulation is always limited by the chosen boundaries of the work, e.g. the National Research Council (1998) studies omitted variable discarding of fish at sea. A more general approach should analyse the suitability of underlying assumptions and alternative models, the fitting process, and the constraints imposed by the quality of data available.

The present review critically analyses the basic estimation tasks of age-based assessments. We find a tendency for estimation methods to become more elaborate than can be justified by the quality of the data, assumptions, and theory that support them. We do not wish to imply by this that all assessments are faulty; in our experience WGs take considerable care to check that results are consistent with accumulated knowledge of the fishery. Our main concern is that advanced mathematical methods are more reliant on assumptions, less intelligible, harder to verify, less open to criticism, and therefore less trustworthy by the primary customers, i.e. fishers, fishery managers, and others with direct economic, social or conservation interests in the stock. In addition, stock assessments, and the data collection for them, absorb much scientific effort. Simplified procedures could allow more study of sampling methods, commercial fishing practices and the biology of the species concerned, thereby allowing management of stocks to be improved in other ways. Similar views were expressed by Hilborn (2003).

The review is centred on North Sea demersal stocks because their situation is familiar to us, they have been studied extensively, and because they are typical of many shelf-sea stocks in relation to the assessment task. First, well-known weaknesses in the various types of data used in North Sea assessments are summarized. Next, ideals for a stock assessment are proposed as a set of common standards for comparing processes and methods. We draw attention to the difficulties associated with the combination of data from different sources, and then critically assess a selection of well-known assessment methods. The selection reflects recent European preferences but is mainly designed to show how data and modelling can interact under a variety of mathematical approaches. [A complementary set of assessment methods preferred in North America is reviewed along with other aspects of stock management by National Research Council (1998).] We conclude that one of the simplest and oldest methods, that based on 'catch-curves' re-formulated to apply to year classes (cohorts) over time, is hard to beat in many respects.

Data types used in North Sea assessments

Landed quantities

Landed quantities are a primary source of information for North Sea stock assessments (Cook 1997) but may be inaccurate because of politically or financially motivated under-, and sometimes over-reporting. Further variance arises when converting quantities to numbers-at-length and numbers-at-age on the basis of samples of length measurements and subsamples of otoliths used to create age-length keys (ALKs). The resulting length and age distributions are properties of the total landings, not of the actual fish stock (Hilborn 1992; Fabrizio and Richards 1996), because fish and fishing activities both tend to be clustered in space and time, and because gear selectivity and discarding (Kell and Bromley 2004) affect the relationship between landings and the stock. Uncertainty over whether or not the stock itself is a discrete unit (Smith *et al.* 1990; Kell *et al.* 2004) further weakens the information conveyed by landed quantities about the set of fish assumed to be the subject of the assessment.

Sampling of landings may be restricted for reasons of operational economy to short time periods and to small subsets of vessels in the fishing fleet (Fournier and Archibald 1982). This can be expected to reduce sampling variance by causing bias. Recent bootstrap studies of North Sea cod and plaice landings data (Pastoors *et al.* 2001) indicate varying coefficients of variation (CVs) with age, ranging approximately from 5% for 3-year olds, to 50% for older fish. Such estimates include measurement noise but cannot allow for geographical or temporal restrictions on the original samples that were bootstrapped.

Discards

Landings have traditionally been referred to as 'catch' in stock assessments but a more rigorous distinction between the two terms is needed (Fabrizio and Richards 1996) when discarding at sea is known to cause significant losses of fish. Table 1

Table 1 Numbers of cod, haddock and whiting caught and discarded by the European otter trawl fleet fishing the North Sea (ICES Area IV) during 2000 as estimated from catch sampling data and fleet raising factors provided by Danish, English, German, Norwegian and Swedish partners in EC project 98/097 (with corrections to original estimates).

	Cod			Haddock			Whiting		
Age (years)	Catch thousands	Discard thousands	%	Catch thousands	Discard thousands	%	Catch thousands	Discard thousands	%
0	304	304	100	28	28	100	0	0	0
1	54816	54398	99	37072	36709	99	12430	12263	99
2	43509	5590	13	39073	15301	39	15314	7489	49
3	6604	67	1	7110	788	11	18727	1975	11
4	21831	28	0	12296	1414	12	9092	1244	14
5	2180	1	0	13232	1499	11	11626	1780	15
6	189	0	0	4508	449	10	4112	630	15
7	459	0	0	667	58	9	1599	154	10
8	51	0	0	534	36	7	995	215	22
9	11	0	0	10	0	0	187	36	19

%, percentage discarded by number.

shows sampling information from the international otter trawl fishery in the North Sea during 2000. More than 90% of the many 1-year old cod, haddock and whiting caught were discarded and almost certainly died because of suffocation, damaged swim-bladders, etc. As discards are mostly smaller than the minimum landing size (MLS), numbers-at-length *caught* will, in many circumstances, only be well estimated by numbers-at-length *landed* for length categories over the MLS. However, even in this case, landings may not be a good measure of catch, for example, when quota or licensing restrictions are causing discarding of otherwise marketable fish, or when the value of the fish is too low for fishers to bother with them (e.g. North Sea whiting and small plaice). Examples of discarding of North Sea species at lengths over the MLS are shown in Fig. 1 but this does not give a complete impression of the mismatch between landings and catch because, as shown in Table 1 for North Sea cod, haddock and whiting, the ogives of discarding span an age range of several years, particularly for whiting. The proportions of all of these age-classes in the stock will therefore be variably under-represented in landings depending on conditions for growth in each year, the exact ages (in years and months) of the fish landed, and the selectivities of the fishing gear in use commercially.

Discarded quantities for North Sea fisheries are mostly estimated by observers sailing on a sample of vessels in the fishery (van Beek 1998; Stratoudakis *et al.* 1998). Usually the catches themselves must also be sampled by the observer (Tamsett *et al.* 1999a), and sometimes the hauls within each sampled trip, making a multilevel sampling scheme (Tamsett and Janacek 1999; Allen *et al.* 2002). The reliability and precision of observer data depend on the survey design, the number of fishing trips observed and the variability among them, how the ALKs were formulated, and the method of raising survey results to the fleet level (Stratoudakis *et al.* 1999; Anonymous 2000). One must also trust fishers to process their catches in the normal way when an observer is present (Liggins *et al.* 1997). Random selection of vessels for sampling can diminish sampling bias and permit estimation of standard errors but is difficult to achieve well in practice (Cotter *et al.* 2002). The high expense and practical difficulties of catch sampling keep the numbers of observed trips small in North Sea fisheries. For these reasons, sampling CVs tend to be high, up to 50% (Stratoudakis *et al.* 1999; Cotter

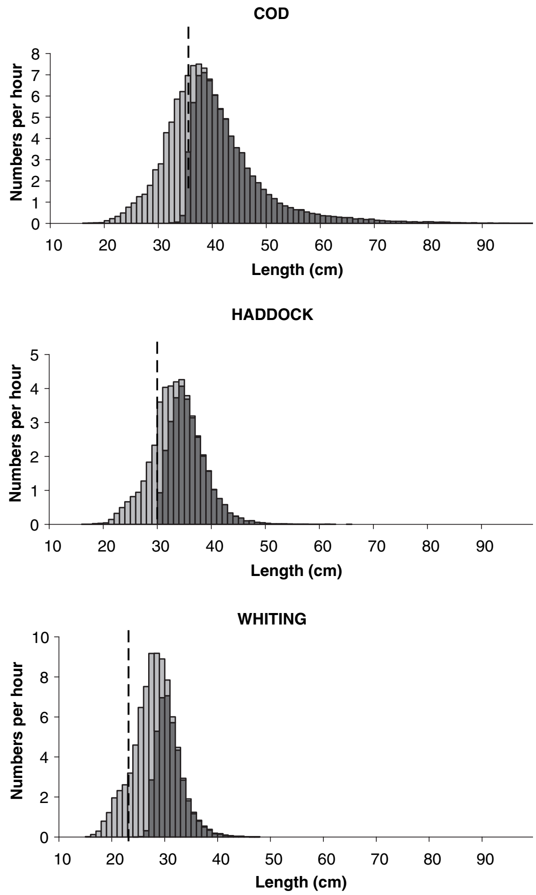

Figure 1 Length frequency distributions for cod, haddock and whiting discarded and retained per hour by the English North Sea trawl fishery between May 1997 and September 1998, as found from catch-sampling at sea by CEFAS, Lowestoft. Pale columns, discards; dark columns, retained. Note different vertical scales for each species. Dotted vertical lines show minimum landing sizes at the time.

et al. 2002; Rochet *et al.* 2002) and may, furthermore, be underestimated if sampling was restricted for practical reasons, or if the CV does not fully allow for variance arising from multiplication by raising factors.

Groundfish survey indices

Catch per unit effort (CPUE) series, often known as 'abundance indices', obtained from groundfish surveys are important in stock assessment because they are independent of commercial fishing and because they are assumed to be proportional to stock abundances by virtue of being collected using a constant survey design from year to year. However, there are reservations about this assumption (Paloheimo and Dickie 1964; Myers and Stokes 1989; Swain and Sinclair 1994; Pennington and Godø 1995; Rose and Kulka 1999). One is that stock distribution in relation to the survey area, or to fixed-position fishing stations within it, may itself vary with abundance (Hutchings 1996) or with changing environmental factors, leading to non-proportionality between abundance and CPUE. Another is that catchability of fish may vary with season, location and other factors (Godø 1994; Hjellvik *et al.* 2002a). A third arises when the catching power of a survey is affected by operational changes to the vessel or gear, although factors can be estimated to compensate for this (Pelletier 1998; Cotter 2001). Trawl surveys are thought to under-estimate the numbers of young fish (Godø 1994) despite the use of standard, small-mesh codend liners (Anonymous 1999). Unfortunately, estimation of the variances and covariances of survey CPUE series in an absolute sense, i.e. without conditioning on the specific survey design (Nicholson *et al.* 1991), requires the difficult assumption that the design did not bias the survey index or its variance. Bootstrap estimates of precision for three North Sea surveys (Beare *et al.* 2002), i.e. conditional upon their designs, indicate CVs for abundance indices for ages 2–5 of the order of 10–40% for cod and whiting, and slightly better for haddock, depending on the survey. A practical estimate of survey variance was obtained by Hjellvik *et al.* (2002b). They used parallel trawling experiments in the Barents Sea and found that measurement error variance was approximately 2–5% of the total variability in the survey catches.

