

Report of the Retrospective Working Group

A Working Paper in Support of Term of Reference 4

This information is distributed solely for the purpose of pre-dissemination peer review. It has not been formally disseminated by NOAA. It does not represent any final agency determination or policy.

GARM 2008 Methods Meeting
Woods Hole, MA
25-29 February 2008

Table of Contents	
Executive Summary	3
Introduction.....	4
Measuring Retrospective Patterns.....	5
Mohn rho.....	5
Woods Hole rho	5
Sufficiency of Measurement.....	5
Causes of Retrospective Patterns	6
Catch	7
Natural Mortality	7
Survey Catchability.....	7
Random Noise.....	8
Closed Areas	8
Pulse vs Shift.....	9
Implications of Retrospective Patterns in Stock Assessment	9
Identifying the Source of a Retrospective Pattern	11
Local Influence Surface	11
Moving Window	11
q Surface	12
Retrotension	12
Fixing Retrospective Patterns	12
Data or Model Modifications.....	13
Local Influence Surface	13
Split Series	13
Adjust Population Estimates for Projections	14
Alternative States of Nature.....	15
Discussion.....	15
Conclusions.....	16
Recommendations.....	17
Literature Cited.....	18
Appendix: List of Working Group Participants.....	34

Executive Summary

This report summarizes a wide range of work related to retrospective patterns in stock assessment, culminating in conclusions and recommendations. A retrospective pattern is a systematic inconsistency among a series of estimates of population size, or related assessment variables, based on increasing periods of data (Mohn 1999). This pattern of change in estimated values can have severe consequences for management of a stock, potentially resulting in depletion of a stock even though the assessments indicate the targets are being met. Retrospective patterns have been observed in some but not all of the stocks in New England, as well as other stocks around the world. Retrospective patterns are not limited to virtual population analysis, having been observed in a wide range of models including statistical catch-at-age models. Instead retrospective patterns are an indication something is inconsistent in the data or model assumptions. However, retrospective patterns are just one diagnostic for stock assessments and lack of a retrospective pattern does not necessarily imply that all is well.

Simulation analyses have demonstrated a number of sources for retrospective patterns, including, missing catch, an increase in natural mortality rate, or a change in survey catchability. The working group examined a number of potential methods to determine the source of a retrospective pattern using simulated data, but was unable to do so. However, the working group found it does appear possible to identify the timing of an intervention which leads to the retrospective pattern in some cases. Similarly, a number of methods were examined to fix retrospective patterns. While the fixes did in fact remove the retrospective pattern, the new assessment was not always closer to the truth than the original assessment, even though the diagnostics of the new model were good. This means that caution must be exercised when applying any fix to an actual assessment to remove the retrospective pattern.

The working group recommends that stock assessment scientists always check for the presence of a retrospective pattern and that a strong retrospective pattern is grounds to reject the assessment model as an indication of stock status or the basis for management advice. The working group also recommended future research to be conducted on the topic to define objective criteria for acceptance of an assessment with retrospective patterns and to determine what type and level of adjustment in management advice is appropriate through management strategy evaluations.

Introduction

This report summarizes a wide range of work related to retrospective patterns in stock assessment. A number of participants in the working group (see Appendix) have conducted analyses over the past two years culminating in a working group meeting held in Woods Hole 14-16 January 2008. This report presents summaries of many of the analyses, along with conclusions and recommendations agreed upon at the January 2008 working group meeting.

One of the first definitions of a retrospective pattern was given in Mohn (1999) as

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.

The problem had been identified previously, for example in the southwest African pilchard stock (Butterworth 1981), Canadian groundfish (Sinclair et al. 1991), Pacific halibut (Parma 1993). Additionally, the ICES Working Group on Stock Assessment Methods has addressed the topic of retrospective patterns since 1991. More recently, the last GARM assessment in 2005 identified a number of New England groundfish stocks which exhibited retrospective patterns. These retrospective patterns have typically persisted for many years, although there are cases when the direction of the pattern has changed suddenly (e.g. Georges Bank cod and summer flounder). The patterns have caused difficulty for the management of these stocks. Although most work with retrospective patterns has employed virtual population analysis (VPA) as the stock assessment model, retrospective patterns have been observed in a wide range of models, including statistical catch-at-age models. In many stock assessments where statistical catch-at-age models are utilized, retrospective analyses are not conducted.

