The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data

R. Mohn

The retrospective problem is a systematic inconsistency among a series of estimates of population size, or related assessment variables, based on increasing periods of data. In some stocks, this problem is of such magnitude that sequential population analyses (SPA) are deemed inapplicable. The eastern Scotian Shelf (ESS) cod fishery, which displays the retrospective problem, and simulated data are analysed to provide insight into the causes and potential solutions to this problem. The retrospective problem is shown to be a result of the traditional analysis techniques and a non-stationarity in the data used in the population analysis. A moving window analysis is developed which allows the non-stationarities to be identified and in some cases rectified. Recommendations are also made for *ad hoc* investigations of the data. The analysis suggests that failure to correct the retrospective problem for a stock with data like ESS cod could lead to catch-level advice that would be twice or more the intended level.

R. Mohn: Department of Fisheries and Oceans, PO Box 1006, Dartmouth, Nova Scotia, Canada, B2Y 4A2. Tel: +1 902 426 4592; fax: +1 902 426 1506; e-mail: mohnr@mar.dfo-mpo.gc.ca

Introduction

Retrospective analysis denotes an investigation of successive estimates of fishery variables, typically population number, biomass, or fishing mortality estimates, as additional years of data (data windows) are added. The retrospective problem is said to exist when the various estimates for a given year do not agree and the disagreement is a systematic pattern of biases rather than a random scattering. Estimates of population biomass for the eastern Scotian Shelf (ESS) cod from nine assessments covering the period 1982–1995 are shown in Figure 1. Subsequent estimates of biomass in the early 1980s from the various assessments are lower than their predecessors. All the estimates in Figure 1 are from sequential population analysis (SPA) although the tuning criteria and model descriptions will have varied from author to author. If a consistent analysis is used on data windows of increasing duration the retrospective problem in ESS cod is even more pronounced (Figure 2a,b). The initial biomass estimates in the mid-1980s are more than 100% greater than the estimates from the entire time series.

Aspects of the retrospective problem have been observed for almost two decades. Butterworth (1981, his Figure 4), showed successive estimates of biomass for the southwest African pilchard stock for the period 1974–1979 which consistently overestimated the terminal year, a similar, though less pronounced, pattern to ESS cod in Figure 1. More recently, the retrospective problem has been reported in Sinclair *et al* (1991) and ICES (1991). In both cases recommendations were vague as the causal mechanisms remained obscure, Parma (1993), with a data series that extended back to 1944, used a moving window approach and concluded that the retrospective pattern "may result from an increasing trend in catchability". Her tuning data were effort and standardized commercial catch rates, for which trends in catchability may be expected as a result of improvements in technology or spatial changes in the halibut resource. This is in contrast to the ESS cod SPA which is tuned to a research survey index of abundance for which the catchability would not be expected to change over time. Finally, Evans (1996) presents a diagnostic for the retrospective problem and again concludes that the problem "probably involves a change
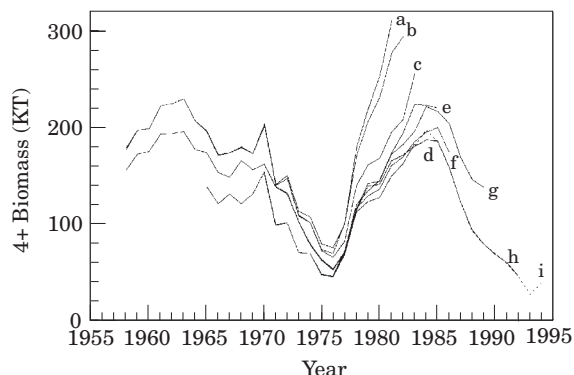
Figure 1. Retrospective pattern of 4+biomass for eastern Scotian Shelf cod from various analyses 1982–1995. a, Maguire *et al.*, 1982; b, Gagne *et al.*, 1983; c, Gagne *et al.*, 1984; d, Sinclair and Gavaris, 1985; e, Sinclair and Annand, 1986; f, Sinclair and Smith, 1987; g, Fanning and MacEachern, 1990; h, Mohn and MacEachern, 1992; i, Fanning *et al.*, 1995.

over time of a quantity that has been assumed constant". It shall be demonstrated that the critical quantity is the catchability coefficient which links the estimated population numbers to the tuning data.

Sequential population analysis may be thought of as a tuning, or calibration, exercise linking catch data to abundance indices via a model and scaling parameters, i.e. the catch equation and catchabilities (qs). Numbers at age in a given year, y, ($N_{a,y}$) are estimated from the SPA equation which scales the catch-at-age data ($C_{a,y}$) with an assumed natural mortality (M) and adds it to the next older age in the cohort, $N_{a+1,y+1}$:

$$N_{a,y} = SPA(C_{a,y}, M, N_{a+1,y+1}). \tag{1}$$

The SPA model also requires assumptions about fishing mortality in the most recent year and oldest ages (Mohn and Cook, 1993). An example of perhaps the most commonly used SPA equation is Pope's (1972) cohort approximation which corrects both the catch and older numbers at age for natural mortality:

$$N_{a,y} = N_{a+1,y+1} e^{M} + C_{a,y} e^{M/2}. \tag{2}$$

The other fundamental relationship of SPA is between a given abundance index at age, $I_{a,y}$, and $N_{a,y}$. We shall assume that they are linear related, as is the most common approach and that the error, ε, is normal with zero mean:

$$I_{a,y} = q_a N_{a,y} * \exp(\varepsilon_{a,y}), \tag{3}$$

where qs are tuning or calibration coefficients and are modelled as functions of age and hence assumed to be invariant in time. Traditionally, these coefficients are also known as catchabilities or gear efficiencies as they
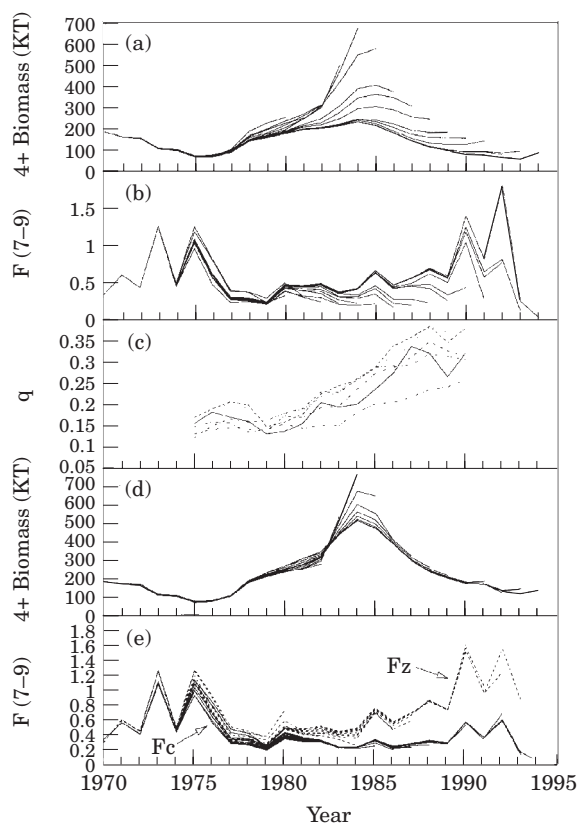

Figure 2. Retrospective analysis for eastern Scotian Shelf cod. Panel (a) is the retrospective pattern for 4+biomass. Panel (b) is the pattern for the average fishing mortality over ages 7–9. Panel (c) displays the estimated calibration factors, q, for the ages 3–8 which are used in tuning the sequential population analysis. Panel (d) is the retrospective pattern of the estimated relative q trend-corrected biomass estimates. Panel (e) shows the retrospective patterns for the two estimates of fishing mortality from the q-corrected population estimates. Fc is derived using catch data [Equation (8)] and Fz is derived from cohort survivorship [Equation (9)].

scale the abundance index to the estimated population numbers. If the catch data, the abundance index, or the model linking them were to change in time, the assumption of a time invariant q would be violated. The q_a may then be estimated with a non-linear least squares (NLLS) procedure (Mohn and Cook, 1993). Some work has been reported in fitting a trend in q within the tuning process. A number of such models were reviewed in ICES (1983) or Armstrong and Cook (1983).

The analysis proceeds by perturbing various aspects of a simulated data set and viewing the effects on retrospectivity. The qs are then estimated using a moving window approach similar to the method described in Parma (1993). It was found that all perturbations which cause a retrospective pattern, also show a time trend in q. If the trend in the estimated qs is not caused by the

Table 1. Summary of perturbations to the simulated data to cause retrospective patterns. The duration of the perturbations are given by the year of onset and duration. The magnitude is the factor by which the unperturbed data are multiplied.

Simulation	Variable	Onset	Duration	Magnitude	Comment
Discarding	Catch	1980	15	0.40	Age dependent
M shift	Natural mortality	1980	15	2.00	Age dependent
Partial recruitment	Fishing mortality	1980	15	0.60	Oldest age
Vessel effect	Abundance index	1980	15	1.20	All ages
Year effect	Abundance index	1980	1	2.00	All ages
Mis-corrected discard*	Catch	1980	15	0.85	Age dependent

*Correcting for discarding induced retrospectivity by erroneously rescaling the abundance index.

changes in abundance estimate, a correction for the q trend can be made, and new numbers and fishing mortalities, F at age generated. If the abundance index is the source, a second approach is proposed, which is an *ad hoc* investigation of the relationship between the index and the SPA-generated numbers at age. The investigation includes an analysis of the residuals from the SPA model and contour surfaces of retrospective metrics and goodness of fit indices as functions of trial perturbations to the data. Finally, the repercussions of incorrectly identifying the source of retrospectivity are assessed.

