The Catch-Survey Analysis (CSA) method of fish stock assessment: an evaluation using simulated data

Benoit Mesnil

Laboratoire MAERHA, IFREMER, B.P. 21105, Nantes F44311, France

Tel: +33 240 37 40 00; Fax: +33 240 37 40 75; e-mail: Benoit.Mesnil@ifremer.fr

Abstract:

The Catch-Survey Analysis method of assessment aims at estimating stock abundance from relative indices by filtering measurement error in the latter through a simple two-stage population dynamics model. The method is not widely used and the associated literature is still limited. The objective of this work is to improve current understanding of the properties of the method, using data generated from a known fully age-structured population. A sensitivity analysis confirms that CSA is capable of providing reliable information about general stock trends. However, absolute estimates are sensitive to the parameter relating survey catchabilities of the two stages, which needs to be estimated externally with methods that warrant further validation. Biased results are also obtained when changes in the catchability of the fleet providing indices are not corrected for during data preparation. Conventional diagnostics fail to detect violations of the constant-q assumption and improved diagnostics are thus needed. CSA is proved to be a very valuable method to support management advice in data-limited contexts in which age data are lacking or uncertain. Subject to some refinements, it can even challenge VPA-based assessments, notably when the latter use uncertain age data.

Keywords: stock assessment; catch-survey analysis; Collie-Sissenwine model; data-limited environments.

1. Introduction

The fish stock assessment community generally proclaims strong preference for age structured models and strives to use some variant of Sequential Population Analysis methods to estimate the parameters of these models. In many instances, age structured models are effectively superior in that they give detailed insight into the dynamics of stocks and provide means to explore a large variety of management measures. However, this advantage is obtained at considerable cost since the continuous provision of reliable age data over long periods is required. Estimating the age composition of catches usually involves age-length keys, which need to be rebuilt anew every year. Processing the hundreds of otoliths (or other hard pieces) required is often the most expensive component of data preparation for assessments. Moreover, there are many species for which the determination of age is still an unresolved issue or is at best uncertain, making age-based approaches inappropriate. For assessing such stocks a classical alternative is to use surplus production models. However, fitting these apparently simple models raises arduous statistical problems, and they often fail to produce reliable estimates for management purposes, notably when there is insufficient contrast in the data (Hilborn and Walters, 1992) or when variations in stock abundance are driven more by fluctuations in recruitment than by changes in harvesting intensity (NRC, 1998). In addition, even though some information additional to catch and cpue might be available, for example about recruitment strength from surveys, surplus production models are unable to utilise it. Length-based methods may be another option but most require that the growth pattern can be modelled properly, which again poses the problem of age determination in the first place. The same constraint holds for applications of the delay difference model (Deriso, 1980) in which the growth coefficient is an important parameter. This means that there is currently a lack of operational methods to handle stocks for which age structures are uncertain, such as crustaceans, or more generally for the so-called "data-limited" situations. Nevertheless, fisheries scientists are requested to provide management advice and giving no response is detrimental to their credibility if not, at times, to the welfare of the fisheries.

With its moderate demand in data, the Catch-Survey Analysis (CSA) method based on the two-stage model first proposed by Collie and Sissenwine (1983) fits well in that niche. In essence, it aims at extracting the real stock abundance signal from often noisy survey data by smoothing the latter through a simple dynamic model. The approach is particularly designed for cases in which a full age structure is lacking, but where a "recruits" stage can be easily distinguished from older ages lumped into a "fully-recruited" component (e.g. a distinct threshold can be seen in length compositions plots), and where time series of abundance indices for each stage are available (from commercial cpue or preferably from surveys). The method seems to be used regularly in North America, notably for assessments of shellfish stocks (e.g. Conser, 1991; Conser and Idoine, 1992; Collie and Kruse, 1998; Cadrin et al., 1999; Zheng et al., 1997, 1998), but has not attracted much attention in Europe despite the positive appreciation by the ICES Methods Working Group (ICES, 1995). However, the associated literature is still very limited. Most papers deal with specific applications and are confined in rather confidential "grey literature". Understandably they tend to focus on those results of immediate relevance for management advice for the particular fishery, rather than on the inherent behaviour of the method, and there is some divergence between authors regarding the consequences of changes in some settings upon various results.

The main objective of this paper is thus to improve the current understanding of the properties and limitations of this method. It tries to clarify the relative importance and the effects of varying key input parameters. It also examines the extent to which standard diagnostics may help to guide the choices or to detect problems. Like Collie and Kruse (1998) and Cadrin (2000), this evaluation uses simulated data from a known underlying population. Moreover, the data were generated with a very different, fully age-structured model then aggregated into one (catch) or two (indices) stages. Deterministic and stochastic simulations were carried out. The latter involved a Monte Carlo approach whereby multiple sets of abundance indices were generated, with variable amount of observation error, and input to CSA yielding series of estimates whose distributions were examined for dispersion and bias. However, the outcomes of this trial were qualitatively the same as those found in the literature, notably in the last two cited papers. Therefore, only the results of the deterministic simulations (data sets without stochastic error, except in bootstrap runs used to assess confidence bounds) are presented and discussed in this paper.

The essentials of the model and of the estimation procedure are first recalled. Then results of a sensitivity analysis and outputs of common diagnostics are presented. Finally, implications of the results are discussed.

2. Methods

2.1. The two-stage model

Following Collie and Sissenwine (1983), it is assumed that the population consists of two stages: the recruits, and the fully recruited animals. The time step is annual, with years defined either on a calendar basis or as the interval between regular surveys. The year range is [1, T].

The population dynamics is described by a discrete difference model:
$$N_{t+1} = (N_t + R_t)e^{-M} - C_t e^{-M(1-\tau)} \qquad [1]$$
where N_t is the population size, in number, of fully recruited animals at start of year t; R_t the population size, in number, of recruits at start of year t; C_t the catch in number during year t (known); M the instantaneous rate of natural mortality (equal for both stages, assumed); τ the fraction of the year when the catch is taken, e.g. 0 if the fishing season is early in the year, or 0.5 if the catch is taken midway through the year or, by resemblance with Pope's (1972) cohort approximation, evenly over the year.

Estimating the time series of N_t and R_t given the catches is the basic task of any assessment but, as with other methods, this requires additional information in the form of relative indices n_t and r_t of abundance for each stage, typically from surveys, which are assumed to be proportional to absolute population sizes N_t and R_t. The indices are deemed to be measured with some (log-normal) observation error:
$$n_t = q_n N_t \exp(\eta_t) \quad ; \text{t= 1, T} \qquad [2]$$
$$r_t = q_r R_t \exp(\delta_t) \quad ; \text{t= 1, T-1} \qquad [3]$$
where q_n and q_r are the catchability coefficients of fully recruited and recruits, respectively, in the survey, supposed to be constant with time; η and δ are normally distributed random variables.

Whereas Collie and Sissenwine (1983) considered a unique catchability coefficient for both stages, Conser (1994) has added the specification that the catchability of the recruits is a fraction *s* of that of the fully recruited:

$$s = q_r / q_n \qquad [4]$$

This addition is essential because *s* is often less than unity and the estimation would fail without the possibility of adjusting it. The difficulty, though, is that *s* has to be pre-set based on external information. Although it is not formally correlated with other parameters in the equations above and is estimable in principle, *s* is in practice strongly (negatively) correlated with q_n (Collie and Kruse, 1998). If the necessary information is available, year-specific values can be set for *s*.

2.2. Estimation

Two approaches for estimation are mentioned in the literature depending on the assumed error structure, which can be either a mix of observation and process errors, or only observation errors. In both cases the estimation equation is arrived at by substituting the expectations of [2] and [3] into [1], yielding :

$$n_{t+1} = \left[\left(n_t + \frac{r_t}{s} \right) e^{-M} - q_n C_t e^{-M(1-\tau)} \right] e^{\varepsilon_t} \qquad [5]$$

where, if relevant, ε is a normally distributed process error, which can be multiplicative as used by default here but is sometimes considered additive (Mendelssohn, 1988; Cadrin, 2000).

In a least-squares approach, the parameters are estimated by minimising an objective function of the general form :

$$SS(\theta) = \lambda_\varepsilon \sum_{t=2}^{T} \varepsilon_t^2 + \sum_{t=1}^{T} \eta_t^2 + \lambda_\delta \sum_{t=1}^{T-1} \delta_t^2 \qquad [6]$$

where λ_ε and λ_δ are the (user defined) relative weights of the process error and of the observation error on recruits, relative to the observation error on the fully recruited, and θ is the set of parameters. The function SS is non-linear with respect to the parameters and may be minimised with any suitable NLLS algorithm such as Levenberg-Marquardt or Nelder-Mead simplex.

With a mixed-error structure, given time series of catches C and of abundance indices r and n for T years, 2T parameters have to be estimated: $q_n + \{r_1 \dots r_{T-1}\} + \{n_1 \dots n_T\}$, using a set of T equations [2], T-1 equations [3] and T-1 equations [5] (i.e. 3T-2 residual terms, leaving T-2 degrees of freedom). When only observation errors are considered, the first term in [6] is ignored and there are also T-2 degrees of freedom but only T+1 parameters to estimate: $q_n + \{r_1 \dots r_{T-1}\}+$ the index n_1 for the first year. In either case the catch and the recruitment index in the terminal year are not used. Use of equation [5] implies that indices must be available for all years (no missing data).

The procedure yields the set of estimated parameters (noted with a hat), from which the time series of population sizes can be reconstructed using the equations:

$$N_t = \frac{\hat{n}_t}{\hat{q}_n} \qquad t = 1,T \qquad\qquad [7]$$

$$R_t = \frac{\hat{r}_t}{s\,\hat{q}_n} \qquad t = 1,T\text{-}1 \qquad\qquad [8]$$

The procedure does not estimate a recruitment index for the final year. The most recent recruitment is thus estimated with equation [8], but using the observed index (r_T).

Total biomass is derived in the usual way, by multiplying the population sizes by the observed mean weights in each stage and year, and summing:

$$B_t = R_t \overline{W}_{R,t} + N_t \overline{W}_{N,t} \qquad\qquad [9]$$

By analogy with an age structured model, Conser (1994) suggested to measure the overall fishing mortality (here noted F*) using:

$$F_t^* = Z_t - M = Log\left(\frac{R_t + N_t}{N_{t+1}}\right) - M \qquad t < T \qquad [10]$$

F* is not defined for the final year, which may be annoying since this is often the year of major interest for managers. An alternative measure of fishing pressure is the harvest rate which, in its simplest form, is calculated in each year as the catch divided by the sum of recruits and fully recruited. The variant proposed by Collie and Kruse (1998) is simply scaled by $e^{M\tau}$ and therefore shows the same trends with time.