There are two types of retrospective patterns: historical and within model. The historical retrospective analysis is conducted by examining the results of each final assessment for a number of years in a row and determining whether there was a consistent pattern of overestimating or underestimating assessment values in successive years. This type of retrospective pattern can be caused by changes in the data, type of assessment model, or assessment model formulation and is most important to managers because it relates directly to the management choices made in the past based on the information available at the time. In contrast, the within model retrospective analysis uses the same data, type of assessment model, and assessment model formulation and trims the most recent year's data in successive model runs. The within model retrospective patterns are most useful for determining an internal inconsistency in the data because the only changes in the different runs are the number of years of data in the model. This document focuses almost entirely on the within model retrospective problem.

The Retrospective Working Group was formed to examine hypotheses regarding the causes of retrospective patterns and the potential to fix the problem. The working group

focused effort on simulation approaches where known sources caused a retrospective pattern instead of actual assessments where there may be multiple true causes and their identification may never be possible. Although the working group did not find “the answer” it did learn a great deal about the issue and arrived at a number of conclusions and recommendations that may be helpful to scientists and managers.

Measuring Retrospective Patterns

Mohn rho

The rho statistic of Mohn (1999) has been commonly used to measure the retrospective pattern. It is defined as the sum of relative difference between an estimated quantity from an assessment with a reduced time series and the same quantity estimated from the full time series:

$$\rho = \sum_y \frac{X_{y,tip} - X_{y,ref}}{X_{y,ref}} \quad (1)$$

where X denotes some variable from the stock assessment such as F or SSB, y denotes year, tip denotes the terminal estimate from an assessment with a reduced time series, and ref denotes the assessment using the full time series (Figure 1). This calculation will be zero when the peeled assessments (meaning years are removed one at a time back from the most recent year) match exactly with the full time series assessment, or when the differences between the peeled assessments and full time series assessment are balanced both positive and negative. The former case has no change from year to year, while the latter case would be characterized as exhibiting noise but not a retrospective pattern. The Mohn rho will become large, either positive or negative, when there is a consistent pattern of change in the peeled assessments relative to the full time series assessment. Although it is a relative measure, there have not been rules of thumb developed regarding how large in absolute value Mohn rho must be before an assessment is declared to exhibit a retrospective pattern.

Woods Hole rho

The Woods Hole rho is a slight modification of Mohn rho which computes the sum over all years in each peeled assessment instead of just the terminal year (Figure 2). Although slightly different numeric values result from a given dataset, differences between the Mohn rho and Woods Hole rho were not found to be important in any analyses.

The two rho calculations are used interchangeably throughout the rest of this document and referred to generically as the retrospective statistic rho.

Sufficiency of Measurement

As mentioned above, there is currently not an accepted level of rho beyond which an assessment is deemed to exhibit a retrospective pattern. One approach to address this point is to conduct a simulation exercise specifically designed to mimic the level of

uncertainty in the assessment data and see how often a pattern as large as that seen in the real assessment arises. This approach is more fully described below in the section on random noise, but is quite labor intensive and must be done for each model formulation of a specific assessment.

Instead, the typical approach is to look at the plots and make a subjective decision based on the number of years which deviate from the full time series assessment in the same direction. For large retrospective patterns, this is obvious and easy to do (Figure 3). However, for noisy stock assessments, the distinction between uncertainty and retrospective pattern can be difficult to make. The eye can also be fooled with respect to the presence or absence of a retrospective pattern, especially when the recent years have a steep slope in the plot (Figure 4). So care must be taken when determining whether a retrospective pattern is present or not.