As well as analysis of simulated data, a case study will be presented using catch and survey data from the ESS cod fishery. Sequential population analyses for this stock have exhibited a strong retrospective pattern. Indeed, because of the magnitude of the retrospective pattern, the SPA of the 1990 fishery was not used as the basis of stock status advice (Fanning and MacEachern, 1991).

Methods

Fishery data

A recent assessment of ESS cod (Fanning *et al.*, 1995) uses catch at age and abundance indices from two surveys for tuning. The catch data and the summer research survey index cover 15 ages and the 25 years from 1970 to 1994. Although a March survey, which began in 1979, is used in the annual assessment of this resource, it will not be used in this study. The exclusion of these data simplifies the analysis and presentation, and does not affect the nature of the retrospective problem seen in ESS cod data or the subsequent analysis. As in Fanning *et al.* (1995), the data are analysed with a standard SPA package which fits terminal Fs and qs for ages 3–8. The software fits Fs at ages 3–8 in the terminal year using a Marquardt algorithm and calibration qs for ages 3–8 algebraically (Mohn and Cook, 1993; Black and Mohn, 1993). The other ages in the terminal year are determined from model assumptions about the pattern of removals at age which is known as

the partial recruitment (PR) pattern. In years before the last year, F on the oldest age is determined as an average of younger ages from cohorts that have already been filled. M is assumed to be 0.2 for all years and ages and it is further assumed that the partial recruitment attains a plateau at age 8. Retrospective analysis will be based on estimates of annual biomass over the ages 4–15 and average F over ages 7–9.

Simulated data

Stock projections are performed to provide a data set for analysis in which various assumptions may be violated in attempts to induce retrospective patterns. The simulations are for 25 years and presumed to run from 1970 to 1994 to match the ESS, cod data period. The simulation begins with a stable age distribution and an assumed sequence of recruitments (the number of youngest aged fish entering the population each year) and a constant fishing mortality. The qs at age for the simulated survey are 0.02, 0.06, 0.08, 0.1, 0.1, 0.1, 0.1 for the 7 ages and the true partial recruitment is 0.05, 0.15, 0.5, 0.8, 1, 1, 1. Catches are determined by the catch equation and exponential survivorship is assumed. The baseline natural mortality is 0.2 for all ages and years. As the simulation has only 7 ages, the variables used for retrospective analysis are the estimated 1+ biomass and the average F over ages 4–6; the true values of which are 3500 and 0.56, respectively. The simulated catch and abundance index data are analysed using the same software as was used for ESS cod. Terminal Fs in 1994 are estimated for ages 2–6 and qs for ages 2–5 in the non-linear fit of the SPA model.

The effects of five perturbations (Table 1), introduced to the simulated data, are examined. With one exception, the perturbations are steps with an onset in 1980 and which last for the duration of the simulation. The first perturbation is one in which the catch data are contaminated by discarding; that is the reported catch is some fraction of the true catch. The discarding is modelled with an onset in 1980 and then continued for

the duration of the simulation. The maximum rate of discarding is 40% on age 1 fish. That is, the reported catch of 1 year olds is 60% of the true catch. The rate of discarding decreases linearly as a function of age to 6% at age 7. The magnitude and age dependency of this perturbation are chosen to approximate the practice in some fisheries of preferentially discarding smaller fish. The SPA model used in subsequent analysis does not explicitly incorporate discarding.

A similar simulation is done in which the natural mortality is shifted in a step perturbation (denoted as M shift in Table 1). Natural mortality is perturbed in 1980, and again is a linear function of age. M at age is perturbed by the addition which ranges from 0.2 for age 1 to 0.03 for age 7. In contrast to the discarding experiment in which the unperturbed catch data could be simply partitioned into reported landings and discards, the stock projection had to be re-run with the perturbed M. This age dependent pattern is chosen to approximate the patterns used in multispecies SPAs which have age dependent natural mortalities (ICES, 1993). In the subsequent analyses M is assumed to be 0.2.

A model mis-specification error is simulated by perturbing the partial recruitment for the oldest age. Such a change might be introduced as the result of a change in fishing practice or the introduction of a new gear. Instead of the true asymptotic value of 1.0 for the oldest age, in 1980 and thereafter the PR on the oldest age is 0.6. Due to constraints imposed in the SPA model, this will affect the starting value for each of the complete cohorts. (Mohn and Cook, 1993) The model used in this analysis assumes that the asymptotic PR applies to all years.

Two types of errors in the abundance index data are simulated. The abundance index is modelled as a survey. The first type of perturbation is denoted as a "vessel effect" in which all the survey abundances are increased by 20% starting in year 1980 and thereafter. This perturbation simulates an uncorrected change to a more efficient research vessel in 1980. The second type is a "year effect" in which the simulated survey abundances at age for only the year 1980 are doubled. Such year effects are often seen in fisheries survey data, especially when the fish population is spatially aggregated and the number of tows in a stratum is low.

Retrospective analysis

The conventional retrospective analysis is a series of SPAs all starting with the same year and progressing for various durations. See for example Figure 2a,b for ESS cod data, Figure 3a,b for simulated data, or for a number of Canadian fishery examples, Sinclair *et al.* (1991). In panels a and b of Figure 2 an SPA is performed with the entire 25 year data series, a second one with the first 24 years of data, the next with the first
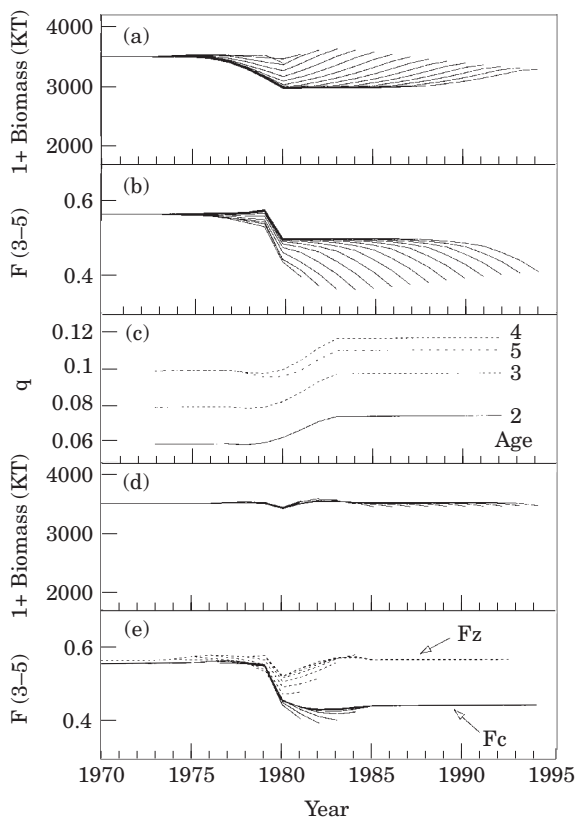

Figure 3. Retrospective analysis of discarding perturbation experiment. The format is the same as for Figure 2 except that biomass and F are 1+ and 3–5, respectively.

23 years etc. The most recent year's data is "peeled off" recursively N_{Peel} times. N_{Peel} for both the real and simulated data sets is 14 years. When the biomass estimates are plotted from the $N_{Peel}+1$ SPAs, a fanning of the successive estimates is seen which is called a retrospective pattern. These figures are often called "spaghetti plots". It will be useful to define a notation to handle data from retrospective analysis. Let $X_{(y1:y2),y}$ denote a stock variable, say 1+ biomass or fully recruited fish mortality, for year y that is estimated from data spanning from year y1 to year y2. Further, we define a metric, ρ, to describe the degree of retrospectivity for a given variable which is the difference between the most recent year's estimate in a data window and that from the entire data set for the same year. The metric is made relative by dividing by the estimate from the entire data series and for convenience will be expressed as a percentage:

$$\rho = \sum_{y} \left[X_{(y1:y),y} - X_{(y1:y2),y} \right] / X_{(y1:y),y}, \qquad (4)$$

where y1 and y2 span the entire set (1970 and 1994 in our data). The metric is the relative vertical distances
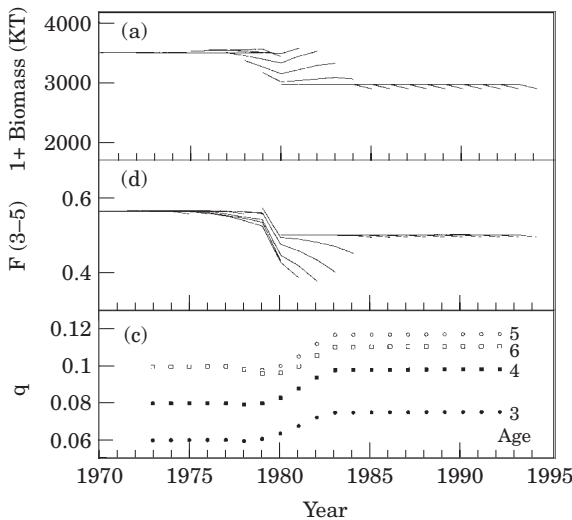
Figure 4. Moving windows estimates of (a) total biomass, (b) average F over ages 3–5, and (c) the estimated q for ages 3–6 from each of the moving windows.

from all the ends in the spaghetti plot (e.g. Figure 2a) to the line from the entire data series. Although measuring a similar quantity, this metric is not the same as the mean square error formulated by Parma (1993, p. 259). The differences being that as well as converting to a logarithmic scale, she accumulated differences from the equal duration window estimates to a reference value whereas we are accumulating the differences in successive retrospective estimates to those from the entire data set. Also, as her metric uses a mean square it cannot discriminate between a random pattern in the spaghetti plot and errors of a consistent sign, i.e. the distinction between error and retrospectivity. As the metric in defined herein is a signed sum, a random pattern will have a low value.