2.3. Data simulation

The basic data required by CSA are time series of total catch in number, and of "survey" indices for the recruits and for the fully recruited. Annual mean weights for each stage are also needed for translating stock numbers into biomass, but play no role in model fitting. These data were generated from an artificial population of known characteristics, comprising 15 age groups, simulated over 40 years of exploitation using an age-structured model. The biological parameters roughly mimic those of Atlantic cod. Only the data for the final 25 years were retained. Details of the simulation and input data preparation are presented in Appendix A. All results considered hereafter are based on true values of catches and of indices obtained from deterministic data generations. Results are displayed for one survey selection pattern only since the findings regarding CSA behaviour were qualitatively the same when input data were generated with the alternative pattern.

In this approach, the true population numbers and the indices for each stage are known in each year, thus the catchabilities for each stage can be easily obtained. Then the ratio of the two catchabilities gives the true value of the s parameter (equation [4]) in each year. Although the selection parameters of the survey were constant over the whole period, the age composition of the underlying population changed considerably in response to recruitment variability and fishing, which resulted in large fluctuations of the true s as shown in Fig. 1. Since this evaluation of the CSA method considers only the simple case where a single constant s is assumed in the estimation, these variations in the true s are akin to a process error in the data passed to CSA. This was not the intention, and is caused by the vagaries of the age structure of the population, not by changes in the survey protocol. In contrast, the model specification error due to the generation and estimation models being different is intentional.

In this instance, the true catchability of the recruits is practically constant over the period (Table 4). Therefore, the calculated s varies inversely with changes in the catchability of the fully recruited. This creates a difficulty for this evaluation since catchability is also a key parameter in CSA and departures from a constant-q assumption are confounded with departures from the constant-s assumption. To try to resolve this ambiguity, indices were generated with and without a steady trend (+3% per annum) in survey efficiency over the final 15 years without changing the selectivity parameters. This ensured that the annual true s remained the same in both scenarios.
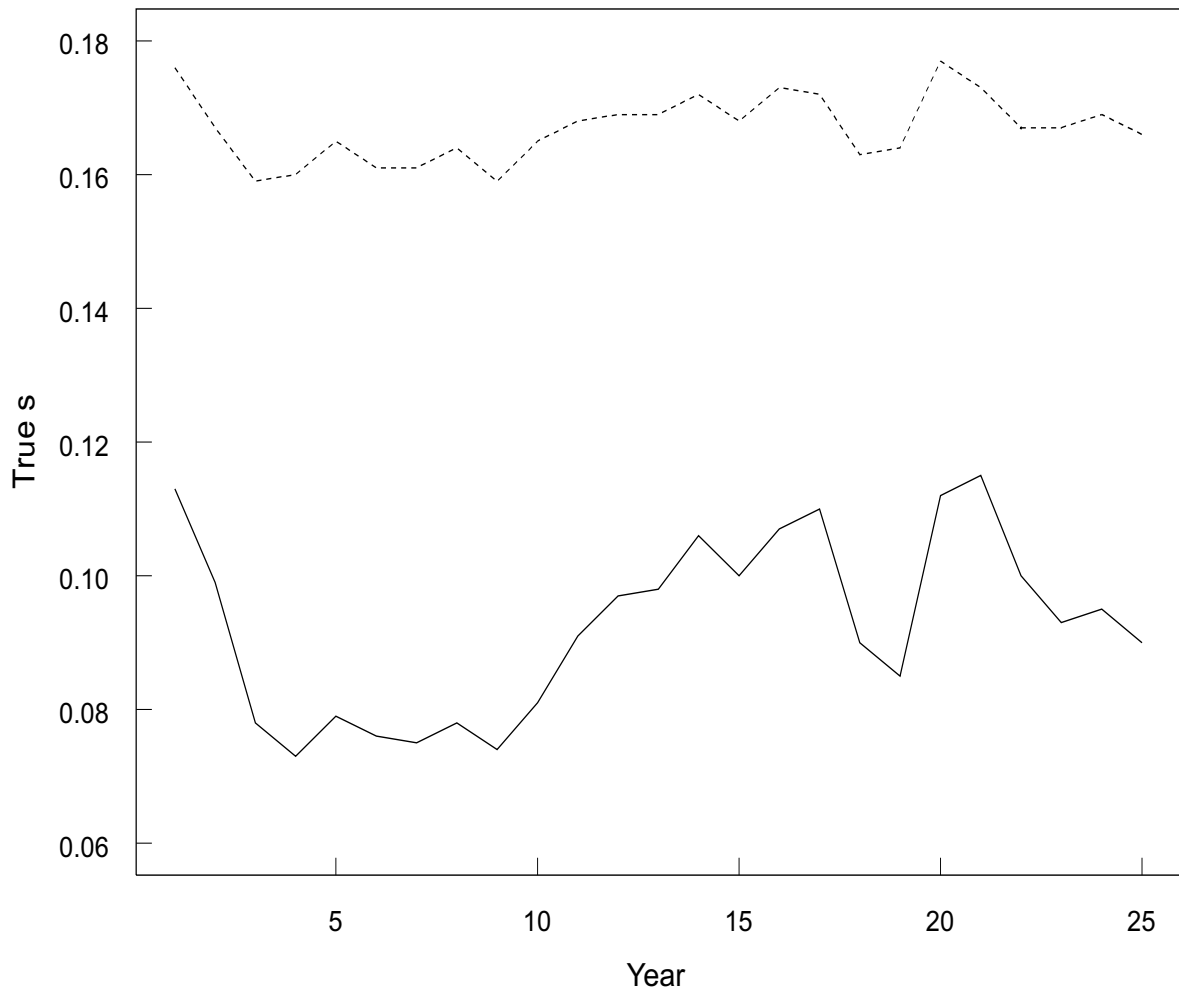

Fig. 1. Time trajectory of the true value of the catchability ratio s between the simulated indices. Solid line: base selection pattern; dashed line: alternative selection pattern.

2.4. Diagnostics

Modern standards require that all assessment methods be accompanied by adequate sets of diagnostics to evaluate the quality of the results and identify potential shortcomings in the data and/or the model, notably when the outcome serves to support management advice. Diagnostics may apply to the estimation of individual "primary" parameters (those of equation [5]) or to derived quantities (e.g. biomasses) of

management interest. The discussion here concentrates on a small set of simple statistics commonly mentioned in the CSA literature.

The overall goodness-of-fit is evaluated with the residual mean square error (RMSE) which is the value of SS (equation [6]) computed at the minimum divided by the number of degrees of freedom (T-2 for both error structures, where T is the number of data years). The contribution of each residual can also be examined to look for patterns or trends. The residuals considered here are log-residuals, in line with the log-normal type of error assumed in equations [2], [3] and [5], standardised to the square root of the RMSE, e.g. for observation residual on recruits δ:

$$\delta_t = \sqrt{\lambda_\delta} \, \frac{(\log \hat{r}_t - \log r_t)}{\sqrt{RMSE}} \qquad [11]$$

The share of each residual term (squared and weighted by the appropriate λ) in the total sum of squares can also be examined to identify influential data and to judge the relative importance of each source of error.

Some minimisation routines estimate the variance or CV of each parameter assuming linearity of the objective function SS near the minimum. However, the approximation is often poor, resulting in inflated variances, and it may not be easy to compute the variance and bias of management quantities derived from the parameters. Bootstrap analysis is now a well established way of evading such problems. The option for this study was to use a non-parametric model-conditioned bootstrap. For each set of options a base fit of the model provided reference estimates and raw residuals. Artificial survey indices were generated by applying randomly selected residuals, drawn within each error type separately, to the reference estimates. Five hundred bootstrap runs were made and empirical percentiles of annual biomasses were extracted. No bias correction was applied to the percentiles. Bootstrap estimates of bias and CV were computed only for the fully recruited catchability q_n which is a key parameter in this method.

Retrospective analyses are now standard in assessment circles and are quite easy to implement with CSA. They are used to check the consistency of past estimates as the data series is augmented. Large and systematic year-to-year revisions of estimates are indicative of violations of some model assumptions, notably constancy of parameters over time. In this instance, retrospective runs were made starting from the complete 25-year data set and removing the terminal year recursively until there were at least 10 years of data left to fit the model. The deviations in successive estimates of biomass, recruitment and fully recruited abundance in the terminal year of each subset were summarised using Mohn's (1999) ρ statistic, e.g. for biomass:

$$\rho = \frac{1}{m} \sum_{t=10}^{24} (B_{(1-t),t} - B_{(1-25),t}) / B_{(1-25),t} \qquad [12]$$

where m is the number of years removed (15 in this instance: years 11-25) and $B_{(1-x),t}$ denotes the result for year t using all data until year x. This statistic has the advantage of taking into account the sign of the updates. A positive ρ indicates that the addition of new data has mostly implied downward revisions of the estimates.