Commercial fleet landings per unit effort indices

Time-series of landings per unit effort (LPUEs) of commercial fishing fleets (or, in a few cases, CPUE series with discard estimates included) have been used in the North Sea and in many other ICES stock assessments because of assumed proportionality with stock abundances. As commercial fleets catch many times more fish than research vessel (RV) surveys, their LPUEs appear to be more precise statistically than the latters' CPUEs. However, this may belie a dependence on poor sampling of landings for length and age determinations. All of the reservations about landings data, discussed above, apply equally to commercial LPUE data.

Commercial LPUEs are more likely to suffer from varying bias over time than RV CPUEs (Collie and Sissenwine 1983; Hilborn and Walters 1992; Quinn and Deriso 1999) because of changing markets, vessels, discarding practices, fish finding techniques, gear, fishing tactics and stock distributions (Fréon and Misund 1999; Rose and Kulka 1999). Commercial LPUEs may therefore reflect all of these factors more than they reflect stock abundance. They are obviously of no use for the age-classes that are variably selected and discarded. Harley *et al.* (2001) found that many commercial LPUE indices for flatfish and gadoids taken from ICES stock assessments remained high when abundance, as indicated by RV surveys, was declining.

Age-length keys

Age-length keys are central to the age-structured stock assessments used for North Sea demersal species, yet the benefits of this two-staged sampling process (Thompson 1992; Lai 1993) are questionable when, as is common practice, ALKs are compiled for large geographical areas and time periods. This is because variations of length-at-age across the time–space domain of an ALK may cause bias, particularly if the length and age samples were collected with differing regional or temporal intensities, something which is quite difficult to avoid in practice if samplers tend to take advantage of a few large catches to collect their quota of otoliths. Sen (1986) noted that variation of length-at-age generally was greater among fishing vessels than within the catches of individual vessels, and evidence for regional differences in length-at-age for North Sea species exists for cod (Graham 1933; Berg and Albert 2003; Vinje *et al.* 2003), haddock (Thompson 1928), and plaice (Bromley 2000). Temporal variation occurs particularly among young fish that grow appreciably over the typical quarterly period of an ALK. Cotter (1998) argued that use of sparse ALKs specific for each fishing station, followed by aggregation of the estimated numbers-at-age by area or period would help to avoid these types of error in groundfish survey results.

Age-length keys are of course vulnerable to age determination errors (Bradford 1991; Richards *et al.* 1992) which not only alter relative frequencies of a cohort but may also distort estimates of fishing mortality (Fournier and Archibald 1982). Estimation of sampling errors in ALKs is greatly complicated by the usual two stage sampling process.

[Simple random sampling for age composition would be much easier to analyse statistically and may in some situations be more efficient than two-stage sampling (Smith 1989).] Other statistical complications with ALKs are caused by correlations between numbers of fish occurring in different length and age groups of catches (Tanaka 1953; Kimura 1977). The correlations arise because of related geographical distributions, related swimming behaviour and vulnerability to the gear, and catch sampling (Cotter 1998; Pennington *et al.* 2002).

Coefficient of natural mortality, M

The coefficient of natural mortality, M, when subtracted from total mortality, Z, yields fishing mortality, F, an important output of stock assessments used for predicting future catch sizes, etc. Few stock assessment WGs would claim that M is accurately known (Sparholt 1990). Beverton and Holt (1957; section 14.3.2) and Beverton (1964) describe estimation of M for North Sea species. The most secure estimates were from small numbers of surviving plaice following cessation of fishing during the Second World War. Estimates of predatory M have been attempted from analysis of stomach contents for several North Sea species caught on RV surveys, followed by application of multispecies virtual population analysis (VPA) (Pope 1989). Sparholt (1990) added to the estimates of M for North Sea species. Of related interest are estimates of M by Sinclair (2001) for cod in the Southern Gulf of St Lawrence following closure of that fishery in 1993, and a general review of estimation of M by Vetter (1988). Sampling variability and expense are major problems for estimation of M when commercial fishing has not conveniently stopped.

Rivard (1989) points out that fixing M at a more or less arbitrary value in the absence of trustworthy scientific measurements requires that stock size estimates must be seen as relative indices rather than absolute values. Schnute and Richards (1995) and Clark (1999) report bias from erroneous M while Gudmundsson (1998) notes that variation of assumed M from 0.1 to 0.3 led to a 20–30% change in stock weights estimated by a time-series method. Pope (1979) points out that variation of M with age is confounded with exploitation pattern with age, i.e. with catchability, selectivity and discarding, while variation with time is confounded with variation of F. Vetter (1988) reviews numerous examples of fishery models that are sensitive to M, and numerous reasons for it to vary.

A relationship between M, estimated cohort size, and F is readily understandable. Since $Z = F + M$, M must lie between the limits 0 and Z. If $M = 0$, $F = Z$, and the initial numbers in each year class would only be the sum of the numbers caught until the year class becomes extinct through fishing alone. If $M = Z$ on the other hand, $F = 0$, implying that the numbers of fish caught have negligible influence on the total population size. Therefore the population must be infinite. Thus small M lead to large estimates of F and small estimates of cohort size, and vice versa.

Ideals for stock assessment

We suggest below several ideals for a stock assessment that are intended to serve as standards for comparison of the different types of data and methods. We are not suggesting that these ideals are widely accepted or easy to achieve. Most obviously, an ideal stock assessment should produce accurate results with appropriate measures of confidence. Additionally, we propose that it should:

1. *Be explicable to the fishing industry and other interested, non-mathematical parties*. Advanced mathematical methods may be acceptable if the effort is made to explain them (Richards *et al.* 1997). Smith and Punt (1998) describe a Bayesian stock assessment in which industry and fishery management personnel both took part in the estimation process.
2. *Weight each data set in the assessment in relation to how much independent information it contributes, and to how much information it shares with other data sets being used*. The information in a data set depends on accuracy, i.e. the inverse of variance around the true values. A data set is independent if no component of its variance is shared with any other data sets being used, e.g. through dependence on a common sample, sampling design, fitted model, ALK or estimated raising factor. Shared information should be weighted once, not once in each data set containing it.
3. *Use no sampling data or estimates more than once*. Repeated use of the same data set in an assessment does not add information but merely increases the weighting being given to that set, and in particular, to its bias and imprecision.

When data of low accuracy are incorporated wholly or partly into two different data sets being applied to an assessment, a spurious concordance arises from the common source of errors. This might be mistaken as evidence of a functional relationship, or of good agreement between the two sets.

4 *Make due allowance for dependence among data when estimating standard errors and statistical significance*. Correlations *within* fisheries data sets are typical, e.g. in matrices of numbers-at-length and numbers-at-age. As a result, degrees of freedom (d.f.) after fitting a model are overestimated by the usual method of subtracting the number of parameters fitted from the number of observations. Standard errors are then underestimated with the consequence that terms in the model may be judged statistically significant when they are not. Collie and Sissenwine (1983), Rivard (1989) and Shepherd (1999) previously noted the importance of independence *among* the different data sets put into a stock assessment. Allowing for dependence in data is not always straightforward. Examples of statistical methods are discussed below.

5 *Use parsimonious models*. A simple model with relatively few parameters (the principal of parsimony, see Burnham and Anderson 2002) is likely to be more durable, more precisely estimated for a given number of data points, easier to justify by prior scientific reasoning, easier to fit, and less likely to imply fallaciously that a component of sampling variability was caused by some factor relevant to the assessment. Pope (1977) and Walters and Martell (2002) remind us of the importance of having reasonable numbers of d.f. for each parameter estimated for a fisheries model. Richards and Schnute (1998) show that much can be achieved with a simple assessment model having few parameters.

6 *Depend upon the minimum of subjective decisions*. Subjective decisions reduce the scientific repeatability of an assessment, undermine theory, and make the assessment harder to explain (ideal 1). Hilborn (2002) notes that modern fish stock assessments (in the USA) rely on 'dozens, sometimes hundreds, of individual judgements'.

7 *Be calculable in less than, say, 1 h*. Some assessment methods perform computations in seconds. Others use iterative methods that take hours or days. This would tend to discourage repeat runs to check that the solution is truly optimal and not just one peak of likelihood in a mountain range, or to check sensitivity of the solution to minor changes in input parameters or data.

Combining different types of data in a stock assessment

Many assessments use data of more than one type and from more than one source. They should therefore allow for different error structures in each set. The statistical implications vary according to whether one joint model is being fitted to all data sets together, or whether the different sets are being modelled separately followed by averaging of the different estimates of the common parameters.

Using a single, joint model for all data sets

Methods for objectively weighting different data sets used to fit one, joint model in accordance with ideal (2) are not well established in fish stock assessment. The problem is worse if the data sets are not independent because the independent and dependent components of information call for separate weightings individually and jointly, respectively, relative to other data sets used in the assessment. The implicit complications are a strong argument for maintaining data sets independent, so far as possible.