Similar to Parma (1993), we wish to consider the catch and abundance index data in a series of moving windows. For example SPA would be performed on data from years 1 to 10, then 2 to 11, 3 to 12, etc. See Figure 4. This approach provides estimates of the scaling factor for age, a, for each data window, w, as:

$$q'_{(w),a} = GM(I_{a,y \in w})/GM(N_{a,y \in w}),$$ (5)

where GM denotes the geometric mean (Mohn and Cook, 1993). If y_{win}, is the number of years in the moving window, the first year of the moving window is incremented from the first year of data, y1, to $y2 - y_{win}$, then the number of windows n_{win} is:

$$n_{win} = (y2 - y1 - y_{win}) + 2,$$ (6)

and a q estimate is produced for each window and age under consideration, Figure 4c. In summary, the

traditional retrospective analysis has windows of varying duration but fixed first year while the proposed method has windows of fixed duration but with a varying first year.

Analytical considerations when q-trends are available

During the SPA, the investigator chooses a range of ages over which the sum of squares in minimized using the relationship in Equation (3) and qs are explicitly calculated. We denote these as the index ages and for the simulated data they are arbitrarily chosen as 2–5. This range is chosen to reflect the usual practice with fisheries data in which the youngest and oldest observed ages are excluded because the estimates of the catch or abundance index are often less reliable. Once a fit is obtained using a non-linear least squares method (NLLS), qs may also be estimated for the ages outside the range of index ages by applying Equation (5), assuming abundance index data are available. Once qs are available for all ages, a trend in qs is estimated by scaling the qs relative to reference values, q_{ref}, either from a reference year or the mean over a period of years, which we denote as the ERQT, the estimated relative q trend. All the results reported here use the first window as the reference. The SPA numbers in each window [$N_{(w),a,y}$] are then multiplied by the ERQT, which implicitly assumes that the survey is correct and SPA numbers in error. The ERQT corrected numbers at age, $N^q_{a,y}$, may then be estimated in a number of ways. We chose an approach based on the observation that for all but the first and last years there will be multiple estimates of the numbers at age from the various windows passing by. These estimates are averaged after ERQT scaling:

$$N^q_{a,y} = \sum_{w \subset y} N_{(w),a,y \in w} \times q_{(w),a} \div \left(q_{(ref),a} \times \sum_{w \subset y} 1 \right).$$ (7)

Once an ERQT corrected set of numbers at age has been estimated, the fishing mortality can be calculated in two ways. The first method of estimation is to re-arrange the catch equation, with subscripts dropped for clarity, into:

$$Fc = C \times (Fc + M) \div \{N^q \times (1 - e^{-Fc - M})\},$$ (8)

and solve iteratively for Fc. Z is the total mortality rate, M+Fc. Because the catch appears explicitly in this estimation, the result will be called Fc. The second method of estimating F comes from the total mortality along a cohort in one year and is called Fz:

$$F_{Z_{a,y}} = Log(N_{a+1,y+1} + N_{a,y}) - M.$$ (9)

As this method looks at the survivorship from one year to the next, Fz is not estimated for the last year or

oldest age. If q-trend correction is not applied, or if there is no trend in q, Fc and Fz are identical to one another and to the conventional F estimated in an SPA.

Ad hoc diagnostics

Implicit in the ERQT corrections described above is the assumption that the surveys are not the source of the non-stationarity. If they are, q correction to scale the SPA output to the surveys will not be useful. Two *ad hoc* methods are proposed to help in discriminating the source of the non-stationarity. The first diagnostic method is a visual inspection of residuals from fitting Equation (3) to detect unusual patterns, especially year effects in the survey data. As the model being fit assumes a lognormal error, the residuals at each age and year are:

$$\text{Resid}_{a,y} = \text{Log}(I_{a,y} + q_a N_{a,y}). \tag{10}$$

Positive residuals are shown as plus signs with larger symbols depicting larger residuals. Similarly, the dots are scaled to the negative residuals. Residuals are displayed as surfaces over the years and the ages which are used to fit the SPA. These plots are useful for discriminating among year effects, cohort effects, or age effects (vertical, diagonal, or horizontal bands) in the fit of the survey data to the cohort analysis. Anomalous years can be either removed or weighted (e.g. inverse variance) to reduce their influence.

The second *ad hoc* diagnostic tool investigates contour plots in response to a step pertubation applied to the survey data of varying onsets and magnitudes. The purpose of the trial step factors is to see if a simple step can correct the retrospectivity. The most probable cause of a step is an uncorrected, or improperly corrected, change in survey design or execution. However, environmental or fishing practices could be the cause. The dependent variables chosen for contouring are the retrospectivity metric, ρ, and the mean of the squared residuals defined by Equation (10) which is an index of the overall goodness of fit of the NLLS model. The axes are the magnitude of the step factor and the year of its onset. If the plots show a clear minimum, the factor and onset year may be used to scale the abundance index data.

The implications of using the second *ad hoc* technique and improperly ascribing the cause of the retrospective pattern are assessed by a final simulation. A change in discarding is used to cause the pattern but it is "corrected" by assuming that the abundance index is in error, finding the minimum from the *ad hoc* analysis, and re-scaling the survey. This simulation experiment is called Mis-corrected Discarding in Table 1 and subsequently.
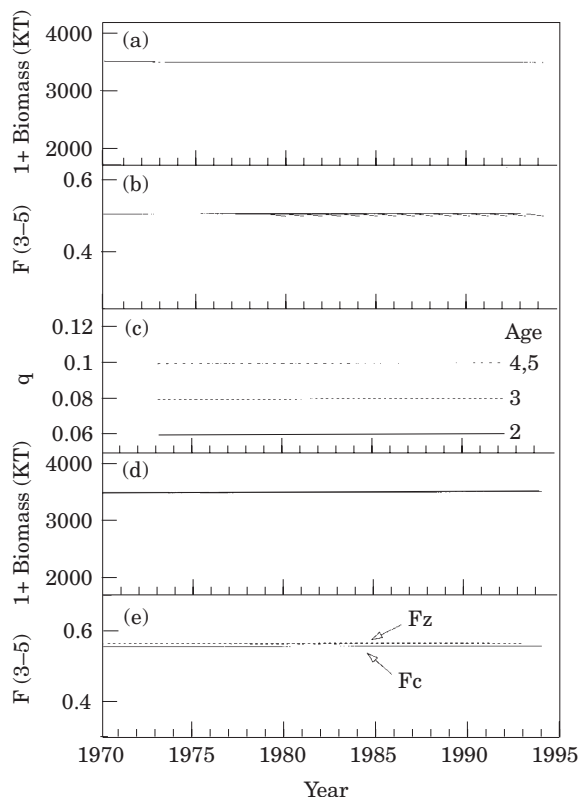

Figure 5. Retrospective analysis of unperturbed simulated data. The format is the same as for Figure 2 except that biomass and F are 1+ and 3–5, respectively.

Results

Figures 3, and 5–9 and Tables 2 and 3 summarize our retrospective analyses in a common format. Each plot has five panels. The upper two are conventional retrospective plots using increasing duration windows of estimated biomass and fishing mortality, respectively. The third panel contains the moving window estimates of q at age. The fourth panel contains retrospectively derived ERQT-corrected biomass. This analysis was performed retrospectively to show how the data discontinuity was handled by ERQT correction. The bottom panel shows retrospective estimates of Fc and Fz. Table 2 compares the accuracy of 1+biomass estimates and the three average Fs over the 25 year simulation to the true values using the mean square residual as an index of accuracy. Table 3 summarizes the retrospective metric for the same biomass and F estimates as in Table 2. Finally, Table 4 contains the retrospective metrics for a TAC based on a fixed fishing mortality target ($F_{0.1}$) and the uncorrected and corrected biomass estimates for each simulation.

Figure 5 contains the results from the Control run in which the simulated data are not perturbed. The Control

Table 2. Mean square residuals between sequential population analyses estimates and true values for uncorrected biomass (B) and F and ERQT-trend corrected biomass (B^q) and Fc and Fz series.