A strong retrospective pattern can be defined by the lack of overlap in the distributions about the point estimates in the retrospective plots. For example, when the Georges Bank yellowtail flounder Base Case assessment had each peel of the retrospective bootstrapped, the distributions for fishing mortality do not overlap when more than two years of peels are compared (Figure 5). Any form of uncertainty estimation, such as bootstrap, MCMC, asymptotic variance, etc., can be used to make the comparison. A strong retrospective pattern causes change in a population estimate beyond what would have been expected due to the uncertainty in the assessment itself.

One measure of the sufficiency of the retrospective statistic ρ was examined by creating a dataset with no bias and then introducing a bias that causes retrospective patterns at different levels. As the level of introduced bias increased, the ρ became larger in absolute value for a limited number of cases examined (e.g. Figure 6). This monotonically increasing relationship demonstrates, but does not prove, that larger interventions causing retrospective patterns will result in larger absolute values of ρ .

Causes of Retrospective Patterns

Retrospective patterns can be caused by a number of factors, but all require a change in parameter value or assumed model value over time. For example, using a wrong constant natural mortality rate in an assessment will not produce a retrospective pattern, instead the results will simply be scaled up or down relative to the true values. The three main causes of retrospective patterns have been known for many years: changes in the level of catch accounted for in the assessment, changes in the natural mortality rate (M), or changes in the survey catchability (q). The working group focused much of the simulation analyses on these three sources. The ability of random noise or the implementation of a closed area to produce retrospective patterns was also addressed by working group members. Comparisons were made between retrospective patterns caused by pulse interventions, which occur in only one or very few years, and step interventions, when a change occurs and is maintained for many years.

Catch

A change in the proportion of actual catch input to the stock assessment model will generate a retrospective pattern. The direction and type of retrospective pattern is determined by when the correct and incorrect catches occur during the time series and whether the incorrect catches are greater or less than the actual values. Catch is most often assumed to be underestimated, due to illegal or unreported landings or else due to underestimation of discards. If early in the time series the catch was correct and then underestimation occurs in recent years, the retrospective pattern will decrease SSB estimates and increase the F estimates as more years of data are added, the most common retrospective pattern observed in actual assessments (Figure 7 left panels). If instead, the recent catch is estimated well, but the early catch is underestimated, the direction of the retrospective pattern changes and the SSB estimates in the early years are biased (Figure 7 right panels). Overestimation of catch will reverse the direction of the retrospective pattern and bias (results not shown).

Natural Mortality

Since natural mortality is an input parameter to most stock assessment models, there are four possible combinations of misspecification that can lead to retrospective patterns: the true M can increase or decrease and the M assumed in the stock assessment model can match either the early or recent period (Figures 8a-b). The directionality of the retrospective pattern is determined by the true M change; increasing M will decrease the SSB estimate as more years are added, while decreasing M will increase the SSB estimate as more years of data are added. The early part of the assessment will be unbiased when the stock assessment M matches the early M, but will be biased when the stock assessment M matches the recent M. In actual assessments the true value of M is unknown, and thus the direction of retrospective bias alone cannot be used to infer if any part of the time series is unbiased.

Survey Catchability

Tuning indices are utilized in most stock assessment models to provide information regarding the trends in stock abundance. These can come from fishery independent surveys, for example the NEFSC bottom trawl survey, or from fishery dependent catch per unit effort (CPUE) time series (hereafter referred to generically as surveys). In either case, stock assessment models commonly assume that the catchability of an index does not change over time. If the true catchability changes during the assessment period, a retrospective pattern will be formed. Increasing the catchability in recent years causes SSB estimates to decrease as more years are added while decreasing the catchability in recent years causes the SSB estimates to increase as more years of data are added (Figure 9). The catchability value described here is the scalar that relates the index to the modeled population, not the ability of the survey gear to retain fish that are encountered. The model catchability can change, for example if the proportion of the stock surveyed changes, even though the gear catchability remains constant.