3. Results

3.1. Sensitivity analysis

Sensitivity of the CSA results to various optional settings was explored by varying input parameters one at a time from a base case where the following were assumed: s = 0.09 (close to the mean of the true s, 0.093); τ = 0.5 (consistent with the Baranov catch equation used in generating the data); mixed error structure with equal weight for the process error and the observation error on indices for the fully recruited (λ_ε = 1). Throughout this study no attempt is made to vary the weight of the observation error on recruitment indices (default λ_δ = 1) since these were generated with the same procedure as the fully recruited indices. Comparisons among options and with the true values are displayed in Fig. 2.
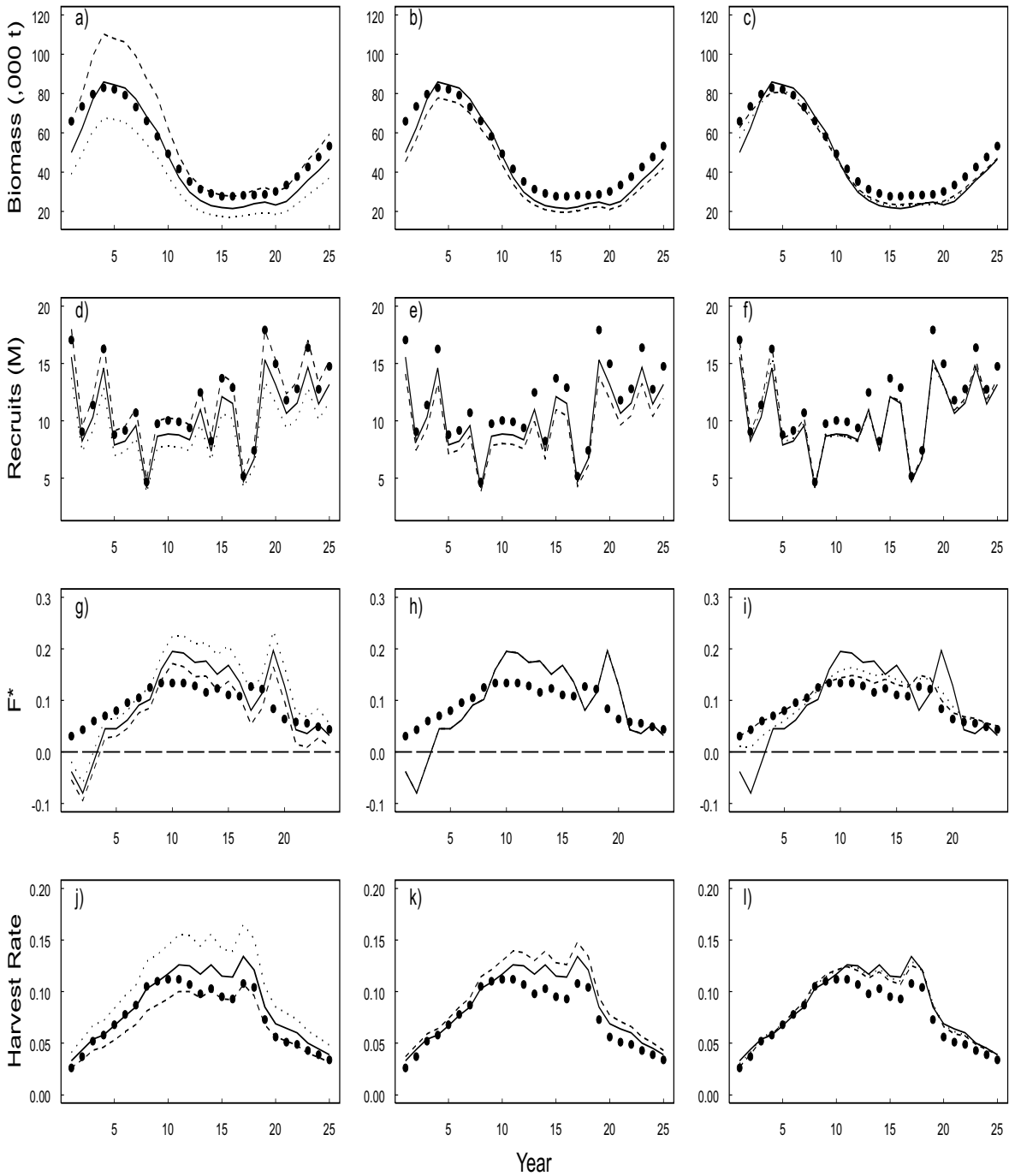

8

Fig. 2. Sensitivity analysis of CSA estimates, from top to bottom: biomass, number of recruits, fishing mortality F* and harvest rate. Left: sensitivity to s (dashed line: $s = 0.10$; solid line: $s = 0.09$; dotted line: $s = 0.08$). Middle: sensitivity to τ (dashed line: $\tau = 0.0$; solid line: $\tau = 0.5$). Right: sensitivity to assumed error structure and weighting (solid line: mixed error, $\lambda_\varepsilon = 1$; dotted line: mixed error, $\lambda_\varepsilon = 10$; dashed line: observation error only). In all panels, the solid circles represent the true values and the solid lines the base case.

From a glance over Fig. 2 it is apparent that the general pattern in the trajectory of each variable is fairly similar under all options and broadly follows that of the true values. High or low estimates consistently occur in the same years and there is only one exception (in panel 2.i) where an option indicates a decrease while others show an increase or vice versa. Nevertheless, in terms of absolute estimates, the analysis shows that the single most influential parameter is the catchability ratio s (panels on the left) which acts as a scaling factor. As s is increased, all estimates of biomass and recruitment are inflated and this mechanically results in lower estimates of fishing mortality F* and harvest rate. Although the variations in s considered are relatively small (±11%), well within the range of uncertainty a user would have to confront in a field case, the effects can be quite large: absolute biomass, for example, is estimated to be 28% larger in the peak year (year 4) or 27% larger in the terminal year when s is varied from 0.09 to 0.1. However, estimates of year-to-year relative changes in biomass are practically the same under the three s options. Still larger sensitivity coefficients were obtained when s was varied about its mean value under the alternative selection pattern, indicating that the results above are not simply an artefact related with the very low s value in the base case. Because the true s has been very variable (Fig. 1) but a single constant value is assumed in the estimation, estimates deviate from the true values in one direction or the other depending on the year. Biomass estimates are generally closer to the truth for $s = 0.09$ in the earlier years and for $s = 0.1$ in the recent period. The latter choice seems more appropriate for recruitments, which are systematically underestimated with the other options, but note that the method is more accurate for very poor than for strong year classes. There is a clear problem with fishing mortality, as measured by F* (panel 2.g and also 2.h-i), in that senseless negative estimates are obtained for the start of the series, a period when the true s fell sharply. Fishing mortalities are also very close to zero in the terminal years for $s = 0.1$. A lower s would remedy these problems but would amplify the discrepancies for other years and other quantities. In brief, no single value of s is adequate for all variables and years. Options that result in negative F* mark the upper bound of the feasibility domain for s, at least for years when this occurs. A lower bound for s is signalled at the extreme when the estimated population is less than the catch (and more evidently by a program crash).

The next option considered is the parameter τ which adjusts for the timing of the fishery within the year (panels in the middle column). As pointed out by Collie and Kruse (1998), the effects are predictable because q_n in equation [5] is simply scaled down by a constant factor $\exp(-\tau M)$ and thus all estimates of biomass and recruitment inflated by $\exp(\tau M)$, i.e. +10.5% for $\tau = 0.5$ compared to $\tau = 0$ when $M = 0.2$. This is relatively minor except when it adds up to the effects of setting s too high. Note in panel 2.h that this option has no effect at all on the estimates of fishing mortality F*.

The choice of error structure is an important issue in modelling and the consequences for this method are examined in the rightmost panels of Fig. 2. For the options considered, the estimates are practically unaffected. Compared to the base case

(mixed error, equal weight), a better match with the true values is obtained for two seemingly opposite options: either observation error only, or mixed error with a comparatively large relative weight (10) for the process error. The improvement is visible for biomass at the start of the period (panel 2.c), and for fishing mortality (2.i) also at the start of the period, where both remove the problem of negative F* discussed above, and in years 17-20 where they smooth out the erroneous sequence of drop and rise. The latter is caused by a fall in true *s* following the recruitment of a very poor year class and is akin to a process error. So, even though process error does exist in the data, it may be better to ignore it altogether, either by using an all-observation-error approach or, equivalently, by weighting it strongly in the mixed error option.

Since the results are so sensitive to errors in *s*, it was felt worthy to check whether results would improve if the true annual *s* were input, although knowing these would be impossible in real applications. Results of this trial are shown in Fig. 3 for biomass and recruitment estimates only. Compared to runs with a constant *s* close to the average, improvements are fairly minimal, if any, and the method does not do much better in approaching the true values when it is provided with the true *s*. This indicates that it is more important to set the overall magnitude of *s* right than to refine the details of its variations through time.
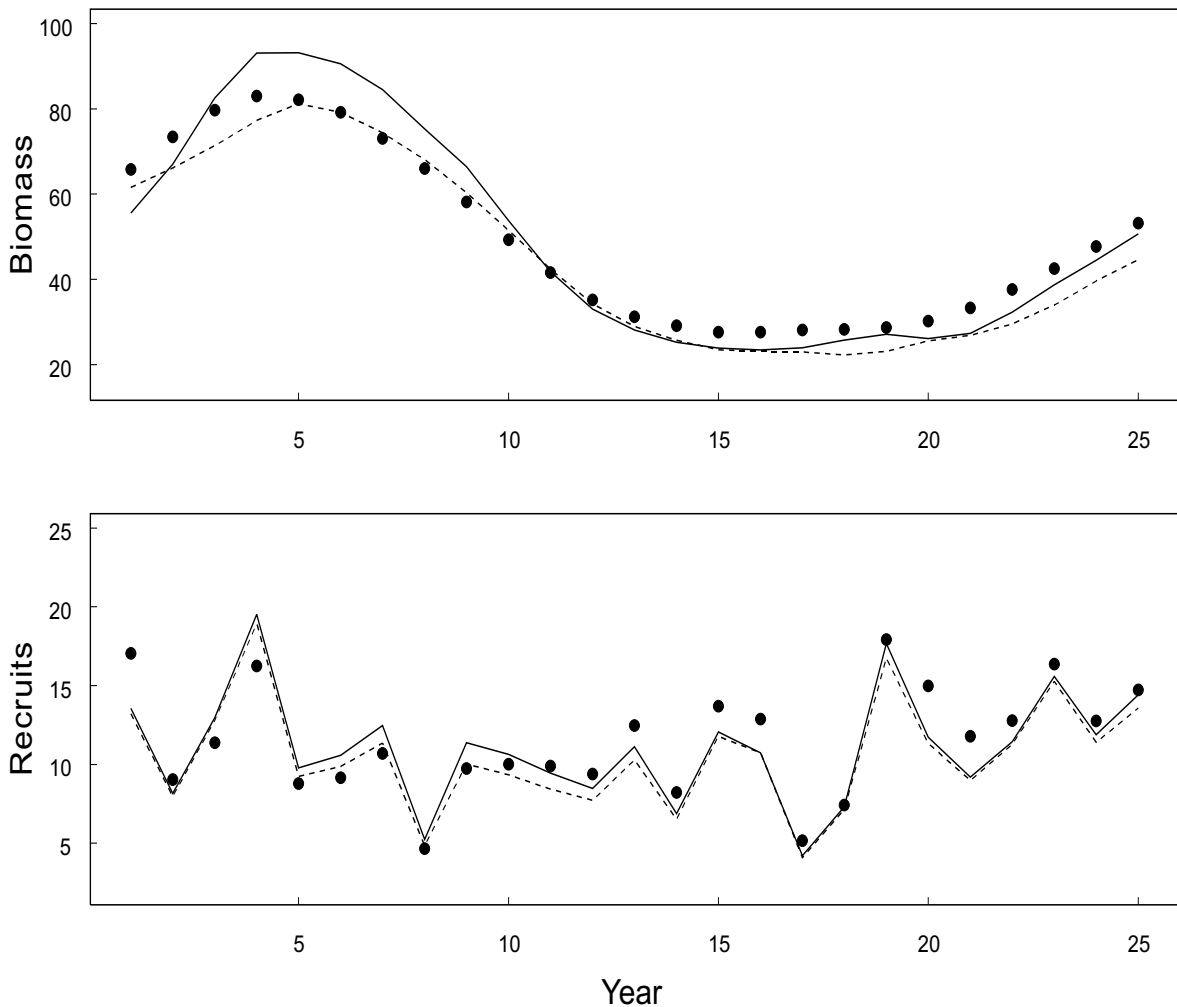
Fig. 3. CSA estimates of biomass (top) and recruitment (bottom) using the true annual catchability ratios s. Solid line: assuming a mixed error structure with $\lambda_\varepsilon = 1$; dashed line: assuming observation error only; solid circles: true values.