Sampling variances as reciprocals (invariances), if available, are likely to be poor for estimating the weights to be applied to different data sets because they do not allow for the variability explained by the model. Furthermore, the situation is multivariate because of correlations between variables, e.g. across-ages-within-a-year, as found for RV survey indices (Myers and Cadigan 1995). This suggests weighting in relation to the sample variance–covariance matrix but this, in turn, raises a special estimation problem because of the shortness and lack of replication of most fisheries data series. At least one independent multivariate observation is needed to estimate each different element of the matrix (Rencher 1995), and many more to produce stable estimates. For example, a 4×4 covariance matrix has 10 different elements to be estimated, implying that at least 10 years of annual observations are needed before we can even start to estimate the matrix without algebraic dependences among the elements.

Quinn and Deriso (1999, p. 339) recommend various diagnostics, sensitivity checks and repeat runs to determine the robustness of estimates of assessment parameters to variably weighted data sets. A concern then is that subjectivity, contrary to ideal (6), may creep in. Statistical literature (Aitkin 1987; Verbyla 1993) includes maximum likelihood and restricted maximum likelihood methods suitable for weighting different data sets when fitting a linear model. We are not aware of trials of these methods by stock assessment scientists. Similar results can be achieved by applying generalized least squares iteratively in the same way, a process referred to as iteratively weighted least squares (IWLS) (Carroll and Ruppert 1988) which is itself an application of the expectation–maximization algorithm (Dempster *et al.* 1977). Cotter and Buckland (2004) show how this procedure may be applied to fit year-class curves (see below) to different types of data with a small number of iterations. Quinn and Deriso (1999; p. 353) point out that IWLS may result in biased data with low variance receiving excessive weighting. However, no objective weighting method can be expected to allow for such a deficiency which probably stems from restrictions on sampling.

Using separate models for each data set

An immediate advantage with respect to ideal (1) of fitting models separately to each data set is easier visualization, verification and interpretation of the performance of each set individually. Separate modelling produces separate residual error variances and/or separate likelihoods or posterior distributions which could subsequently be used to combine the multiple estimates of each parameter that is common to more than one of the models into a single, pooled estimate (Hilborn and Walters 1992). Independence of the data sets is then crucial to avoid difficulties with ideals (3) or (4). Despite the advantages, separate modelling of different data sets is not a common feature of current stock assessment methods.

Review of selected stock assessment methods

VPA and related methods

Basic VPA

Virtual population analysis, also referred to as sequential population analysis, and in an approximate form (Pope 1972, 1982; Siddeek 1982), as cohort analysis, has a long history (Megrey 1989) and is widely used by ICES WGs, usually in one of its developed forms, e.g. XSA, to be discussed below. Basic VPA involves solving

$$\frac{C_{a,y}}{N_{a+1,y}} = \frac{F_{a,y}}{Z_{a,y}}\{\exp(Z_{a,y}) - 1\} \qquad (1)$$

for each age a in year y for every year class. C is nominally 'catch' but it may in practice be landings; Z is the coefficient of instantaneous total mortality where $Z = F + M$ as noted above. M takes assumed, fixed values while a terminal F is supplied by the analyst for the oldest age category, A, of the year class. It is easily found that estimates of F and N in middle and young age classes are little affected by the chosen value of terminal F. This robustness was an important reason for the widespread adoption of VPA for stock assessments.

Virtual population analysis fails the basic requirement of an assessment to provide confidence limits for its estimates. This is because the data and Equation (1) are both assumed to be exact, a subjective decision contrary to ideal (6). One consequence is that trivial variations of F and recruitment may attract comment in an assessment report when they would be more securely thought of as consequences of sampling error. VPA is generally regarded as a black box by non-scientists probably because of the need for a numerical method of solution and, typically, a lack of illustrated outputs. This gives a low score against ideal (1). Analyses of the effects on VPAs of many types of errors in landing or catch values, M, and terminal F are now available (e.g. Saila *et al.* 1985; Sampson 1987; Rivard 1989; Bradford 1991; Mertz and Myers 1997). These errors, most notably inaccurate landings data, lack of, or biased discard estimates in presumed 'catch', and poorly known M, clearly show that VPA cannot be considered to provide error free or absolute estimates of recruitment, population sizes, or F.

Ad hoc tuning

The main procedural problem with VPA is that it operates retrospectively and provides least information about stock sizes and values of F in the final (terminal) year, that of most interest for predicting future performance of the fishery. Tuned VPAs were developed to rectify this. Ten tuning methods were evaluated by Pope and Shepherd (1985) but none reduces the opacity of VPA so as to permit warm approval against ideal (1). Tuned VPAs have generally been replaced at ICES by XSA but a brief analysis of one well-used example, namely the Laurec–Shepherd (L-S) method (Laurec and Shepherd 1983), is instructive about similar problems arising with XSA. The term 'fleet' below may refer to a commercial fishing fleet or an RV survey.

Laurec-Shepherd tuning, as set out by Darby and Flatman (1994), involves iteration of VPA and tuning steps until stable values for terminal F-at-age are found. The VPA step produces a matrix of $F_{y,a}$ for years, y, and age, a. These are partitioned amongst each tuning fleet, f, according to their respective catches (or landings), C, in that year, i.e.

$$F_{y,a,f} = \frac{C_{y,a,f} F_{y,a}}{C_{y,a}}$$

The 'catchability' of each fleet, defined as the ratio of partial fishing mortality and effort, i.e.

$$q_{y,a,f} = \frac{F_{y,a,f}}{E_{y,f}}$$

is assumed to vary randomly from year to year around a constant value, $q_{a,f}$, estimated as the geometric mean of catchability over years 1 to $Y - 1$. Applying the mean to the terminal year, Y, of the assessment together with new effort and catch information in that year provides a single-fleet estimate of terminal F for the whole fishery. These estimates are weighted by the invariance of $\log q_{y,a,f}$ to find the all-fleet estimate of terminal F, denoted $F_{y,a}$, which is then used to start a new VPA, and so on. Having found a stable solution, terminal population numbers and future catch (or landings) projections can be calculated and standard errors for the terminal Fs are available. Note that the L-S method does not diminish reliance on the assumed accuracy of the VPA, contrary to ideal (6).

Now, suppose that there are statistical errors, $e_{y,a,f}$ and $e_{y,a}$, to be added to the true fleet catch-at-age, $C_{y,a,f}$, and to the true total catch-at-age by all fleets, $C_{y,a}$. Adding these errors to the estimator of catchability given by Darby and Flatman (1994) and taking logs, as needed for the L-S method, gives:

$$\log(q_{y,a,f}) = \log \frac{F_{y,a}}{E_{y,f}} + \log \frac{C_{y,a,f} + e_{y,a,f}}{C_{y,a} + e_{y,a}} \tag{2}$$

The variance of $\log q_{y,a,f}$ will include the variances of each of the terms on the right. Variance of the second will depend on the accuracy of the catch-at-age matrix plus an additional component as a result of error if $e_{y,a,f}$ varies independently of $e_{y,a}$. On the other hand, if the two error terms are positively correlated, they will reduce the variance of the second term by smoothing annual fluctuations of the ratio of numerator to denominator. Positive correlation can be expected when:

- The landings of fleet, f, usually a commercial fleet, have been included in the landings for the total fishery. Misreported tonnages, and inaccurate ageing are two examples of errors that would be common to both sets of data.
- Poor or absent estimates of discarding occur in both $C_{y,a,f}$ and $C_{y,a}$.

The variance of $\log q_{y,a,f}$ will also include covariance between the two terms on the right of Equation (2) because the Fs and Cs are both derived from the C matrix in VPA, a double use of data, contrary to ideal (3) if sampling errors are acknowledged. This analysis casts doubt on the reliability of the fleet weighting procedure of the L-S method and may mean that estimates of terminal F are not based on those fleets characterized by the most consistent catchabilities, as intended.

More recently, Myers and Cadigan (1995) described models that revive the *ad hoc* tuning approach. By using RV survey indices rather than commercial LPUE series they avoided dependence between the landings and the tuning data. In addition, by using mixed models with a random *year* effect in the survey indices they allowed for correlations across ages within years and reduced the variability and bias of the estimates of terminal year survivors. However, their method does not drop the strong assumption that VPA data are perfectly known.

Extended survivors analysis

Darby and Flatman (1994) and Shepherd (1999) state that the performance of *ad hoc* tuning methods is unsatisfactory because they unrealistically treat abundance indices for the final year as exact, and disregard estimates of year-class strength available from the sequence of catches taken from a cohort by the tuning fleets. This led to the development of a more elaborate method called 'extended survivors analysis' (Darby and Flatman 1994; Shepherd 1997, 1999), or XSA for short. XSA can produce standard errors for terminal population estimates but they do not allow for uncertainties in the landings- or catch-at-age matrix. XSA, being mathematically more intricate than VPA and tuned VPA, rates poorly against ideal (1).

Described briefly, tuning series of LPUE or CPUE indices for commercial fleets and RV surveys are first evaluated against a standard VPA in a screening process. This involves a catchability model relating the tuning fleet indices to the VPA populations at each age. Fleets whose residuals show pronounced year effects or trends over time are rejected from the assessment. CPUE or LPUE information is then used to calculate fleet-specific estimates of survivors of each year class at the end of the terminal year of the assessment. Each estimate is weighted by its residual invariance in the overall geometric mean estimate which may also include geometric mean VPA population at the age of the year class in the terminal year, weighted by its invariance, a process referred to as 'shrinkage' to the mean population. Knowing the final survivors, new estimates of terminal F are possible for the VPA, allowing iteration of the process until terminal Fs change by less than a small amount. XSA has several options such as alternative catchability models, down-weighting of historical data, and inclusion of one further estimate of survivors in the terminal year, that found by using F-at-age averaged over several previous years as a smoother for year-by-year estimates of F. These options allow a skilled analyst to bring the computed results into line with perceptions of the fishery but, as they are subjective, highly technical judgements, they score low against ideals (1) and (6).