Random Noise

A large simulation study was conducted to examine the ability of random noise to generate retrospective patterns. Simulated populations were created based on fluke-like biology and fishery characteristics. Random noise was added to each simulated dataset at levels that correspond to the estimated levels of uncertainty in catch at age and survey indices. The random noise was completely white, there were no correlations among any of the random deviates applied to the indices, while catch was simulated following a sub-sampling scheme that mimics the actual data collection for this species. The simulations found that while retrospective patterns could be produced, they were quite minor in appearance, especially relative to the strong retrospective pattern observed in the fluke assessment. For example, when the simulated results were ordered according to their retrospective statistic value, the 5th percentile had only a minor retrospective pattern (Figure 10). Even when the level of noise added to the data was increased in the simulations, datasets with retrospective patterns were rare and none were strong. This led the working group to conclude that interventions (correlated errors) are more likely to cause retrospective patterns than random noise.

It was suggested that this simulation approach could be used to measure the distribution of the rho statistic under the null hypothesis of only white noise. If the actual assessment had a rho outside this distribution, then it would be concluded that an intervention had occurred to generate the observed retrospective pattern. While an appealing concept, it requires a large number of simulations to generate the distribution of the rho statistic under the conditions corresponding to a particular assessment, and so may not be practical in many cases. If this approach is pursued in the future, characteristics of the rho distribution could be developed and compared between situations with white noise and interventions.

Closed Areas

Simulations have demonstrated that implementing a closed area during the assessment time series can create retrospective patterns for sessile organisms (Figure 11). In these simulations, the stock was assumed to be fully available to the fishery for the first half of the time series. Half way through the time series, a closed area was implemented and organisms in the closed area remained in the closed area while those in the open area remained in the open area, as would happen for sessile organisms. Recruitment was divided between the two areas proportionally. Fishing effort could be either shifted to the open area or removed from the simulations. Simulated surveys estimated the stock abundance from all areas while catches were taken only from the open areas. Due to differential mortality in the two areas, the population increased rapidly in the closed area relative to the open area. Since the fishery was effectively operating on only a portion of the total stock, while the surveys were indexing the total stock, a retrospective pattern emerged when the data were assessed in a VPA which assumed one homogenous area for the stock. Further investigation of this effect with simulated organisms that could move between areas reduced or eliminated the level of retrospective patterns, depending on the exact configuration. This topic remains an area for future research.

Pulse vs Shift

Regardless of the source of the retrospective pattern, there remains two ways that an intervention can occur: a pulse event of one or very few years or a shift where a change happens and becomes fixed into the future. These two types of interventions produce different retrospective patterns. A retrospective pattern generated by a pulse shows a bias in one direction followed by a flip to a bias in the other direction (e.g. Figure 1). In contrast, a retrospective pattern generated by a shift shows a bias in only one direction (e.g. Figure 7). Random variability in the data may cause detection of a pulse event to be hampered, while a shift event produces a much longer lasting and generally stronger retrospective pattern even under additional variability in the data. In actual assessments, both pulses and shifts may occur for different sources, making detection extremely difficult.

Implications of Retrospective Patterns in Stock Assessment

If retrospective patterns were merely an interesting statistical oddity it is unlikely so many people would have spent so much time studying this topic, both within and outside this working group. However, there are severe implications for both managers and the stock itself when stock assessments exhibit strong retrospective patterns. Management advice will be biased and could lead to continued overfishing of the stock, inability to achieve rebuilding targets, and loss of potential yield. These problems have been demonstrated in a number of actual stocks, notably Georges Bank yellowtail flounder.

As an example, consider an initial stock assessment for a species occurring in 2001 based on data through 2000. Let some mechanism occur such that the stock assessment, and all successive stock assessment, contains a retrospective pattern which reduces estimated stock size as additional years of data are added. The 2001 assessment will overestimate stock size at the start of 2001. Managers need advice for 2002, based on knowledge of the catch that will occur in 2001, an assumed recruitment level in 2001 and 2002, and a desired fishing mortality rate in 2002. In the projections, when the catch in 2001 is removed from the overestimated stock, the resulting F will be underestimated and the stock size in 2002 overestimated. Application of the management F in 2002 will result in a catch that is too high relative to the true stock abundance. This can be compounded if the management F depends on the underestimated 2000 F or the overestimated 2000 stock size, as many control rules do.