	MSR				
Simulation	B	B^q	F	Fc	Fz
Control	24	27	0.004	0.005	0.003
Discard effect	350	33	0.065	0.094	0.012
M effect	476	66	0.053	0.025	0.067
Partial recruitment	231	30	0.126	0.008	0.017
Vessel effect	166	553	0.031	0.093	0.017
Year effect	47	311	0.016	0.048	0.036
Mis-corrected discard*	413	384	0.047	0.019	0.018

*Correcting for discarding induced retrospectivity by erroneously rescaling the abundance index.

Table 3. Retrospective metric rho (%) for sequential population analyses estimates for biomass (B) and F and ERQT-trend corrected biomass (B^q) and Fc and Fz series.

	%				
Simulation	B	B^q	F	Fc	Fz
Control	0	0	-1.4	0	-0.3
Discard effect	14.7	-1.1	-21.1	-2.1	-2.8
M effect	33.9	7.2	-26.6	-2.4	-2.8
Partial recruitment	3.2	-1.9	-14.9	0.1	2.6
Vessel effect	19.2	1.0	-24.6	-1.9	-3.3
Year effect	5.2	0.0	-3.3	1.7	0.2
Mis-corrected discard*	-1.6	-1.9	-0.2	-0.3	-0.2

*Correcting for discarding induced retrospectivity by erroneously rescaling the abundance index.

is a check of internal consistency in the analysis and to provide an upper level of performance for comparison with the other simulations. The figure shows the expected results, the true underlying population up to the resolution of the models and programs used. The techniques show a slight retrospectivity with an underestimation in the average F in the final year of each decreasing time window (Figure 5b), due primarily to the use of the cohort approximation (Pope, 1972). Also

Table 4. Restrospective metric ρ (%) for TAC estimates using a fixed target fishing mortality (F$_{0.1}$) and either biomass (TAC) or ERQT-trend corrected biomass (TAC^q).

Simulation	TAC	TAC^q
Control	0	0.1
Discard effect	17.7	1.1
M effect	24.4	2.4
Partial recruitment	5.9	-0.4
Vessel effect	19.8	1.2
Year effect	5.0	1.1
Mis-corrected discard*	0.2	0.1

*Correcting for discarding induced retrospectivity by erroneously rescaling the abundance index.

as expected, the qs are constant through time and those for ages 4 and 5 are superimposed. The ERQT correction therefore does not affect the biomass estimate (Figure 5d). The 15 ERQT corrected estimates for the 15 data windows superimpose. There is a slight divergence in the estimates of Fc and Fz (Figure 5e), again due to the use of the cohort approximation (Pope, 1972). Also as expected, the control row of Table 2 shows good agreement between the control estimates and the true values, columns B and F.

Figure 3a,b shows the retrospective patterns resulting from the initiation of age dependent discarding in 1980. Biomass appears to be overestimated in the terminal year of each retrospective data window, while the average F is considerably underestimated. The estimates of q (Figure 3c) show the change as the moving windows pass through the onset of the discarding. The ERQT-corrected biomass in Figure 3d estimates the true biomass very well and shows very little retrospectivity. The ERQT correction reduces the degree of retrospectivity (ρ) by more than an order of magnitude for biomass. More importantly, the mean square residual between the estimates and the true values is also improved by more than a factor of 10 (Table 2, column B vs. B^q) by ERQT correction. The ρ for either Fc or Fz is much better than
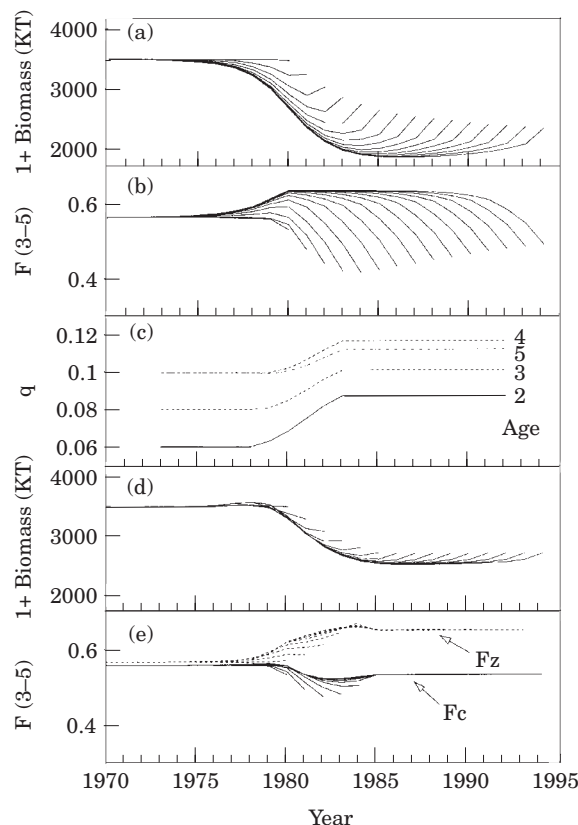
Figure 6. Retrospective analysis of natural mortality perturbation experiment. The format is the same as for Figure 2 except that biomass and F are 1+ and 3–5, respectively.
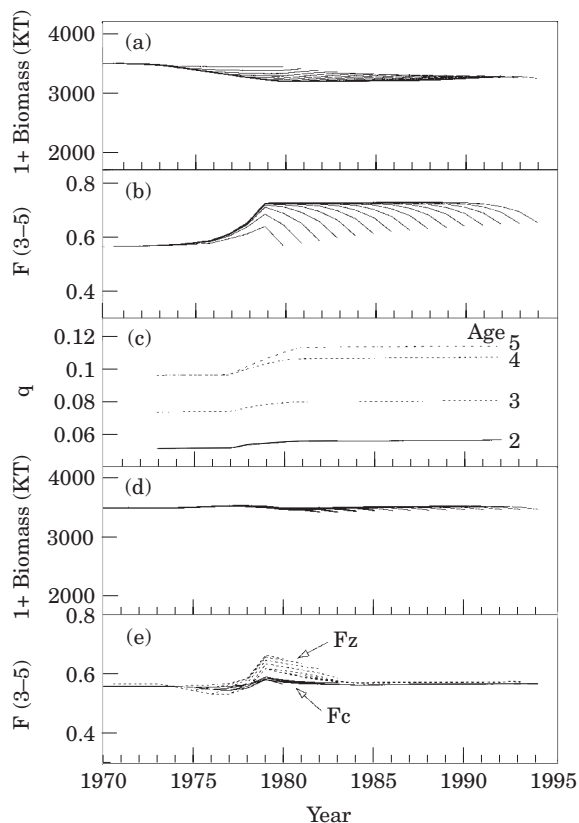

Figure 7. Retrospective analysis of partial recruitment perturbation experiment. The format is the same as for Figure 2 except that biomass and F are 1+ and 3–5, respectively.

the uncorrected F estimate. However, Table 2 shows Fz producing more accurate estimates than the uncorrected F while Fc is worse that the uncorrected values. In Figure 5e the discrepancy between Fc and Fz shows the loss due to discarding in Fc, which is derived from the simulated reported catch. The step down in F in 1980 is also evident in the uncorrected estimate (Figure 3b).

The results of the retrospective analysis of a shift in natural mortality are shown in Figure 6. The effects on the estimated q are quite similar to the discarding experiment and the improvements from q-correction are also of a similar magnitude. The improvement in terms of ρ on biomass is about a factor of 4 and about 10 for average F. Even in the early 1980s the Fc and Fz estimates are quite stable. Because the discrepancy between Fc and Fz is now due to the change in natural mortality, Fc is closer to the true values than Fz, with the uncorrected F being between the two in terms of MSR (Figure 6e or Table 2, 3rd row).

The third experiment is a change in the partial recruitment (Figure 7). The retrospective pattern is not as pronounced as in the previous two simulations,

especially in terms of biomass. The q estimates are less affected than in the previous two simulations, and the effect increases with age. The ERQT-corrected biomass is much closer to the true values (viz. Table 2, MSR = 30 vs. 231). The retrospectivity of the biomass is less than in the previous two simulations and the improvement with ERQT correction is also less. Fz and Fc are improved significantly over F and both are closer to the true values.

A step change in the tuning data (vessel effect) produces a strong retrospective pattern in both biomass and F estimates (Figure 8). ERQT correction essentially removes the retrospectivity (Fig. 8d) from the biomass estimates but moves them away from the true value. Table 2 shows that the mean square residual is more than three times larger for the ERQT-corrected biomass estimate. Fc and Fz display a pattern which is quite similar to the discarding experiment (compare Figures 3e and 8e).