In common with most surplus production and VPA tuning methods, the basic form of CSA makes the strong assumption that catchability in the survey or fleet providing the indices has been constant over the period of interest. Violations of this assumption are the major cause of blunders in assessments and advice (e.g. ICES, 2002a). The effects of a trend in catchability, when it is not corrected for in the cpue calculations, have been explored using data generated with a 3% increase in survey efficiency during the last 15 years (this implies that effective catchability of both stages in year 25 is about 1.5 times that in year 10). The outcome under two error structures is shown in Fig. 4, again for biomass and recruitment estimates only. Other settings are as in the base case (s = 0.09, τ = 0.5). As expected, the trajectories of both quantities diverge gradually from those of the true values in recent years, resulting in a gross overestimation of terminal stock abundance. More troubling is the observation that the method estimates smaller catchabilities when an increasing trend in the latter is simulated (Table 1, right): the only mechanism it knows for increasing stock size estimates given a series of indices is to reduce the catchability estimate. CSA definitely has the same problem with catchability trends as other stock assessment methods.
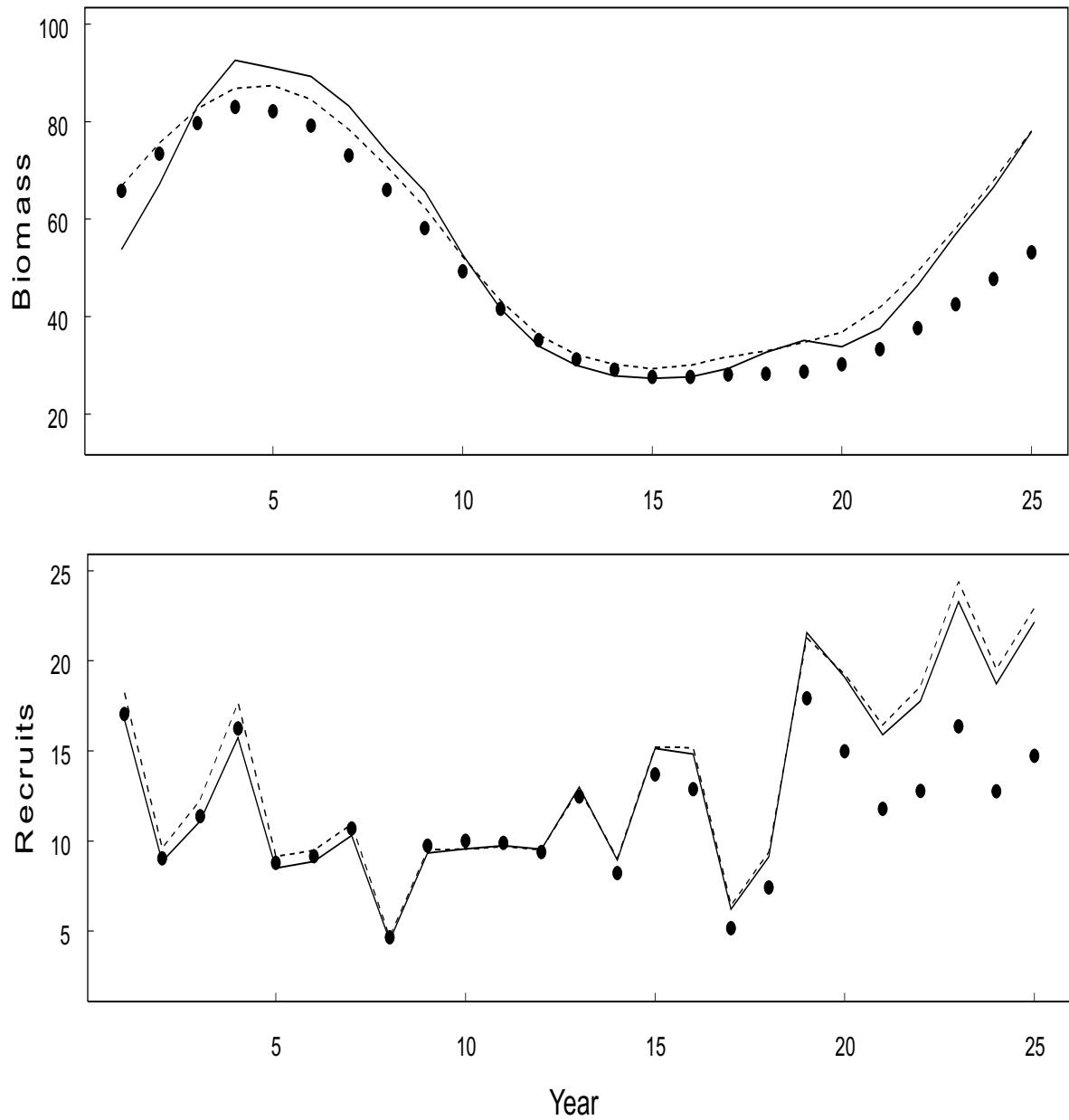
Fig. 4. CSA estimates of biomass (top) and recruitment (bottom) when indices are subject to a catchability trend of 3% per annum in years 10-25. Solid line: assuming a mixed error structure with $\lambda_\varepsilon = 1$; dashed line: assuming observation error only; solid circles: true values.

Table 1. Estimates of fully-recruited catchability and diagnostics for various combinations of input parameters (top panel) considered in the sensitivity analysis.

s	0.10	0.09	0.08	0.10	0.09	0.09	0.10	0.09	0.08	0.09	0.09
Error type	mix	mix	mix	mix	mix	mix	obs	obs	obs	mix	obs
Process wt $\lambda\varepsilon$	1	1	1	1	1	10	-	-	-	1	-
τ	0.5	0.5	0.5	0	0	0.5	0.5	0.5	0.5	0.5	0.5
q trend	no	no	no	no	no	no	no	no	no	+3%	+3%
Estimated q_n	4.83E-4	6.21E-4	7.91E-4	5.34E-4	6.86E-4	6.15E-4	4.74E-4	6.03E-4	7.64E-4	5.75E-4	5.56E-4
RMSE	4.90E-3	4.99E-3	5.35E-3	4.90E-3	4.99E-3	1.16E-2	1.59E-2	1.51E-2	1.50E-2	4.63E-3	1.33E-2
% process error	57.24	54.15	51.21	57.24	54.15	20.49	-	-	-	53.35	-
% obs. error on recruits	3.36	4.19	5.44	3.36	4.19	11.46	14.92	17.99	22.70	4.09	16.40
% obs. error on fully recruited	39.40	41.65	43.35	39.40	41.65	68.05	85.08	82.01	77.30	42.56	83.60
CV(q) :											
Linear approx.	1.721	1.339	1.057	1.723	1.361	0.614	0.538	0.447	0.380	1.647	0.540
Bootstrap	0.037	0.030	0.025	0.036	0.030	0.038	0.062	0.049	0.041	0.035	0.057
Rel. Bias(q) %	2.4	1.2	0.2	2.7	1.4	0.0	3.1	1.3	0.1	1.5	1.2
Biomass ρ	0.064	0.071	0.080	0.064	0.071	0.149	0.239	0.193	0.172	0.059	0.157
Recruits ρ	0.060	0.066	0.075	0.060	0.066	0.128	0.199	0.160	0.148	0.056	0.134

3.2. Diagnostics

The standard estimation diagnostics for a few variations in the settings are summarised in Table 1 which also shows the corresponding estimates of catchability q_n. Unfortunately, most of these diagnostics are unhelpful to decide about a "best" choice, and some are actually misleading since they indicate an improvement in the fit whereas the plots in the previous figures show more divergence from the true values, which we are fortunate to know in this case. RMSE, for example, is smaller for a mixed error structure (except of course when the process weight is increased) than when only observation error is assumed, or when there is a trend in catchability. For reasons already given, it is unchanged when τ is varied. It is right in indicating that an s value of 0.08 is worse than 0.09 or 0.1 under a mixed error structure, but the opposite would be wrongly concluded under an observation error assumption. Options are ranked differently when their quality is judged by the CV of the catchability estimate, with the additional difficulty that its measure by linear approximation is one or two orders of magnitude larger than the bootstrap estimate.
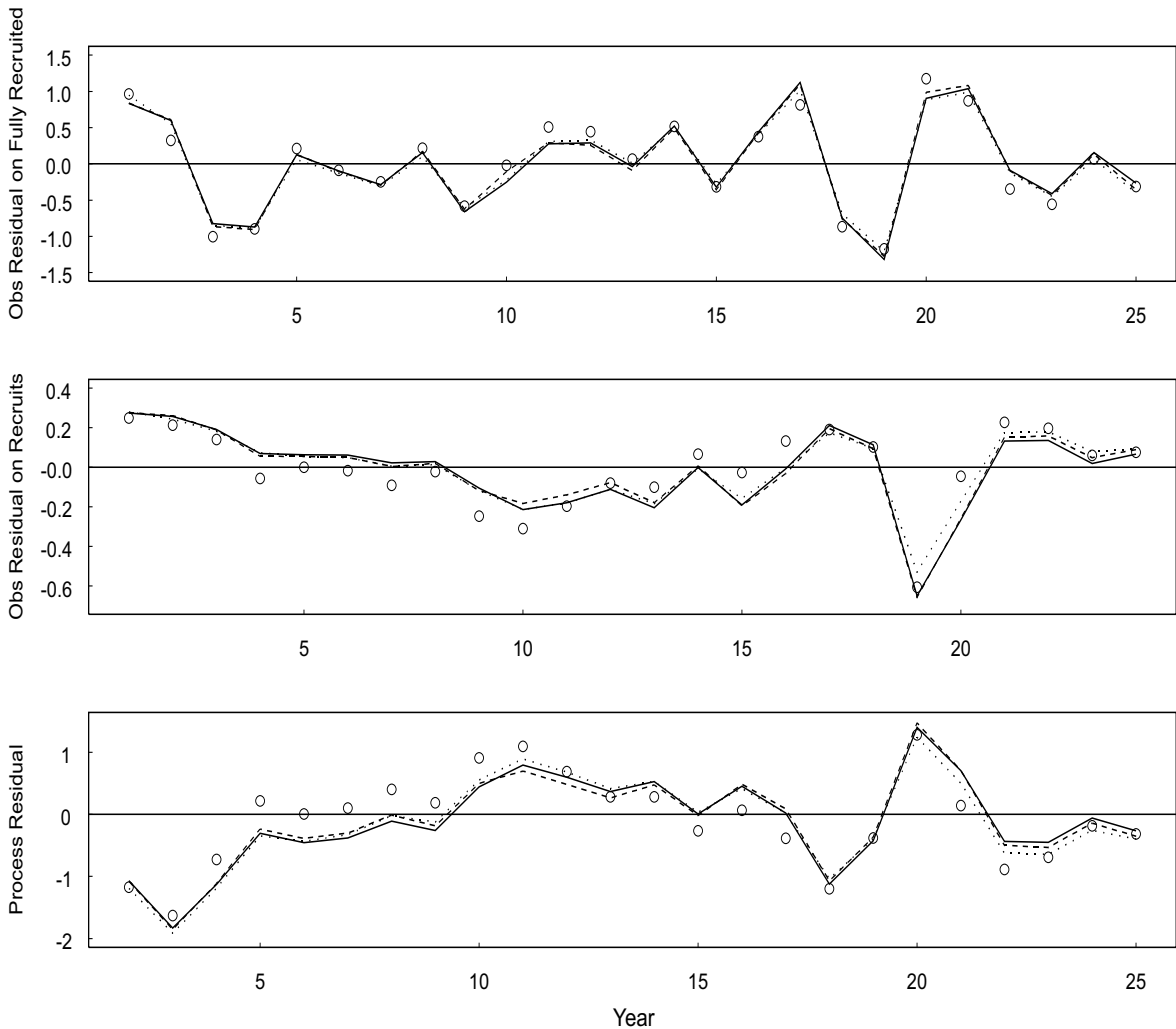