XSA has seen extensive use at ICES WGs but few analyses of its precision or accuracy have been published, perhaps because the many different ways of performing the method greatly complicate that task. Shepherd (1999) noted several necessary assumptions:

- the catch-at-age matrix and VPA are treated as exact
- all CPUE and RV survey data sets are operationally independently collected so that it is reasonable to treat their observation errors as statistically independent
- correlations amongst errors do not occur as a result of mis-reading of ages, or across-ages as a result of year effects (e.g. unusual survey coverage).

This review has identified several situations when these assumptions would be invalid.

Examination of the catchability models used by XSA clarifies the problems that can arise. Catchabilities, $q_{a,f}$, constant over years, y, are initially assumed for all age-classes, a, and fleets, f (Darby and Flatman 1994). The logarithmic form of the basic model is

$$\log U_{y,a,f} = \log q_{a,f} + \log N_{y,a} + e_{y,a,f}$$

where $U_{y,a,f}$ denotes CPUE, $N_{y,a}$ is the VPA estimate of population numbers, and $e_{y,a,f}$ is an error term. The alternative, power model considered for younger ages has an additional coefficient, $\gamma_{a,f}$, for the $\log N_{y,a}$ term to allow for a nonlinear relationship between catchability and stock size, e.g. if larger than average populations-at-age tend to spread to areas or depths not fished by the fleet. Both models are affected if tuning fleet f is a commercial fleet whose catch or landings is included in the total catches or landings of the fishery because any inaccuracies in the landings, ALKs, or discard information would then result in positive correlation between the errors in U and N leading to bias in the estimates of $q_{a,f}$ and $\gamma_{a,f}$. At the same time, the residual variance of $e_{y,a,f}$ is erroneously reduced suggesting that the model is a better fit than it actually is. Further complications could arise from inclusion of discard data for some fleets but not others. The end result may be inappropriate weightings in XSA similar to those already discussed for *ad hoc* tuning. The situation is made worse in XSA because the U and N are used repeatedly, once to estimate fleet catchabilities (Shepherd 1999, Equation 19), the U are used again for fleet estimates of population-at-age (Shepherd 1999, Equation 5), and the U and N are used to weight the fleet estimates to give terminal year survivors [Shepherd 1999, Equation 22 for which Darby and Flatman (1994; equations 12 and 13) provide formulae for the weights]. This multiple use of data, contrary to ideal (3), re-enforces the influence of those tuning data sets that best match a biased VPA because of positively correlated errors. This in turn could lead to erratic results from year to year if different tuning fleets are favoured by changing VPA errors, or to a gradually increasing bias as a result of the VPA and the most heavily weighted tuning fleet series wandering together from the truth. A further complication arises if catchability is not constant over time, as assumed for XSA, but follows a trend (Marchal *et al.* 2002).

Catch-at-age models without process error

Catch-at-age assessment models that acknowledge observation errors but not errors in the processes described by the model are discussed in this section. We distinguish two categories according to the treatment of time and age as either continuous or class variables.

Time and age treated as continuous variables (year-class curves)

It has long been known that a straight or nearly straight relationship is typically found when the log of CPUE or catch is plotted against age (Jensen 1939). The defining equation for total mortality, Z, for a population of N fish

$$\frac{dN}{dt} = -ZN$$

leads straightforwardly to a regression model for log CPUE:

$$\ln U_{a,c} = \ln k \cdot R_c + Z \cdot a + e_{a,c} \quad (3)$$

where k is a constant, R is recruitment at age zero to year class or cohort c, a is age, and e is usually thought of as $\sim N(0, \sigma^2)$, i.e. normally distributed around zero with variance σ^2. Quinn and Deriso (1999), section 8.2.2) show how an equivalent equation may be derived for the total catch of a fishery given that the probability of catching a fish is constant. For this to be true, effective effort of the fishery must be constant.

We call Equation (3) a 'year-class curve' to distinguish it from a 'catch curve' applied to a single catch or to the total catch from a single season. Catch curves make the strong assumptions of constant recruitment (Ricker 1975), fishing mortality and exploitation pattern (Deriso *et al.* 1985; Shepherd and Nicholson 1991), whereas year-class curves allow variable annual recruitments to be estimated and may be supplemented with terms to allow for varying fishing mortality and exploitation over time. Year-class curves do not estimate absolute population sizes directly as VPA is sometimes claimed to do but in view of the strong influence of poorly known values of M, discussed above, this may not be a disadvantage with respect to ideal (6).

Examples of terms that may be added to equation (3) are a^2 or $\ln(a + 1)$. These offer curvature to allow for varying selectivity and/or discarding with age. Specific events, such as a change in mesh size regulations, can be explicitly modelled by adding a term having an indicator variable set to zero before the event and unity afterwards. This method was applied to intercalibrate RV surveys when new nets and other changes were made (Cotter 2001).

Changes of Z over time may not be estimated by simply adding a term in years to equation (3) because the equivalence of years, year class and age prevent a unique solution unless the system is constrained in some way (Shepherd and Nicholson 1986, 1991). The problem goes away if CPUE was observed two or more times per year. Alternatively, time may be reckoned in multi-year periods instead of years. For example, use of a linear or quadratic term in bi-years could permit gradual changes in Z over time to be discerned. Some would argue that this is not good enough because fishing effort changes annually and so F and Z should also be allowed to vary annually. The counter argument is that annual estimates must be shown to contain a signal from the fishery and not merely be the algebraic results of bias and sampling error, or be the result of annual changes in M. This would be a challenging task.

Therefore, despite their stiffness, year-class curves can offer valuable support for stock assessments. They allow estimation of Z and CPUE or LPUE in the terminal year, and for the future, by projection. They are also visually instructive about the different properties of different data sets, recruitment, and the behaviour of Z over time in accordance with ideal (1). In many cases a satisfactory fit will be found with a model having relatively few parameters in accordance with ideal (5).

To illustrate this, Equation (3) was fitted to stock data for North Sea cod taken, slightly modified by adding 10% random variation (because a re-run of the official assessment was not intended), from Anonymous (2001) to compare results with and without English discarding data applied by simply raising landings to catch. Results for a subset of the year classes (selected to prevent overlap of the curves and data in the figures) from total landings, total catch, and the RV CPUE series are shown in Fig. 2a–c. In all cases, the slopes, −Z, were common to all year classes and fixed over time. The data do not indicate a need to allow more variation of Z. The estimated common Z, residual variances (unadjusted for correlated errors), and predicted numbers for each series at 1 January 2000 are given in Tables 2a and 2b. The total catch series estimated a larger value for Z (more negative slope) than the total landings series, and the RV CPUE series estimated a larger value again. The small standard errors imply that this was a real effect. It is readily explained by the presence of more small fish in the total catch series because of inclusion of discards, and more again in the RV CPUE series because of the use of small mesh nets. The exercise demonstrates the explanatory value of looking at different sets of data separately.

Year-class curves allow standard errors to be estimated for all parameters with the usual least

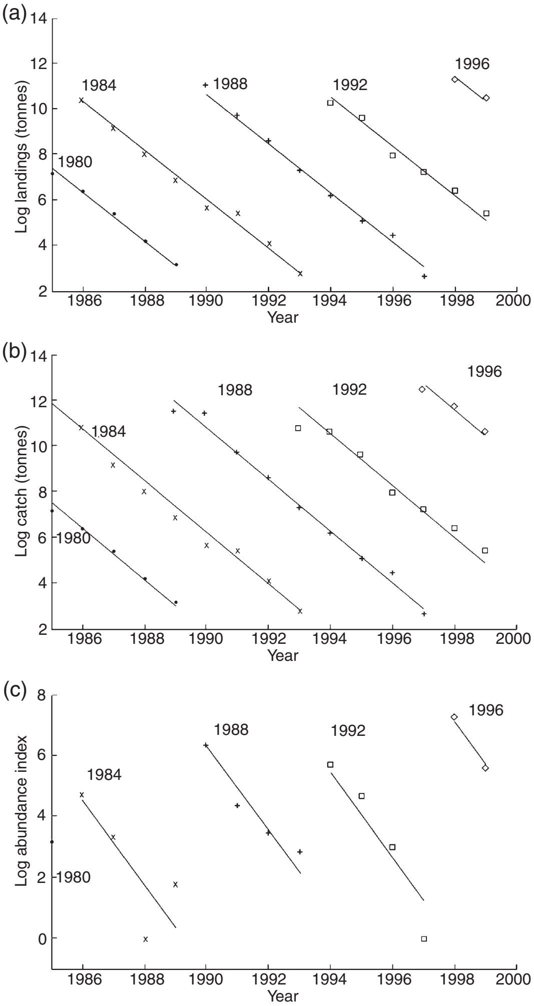

Figure 2 North Sea cod: year-class curves for selected year classes of North Sea cod fitted by ordinary least squares regression to (a) landings, 2 years and older; (b) total catch (i.e. landings + estimated discards), 1 year and older; (c) research vessel abundance indices, 2 years and older. Data slightly modified from ICES NSSK w.g. (Anonymous 2001).

squares formulae but they should be treated as lower limits because non-zero correlations of residual errors across-ages-within-years are likely to occur in any catch-age matrix, a problem noted in connection with XSA by Shepherd (1999). Cotter (2001) put forward a method for deciding an appropriate decrease in the d.f. of the residual variance thereby increasing standard errors to allow for this. Use of a random *year* effect in a mixed model (Myers and Cadigan 1995) is another approach worth considering.