Jumping ahead to 2003 in this example, the 2003 stock assessment uses data through 2002, the year which had a quota set based on the management F from the 2001 assessment. Assuming the survey indices are accurate, the 2002 survey observations should indicate a relative decline in the stock compared to the projections because the projections were made with overestimated stock sizes. However, due to the mechanism causing the retrospective pattern, the estimated 2002 stock size may be equal to the projected value, even though the estimated 2001 stock size is less than the original

estimate. If the 2002 quota was caught exactly, then the actual F in 2002 will be greater than the management F used to set the quota. Due to the overestimation of stock size, the estimated F could match the management value, though. In this example recruitment variability is not considered for ease of understanding. Thus, the conclusion from the 2003 assessment is that the 2001 advice was appropriate and the management targets have been achieved although in reality the 2002 F was too high and stock size was actually lower than the 2003 estimate. As future assessments are conducted, the estimates of 2002 stock size will decrease and F will increase as the retrospective pattern is exhibited for year 2002. This can lead to a cycle of the stock being declared in good condition based on the terminal year in every assessment, yet overfishing continuing in reality. This overfishing keeps stock size lower than desired based on the management goals, and causes yield to be forgone in the long term.

This simple example would be exacerbated if the management measures are implemented relative to current conditions. This is because the most recent estimates of F are underestimated and so the relative change needed to achieve the target F is also underestimated, meaning insufficient management measures will be imposed next year to achieve the target F , resulting in continued overfishing.

This simple example started with the assumption that the retrospective pattern caused the stock size to be reduced towards the true value as successive years are added. As shown above, depending on the source and the timing of the retrospective pattern, the stock size could be reduced away from the true value, increased toward the true value, or increased away from the true value. This means that knowledge of the actual cause of the retrospective pattern is required to determine how to correctly account for the retrospective pattern when setting management regulations (see below).

All stock assessments have some level of uncertainty. The retrospective pattern can be considered an additional source of uncertainty. However, the uncertainty due to random noise in the data is generally symmetric in terms of directional change, while retrospective patterns are by definition directional, and thus need to be considered differently. The risk associated with a noisy but unbiased estimate of current stock size is easy to understand, higher catch now translates directly into a lower probability of achieving a desired F . In contrast, when a stock assessment exhibits a retrospective pattern, even choosing the 1% level of catch associated with the target F (an extremely risk-averse choice) may in fact still lead to overfishing because the entire distribution of stock size was above the true value. This inability to directly relate probabilities of achieving the management target with levels of risk-aversion makes the managing a stock exhibiting a retrospective pattern difficult. The ICES WGMG (2007) recommends precautionary management in this situation, but could not provide advice regarding how precautionary to be.

A retrospective pattern is an indication that something is inconsistent in the data or model. Lack of a retrospective pattern does not mean that all is well with the stock assessment, there are many other diagnostic tools that should be considered when conducting stock assessments. Simulation studies have indicated that when an

intervention is introduced into stock assessment data, all models encounter similar problems when estimating stock abundance and current fishing mortality rates (see the PopSim Working Paper from this meeting). Thus, retrospective patterns will be a problem for any type of stock assessment and care must be taken to determine how to respond to such a situation (see below).

Identifying the Source of a Retrospective Pattern

Many different approaches have been attempted to identify the source of retrospective patterns. While there has been some limited success with identifying the timing of the source of a retrospective pattern in simulated data, there are no methods currently known that can correctly identify the source, even in simulated data sets with little or no noise. Similarly, it has not been possible to definitively identify the source of retrospective patterns in actual stock assessments. Below are some of the approaches examined by the working group describing their uses and limitations.