Fig. 5. Plots of standardised residuals against time assuming a mixed error structure with $\lambda_\varepsilon = 1$. Top: observation residuals on fully recruited; middle: observation residuals on recruits; bottom: process error residuals. Open circles: using the true annual catchability ratios; solid line: assuming $s = 0.09$; dashed line: same with a catchability trend; dotted line: assuming $s = 0.1$.

Fig. 5 shows plots of each type of residuals against time for some variations in the settings. It is apparent that the pattern of residuals is very similar across the options considered, including in the presence of a trend in catchability, with differences which are even not discernible in some instances. One might have expected that informing the method with the right annual s would result in smaller residuals, in particular those related with process error, but this does not occur. As noticed by a referee, the reason is that the fully recruited catchability is not constant. He showed that the indices can be reproduced exactly when a variant of equation [5] is used, in which a correction is made for changes in q_n from one year to the next. This is feasible in a simulation context where the true annual catchabilities are known, but would unfortunately be impossible in practice. He also pointed out the negative correlation between process errors and recruits observation errors in Fig. 5: when indices predicted by equation [5] are less than estimated (negative process residuals), the fitting procedure tries to correct this by increasing the recruits estimates, which results in positive observation residuals on recruits, or vice versa.

Other ways of analysing the residuals were explored, taking advantage that the true values of s or q were known, but the attempts were inconclusive. In particular, plots of the observation residuals for the fully recruited (η_t in equation [2]) against the error in s or in q_n seemed to indicate patterns wherein the largest positive residuals tended to be associated with years when the true s was larger than assumed, and vice versa for large negative residuals, but the same pattern also emerged when the true s were used; it may thus not provide a reliable diagnostic for general cases. Fig. 5 eventually carries bad news in that residuals provide no help to detect errors in the magnitude of s nor violations of the constant-q assumption which are potentially pathological for this method as seen above.

Additional elements of diagnostic are provided by the bootstrap analysis. Fig. 6 shows the estimated 10, 50 (median) and 90 percentiles of the empirical distributions for one key quantity, biomass, under two error structures. Patterson et al. (2001) found that the usual 5 and 95 percentiles may not be reliably estimated with such methods, and this is why less extreme values were considered here, also keeping in mind that the number of residuals to draw from is small. The "confidence interval" is quite narrow in most years and symmetrical around the median, giving an impression that the assessment is of high quality in terms of precision. This is also confirmed by the low estimates of CV and of relative bias in catchability obtained with the bootstrap (Table 1). However, Fig. 6 shows that this impression is false since the interval does not include the true value in most years, a discrepancy which is even worse when a larger s is assumed or when there is a trend in survey catchability. For the case shown in Fig. 6, the observation error assumption again performs somewhat better than the mixed error with respect to this criterion. The conclusion is that the bootstrap analysis is not reliable to rank the various options, nor to detect errors in s or catchability trends.
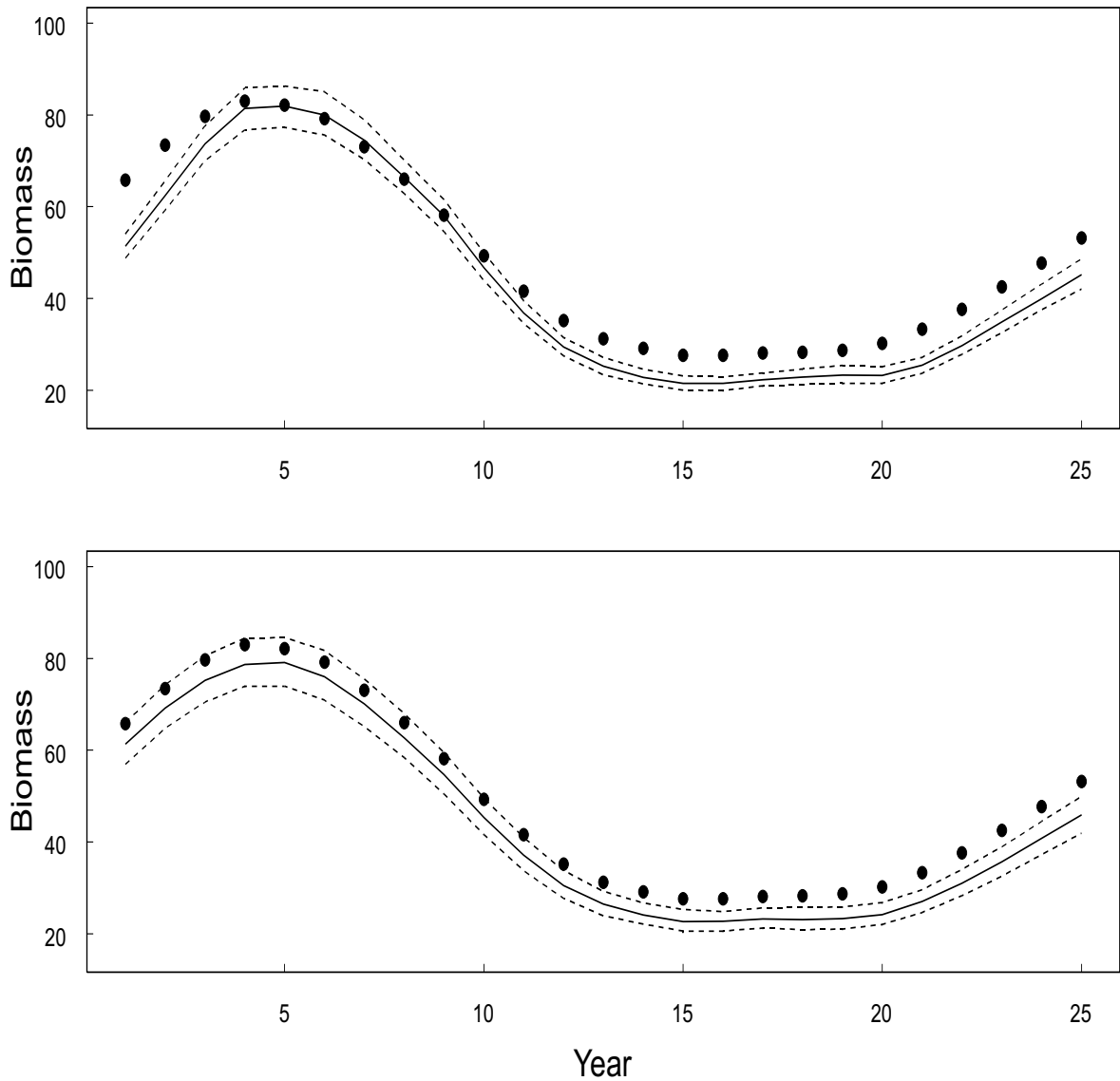
Figure 6. Bootstrap estimates of confidence intervals on biomass assuming $s= 0.09$ and a mixed error structure (top) or observation error only (bottom). Solid line: median; dashed lines: 10 and 90 percentiles; solid circles: true values.

Retrospective analyses were carried out using data with and without catchability trend. For biomass and recruitment, the retrospective plots are nearly identical in either scenario. Therefore, only those for the former (with trend) and for a mixed error structure are presented, with a further restriction to runs ending in the last 10 years (Fig. 7). In general, the revisions due to the addition of later data are fairly small. This is confirmed by the low ρ values shown in Table 1 which, incidentally, are positive for all options. One exception is the run with year 19 as the terminal year, which remains atypical over the whole time series, and this may be associated with the sharp drop in true s in that year. Plots obtained when only observation error is assumed (not shown) display still tighter trajectories for assessments starting in the last 10 years, but larger divergence for some shorter data series and this is the cause for the larger ρ values in Table 1 for this option. There is some indication of systematic upward revisions for recent years, which are also affected by increasing departures from the true estimates (Fig. 4), but this pattern is far from being as dramatic as seen for tuned VPA in presence of catchability trends (e.g.