Time and age treated as class variables

Time and age are treated as class variables rather than as continuous variables in several stock assessment methods. This allows estimation of parameters separately for each combination of year and year-of-age. An example is the catch-age equation as set out by Quinn and Deriso (1999), p. 334):

$$C_{a,y} = F_{a,y} Z_{a,y}^{-1} \left[1 - \exp(-Z_{a,y})\right] \times N_{r,y-a+r} \exp\left(-\sum_{j=1}^{a-r} Z_{a-j,y-j}\right) \quad (4)$$

where N here represents population size at the age r of recruitment to the fishery, and C represents catch (or landings with the definition of F adjusted). Both Z and F are allowed to vary for every age, a, and year, y, in the assessment. Having fewer parameters are the 'separable' models in which F is allowed to vary with year, and exploitation pattern, S, with age while possible interactions between them are ignored. The various published examples (e.g. Pope 1977; Pope and Shepherd 1982; Shepherd and Nicholson 1986; Megrey 1989) differ in method of solution and in how many constraints are applied to find a unique solution (Shepherd and Nicholson 1991). Kimura (1990) considered that these models

Table 2a North Sea cod: year-class curves estimated separately for landings, catch (i.e. landings + estimated discards), and research vessel (RV) CPUE data: estimated slopes (Z), and residual variances (unadjusted for correlations between numbers-at-age).

Estimate	Total landings	Total catch	RV CPUE
Total mortality, Z	1.08 (1.07–1.09)	1.14 (1.12–1.16)	1.39 (1.29–1.50)
Unadjusted residual variance	0.058	0.208	0.345

Estimated range of values corresponding to 1 unadjusted standard error factor shown in brackets. These ranges are likely to be too narrow, as noted in the text. Data for the analysis slightly modified from ICES NSSK w.g. (Anonymous 2001).

Table 2b North Sea cod: year-class curves estimated separately for landings, catch (i.e. landings + estimated discards), and research vessel (RV) CPUE data: Estimated numbers-at-age in the terminal year of the assessment. Last column repeats the RV CPUE results but raised for purposes of comparison.

	Numbers predicted for 1 January 2000			
Age (years)	Total landings	Total catch	RV CPUE	RV CPUE raised to total catch at 4 years
2	–	8400 (4398–16044)	–	–
3	3634 (2861–4615)	2233 (1276–3908)	15.1 (7–35)	2048
4	10007 (8409–11908)	10415 (6163–17601)	76.8 (37–161)	10415
5	1411 (1.15)	1068 (640–1784)	4.1 (2–8)	556
6	589 (521–666)	571 (346–942)	1.1 (<1–2)	149
7	268 (239–300)	231 (141–379)	–	–
8	54 (49–60)	40 (24–66)	–	–
9	28 (25–31)	22 (13–36)	–	–
10	5 (5–6)	4 (2–7)	–	–

Estimated range of values corresponding to 1 unadjusted standard error factor shown in brackets. These ranges are likely to be too narrow, as noted in the text. Data for the analysis slightly modified from ICES NSSK w.g. (Anonymous 2001).

are variants of VPA (SSPA in his paper) in which the catch-at-age matrix is acknowledged to include sampling errors.

Estimation of separate parameters for every year and age, and possibly for every combination, is a natural requirement of an assessment given varying fishing effort and selectivity but, as noted in connection with year-class curves, is difficult to justify statistically unless there is clear knowledge of the limits to accuracy posed by sampling and estimation problems. Furthermore, the large number of parameters is contrary to the principal of parsimony, ideal (5), and typically, a catch-at-age matrix will not have sufficient d.f. for reliable estimation (Pope 1979) with the result that constraints must be assumed (Shepherd and Nicholson 1986) or auxiliary information (Deriso *et al.* 1985), e.g. RV survey indices, must be included in the modelling. This then raises the difficulty of finding relative weights for the different data sets in accordance with ideal (2). Low residual d.f. also means that residual variance is likely to be large and poorly estimated, leading to large standard errors and confidence limits for estimated parameters of the model, if they are estimable at all.

Catch-at-age models with process error

Preliminary

Statistical models having two equations, one describing an unobservable state or process, the other describing observations of that process are called state-space models. They permit two separate sources of error in the model: process error to quantify non-conformity of the true process with the state equation, and observation error to quantify sampling and measurement errors. The state equation of an age-based fish stock assessment describes change in the state of the stock from time to time (usually year to year, so we continue to use the subscript y). We have a vector of stock numbers-at-age ($\mathbf{N}_y$) that changed from year ($y-1$) to y as a function of the probabilities of transferring from one age to the next ($\mathbf{P}$), the probabilities of survival ($\mathbf{S}_y$), and recruitment ($\mathbf{R}_y$):

$$\mathbf{N}_y = \mathbf{P}\mathbf{S}_y\mathbf{N}_{y-1} + \mathbf{R}_y + \mathbf{w}_y \quad (5)$$

where $\mathbf{w}_y$ is a vector of process errors having mean and covariance ($\mathbf{0}$, $\Sigma^{\mathbf{w}_y}$). For the observation equation, we have a vector of catches-at-age ($\mathbf{C}_y$) modelled as proportional to stock numbers:

$$\mathbf{C}_y = \boldsymbol{\mu}_y\mathbf{N}_y + \mathbf{v}_y \quad (6)$$

where $\mathbf{v}_y$ is a vector of observation errors over each age-class having mean and covariance ($\mathbf{0}$, $\Sigma^{\mathbf{v}_y}$). Sullivan *et al.* (1990) and Sullivan (1992) set out standard fishery equations to fill in the $\mathbf{P}$, $\mathbf{S}$ and μ matrices for both length- and age-based models.

While state-space models offer many possibilities for modelling fish stocks, the large number of parameters causes the same problems and infringements of ideals as described above for models without process error. The number of parameters in a state-space model can be reduced by utilizing random walks to model persistence in the system (Gudmundsson 1998) but predictive precision may be reduced as a result. State-space models allow different weightings to be given to different types of auxiliary data by adjusting appropriate diagonal elements of the residual covariance matrix, Σ^{v_y}, associated with the observation equation. However, a satisfactory basis for relative weightings remains difficult to find. Estimation of a residual covariance matrix based on previous modelling work may occasionally be possible. On the other hand, use of a sample covariance matrix obtained by treating each year of data as a replicate observation would overestimate residual variances and underestimate information content, because the sample mean would not explain as much variability as the expectation of the model of Equation (6). If, nevertheless, the Σ^{v_y} is satisfactorily estimated from the data in some way, many d.f. will have been used, thereby substantially reducing those left for estimating the parameters of interest. In addition, the process errors represented in the covariance matrix, Σ^{w_y}, associated with the state equation can have an important bearing on the solution, yet are hard to model or estimate confidently, given that the state equation cannot be sampled from or observed.

Kalman filter

The Kalman filter is a device for estimating the parameters of state-space models. It is highly mathematical, implying a low rating against ideal (1). Meinhold and Singpurwalla (1983) provide a general account while Grønnevik and Evensen (2001) describe developments to deal with nonlinear and high-dimensional fisheries state space models.

Sullivan (1992) describes a basic application. He assumed that w_y and v_y in Equations (5) and (6) are independent and obtained expressions for $\Sigma^{\mathbf{w}_y}$ and $\Sigma^{\mathbf{v}_y}$ based on the MVN approximation to the multinomial distribution. This allowed estimation of parameters using a likelihood function. The Kalman filter involves an updating procedure for **N** at each time step using a gain factor based on $\Sigma^{\mathbf{w}_y}$ and $\Sigma^{\mathbf{v}_y}$ that weights the relative influences on the new **N** of the most recent observed catches (Equation 6 inverted) and the state equation (Equation 5). Thus much depends on the adequacy of the MVN approximations used for the two covariance matrices. For the observation equation, practical difficulties of randomly sampling commercial landings and discards (Tamsett *et al.* 1999a,b), as discussed in 'Data Types Used in North Sea Assessments', suggest that the multinomial would provide only a minimal estimate of $\Sigma^{\mathbf{v}_y}$. A different way of preassigning covariance matrices in a Kalman filter model was given by Gudmundsson (1998). He assumed that the diagonal elements (variances) were proportional to calculated average values, and that off-diagonal elements (covariances) were based on correlations of −0.2 for adjacent ages and zero elsewhere. Both of these approaches to defining covariance matrices are highly technical, influential and subjective, and therefore infringe ideals (1) and (6). Unfortunately, estimation of the two error covariance matrices at the same time as the parameters in the maximum likelihood procedure appears not to be straightforward. Our attempts to do this when fitting Sullivan's model assuming independent errors found that $\Sigma^{\mathbf{w}_y} \to \mathbf{0}$, while $\Sigma^{\mathbf{v}_y} \to \infty$. Schnute (1994) provides another helpful account of the Kalman filter in a fisheries context but notes that 'the outcome of the analysis often depends critically on values of these (correlation and standard error) parameters'. The Kalman filter can provide asymptotic estimates of the variance of estimated parameters by calculation of a Hessian matrix (Pella 1993) but this would need an extraordinarily long time-series of results to hold any credibility when many parameters are being estimated.

Bayesian methods

Bayesian methods of stock assessment applied to state-space models have often been advocated. McAllister and Kirkwood (1998) and de Valpine (2002) review references and provide good introductions. Spreadsheet software is provided by Punt and Hilborn (2001). Bayesian methods differ from those of frequentist statistics in that the parameters of the model are thought of as random variables rather than as fixed quantities. In practice, the two approaches give comparable results for many elementary analyses when starting with no prior information but Bayesian methods can be the more powerful when prior information is available and can be incorporated into a frequency distribution for each parameter, the so-called prior probability distributions. Subjective probabilities and expert opinions can be put to use in this way. On the other hand, there is a danger that vested interests could be given too much weight. Punt and Hilborn (1997) remark that formulation of priors is 'undoubtedly the most controversial aspect of any Bayesian analysis'. The independence of priors is another consideration. Priors can be taken as independent if no combination of parameter values considered jointly 'are assumed more credible than others' (McAllister *et al.* 1994; McAllister and Kirkwood 1998). Alternatively, joint priors incorporating correlations between parameters can be used. In either case, the data should reveal the degree of dependence between the parameters in the joint posterior distributions, those obtained from applying Bayes' formula.