The "year effect" is a one year perturbation to the abundance index data (Figure 9) which is a spike as opposed to the steps used to induce retrospectivity above. It also differs from the above in that the
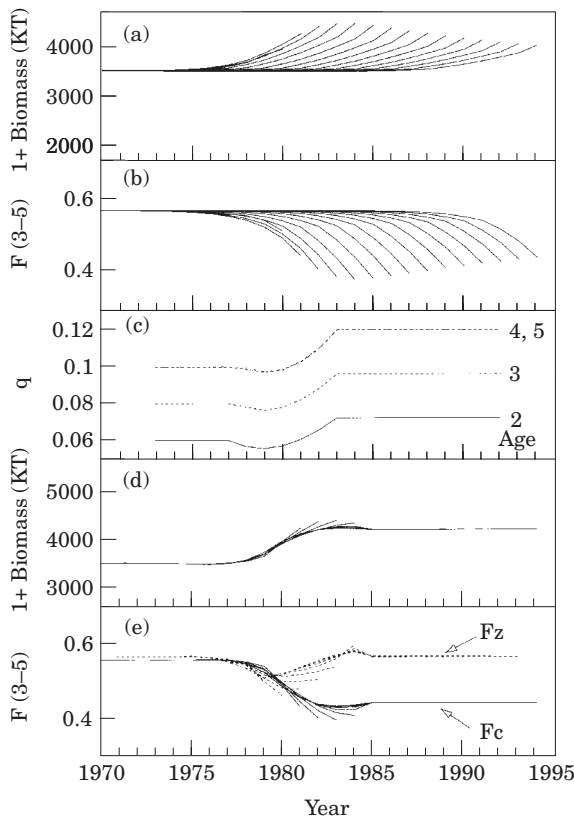
Figure 8. Retrospective analysis of single year perturbation of the abundance index experiment. The format is the same as for Figure 2 except that biomass and F are 1+ and 3–5, respectively.
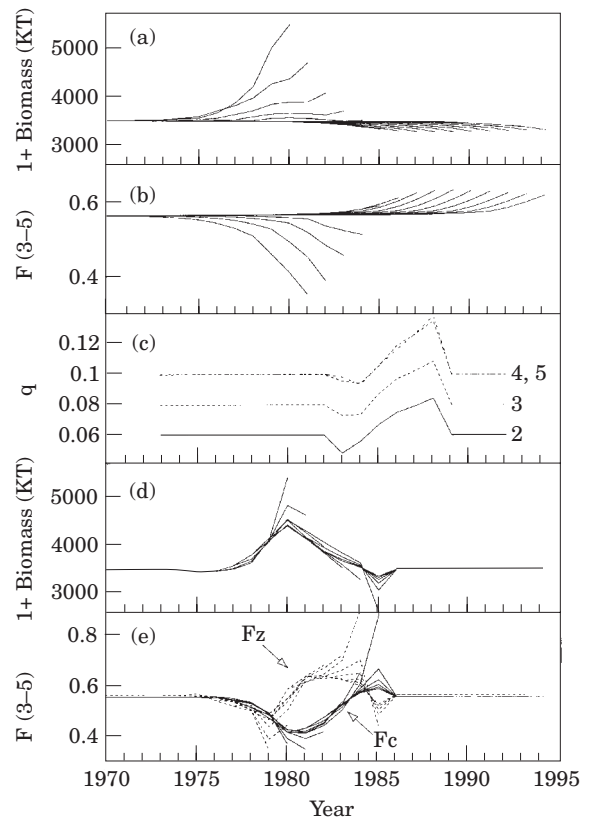

Figure 9. Retrospective analysis of step perturbation of the abundance index experiment. The format is the same as for Figure 2 except that biomass and F are 1+ and 3–5, respectively.

retrospective pattern now has both positive and negative errors. The retrospective pattern for biomass shows a strong overshoot in the early 1980s followed by a period of weaker negative residuals. The F estimates mirror this pattern. The estimates of q produce an undershoot and overshoot pattern as the perturbed survey year affects different portions of the moving windows. ERQT correction improves the retrospectivity of the biomass estimates, but the accuracy of these estimates is seriously degraded (Table 2, MSR = 311 vs. 43). The situation is similar for the F estimates.

The ESS cod data will now be analysed using the same techniques. Figure 2c shows a trend in the catchabilities for ESS cod which is similar for all 6 ages used in tuning. The ERQT-corrected biomass (Figure 2d) has a greatly reduced retrospective pattern but still has a fair degree of overshoot in the mid-1980s. Fc and Fz estimates diverge from one another but neither show much retrospectivity. If the ESS cod survey data are correct, the discrepancy between Fc and Fz could be compared to Figures 3 or 6, suggesting a missing mortality due to discarding or a change in the natural mortality.

ERQT correction improved the accuracy of the biomass estimates in the discarding, M shift and partial recruitment perturbations, but worsened them in the year and vessel effect simulations. An examination of the residual pattern of the SPA may give an insight into when to use ERQT correction. Figure 10 displays the residual pattern from simulated data which have both discarding effect starting in 1980 and a survey year effect in 1985. The effects of the two perturbations are seen as the transition from negative to positive residuals in 1980 and the strong positive column in 1985. The two-dimensional presentation shows that the residuals are not randomly distributed nor do they follow cohorts. A small region of negative residuals is seen in the lower right hand corner which is the result of the NLLS balancing positive and negative residuals down cohorts as it fits the Fs in the most recent year. The residual pattern still does not indicate whether discarding or a vessel effect took place in 1980. Exogenous information, such as knowledge of a change in research survey protocol or perhaps a change in fishery regulations, is required to make this determination.
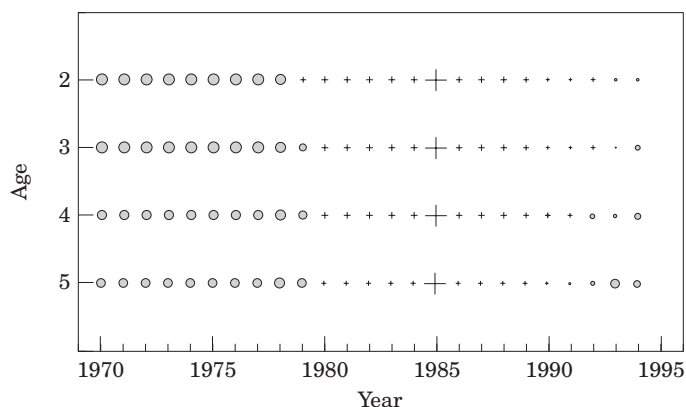
Figure 10. Residual pattern from non-linear least squares fit of combined discard effect beginning in 1980 and single year effect in 1985.
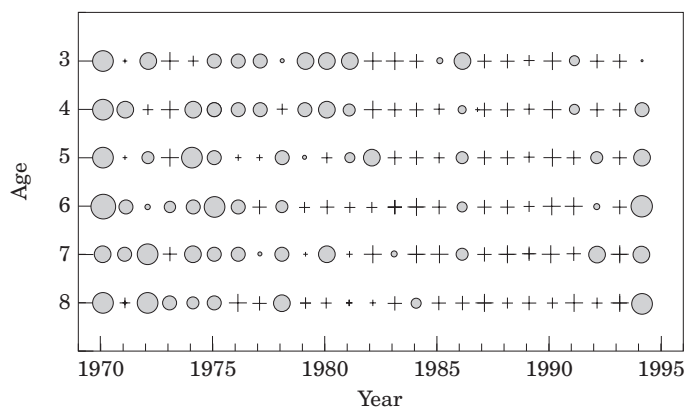

Figure 11. Residual pattern from non-linear least squares fit of eastern Scotian Shelf cod data.

Residuals for the 6 ages used in tuning the ESS cod SPA are shown in Figure 11. The residuals before 1982 tend to be negative while those after 1982, except for 1986 and 1994, tend to be positive. As was seen in the simulated data (Figure 10), the negative residuals in 1994 reflect model constraints within the NLLS minimization which causes the number to be overestimated in the terminal year, except for the youngest age. A year effect is displayed in the 1986 survey as fewer cod are seen over all ages than would be expected when compared to neighbouring residuals.

Figures 12, 13 and 14 are the response surfaces to the *ad hoc* analysis in which trial steps are applied to the survey data for the survey vessel effect, discarding and partial recruitment simulations, respectively. The trial steps ranged in magnitude from 0.7 to 1.1 with onsets ranging from 1976 to 1986. The mean square residual surface (Figure 12a) for the vessel effect simulation data has a minimum in the correct year, 1980, and at the correct magnitude of 0.83, approximately the reciprocal of 1.2 (Table 1). However, the minimum of the ρ surface (Figure 12b) is not well defined and spreads over 8 years

starting with estimates of a correction factor that are too small in 1976. A step correction to the abundance index data is expected to work well because it can completely correct the perturbation which caused the retrospective pattern. Figure 13 shows that the discarding simulation results in very similar surfaces to the vessel effect perturbation; except that the minimum residual is now spread over two years. In practice it would be difficult to discriminate between retrospective patterns caused by shifts in abundance index as opposed to discarding. Figure 14 shows a very different pattern in the mean square residual surface in which there is no clear minimum and a valley spread over the time dimension.