Mohn, 1999; ICES, 2002a). Readers familiar with VPA will also note that estimates for the earlier years always remain distinct since, unlike VPA, CSA does not benefit from a convergence property. Trials reported in ICES (2002a) show that very nice retrospective VPA plots can be obtained when the tuning data are taken from the same source as the total catch data, implying that both are highly correlated and subject to simultaneous errors. To some extent, this explanation might apply here in view of the procedure used to generate the data for most years (survey indices are a constant fraction of the population at each age and the selection curve is the same for the survey and the catch in the base case). However, it does not hold for the recent years, when the catchability trend affects only the indices, and only partially for the runs using indices generated with the alternative selection curve, which also show tight retrospective plots. Thus, it appears that retrospective CSA analyses are not helpful either to detect catchability trends.
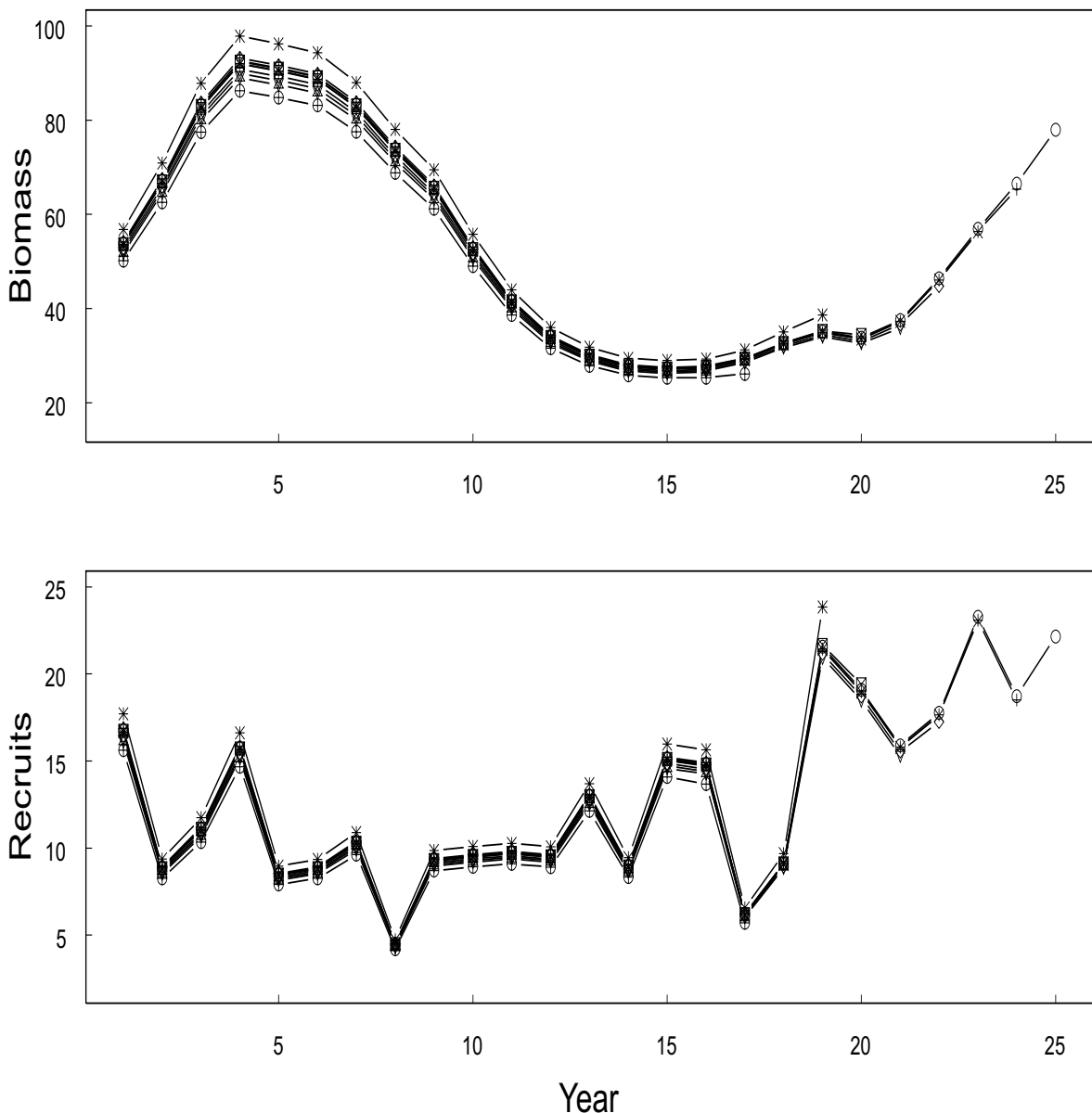

Fig. 7. Retrospective analysis of CSA estimates of biomass (top) and recruitment (bottom) when indices are subject to a catchability trend ($s= 0.09$; mixed error structure).
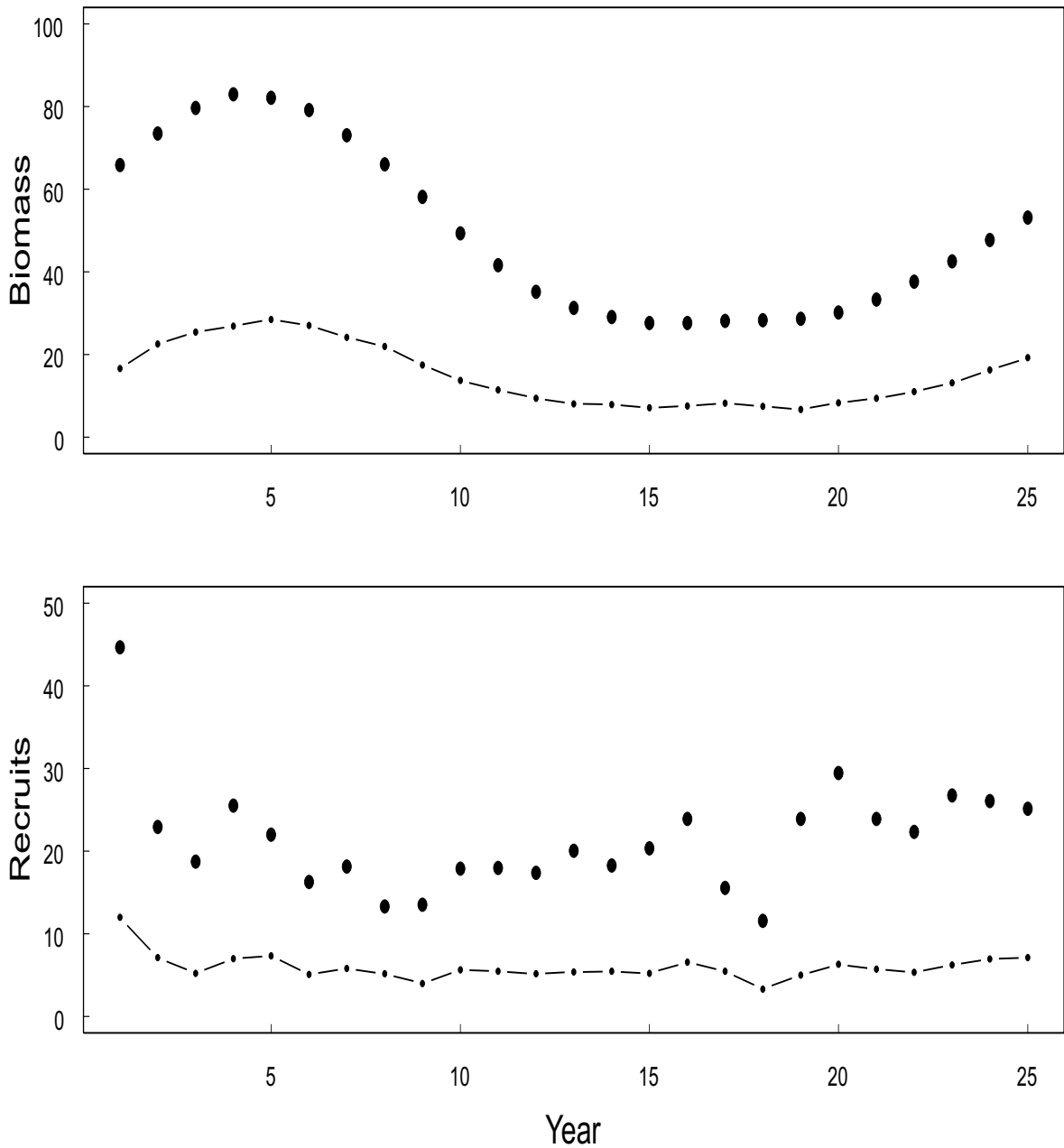
Fig. 8. CSA estimates of biomass (top) and recruitment (bottom) when recruits comprise ages 1 and 2 and the fully recruited are ages 3+. Solid circles are the true values.

3.3. Aging error

The literature is equivocal about the seriousness of aging errors ("staging errors" would be a more appropriate terminology in this case) for CSA results. The simplicity of this simulation framework gave an opportunity to look at this issue with a case which might arise in practice, keeping in mind that CSA is particularly aimed at data-limited contexts. For example, a threshold size may be adopted to partition the survey catch into recruits and fully recruited, but the recruits stage defined in that way may in fact include two year classes. To mimic this, the simulated data for ages 1 and 2 were combined to obtain the indices and mean weights for recruits, and likewise for the fully recruited using

the data for ages 3 and older. The total catches and the true biomasses are the same as in the base case. Results of a run assuming $s = 0.1$ (mean of the true values is 0.103), $\tau = 0.5$ and observation error only (no perceptible difference with mixed error) are shown in Fig. 8. It is clear that such an exaggerated staging error results in a major blunder, with estimated biomasses and recruits abundance consistently reduced to about one third of the true values. There is little effect on estimates of the relative changes in biomass from year to year but, as expected, the recruitment signal is considerably smoothed. Clearly, there is a need to define the stages carefully.

4. Discussion

The foremost conclusion emerging from this study is the confirmation of the broad applicability of CSA for stock assessment. Its ability to capture with a fair degree of exactness the major events in the evolution of a fully age structured population is striking given that it uses only very condensed data. It seems capable of providing quite useful indications about when stock abundance has been high or low, decreasing or increasing, whether it is currently larger or smaller than in earlier years, when strong, poor or average year classes occurred. All too often advisory bodies, particularly in Europe, have been paralysed by their inability to find justifications for sensible management recommendations in absence of VPA based assessments. For many stocks full analytical assessments (i.e. using some variant of VPA) remain a remote prospect due to considerable difficulties in assembling adequate series of age disaggregated data. From scientists the excuse that the data do not fit with conventional assessment tools is not tenable and can only lead to disastrous management. In circumstances where the less elaborate data needed by CSA are available, this method can be regarded as a worthy substitute able to provide grounds for valuable advice on the state of stocks for management decisions. Some institutes possess a wealth of survey data which are under-utilised for species not included in the main stream of assessments, and the method also offers a chance of making good use of such data. Comparative tests using real (ICES, 1995) or simulated (Cadrin, 2000) data indicate that CSA outperforms surplus production approaches, probably due to its ability to use information about recruitment variability. It is certainly not less reliable than a VPA using poor age data. Moreover, many experts (e.g. ICES, 1995; NRC, 1998) sensibly recommend that different assessment models should be applied to the same or similar data as this can shed light on the quality of the results. Thus, even when a VPA is feasible, CSA is a reasonable candidate for alternative assessments to cross-check the results.