Punt and Hilborn (1997) illustrate how posterior probabilities for different stock sizes can be used to estimate expected, long-term future stock sizes under different harvesting policies. The implications would be clear in accordance with ideal (1) provided that the complications of dependence among posterior distributions could be explained, e.g. for stock size and *F*. Myers and Cadigan (1995) make a similar point about the effects of correlated errors across ages on the risks of various management options.

The large number of parameters to be estimated in state-space modelling of fish stocks usually means that posterior distributions can only be found by computationally intensive, numerical methods. Two classes are important, (i) the Markov-Chain Monte Carlo (MCMC) methods such as the Gibbs sampler (Millar and Meyer 2000) and the Metropolis-Hastings (MH) algorithm (Punt and Hilborn 1997), and (ii) various importance sampling algorithms (McAllister *et al.* 1994; Kinas 1996; McAllister and Ianelli 1997; Trenkel *et al.* 2000). These methods all exploit the same Bayesian theory and should, in theory, arrive at the same posterior distributions given the same model, data and priors. However, the computational methods make important assumptions contrary to ideal (6), e.g. concerning burn-in times for MCMC (Gilks *et al.* 1996), the adequacy of initial sampling for importance sampling, and the values for the covariance matrix of the observation equation, i.e. Σ^{v_y}, in order to estimate the importance weights in importance sampling. Punt and Hilborn (1997) and McAllister and Kirkwood (1998) note that a major impediment to the Bayesian approach is computing power, with a typical run of an age-structured model taking from several hours to days. However, McAllister and Kirchner (2002) claim that, using an importance function, convergence to an accurate estimate of the posterior can be achieved in less than 10 min for 80–82 parameters. We conclude that while Bayesian theory has much to offer an assessment, the necessary computational complexity associated with large state-space models calls for considerable skill and does not fit well with ideals (1), (6) or (7).

Discussion

Improving the accuracy and acceptability of stock assessments is a continuing scientific task. We have argued that mathematical developments in age-based assessment methods are undermined when data are imprecise or biased, assumptions are wrong, private decisions are made about important procedures, or when results are difficult to verify or explain. Our arguments also imply that improvements to data collection procedures will not achieve much if the improved data sets are subsequently contaminated with inappropriate or low quality information from another sampling exercise, if the assessment method fails to weight all data in accordance with their contribution of independent information, or if the assessment otherwise causes bias for internal mathematical reasons.

There are several relatively simple ways of refining assessments if problems are thought to exist. One is removal of any unnecessary links, and thus statistical dependence, *between* data sets, for example, by creating independent rather than shared ALKs for each set, or by removing multiple uses of error-vulnerable data, for example, when fleet level catches are estimated as landings plus discards, the latter raised from trip to fleet level also using landings. Removal of unnecessary links facilitates separate analyses of the factors affecting each set, and appropriate weighting – whether of the different sets contributed to a joint model, or of the different estimates of the same parameter obtained from separate models. Removal of dependence *within* sets is another area where improvements may be possible, e.g. by increasing the randomized element of sampling, and/or by improving its temporal and spatial evenness. Reduced bias in both the mean and variance of the data can then be expected; the variance may be larger as a result of the wider spread of observations but knowledge of that will assist correct weighting of the data when the assessment model is fitted, or for other uses.

The choice of assessment method is also important. We found that the simplest of the methods we reviewed, namely year-class curves, not to be confused with catch curves, scored most highly against our ideals. Year-class curves give estimates of precision, are readily explicable, use data only once, and call for a minimum of parameters and assumptions. From standard least squares theory, relative annual recruitments, total mortality with or without gradual changes over time, and other parameters related to selectivity, fleet, gear, and type of data (e.g. landings, discards, abundance indices) are all estimable with variances and without bias. The variances are available to assist with weighted averaging of different estimates of the same parameters derived from separate modelling of different data sets. Poor data or lack of confidence in the chosen model would, of course, damage this pristine statistical image. However, as the curves and data are easily illustrated to reveal goodness-of-fit, discussions about the reasons for inconsistent slopes, curvatures, or other puzzles, can include all people with interests in the assessment, not just the scientists operating the computers. Variable selectivity with age or time, discarding, fishery regulations, movements of fish, a poor model, and biases and dependences in the data, etc. could all be relevant factors, and discussion of them is likely to be greatly enriched by the contributions of fishers. Year-class curves, like other assessment methods, suffer from dependences among observed numbers of fish in different age groups when all were caught with the same fishing operations. However, this can be allowed for by adjusting d.f. or by fitting a random *year* factor. Year-class curves do not estimate absolute stock numbers or F but nor do most other assessment methods if major uncertainties in M are acknowledged.

In contrast to year-class curves, several other methods of assessment cannot estimate precision, either because theory does not allow this (VPA based methods) or because too many d.f. are tied up in estimating large numbers of parameters for the model. Kalman filter and Bayesian methods do estimate precision but this (and the estimates of parameters) depend on model-dependent covariance matrices, derivation of which may depend on informal methods and major assumptions. Concerning the other ideals we proposed, none of the methods apart from year-class curves is easy to explain to non-mathematicians. A likely consequence is that fishers distrust the assessment, particularly when stocks are low and the advice directly impinges on their livelihoods; unfortunately, this is the time when their co-operation is most needed to help restore the stock. Weighting of data sets remains an important issue, especially when all sets are combined in the fit of one joint model. Parsimony is not a feature of VPA-based or other methods that claim to estimate numerous fishing mortalities by age and year as well as recruitments by year. While such models may accurately reflect the real world, estimation of their many parameters with imperfect data may be a highly unrealistic goal. Most of the more elaborate assessment methods make technical assumptions critical to the estimated outputs, e.g. concerning M, covariance matrices and error distributions. Our view is that these several weaknesses in assessment methods become harder to find and evaluate as the mathematics of the method becomes more elaborate.

Finally, we note that, despite several decades of fisheries management by stock assessment and controls on landings in Europe, many difficult problems remain in data collection and analysis. Pope (1979) explains why controls on 'catches' rather than on effort were originally preferred for fisheries management and thus why assessments of fish numbers and mortality became so important. Re-examination of these reasons deserves some priority in the light of the technical difficulties and changed situations in the North Sea and elsewhere. Pragmatic rules to control catches/landings or effort without direct links to annual stock assessments have been proposed (Koeller *et al.* 2000; Parma 2002; Walters and Martell 2002; Shepherd 2003). These alternatives may become more widely acceptable in future. They are likely to continue the requirement for knowledge of the quantitative state of the stock but less accurately and less frequently than is needed for setting annual limits on landings. This could alleviate the pressures on assessment scientists from the many theoretical and practical difficulties that have to be faced.

Acknowledgements

This research was funded by European Commission project EC 98/097, and the Department of Environment, Food and Rural Affairs, UK. JC was seconded to the School of Maths and Stats, by courtesy of the University of St Andrews and CEFAS. He is grateful to many for discussion of fisheries problems, notably to S. Flatman (XSA) and C. Bannister (M) at CEFAS; to T. Quinn II of the Juneau Center, Alaska (weighting); to R. Fryer of FRS, Aberdeen (Kalman filter), to Beare *et al.* (2002) (bootstrap and surveys), and to B Cullis of University of Texas at Arlington (likelihood and weighting). Neither they, nor CEFAS, DEFRA, the EC, nor any partner in project EC98/097 are responsible for any statement in this report.

References

Aitkin, M. (1987) Modelling variance heterogeneity in normal regression using GLIM. *Applied Statistics* **36**, 332–339.

Allen, M., Kilpatrick, D., Armstrong, M., Briggs, R., Course, G. and Pérez, N. (2002) Multistage cluster sampling design and optimal sample sizes for estimation of fish discards from commercial trawlers. *Fisheries Research* **55**, 11–24.

Anonymous (1999) Manual for the International Bottom Trawl Surveys. Rev. VI, *ICES CM 1999/D* **2**, Addendum 2 Ref.:G.

Anonymous (2000) Report of the study group on discard and by-catch information, *ICES CM 2000/ACFM*: **11**.

Anonymous (2001) Report of the North Sea and Skagerrak demersal fish stock working group, *ICES CM 2002/ACFM*: **01**.

Beare, D., Castro, J., Cotter, J. et al. (2002) Evaluation of research surveys in relation to management advice (EVARES). *Final report of FISH/2001/02 - Lot 1*, DGXIV Fisheries, European Commission, Brussels. Available at a.j.cotter@cefas.co.uk.

Berg, E. and Albert, O.T. (2003) Cod in fjords and coastal waters of North Norway: distribution and variation in length and maturity at age. *ICES Journal of Marine Science* **60**, 787–797.

Beverton, R.J.H. (1964) Differential catchability of male and female plaice in the North Sea and its effect on estimates of stock abundance. *Rapports et Proces- verbaux des Reunions, Conseil Permanent International pour L'Exploration de la Mer* **155**, 103–112.

Beverton, R.J.H. and Holt, S.J. (1957) On the dynamics of exploited fish populations. *Fishery Investigations, Series II*, Vol. 19. Her Majesty's Stationery Office, London, 533 pp.

Bradford, M.J. (1991) Effects of ageing errors on recruitment time series estimated from sequential population analysis. *Canadian Journal of Fisheries and Aquatic Sciences* **48**, 555–558.

Bromley, P.J. (2000) Growth, sexual maturation and spawning in central North Sea plaice (*Pleuronectes platessa* L.), and the generation of maturity ogives from commercial catch data. *Journal of Sea Research* **44**, 27–43.

Burnham, K.P. and Anderson, D.R. (2002) *Model Selection and Multimodel Inference*, 2nd edn. Springer, New York.

Carroll, R.J. and Ruppert, D. (1988) *Transformation and Weighting in Regression*. (Monographs on statistics and applied probability). Chapman and Hall, New York.

Clark, W.G. (1999) Effects of an erroneous natural mortality rate on a simple age-structured stock assessment. *Canadian Journal of Fisheries and Aquatic Sciences* **56**, 1721–1731.