The response surfaces for ESS cod (Figure 15) are roughly similar to Figures 12 and 13. The minimum MSR is centred in 1982 with a step magnitude of about 0.47. Again, the retrospectivity surface is minimized over a number of years, 1978–1983. On the basis of this figure alone, it is not possible to discriminate between discarding or a shift in the abundance index or both as the cause of the retrospective pattern in ESS cod.
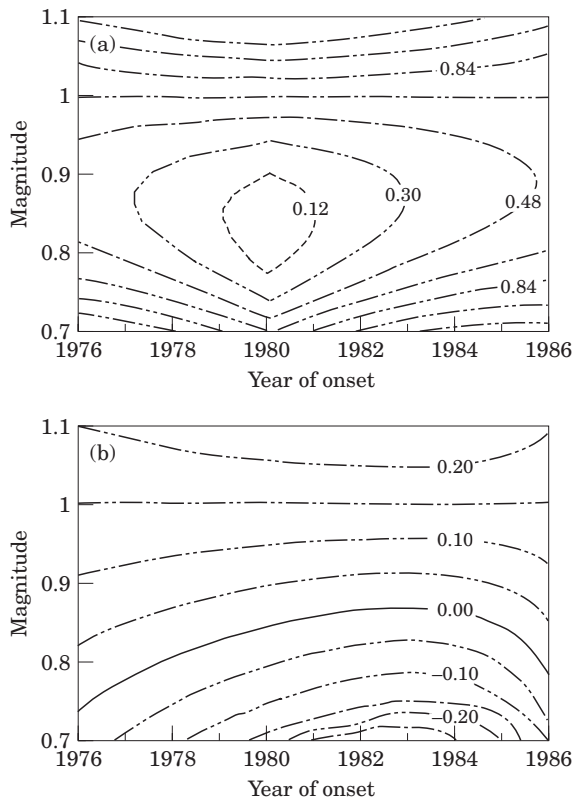
Figure 12. Contour plots of (a) residual sum of squares and (b) retrospectivity index for the vessel effect simulation. The axes are the year of the onset of a trial step perturbation and its magnitude.
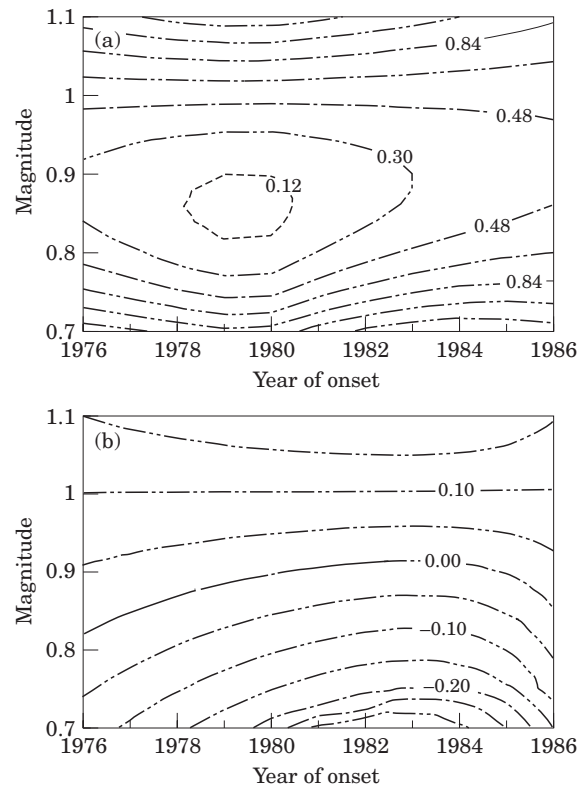
Figure 13. Contour plots of (a) residual sum of squares and (b) retrospectivity index for the discard effect simulation. The axes are the year of the onset of a trial step perturbation and its magnitude.

The location of the minimum in the residual sum of squares surface provides a good correction factor if it is known that the source of retrospectivity is caused by a step perturbation in the tuning data. The effect of erroneously applying such a factor is investigated by taking the minimum from the discarding simulation (Figure 13a) and applying the resultant step of 0.85 in 1980 to the survey data. When the erroneously-scaled data are analysed, Figure 16, it is seen that although the retrospective patterns in biomass and F estimates are removed, the estimates are incorrect. The q estimates are fairly flat so q correction does not affect the biomass estimate. Both Fc and Fz are more sensitive to the q correction and the period after 1980 is better estimated than the uncorrected values, about 0.02 as opposed to 0.05, Table 2 bottom row.

The minimum of the residual sum of squares for ESS cod data, Figure 15a is found when a step is located in 1982 at a magnitude of 0.47. This factor is applied to the summer survey data and a retrospective analysis performed (Figure 17). Figure 17a shows that the retrospec-

tive pattern has been removed from the 4+biomass estimaes: ρ is decreased from 1139 to 129. Although slightly domed, the q estimates are much flatter than the unscaled data; compare Figures 3c and 17c. Fc and Fz are in much closer agreement after scaling the abundance index data. If the retrospective problem in ESS cod were caused by a change in discarding practices in the early 1980s, the q-corrected results from the raw data would be our best estimates of the true time course. If there had been a change in the survey, which produced the abundance indices for tuning, the results after scaling the index by the minimum from Figure 15 would be chosen. In 1982 the summer groundfish survey for ESS cod changed vessels and moved to a larger trawl and then changed vessels again in 1983. Unfortunately, this is the same period that a major change in the fisheries management regime also occurred, the change from a competitive to a "Enterprise Allocation" fishery (Gardner, 1988). However, although introduced in 1982, the transition was protracted over a number of years and unlikely to produce such a sharp minimum as seen in Figure 15a.
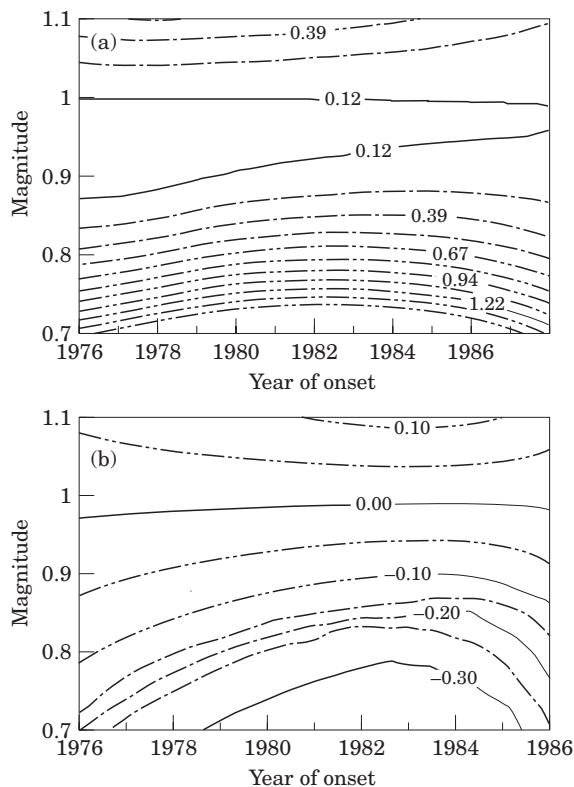
Figure 14. Contour plots of (a) residual sum of squares and (b) retrospectivity index for the partial recruitment effect simulation. The axes are the year of the onset of a trial step perturbation and its magnitude.
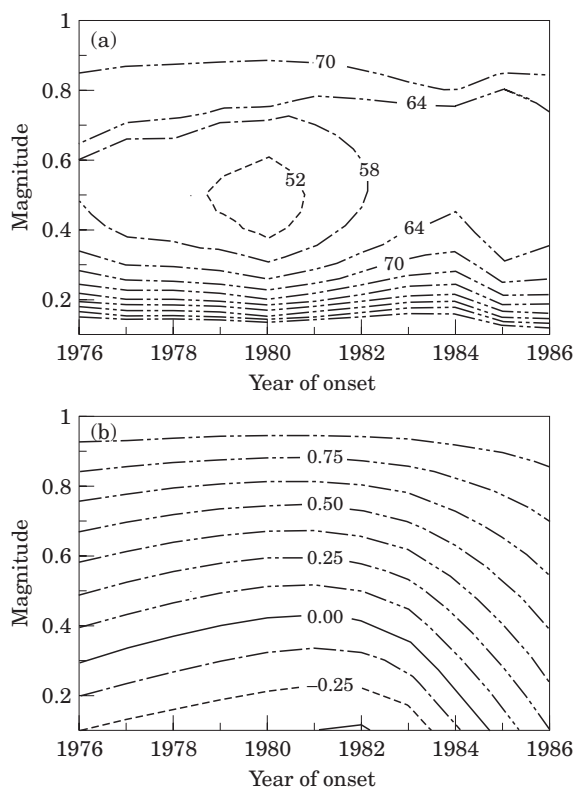
Figure 15. Contour plots of (a) residual sum of squares and (b) retrospectivity index for eastern Scotian Shelf cod. The axes are the year of the onset of a trial step perturbation and its magnitude.