Although straight applications of CSA are fine to elucidate general trends in the stock and fishery, input parameters deserve attention when, as is usually the case, absolute estimates of abundance or of harvest rate are needed for management. This study highlights in particular the large sensitivity of absolute estimates to the choice of recruit-to-fully-recruited catchability ratio s, which is not so apparent in the literature. The pioneering work of Collie and Sissenwine (1983) ignores this parameter, but mentions that higher natural mortality of the recruits may counterbalance the effects of their lower catchability on the estimate of adult q. Zheng et al. (1997, 1998) assume s to be unity on the grounds that recruits are large enough to be equally vulnerable to the survey gear, which seems reasonable in their case, and mostly focus on the inverse relationship between s and q_n or M. The same comment applies to Collie and Kruse (1998) who also assume s to be unity but with the bizarre justification that "*both stages are of legal size*" (implicitly in the fishery, whereas s actually has to do with the survey). Such ambiguity is

also commonplace in Cadrin et al. (1999) where the reader is often in trouble to sort out the exact role of data obtained from the 44 mm mesh commercial fishery or from the 32 mm mesh survey. They find that results, in terms of stock diagnostic and total mortality estimates, are very similar over a range of s. The effects on absolute abundance estimates are not shown. It is to be noted that their assessment takes some account of the temporal variability in s. Cadrin (2000) indicates that CSA results are sensitive to incorrect model assumptions, including errors in setting s, but the latter are not the focus of his paper. At least, all papers agree on the qualitative effects of errors in setting s upon the various CSA estimates, and on the need to estimate s externally using additional information. In principle, s may be estimated (and adjusted annually) on the basis of the mean size in each stage, and of the selection curve of the survey gear. However, the example treated here shows that the actual s is influenced by variability in the age and size composition of the population in addition to gear selectivity. This study also indicates that refining the estimation of s by year is less important than setting its magnitude right. Perhaps the next step will be to develop a simulation with a length- and age-structured artificial population and a length selective survey to examine whether information about the mean sizes in each stage and the gear selection properties is sufficient for adequate estimation of the general level of s.

The literature is also unclear about the issue of aging errors. However, as now recognised, a fundamental assumption of this model is that recruits should consist of those animals that grow to the fully recruited stage in one year. Collie and Kruse (1998) simulate moderate aging errors and find that these result in a small bias in q, but mention that larger bias might be a problem if aging errors are more serious and asymmetric. Cadrin (2000) insists that a correct definition of the recruits and fully recruited stages is of major importance. This is probably why Cadrin et al. (1999) were compelled to go through a rather tedious data processing in order to exclude the pre-recruits. Zheng et al. (1997, 1998) also consider the need for a three-stage model, with pre-recruits, although their concern is mostly the prediction of recruits in the terminal year. Preliminary trials on *Nephrops* by ICES (2002b) indicated that pooling the first two ages as recruits made little difference in absolute abundance estimates compared to using only the first age. The example here effectively shows that a severe and asymmetric staging error may lead to grossly erroneous results. The reason is not so much the switch of age 2 from the fully recruited to the recruits (it makes only 5% in average of the indices for the former) but rather that age 1, which makes 20-56% of the indices for ages 1+2 recruits, stays two years in the recruits stage instead of growing to the fully recruited at the end of each year. This upsets the rationale of the dynamics equation [1]. Thus, it is important to not overrate the simplicity of this model and of the two-stage structure in general, and to carefully allocate members to either stage.

In common with a number of assessment methods which assume constant catchability of the survey or fleet providing abundance indices, CSA is bound to produce a misleading picture of stock development if there is a trend or shift in catchability that is not corrected for beforehand when cpue data are processed. It is also counterintuitive that CSA produces lower estimates of the overall catchability in cases where an increasing trend in q is present. In principle, scientific surveys should not be affected by changes in efficiency, but protocols may vary (if only due to bad weather) and survey catchability may be altered by changes in the fish distribution. Like with tuned VPA, catchability adjustments cannot be dealt with internally by the method but have to be based on independent analyses of the survey or fleet data leading to appropriate selection of the stations or vessels used to construct the indices (e.g. ICES, 2002a). A practical difficulty

highlighted in this study is that conventional diagnostics do not help to detect violations of the constant-q assumption. More work is needed to understand what is effectively reflected in the current diagnostics, and to develop more perspicacious new diagnostics.

Sensitivity of CSA results to errors or variations in natural mortality M was not addressed in this work because the issue is well settled in the literature. M and survey catchability q_n are negatively correlated, implying that underestimating M translates into overestimation of q_n then into underestimation of abundance and overestimation of fishing mortality or harvest rate. There is a broad consensus, already since Collie and Sissenwine (1983), that M cannot be estimated along with the other CSA parameters using only the basic input data (catch and indices). Zheng et al. (1998) add data about commercial effort, and corresponding terms in the objective function SS, but even this is not sufficient to improve the estimation of M. The model is presented here in its simplest form, with M constant by stage and over time, but allowing for variations in M is a simple matter. The real constraint is to have external data and analyses to estimate stage or year specific values. For the latter case, Collie and Kruse (1998) consider a potential relationship with sea temperature but confess mixed results. It should be noted that the definition of fishing mortality F^* as given by equation [10] is valid only if M is the same for both stages. This study indicates that this measure of fishing mortality is problematic anyway, and that other measures of fishing pressure (e.g. harvest rate or variants thereof) have to be considered. There is also consensus that a precise determination of when the catch should be discounted within the year (parameter τ) needs not be a major concern, unless M is large. The effect is a simple scaling which is perfectly predictable.

This work provides another contribution to the debate about the choice of an appropriate error structure for estimation, which is always a difficult question in modelling. Following the original presentation of the method (Collie and Sissenwine, 1983), one school (e.g. Cadrin et al., 1999; Cadrin, 2000) has adopted a mixed error estimation procedure, whereas another school (Collie and Kruse, 1998; Zheng et al., 1997) gives preference to an all-observation-error approach on common sense grounds of parameter parsimony. One may also argue that the process-error residuals should be based on differences between genuine independent observations and model predictions, rather than between two internally computed quantities as currently used. In the case of surplus production models, Polacheck et al. (1993) or Punt and Hilborn (1996) show that estimations based on process error estimators have larger variance and conclude that, if a choice has to be made, the option of observation error only is generally preferable. NRC (1998) expresses another view: " ... *stock assessment methods incorporating both measurement and process error reflect the uncertainty in both the data and the population more accurately*". Incidentally, it is unfortunate that CSA was not selected for the large scale testing exercise conducted for the NRC review. The findings from this study give some support to the second school. At least they indicate that there is no degradation in the assessment if process error is ignored, even though it is known to exist in the data as in this example where a unique value of s is assumed whereas true s varies considerably. If the route of observation error only can be safely taken, a practical advantage is that users elude some difficult choices. One is the choice of relative process weight. The sensitivity analysis above confirms previous findings (Collie and Sissenwine, 1983) that CSA estimates are weakly sensitive to changes in λ_ε within a reasonable range, which is fortunate because finding defensible justifications for a weighting scheme is generally a problem (Thompson and Bakkala, 1990). Another difficult choice is the form of the process error. Collie and Sissenwine (1983) consider an additive normal error, a form also advocated by Mendelssohn (1988), whereas other applications (e.g. Cadrin, 2000)

consider a multiplicative log-normal error as often assumed, by default at times, in fishery science. Since process error combines a variety of factors (M, s, catch), justifying one form or the other is challenging. In addition, if the objective function SS is constructed with additive process errors and logarithms of observation errors, the former are of a much larger magnitude and may drive the estimation entirely.

If we push the idea of an all-observation-error approach further, it may provide an opportunity to recast the estimation differently, such that the unknowns would directly be the absolute stock sizes of recruits in all years except the last, and of the fully recruited in the first year (those for later years being obtained by forward projection with the dynamics equation). A similar evolution occurred for surplus production models (Punt and Hilborn, 1996) and would afford the same advantages. Firstly, this would allow for missing survey years, if very occasional of course, a case which is likely to be encountered in data-limited situations and is a barrier to using the current form of CSA. Secondly, it would be possible to "tune" the estimation with data from several index fleets, when available. An additional advantage is that catchability can be computed outside the NLLS search, which is likely to improve the behaviour of the latter.

This presentation has concentrated so far on the use of CSA for estimation of past abundance. However, a complete fish stock assessment has also to consider prediction of future states under alternative management scenarios (Hilborn and Walters, 1992). Given CSA results, a broad choice of scenarios can be contemplated, including pre-determined catch or harvest rate sequences or target biomasses. Multiple fleet versions can even be envisaged since it is possible to estimate harvest rates by stage and by fleet if catches are known with the same resolution. Like with many other forecasting procedures, the critical issue is the prediction of future recruitments but at least CSA gives no less information on past year classes than a VPA and the same usual inferences can be made (average, random draws, etc.). The two-stage model imposes limitations to the consideration of a spawner-recruit relationship unless one is fortunate that the fully recruited stage coincides with fully mature animals. Otherwise, it would be very daring to assume that spawners are a constant fraction of abundance in that stage (in our artificial population, SSB ranged 40-70% of the biomass of fully recruited depending on years). The same difficulty applies to the definition of precautionary reference points based on spawning biomass.

In summary, Catch-Survey Analysis is a very attractive method of fish stock assessment, notably for data-limited contexts where catch-at-age data are missing or uncertain, and deserves consideration for inclusion in standard toolkits. It provides useful indications about general stock trends, which may already be a clear improvement on the rudimentary information currently passed to managers for some fisheries. More work is needed on the specification of key parameters that influence absolute estimates of abundance and of fishing pressure, and on the provision of adequate diagnostics for identifying violations of the model's basic assumptions. The method is relatively "young" in terms of amount of literature. Perhaps it is worth recalling that it took decades to develop VPA after J. Gulland popularised it in the mid 1960's and that, despite hundreds of publications about the theory and application of VPA, the method does not seem to be mature yet in view of the persisting problems with its use (NRC, 1998; ICES, 2002a).

Table 2. Specifications of the age-structured population and fishery simulation used to generate the CSA input data (the notation 'n*value' means that the value applies to the next n ages).

Natural Mortality	M = 0.2, all ages and years
Growth	K = 0.15; Linf = 100; to = 0
Length Weight	a = 0.00001; b = 3
Maturity at age	4*0.0, 0.3, 0.5, 0.7, 0.9, 0.95, 6*1.0
Recruitment:	
Type	Beverton-Holt: R = S/(α+βS)
Parameters	α = 0.67945; β = 5.6621E-5 (steepness = 0.7)
Variability	Log-Normal; CV = 0.6; auto-correlation ρ = 0.5
Selectivity at age	
Base	0.05, 0.1, 0.3, 0.7, 0.9, 10*1.0
Alt.	0.15, 0.7, 0.9, 12*1.0
Survey q	0.001

Appendix A. Generation of the simulated data.

The data generation procedure used is an adaptation of that used by Restrepo et al. (2000) to evaluate uncertainty estimation methods in the context of tuned VPA. It considers an age structured population comprising 15 ages (1-15, no plus-group: contributions of ages 16 and older are ignored). The specifications are summarised in Table 2. The population structure in the first year is generated under equilibrium and with a recruitment of 14558 animals. This population is then simulated forward over 41 years, with nominal fishing mortality maintained at 0.5*F_{MSY} (F_{MSY} = 0.166) during a burn-in period of 17 years, then increased gradually to twice F_{MSY}, maintained there during years 27-33, and subsequently reduced toward half F_{MSY} ("two-way trip"). Recruitment in each year is stochastic about a Beverton-Holt stock-recruitment relationship. In the base case survey indices at age are generated assuming that the survey gear has the same selectivity pattern as the fishing fleet and a nominal catchability of 0.001. The latter is increased by 3% per annum during the last 15 years in the q-trend scenario. In another alternative, a steeper selection pattern is assumed for the survey (as might be expected

with a small-mesh gear) while the fishery's selection remains the same, in order to check that results are not dependent on the specifics of a single simulation.