Collie, J.S. and Sissenwine, M.P. (1983) Estimating population size from relative abundance data measured with error. *Canadian Journal of Fisheries and Aquatic Sciences* **40**, 1871–1879.

Cook, R.M. (1997) Stock trends in six North Sea stocks as revealed by an analysis of research vessel surveys. *ICES Journal of Marine Science* **54**, 924–933.

Cotter, A.J.R. (1998) Method for estimating variability due to sampling of catches on a trawl survey. *Canadian Journal of Fisheries and Aquatic Sciences* **55**, 1607–1617.

Cotter, A.J.R. (2001) Intercalibration of North Sea International Bottom Trawl Surveys by fitting year-class curves. *ICES Journal of Marine Science* **58**, 622–632 [Erratum, Ibid 58, 1340].

Cotter, A.J.R. and Buckland, S.T. (2004) Using the EM algorithm to weight data sets of unknown precision when modeling fish stocks. *Mathematical Biosciences* **190**, 1–7.

Cotter, A.J.R., Course, G.P., Buckland, S.T. and Garrod, C. (2002) A PPS sample survey of English fishing vessels to estimate discarding and retention of North Sea cod, haddock, and whiting. *Fisheries Research* **55**, 25–35.

Darby, C.D. and Flatman, S. (1994) Virtual population analysis, version 3.1 (Windows/DOS) user guide. *Information Technology Series 1*. CEFAS, Lowestoft, UK.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, series B* **39**, 1–38.

Deriso, R.B., Quinn, T.J. and Neal, P.R. (1985) Catch-age analysis with auxiliary information. *Canadian Journal of Fisheries and Aquatic Sciences* **42**, 815–824.

de Valpine, P. (2002) Review of methods for fitting time-series models with process and observation error and likelihood calculations for nonlinear, non-Gaussian state-space models. *Bulletin of Marine Science* **70**, 455–471.

Fabrizio, M.C. and Richards, R.A. (1996) Commercial fisheries surveys. In: *Fisheries Techniques* (eds B.R. Murphy and D. Willis), 2nd edn. American Fisheries Society, Bethesda, MD, pp. 625–650.

Fournier, D. and Archibald, C.P. (1982) A general theory for analysing catch at age data. *Canadian Journal of Fisheries and Aquatic Sciences* **39**, 1195–1207.

Fréon, P. and Misund, O.A. (1999) *Dynamics of Pelagic Fish Distribution and Behaviour: Effects on Fisheries and Stock Assessment*. Fishing News Books, Blackwell Science Ltd, Oxford, UK.

Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (1996) *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC, London.

Godø, O.R. (1994) Factors affecting the reliability of groundfish abundance estimates from bottom trawl surveys. In: *Marine Fish Behaviour in Capture and Abundance Estimation* (eds A. Ferro and S. Olsen). Fishing News Books, Blackwell Science Ltd, Oxford, UK, pp. 166–199.

Graham, M. (1933) Report on the North Sea cod. *Fishery Investigations, Ser. II* **No. XIII**, 166 pp.

Grønnevik, R. and Evensen, G. (2001) Application of ensemble-based techniques in fish stock assessment. *Sarsia* **86**, 517–526.

Gudmundsson, G. (1998) Joint time series analysis of catch-at-age and CPUE data. *Alaska Sea Grant College Program*, AK-SG-98–01, 199–217.

Harley, S.J., Myers, R.A. and Dunn, A. (2001) Is catch-per-unit-effort proportional to abundance? *Canadian Journal of Fisheries and Aquatic Sciences* **58**, 1760–1772.

Hilborn, R. (1992) Current and future trends in fisheries stock assessment and management. *South African Journal of Marine Science* **12**, 975–988.

Hilborn, R. (2002) The dark side of reference points. *Bulletin of Marine Science* **70**, 403–408.

Hilborn, R. (2003) The state of the art in stock assessment: where we are and where we are going. *Scientia Marina* **67**(Suppl. 1), 15–20.

Hilborn, R. and Walters, C.J. (1992) *Quantitative Fisheries Stock Assessment*. Chapman and Hall, New York, 570 pp.

Hjellvik, V., Godø, O.R. and Tjøstheim, D. (2002a) Diurnal variation in bottom trawl survey catches: does it pay to adjust? *Canadian Journal of Fisheries and Aquatic Sciences* **59**, 33–48.

Hjellvik, V., Godø, O.R. and Tjøstheim, D. (2002b) The measurement error in marine survey catches: the bottom trawl case. *Fisheries Bulletin* **100**, 720–726.

Hutchings, J.A. (1996) Spatial and temporal variation in the density of northern cod and a review of hypotheses for the stock's collapse. *Canadian Journal of Fisheries and Aquatic Sciences* **53**, 943–962.

Jensen, A.J.C. (1939) On the laws of decrease in fish stocks. *Rapports et Proces- verbaux des Reunions, Conseil Permanent International pour L'Exploration de la Mer* **110**, 85–96.

Kell, L.T. and Bromley, P.J. (2004) Implications for current management advice for North Sea plaice (*Pleuronectes platessa* L.): Part II. Increased biological realism in recruitment, growth, density-dependent sexual maturation and the impact of sexual dimorphism and fishery discards. *Journal of Sea Research* **51**, 301–312.

Kell, L.T., Scott, R. and Hunter, E. (2004) Implications for current management advice for North Sea plaice: Part I. Migration between the North Sea and English Channel. *Journal of Sea Research* **51**, 287–299.

Kimura, D.K. (1977) Statistical assessment of the age-length key. *Journal of the Fisheries Research Board of Canada* **34**, 317–324.

Kimura, D.K. (1990) Approaches to age-structured separable sequential population analysis. *Canadian Journal of Fisheries and Aquatic Sciences* **47**, 2364–2374.

Kinas, P.G. (1996) Bayesian fishery stock assessment and decision making using adaptive importance sampling. *Canadian Journal of Fisheries and Aquatic Sciences* **53**, 414–423.

Koeller, P., Savard, L., Parsons, D.G. and Fu, C. (2000) A precautionary approach to assessment and management of shrimp stocks in the northwest Atlantic. *Journal of Northwest Atlantic Fisheries Science* **27**, 235–246.

Lai, H.-L. (1993) Optimal sampling design for using the age-length key to estimate age composition of a fish population. *Fisheries Bulletin* **92**, 382–388.

Laurec, A. and Shepherd, J.G. (1983) On the analysis of catch and effort data. *Journal du Conseil international pour l'Exploration de la Mer* **41**, 81–84.

Liggins, G.W., Bradley, M.J. and Kennelly, S.J. (1997) Detection of bias in observer-based estimates of retained and discarded catches from a multispecies trawl fishery. *Fisheries Research* **32**, 133–147.

Marchal, P., Ulrich, C., Korsbrekke, K., Pastoors, M. and Rackham, B. (2002) Annual trends in catchability and fish stock assessments. *Scientia Marina* **67** (Suppl. 1), 63–73.

McAllister, M.K. and Ianelli, J.N. (1997) Bayesian stock assessment using catch-age data and the sampling-importance resampling algorithm. *Canadian Journal of Fisheries and Aquatic Sciences* **54**, 284–300.

McAllister, M. and Kirchner, C. (2002) Accounting for structural uncertainty to facilitate precautionary fishery management: illustration with Namibian Orange Roughy. *Bulletin of Marine Science* **70**, 499–540.

McAllister, M.K. and Kirkwood, G.P. (1998) Bayesian stock assessment: a review and example application using the logistic model. *ICES Journal of Marine Science* **55**, 1031–1060.

McAllister, M.K., Pikitch, E.K., Punt, A.E. and Hilborn, R. (1994) A Bayesian approach to stock assessment and harvest decisions using the sampling/importance resampling algorithm. *Canadian Journal of Fisheries and Aquatic Sciences* **51**, 2673–2687.

Megrey, B.A. (1989) Review and comparison of age-structured stock assessment models. *American Fisheries Society Symposium* **6**, 8–48.

Meinhold, R.J. and Singpurwalla, N.D. (1983) Understanding the Kalman filter. *American Statistician* **37**, 123–127.

Mertz, G. and Myers, R.A. (1997) Influence of errors in natural mortality estimates in cohort analysis. *Canadian Journal of Fisheries and Aquatic Sciences* **54**, 1608–1612.

Millar, R.B. and Meyer, R. (2000) Bayesian state-space modeling of age-structured data: fitting a model is just the beginning. *Canadian Journal of Fisheries and Aquatic Sciences* **57**, 43–50.

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* **52**, 1265–1273.

Myers, R.A. and Stokes, K. (1989) Density-dependent habitat utilisation of groundfish and the improvement of research surveys. *ICES CM 1989/D* **15**.

National Research Council (1998) *Improving fish stock assessments. Committee on Fish Stock Assessment Methods.* National Academy Press, Washington D.C. Available at http://www.nap.edu/books/0309057256/html/.

Nicholson, M.D., Stokes, T.K. and Thompson, A.B. (1991) The interaction between fish distribution, survey design and analysis. *ICES CM 1991/D* **11**.

Paloheimo, J.E. and Dickie, L.M. (1964) Abundance and fishing success. *Rapports et Proces- verbaux des Reunions, Conseil Permanent International pour L'Exploration de la Mer* **155**, 152–163.

Parma, A.M. (2002) In search of robust harvest rules for Pacific halibut in the face of uncertain assessments and decadal changes in productivity. *Bulletin of Marine Science* **70**, 423–453.

Pastoors, M.A., O'Brien, C.M., Flatman, S. et al. (2001) Evaluation of market sampling strategies for a number of commercially exploited stocks in the North Sea and development of procedures for consistent data storage and retrieval (EMAS). *Final report of Project 98/075*. DG XIV, Fisheries, European Commission, Brussels. Available at a.j.cotter@cefas.co.uk.