Discussion

It has been demonstrated that changes in catch data, abundance index or the model assumptions can cause a retrospective pattern. The pattern is caused by the failure to meet the assumed constancy of the calibration coefficients that link the abundance index and the reconstructed population from the catch. Moving data windows are used to provide an estimate of the time trend in calibration coefficients which may be used as a diagnostic. If it can be determined that a temporal trend in calibration coefficients is not caused by the abundance index, then an improved estimate of the population trends and fishing mortality can be obtained. In the simulated data for discarding, natural mortality, and partial recruitment induced retrospectivity, the retrospective pattern was removed (Figures 3, 6, and 7) and accurate estimates of biomass were produced (Table 2). The ERQT corrected numbers may then be used to estimate Fc and Fz which, when compared, provide a further diagnostic of potential causes. Even after ERQT correction the true population size may still be not known because scaling for the time trend in the calibration coefficient is relative to a reference year which may

not be well determined. For example, Table 2 shows the results when the q-trend is scaled relative to the first window, with the biomass estimate improving by about a factor of 10. If instead the scaling were done relative to the last window, the temporal pattern is correct but the magnitude is too small and the mean square residual for the biomass increases by 50%. Also such a change in reference year affects the pattern of Fc relative to Fz. In Figure 3, which is scaled to the first window, Fc falls below Fz as a result of discarding. If the q are scaled relative to the last window the pattern is reversed and Fc is greater than Fz before 1980. The observation that Fc exceeded Fz would indicate an inconsistency in the analysis and should trigger a search for an alternative reference for ERQT correction.

Estimating the pattern of calibration coefficients by moving windows has the advantage of not requiring a model for the trend. Some of the earlier work on this problem assumed linear or exponential models. See for example ICES (1983) which reviewed several such models and concluded with a warning about the difficulty in estimating a trend in q without independent data. The moving windows approach allows such a trend to be
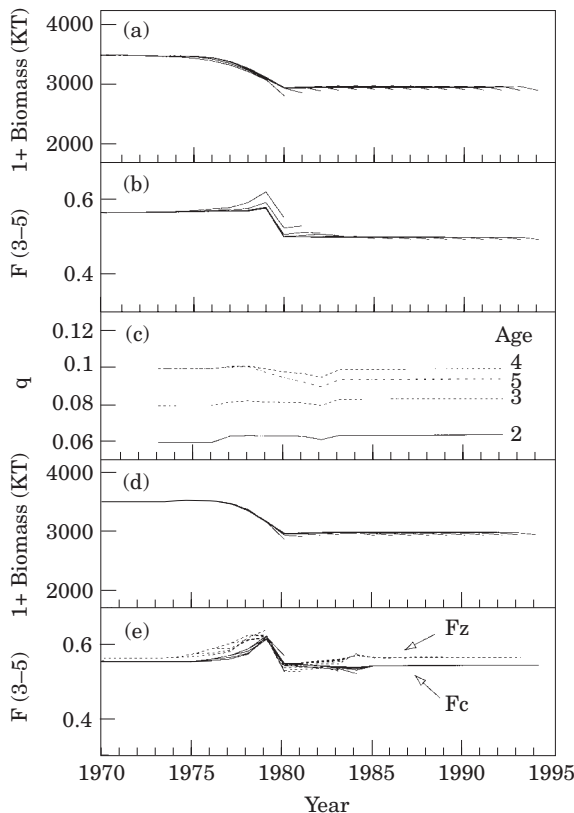
Figure 16. Retrospective analysis of discarding perturbation experiment after scaling the abundance index by the step function which caused a minimum in Figure 13a. The format is the same as for Figure 2 except that biomass and F are 1+ and 3–5, respectively.
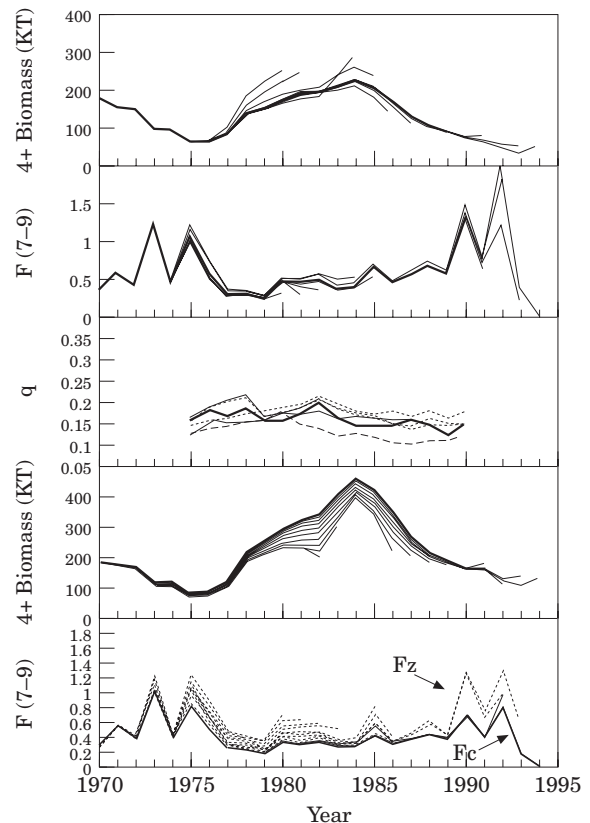
Figure 17. Retrospective analysis of eastern Scotian Shelf cod after scaling the abundance index by the step function which caused a minimum in Figure 15a. The format is the same as for Figure 2.

estimated but cannot discriminate the cause, which is important in deciding whether to correct for the q trend or not.

The moving window approach requires consideration of the best window size. The choice of window size is a compromise between the temporal resolution given by small windows and the need for sufficient data to provide reasonable fits for the SPA within each window. The higher resolution may be of importance if the non-stationarity is seen in the more recent years. However, a rule of thumb derived from observations made during many applications of the technique suggests that time windows of about 1.2–1.5 times the number of ages in the catch data is a good starting value. This may be compared to the 6 year windows used on the simulated data having 7 ages. Because these data are noise-free smaller windows could be used. With noisier data or low exploitation rates, larger data windows would be required. For the ESS cod data, a window duration of 10 years was chosen. A systematic investigation of these trade-offs is recom-

mended when applying the moving window methods to data sets from specific fisheries.

Except for the year effect, our simulations used step discontinuities and did not have either process or sampling noise and thus represent easy signals to detect. For those cases where applicable, the ERQT correction would be expected to perform well for more complicated temporal patterns as it has made no assumptions about the form of the time trend in q. On the other hand, the above *ad hoc* analysis of residual and ρ contour surfaces assumes a step function and may not perform well for other signals. Also, it was observed (Figures 12 and 13) that the *ad hoc* analysis has a very low power of discrimination between a step in the abundance index and discarding, even with noise-free data. The same analysis was applied to the step discontinuity in natural mortality and partial recruitment (results not shown) induced retrospectivity. For the natural mortality case, the contour surfaces were quite similar to those in Figures 12 and 13. The partial recruitment discontinuity (Figure 14) produces very different contours which do

not have a minimum point in the RSS contour but rather are spread across all years with a magnitude of about 0.9. This pattern is sufficiently different that it may be of diagnostic use with real fishery data.

Although the ERQT correction worked well for most of the simulations, it seriously failed with the year-effect data. While ERQT removed the retrospective pattern it also greatly degraded the accuracy of the B and F estimates. As the year affect is easily detected by an examination of residuals, it should be dealt with in another manner, presumably with some sort of down weighting or perhaps removal of the discrepant year(s). The situation is most serious when the inconsistent year is in the most recent portion of the data.

As is seen in the ESS cod fishery data, the retrospective effect can be large and important. Figure 1 shows that the 1981 estimate of the 1981 4+biomass is 1.9 times the estimate made with the entire data series. The degree of overshoot decreased to 10–20% in the mid-80s. The retrospective analysis using consistent data and methods (Figure 2) has overshoots of 110–200% compared to the estimate produced from the entire data series. If a catch quota were derived from standing stock estimates and a target fishing mortality, the quota also would be too large by approximately the same factor. It is interesting to note that the overshoot from the various assessments (Figure 1) and the consistent assessments (Figure 2a) peaked in different years. Resolution of the discrepancy is not undertaken because of the variety of methods and data used. Because of the potential magnitude and importance of retrospective inconsistencies, it is recommended that retrospective analysis be performed on a routine basis.

This analysis suggests a sequence of steps that should be applied to those stocks which display a retrospective pattern. First, a moving window analyses of the temporal pattern of calibration coefficients should be undertaken. Although our simulations used only one source of tuning data for simplicity, the technique is not limited to a single index. If a temporal pattern is found, then it must be determined if the non-stationarity is in the tuning index or not. If it is not, then the ERQT correction method described above works well. The accompanying F_c and F_z estimates should also be calculated; their comparison may give insight into discarding practices and unmodelled natural mortality trends. If the tuning index is the source of the problem the solution is less clear. If it is caused by one or more bad years, then a residual analysis (Figures 10 and 11) should aid identification and the data could be removed or down-weighted. If the cause is a step as might result from change in survey gear, then the *ad hoc* step search (Figures 12–15) is warranted. If the abundance index were affected by a slow trend, say caused by a shift in the resource by environmental warming, it will be very

difficult to discriminate it from a gradual change in fishing practice. The *ad hoc* methods are not meant to be a substitute for a careful analysis of residuals, but rather a potential adjunct. Indeed, the year effect simulation would be obvious from a simple examination of residuals.

The most commonly used techniques for handling retrospective patterns tend to be *ad hoc* also. Often the data are to truncated or split until the pattern is removed or reduced. Our moving window analysis should help in defining appropriate splitting points if this approach is desired. Also, the methods used in many ICES assessments employ a arbitrary parameter to induce shrinkage (see Mohn and Cook, 1993 for more details) in which the most recent estimates are weighted with the recent mean. This technique does reduce the retrospective pattern but does not yield any insights as to the temporal pattern or potential cause.