For the analysis, only the final 25 years are retained (re-coded 1 to 25) and data are aggregated in line with CSA requirements. Annual catches are the sums of catch numbers at age. In the base case, recruit indices are directly the age 1 indices, whilst the indices for the fully recruited are the sums over ages 2-15. In the "staging error" scenario, recruits are a combination of ages 1 and 2 (it make sense to sum up indices for these ages in each year because they belong to distinct cohorts), and the fully recruited are age 3 and older. Mean weights for each stage are weighted means, taking account of population number at each age. The CSA input data for the base case are shown in Table 3. All estimation runs assumed $M = 0.2$. Table 4 shows the true values of key variables for the base selection pattern. True catchabilities by stage are the survey indices divided by the true population numbers in each year; their ratio gives the true s.

Table 3. Simulated CSA input data (U: survey indices; W: mean weights by stage).

Year	CatchN	Urec	Ufull	Wrec	Wfull
1	2078.7	0.852	27.598	0.027	1.050
2	2681.0	0.451	31.734	0.027	1.163
3	3492.2	0.568	35.927	0.027	1.407
4	3969.8	0.812	35.739	0.027	1.578
5	4156.7	0.438	33.114	0.027	1.564
6	4339.1	0.457	30.181	0.027	1.709
7	4528.1	0.534	27.546	0.027	1.768
8	4505.4	0.232	24.340	0.027	1.722
9	4488.6	0.487	21.060	0.027	1.866
10	4393.9	0.501	18.090	0.027	1.680
11	4268.4	0.495	15.365	0.027	1.472
12	3912.2	0.469	13.971	0.027	1.286
13	3813.1	0.623	13.396	0.027	1.173
14	3744.2	0.411	13.352	0.027	1.021
15	3806.9	0.685	13.269	0.027	1.032
16	3921.0	0.644	13.736	0.027	0.928
17	3905.7	0.258	14.068	0.027	0.901
18	3484.3	0.370	14.446	0.027	1.074
19	3074.7	0.895	14.315	0.027	1.160
20	2601.2	0.748	14.211	0.027	0.937
21	2424.1	0.589	15.584	0.027	0.919
22	2452.3	0.638	18.443	0.027	1.012
23	2370.2	0.818	20.714	0.027	1.096
24	2152.6	0.637	22.440	0.027	1.109
25	2000.5	0.736	24.074	0.027	1.215

Table 4. True values of selected state variables.

Year	Biomass	Recruits R	Fully Recr. N	F*	C/N	Catchability		s
						qr	qn	
1	65 808.0	17 043.9	62 220.0	0.031	0.026	5.00E-5	4.44E-4	0.113
2	73 430.6	9 032.3	62 929.2	0.043	0.037	4.99E-5	5.04E-4	0.099
3	79 690.4	11 378.3	56 415.3	0.060	0.052	4.99E-5	6.37E-4	0.078
4	82 969.5	16 257.0	52 287.4	0.070	0.058	4.99E-5	6.84E-4	0.073
5	82 109.5	8 777.3	52 347.0	0.080	0.068	4.99E-5	6.33E-4	0.079
6	79 184.3	9 146.8	46 191.1	0.096	0.078	5.00E-5	6.53E-4	0.076
7	73 080.7	10 694.2	41 179.4	0.105	0.087	4.99E-5	6.69E-4	0.075
8	65 989.6	4 655.2	38 253.7	0.125	0.105	4.98E-5	6.36E-4	0.078
9	58 126.9	9 745.3	31 002.7	0.134	0.110	5.00E-5	6.79E-4	0.074
10	49 318.2	10 020.5	29 186.3	0.134	0.112	5.00E-5	6.20E-4	0.081
11	41 577.4	9 900.2	28 062.4	0.134	0.112	5.00E-5	5.48E-4	0.091
12	35 182.8	9 386.0	27 170.0	0.128	0.107	5.00E-5	5.14E-4	0.097
13	31 245.1	12 466.6	26 343.0	0.116	0.098	5.00E-5	5.09E-4	0.098
14	29 100.8	8 227.7	28 288.3	0.123	0.103	5.00E-5	4.72E-4	0.106
15	27 648.8	13 703.4	26 429.9	0.111	0.095	5.00E-5	5.02E-4	0.100
16	27 622.9	12 883.4	29 393.3	0.108	0.093	5.00E-5	4.67E-4	0.107
17	28 131.7	5 168.0	31 068.1	0.127	0.108	4.99E-5	4.53E-4	0.110
18	28 270.3	7 416.4	26 135.4	0.122	0.104	4.99E-5	5.53E-4	0.090
19	28 673.0	17 916.0	24 310.7	0.084	0.073	5.00E-5	5.89E-4	0.085
20	30 204.3	14 972.3	31 791.6	0.064	0.056	5.00E-5	4.47E-4	0.112
21	33 337.9	11 786.4	35 930.9	0.058	0.051	5.00E-5	4.34E-4	0.115
22	37 646.0	12 777.8	36 866.8	0.056	0.049	4.99E-5	5.00E-4	0.100
23	42 541.2	16 372.6	38 426.4	0.049	0.043	5.00E-5	5.39E-4	0.093
24	47 709.2	12 751.1	42 708.4	0.044	0.039	5.00E-5	5.25E-4	0.095
25	53 159.9	14 734.3	43 439.4		0.034	5.00E-5	5.54E-4	0.090

References

Cadrin, S.X., 2000. Evaluating two assessment methods for Gulf of Maine northern shrimp based on simulations. J. Northw. Atl. Fish. Sci. 27, 119-132.

Cadrin, S.X., Clark, S.H., Schick, D.F., Armstrong, M.P., McCarron, D., Smith, B., 1999. Application of catch-survey models to the northern shrimp fishery in the Gulf of Maine. N. Am. J. Fish. Man. 19, 551-568.

Collie, J.S., Kruse, G.H., 1998. Estimating king crab (*Paralithodes camtschaticus*) abundance from commercial catch and research survey data. In: Jamieson, G.S., Campbell, A. (Eds.), Proceedings of the North Pacific Symposium on Invertebrate Stock Assessment and Management. Can. Spec. Publ. Fish. Aquat. Sci. 125, 73-83.

Collie, J.S., Sissenwine, M.P., 1983. Estimating population size from relative abundance data measured with error. Can. J. Fish. Aquat. Sci. 40, 1871-1879.

Conser, R.J., 1991. A DeLury model for scallops incorporating length-based selectivity of the recruiting year-class to the survey gear and partial recruitment to the commercial

fishery. Research Document SAW 12/2. Appendix to CRD-91-03, Northeast Regional Stock Assessment Workshop Report, Woods Hole MA, 18 pp.

Conser, R.J. 1994. Stock assessment methods designed to support fishery management decisions in data-limited environments: development and application. PhD dissertation. School of Fisheries, University of Washington, Seattle, 292pp.

Conser, R.J., Idoine, J., 1992. A modified DeLury model for estimating mortality rates and stocks sizes of American lobster populations. Research Document SAW 14/7. Appendix to CRD-92-07, Northeast Regional Stock Assessment Workshop Report, Woods Hole MA, 28 pp.

Deriso, R.B., 1980. Harvesting strategies and parameter estimation for an age-structured model. Can. J. Fish. Aquat. Sci. 37, 268-282.

Hilborn, R., Walters, C.J., 1992. Quantitative Fisheries Stock Assessment. Choice, Dynamics and Uncertainty. Chapman & Hall, New York, 570 pp.

ICES, 1995. Report of the Working Group on Methods of Fish Stock Assessments, Copenhagen, 6-14 February 1995. ICES CM 1995/Assess:11, 215 pp.

ICES, 2002a. Report of the Working Group on Methods of Fish Stock Assessments, Copenhagen, 3-7 December 2001. ICES CM 2002/D:01, 102 pp.

ICES, 2002b. Report of the Working Group on *Nephrops* Stocks, Lorient, 3-9 April 2002. ICES CM 2002/ACFM:15, 258 pp.

Mendelssohn, R., 1988. Some problems in estimating population sizes from catch-at-age data. Fish. Bull. 86, 617-630.

Mohn, R. 1999. The retrospective problem in sequential population analysis: an investigation using cod fishery and simulated data. ICES J. Mar. Sci. 56, 473-488.

NRC (National Research Council), 1998. Improving Fish Stock Assessments. National Academy Press, Washington D.C., 177 pp.

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

Polacheck, T.R., Hilborn, R., Punt, A.E., 1993. Fitting surplus production models: comparing methods and measuring uncertainty. Can. J. Fish. Aquat. Sci. 50, 2597-2607.

Pope, J.G. 1972. An investigation of the accuracy of Virtual Population Analysis using cohort analysis. ICNAF Res. Bull. 9, 65-74.

Punt, A.E., Hilborn, R., 1996. BIODYN: biomass dynamic models. User's manual. FAO Computerized Information Series (Fisheries) 10, 62 pp.

Restrepo, V.R., Patterson, K.R., Darby, C.D., Gavaris, S., Kell, L.T., Lewy, P., Mesnil, B., Punt, A.E., Cook, R.M., O´Brien, C.M., Skagen, D.W., Stefánsson, G., 2000. Do different

methods provide accurate probability statements in the short term? ICES CM 2000/V:08, 18 pp.

Thompson, G.G., Bakkala, R.G., 1990. Assessment of the eastern Bering Sea Pacific cod stock using a catch-at-age model and trawl survey data. Int. North Pac. Fish. Comm. Bull. 50, 215-235.

Zheng, J., Murphy, M.C., Kruse, G.H., 1997. Application of a catch-survey analysis to blue king crab stocks near Pribilof and St Matthew Islands. Alaska Fish. Res. Bull. 4(1), 62-74.

Zheng, J., Murphy, M.C., Kruse, G.H., 1998. Abundance estimation of St. Matthew Island blue king crab using survey and commercial catch and effort data. In: Funk, F., Quinn II, T.J., Heifetz, J., Ianelli, J.N., Powers, J.E., Schweigert, J.F., Sullivan, P.J., Zhang, C.-I. (Eds), Fishery Stock Assessment Models. Alaska Sea Grant College Program Report No AK-SG-98-01, University of Alaska Fairbanks, pp. 575-589.