Pella, J.J. (1993) Utility of structural time series models and the Kalman filter for predicting consequences of fishery actions. Proceedings of the International Symposium on Management strategies for exploited fish populations. *Alaska Sea Grant College Program* AK-SG-93-02, 571–593.

Pelletier, D. (1998) Intercalibration of research survey vessels in fisheries: a review and an application. *Canadian Journal of Fisheries and Aquatic Sciences* **55**, 2672–2690.

Pennington, M. and Godø, O.R. (1995) Measuring the effect of changes in catchability on the variance of marine survey abundance indices. *Fisheries Research* **23**, 301–310.

Pennington, M., Burmeister, L.-M. and Hjellvik, V. (2002) Assessing the precision of frequency distributions estimated from trawl-survey samples. *Fishery Bulletin* **100**, 74–80.

Pope, J.G. (1972) An investigation of the accuracy of virtual population analysis using cohort analysis. *ICNAF Research Bulletin* **9**, 65–74.

Pope, J.G. (1977) Estimation of fishing mortality, its precision and implications for the management of fisheries. In: *Fisheries Mathematics* (ed. J.H. Steele), Academic Press, London, pp. 63–76.

Pope, J.G. (1979) Population dynamics and management: current status and future trends. *Investigación Pesquera* **43**, 199–221.

Pope, J.G. (1982) Response to the note by M.S.M. Siddeek. *Journal du Conseil international pour l'Exploration de la Mer* **40**, 306.

Pope, J.G. (1989) Multispecies extensions to age-structured assessment models. *American Fisheries Society Symposium* **6**, 102–111.

Pope, J.G. and Shepherd, J.G. (1982) A simple method for the consistent interpretation of catch-at-age data. *Journal du Conseil international pour l'Exploration de la Mer* **40**, 176–184.

Pope, J.G. and Shepherd, J.G. 1985. A comparison of the performance of various methods for tuning VPAs using effort data. *Journal du Conseil international pour l'Exploration de la Mer* **42**, 129–151.

Punt, A.E. and Hilborn, R. (1997) Fisheries stock assessment and decision analysis: the Bayesian approach. *Reviews in Fish Biology and Fisheries* **7**, 35–63.

Punt, A.E. and Hilborn, R. (2001) BAYES-SA: Bayesian stock assessment methods in fisheries. User's manual. *FAO Computerized Information Series (Fisheries)* **12**, FAO, Rome.

Quinn, T.J. II and Deriso, R.B. (1999) *Quantitative Fish Dynamics*. Oxford University Press, Oxford.

Rencher, A.C. (1995) *Methods of Multivariate Analysis*. J. Wiley & Sons, New York.

Richards, L.J. and Schnute, J.T. (1998) Model complexity and catch-at-age analysis. *Canadian Journal of Fisheries and Aquatic Sciences* **55**, 949–957.

Richards, L.J., Schnute, J.T., Kronlund, A.R. and Beamish, R.J. (1992) Statistical models for the analysis of ageing error. *Canadian Journal of Fisheries and Aquatic Sciences* **49**, 1801–1815.

Richards, L.J., Schnute, J.T. and Olsen, N. (1997) Visualizing catch-age analysis: a case study. *Canadian Journal of Fisheries and Aquatic Sciences* **54**, 1646–1658.

Ricker, W.E. (1975) Computation and interpretation of biological statistics of fish populations. *Bulletin of the Fisheries Research Board of Canada* **191**, 382 pp.

Rivard, D. (1989) Overview of the systematic, structural, and sampling errors in cohort analysis. *American Fisheries Society Symposium* **6**, 49–65.

Rochet, M.-J., Péronnet, I. and Trenkel, V.M. (2002) An analysis of discards from the French trawler fleet in the Celtic Sea. *ICES Journal of Marine Science* **59**, 538–552.

Rose, G.A. and Kulka, D.W. (1999) Hyperaggregation of fish and fisheries: how catch-per-unit-effort increased as the northern cod (*Gadus morhua*) declined. *Canadian Journal of Fisheries and Aquatic Sciences* **56**, 118–127.

Saila, S.B., Lorda, E. and Walker, H.A. (1985) The analysis of parameter error propagation in simple fishery models. *Marine Resource Economics* **1**, 235–246.

Sampson, D.B. (1987) Variance estimators for Virtual Population Analysis. *Journal du Conseil international pour l'Exploration de la Mer* **43**, 149–158.

Schnute, J.T. (1994) A general framework for developing sequential fisheries models. *Canadian Journal of Fisheries and Aquatic Sciences* **51**, 1676–1688.

Schnute, J.T. and Richards, L.J. (1995) The influence of error on population estimates from catch-age models. *Canadian Journal of Fisheries and Aquatic Sciences* **52**, 2063–2077.

Sen, A.R. (1986) Methodological problems in sampling commercial rockfish landings. *Fishery Bulletin* **84**, 409–421.

Shepherd, J.G. (1997) Prediction of year-class strength by calibration regression analysis of multiple recruit index series. *ICES Journal of Marine Science* **54**, 741–752.

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

Shepherd, J.G. (2003) Fishing effort control: could it work under the common fisheries policy? *Fisheries Research* **63**, 149–153.

Shepherd, J.G. and Nicholson, M.D. (1986) Use and abuse of multiplicative models in the analysis of fish catch-at-age. *The Statistician* **35**, 221–227.

Shepherd, J.G. and Nicholson, M.D. (1991) Multiplicative modelling of catch-at-age data, and its application to catch forecasts. *Journal du Conseil international pour l'Exploration de la Mer* **47**, 284–294.

Siddeek, M.S.M. (1982) A note on Pope's cohort analysis. *Journal du Conseil international pour l'Exploration de la Mer* **40**, 209–210.

Sinclair, A.F. (2001) Natural mortality of cod (*Gadus morhua*) in the Southern Gulf of St. Lawrence. *ICES Journal of Marine Science* **58**, 1–10.

Smith, P.J. (1989) Is two-phase sampling really better for estimating age composition? *Journal of the American Statistical Association* **84**, 916–921.

Smith, A.D.M. and Punt, A.E. (1998) Stock assessment of gemfish (*Resea solandri*) in eastern Australia using maximum likelihood and Bayesian methods. *Alaska Sea Grant College Program*, AK-SG-98–01.

Smith, P.J., Jamieson, A. and Birley, A.J. (1990) Electrophoretic studies and the stock concept in marine teleosts. *Journal du Conseil international pour l'Exploration de la Mer* **47**, 231–245

Sparholt, H. (1990) Improved estimates of the natural mortality rates of nine commercially important fish species included in the North Sea Multispecies VPA model. *Journal du Conseil international pour l'Exploration de la Mer* **46**, 211–223.

Stratoudakis, Y., Fryer, R.J. and Cook, R.M. (1998) Discarding practices for commercial gadoids in the North Sea. *Canadian Journal of Fisheries and Aquatic Sciences* **55**, 1632–1644.

Stratoudakis, Y., Fryer, R.J., Cook, R.M. and Pierce, G.J. (1999) Fish discarded from Scottish demersal vessels: estimators of total discards and annual estimates for targeted gadoids. *ICES Journal of Marine Science* **56**, 592–605.

Sullivan, P.J. (1992) A Kalman filter approach to catch-at-length analysis. *Biometrics* **48**, 237–257.

Sullivan, P.J., Lai, H.-L. and Gallucci, V.F. (1990) A catch-at-length analysis that incorporates a stochastic model of growth. *Canadian Journal of Fisheries and Aquatic Sciences* **47**, 184–198.

Swain, D.P. and Sinclair, A.F. (1994) Fish distribution and catchability: what is the appropriate measure of distribution? *Canadian Journal of Fisheries and Aquatic Sciences* **51**, 1046–1054.

Tamsett, D. and Janacek, G. (1999) Sampling trips for measuring discards in commercial fishing based on mulitlevel modelling of measurements in the North Sea from NE England. *Fisheries Research* **42**, 103–115.

Tamsett, D., Janacek, G. and Emberton, M. (1999a) A comparison of methods for onboard sampling of discards in commercial fishing. *Fisheries Research* **42**, 127–135.

Tamsett, D., Janacek, G., Emberton, M., Lart, B. and Course, G. (1999b) Onboard sampling for measuring discards in commercial fishing based on multilevel modelling of measurements in the Irish Sea from NW England and N Wales. *Fisheries Research* **42**, 117–126.

Tanaka, S. (1953) Precision of age-composition of fish estimated by double sampling method using the length stratification. *Bulletin of the Japanese Society of Scientific Fisheries* **19** 657–670.

Thompson, H. (1928) Haddock biology (North Sea). A brief survey of recent data, methods and results. *Rapports et Proces- verbaux des Reunions, Conseil Permanent International pour L'Exploration de la Mer* **54**, 135–165.

Thompson, S. (1992) *Sampling*. J. Wiley, New York.

Trenkel, V.M., Elston, D.A. and Buckland, S.T. (2000) Fitting population dynamics models to count and cull data using sequential importance sampling. *Journal of the American Statistical Association* **95**, 363–374.

van Beek, F.A. (1998) Discarding in the Dutch beam trawl fishery, *ICES CM 1998/BB*: **5**.

Verbyla, A.P. (1993) Modelling variance heterogeneity: residual maximum likelihood and diagnostics. *Journal of the Royal Statistical Society, Series B* **55**, 493–508.

Vetter, E.F. (1988) Estimation of natural mortality in fish stocks: a review. *Fishery Bulletin* **86**, 25–43.

Vinje, F., Heino, M., Dieckmann, U., Godø, O.R. and Mork, J. (2003) Spatial structure in length at age of cod in the Barents Sea. *Journal of Fish Biology* **62**, 549–564.

Walters, C. and Martell, S.J.D. (2002) Stock assessment needs for sustainable fisheries management. *Bulletin of Marine Science* **70**, 629–638.