As expected, the degree of retrospectivity seen in the TAC estimates in Table 4 show a similar pattern to the biomass estimates seen in Table 3. This is because the target fishing mortality is fixed and the TACs are proportional to the fishable biomass (the product of a biomass at age and the partial recruitment pattern). If the TAC were proportional to the recent fishing mortality, a so-called *status quo* TAC, the effect of ERQT correction would be much less, as the errors in the standing stock biomass would be balanced to a large degree by compensatory errors in the estimated recent F. See for example panels a and b of Figure 2.

As is seen in the ESS cod fishery data, it is not possible to ascribe a unique cause for the temporal pattern in qs from a moving window analysis alone. External data or information are needed. It is unfortunate that in the case of ESS cod the research vessel which produces the tuning index data was changed in the same year that a major change was initiated to the management regime. The temporally well-defined minimum in Figure 15a suggests a single event and therefore favours the survey as the principal source of the retrospective pattern. Although Enterprise Allocations were introduced in 1982, they were not fully adopted until 1985 at which time they accounted for 60% of the total catch. Fishing practices also changed in the early 1990s as the resource contracted. Therefore, on balance, the well-defined minimum in the step function search is more likely to have resulted from the changes in research vessel. After the tuning data are scaled by the minimum in Figure 15a there is still a small dome seen in the pattern of calibration coefficients (Figure 17c) suggesting that other factors are also non-stationary. One possibility is that natural mortality could have been affected by an environmental change or by changes in predation, notably an increase in seals (Mohn and Bowen 1996). And of course, a combination of effects, rather than a single cause, may have occurred.

The change in underlying partial recruitment experiment (Figure 7) represents a serious problem in SPAs. Partial recruitment is used in ours, and many, versions of SPA to assign a fish mortality, and hence population size, to the oldest age in a cohort (Mohn and Cook, 1993). As Sinclair *et al.* (1991) observed, the assumed shape of the partial recruitment pattern affects the degree of retrospectivity and such patterns are difficult to determine from fisheries' data. Our simulation, which assumed a constant asymptotic pattern for the SPA, is used as an example of model mis-specification. It is conceivable, though less likely, that an investigator might erroneously impose a change in partial recruitment because of a change in fishing practice. Either way, a change is difficult to detect but may have considerable impact on retrospectivity and overall accuracy of an SPA.

The discarding simulation (Figure 3) is very similar to a discarding simulation reported in Sinclair *et al.* (1991). From their Figure 9, which is similar to the upper-most panel of Figure 3, they made the observation that the terminus of each estimation series is closer to the truth and then moved away to a less correct lower level. From this the conclusion is drawn that "estimates of population size from the converged part of the SPA do not necessarily represent the true population size for those years". While this is true, it is incomplete and is potentially dangerous. They did not include the accompanying pattern of the estimated Fs as we do in Figure 3b. This figure shows that although the population size is better estimated at the ends of each series, the F estimates are worse that the "converged" region and are even further from the true values. Basing the stock status on the contemporaneous estimate would lead to an overly optimistic view of exploitation and could in turn lead to setting catch levels too high.

Mohn (1993) analysed the ESS cod data for retrospective patterns using bootstrap techniques. He reported that when the residuals [Equation (10)] were resampled to produce bootstrapped abundance index data, the mean of the estimates did not display a retrospective pattern nor were the bootstrap estimates centered on the input data. No explanation was given for these observations. Figure 11 shows that the residuals for ESS cod are not randomly distributed as is assumed in the bootstrap analysis. The present study shows that a temporal discontinuity is an important factor in causing a retrospective pattern. Thus, the bootstrapping broke the temporal pattern and removed the retrospectivity from the ESS data.

Recently, Evans (1996) derived a metric, essentially a ration of two estimates of population numbers at a given age and year, which is found to mirror retrospectivity in fisheries data. As have we, he observes that the cause can be either problems in the number of deaths (natural mortality or discarding) or in the survey indices. His index is related to the inverse of the ERQT factor we derive above. Evans uses un-tuned SPAs to estimate his retrospective metric which has the advantage of being more easily performed than the conventional series of SPAs. However, this might be accomplished more directly by using Pope's cohort approximation and trial starting values for each cohort and dividing the resultant numbers at age into the survey index at age. The resultant q(y,a)s could then be normalized by a reference year or cohort to directly show relative q over the data period. Of course, the un-tuned numbers at age would be particularly weak in the most recent years and oldest ages, but general patterns should be evident.

Acknowledgements

References

Armstrong, D. W., and Cook, R. M. 1983. A proposal for the use of catch per unit effort data to assess terminal F in VPA. ICES CM 1983/D:15, 7 pp (mimeo).

Black, G. A. P., and Mohn, R. 1993. Interactive Data Visualization of Fisheries Management Systems. ICES CM 1993/D:6, 17 pp.

Butterworth, D. S. 1981. The value of catch-statistics-based management techniques for heavily fished pelagic stocks, with special reference to the recent decline of the Southwest African pilchard stock. Proceedings NATO Conference Applied Operations Research in Fishing, Trondheim (August 1979) NATO Conference Series II Vol. 10: 441–464.

Evans, G. T. 1996. Using the elementary operations of sequential population analysis to display problems in catch or survey data. Canadian Journal of Fisheries and Aquatic Science, 53: 239–243.

Fanning, L. P., and MacEachern, W. J. 1990. Stock status of 4VsW cod in 1989 using a half-year SPA formulation. CAFSAC Research Document 90/88. 52 pp.

Fanning, L. P., and MacEachern, W. J. 1991. Assessment of 4VsW cod in 1990. CAFSAC Research Document 91/44. 42 pp.

Fanning, L. P., Mohn, R. K., and MacEachern, W. J. 1995. Assessment of 4VsW cod in 1994 with consideration of ecological indicators of stock status. DFO Atlantic Fisheries Research Document 95/73. 29 pp.

Gagné, J. A., Sinclair, A. F., and Currie, L. 1983. Status of the 4VsW cod complex. CAFSAC Research Document 83/56. 56 pp.

Gagné, J. A., Sinclair, A. F., and Dale, C. 1984. The 1984 assessment of 4VsW cod: a completely revised procedure. CAFSAC Research Document 84/78. 60 pp.

Gardner, M. 1988. Enterprise allocation system in the offshore groundfish sector in Atlantic Canada. Marine Resource Ecosystems, 5: 389–414.

ICES 1983. Reports of the *ad hoc* working group on the use of effort data in assessments and of the Report of the Working Group on Methods of Fish Stock Assessment. ICES Co-operative Research Report 129. 134 pp.

ICES 1991. Report of the Working Group on Methods of Fish Stock Assessment. ICES CM 1991/Assess:25, 147 pp (mimeo).

ICES 1993. The Working Group Report on the assessment of demersal stocks in the North Sea and Skagerrak. ICES CM 1993/Assess:5.

Maguire, J. J., Young, G. A., and Sinclair, A. F. 1982. The 1982 assessment of the Eastern Scotian Shelf (4VsW) cod stock complex. CAFSAC Research Document 92/40. 43 pp.

Mohn, R. K. 1993. Bootstrap estimates of ADAPT parameters, their projection in risk analysis and the retrospective patterns. *In* Risk Evaluation and Biological Reference Points for Fisheries Management, pp. 173–184. Ed. by S. J. Smith, J. J. Hunt, and D. Rivard. Canadian Special Publication Fisheries and Aquatic Science 120.

Mohn, R., and Bowen, W. D. 1996. Grey seal predation on the eastern Scotian Shelf: modelling the impact on Atlantic cod. Canadian Journal of Fisheries and Aquatic Science, 53: 2722–2738.

Mohn, R., and Cook, R. 1993. Introduction to Sequential Population Analysis. Northwest Atlantic Fisheries Organisation Science Council Study No 17. 110 pp.

Mohn, R., and MacEachern, W. D. 1992. Assessment of 4VsW cod in 1991. CAFSAC Research Document 92/54. 37 pp.

Parma, A. N. 1993. Retrospective catch-at-age analysis of pacific halibut: implications on assessment of harvesting policies. *In* Proceedings of the International Symposium on Management Strategies for Exploited Fish Populations, pp. 247–265. Ed. by G. Kruse, D. M. Eggers, C. Pautzke, R. J. Marasco, and T. J. Quinn II. Alaska Sea Grant College Program.

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

Sinclair, A., and Annand, C. 1986. Assessment of the 4VsW cod management unit following the 1985 fishery. CAFSAC Research Document 86/46. 45 pp.

Sinclair, A., and Gavaris, S. 1985. Sequential population analysis of 4VsW cod following the 1984 fishery. CAFSAC Research Document 85/4. 28 pp.

Sinclair, A., and Smith, S. J. 1987. Assessment of 4VsW cod. CAFSAC Research Document 87/72. 62 pp.

Sinclair, A., Gascon, D., O'Boyle, R., Rivard, D., and Gavaris, S. 1991. Consistency of some northwest Atlantic groundfish stock assessments. NAFO Scientific Council Studies, 16: 59–77.