Local Influence Surface

Local influence surfaces (LIS) are computed by making small perturbations to input matrices and plotting the change of a response variable. They were first described for use in retrospective analyses by Cadigan and Farrell (2002, 2005) using the retrospective statistic ρ as the response variable and examined changes to catch, natural mortality, and survey catchabilities. The surfaces indicated the age and year combinations which had the largest effect on ρ . This was hypothesized to be useful as a diagnostic for the timing of the intervention that produced the retrospective pattern. Additionally, by comparing the magnitude of the three surfaces (catch, M , and q), it was hypothesized that the source could also be identified. When applied to simulated data, neither of these hypotheses was found to be true (e.g. ICES 2007). It was determined that the ρ statistic was not an appropriate response variable. The LIS depended only on the number of peels (Figure 12) and always determined survey catchability to be the most likely cause of the retrospective analysis.

However, the use of other response variables with LIS, such as mean square residual, has been able to find the timing of interventions leading to retrospective patterns in simulated datasets with relatively little noise (Figure 13). The mean square residual LIS uses only the full time series assessment, not any peeled assessments, and thus cannot depend on the number of peels. To date no response variable for LIS has been found which can correctly identify the source of retrospective patterns in simulated data.

Moving Window

An alternative approach to the standard retrospective calculation, when years are removed (peeled) from the end of the time series one at a time, is to conduct assessments using a moving window of years. This means that the early years in the assessment no longer influence the results from later years in the assessment. Plotting average annual

catchability coefficients from the multiple moving windows that overlap each year can provide an indication of when an intervention occurred that led to the retrospective pattern in the standard calculations, although noise in the dataset can make this detection difficult (Figure 14). While this approach shows potential for identifying the timing of an intervention, it is not possible to use it to identify the source of the problem.

q Surface

When surveys are split into age-specific tuning indices, a catchability coefficient (q) is estimated for each age. This scalar is used to relate the observed index values to the associated model estimates of stock abundance through the equation $I=qN$. The resulting population abundance estimates from the stock assessment can be used to generate derived catchabilities by dividing the observed index values by their associated population abundance estimates. Standardizing the values in each index allows plotting of the q surface (e.g. Figure 15). These plots are similar to residual plots except that all ages for a given survey are plotted at once to allow easier detection of year or cohort specific changes.

Retrotension

Conflicts within the data themselves can often be observed prior to use of a stock assessment model when retrospective patterns are present. For example, the Georges Bank yellowtail flounder data have a survey Z time trend that is relatively flat while the relative F (catch/survey biomass) shows a strong decline since 1994. Since M and survey catchability are assumed constant, these two time series of basic data are not consistent and lead to the Base Case assessment showing a retrospective pattern. This conflict has been termed “retrotension” by members of the working group.

Another demonstration of retrotension in an assessment is to plot the estimated survey catchabilities from each of the peeled assessments. When retrospective patterns are present, these catchability plots show strong directional change, meaning each additional year of data changes the catchability in a consistent direction. Assessments which do not exhibit a retrospective pattern do not show consistent changes in survey catchabilities as years are peeled.

It has been hypothesized that some features of the basic data could be described that would allow prediction of when a retrospective pattern will emerge and how long it will be expected to last. To date, these features have not been determined, but remain a topic for future research.

Fixing Retrospective Patterns

When a retrospective pattern has been identified in a stock assessment, the question arises, can it be fixed? A number of methods have been examined to determine if this is possible. Generally, there are many ways that the data or model can be adjusted to

remove the retrospective pattern. However, simulations have demonstrated that none of these approaches guarantee that the new assessment will be any closer to the truth than the original assessment, unless the source and timing of the cause of the retrospective pattern has been correctly identified and exactly adjusted. Since it is not possible currently to definitively identify the source(s) of a retrospective pattern in any actual stock assessment, these fixes should be considered with caution.

Data or Model Modifications

There are a number of methods that can be applied to remove retrospective patterns in stock assessments when the true source is assumed known. These methods will always work to remove the retrospective pattern, but may produce stock and F estimates that are farther from the truth than the original assessment. For example, if missing catch in recent years is assumed to be the cause of the retrospective pattern, Badapt is an approach which estimates the level of missing catch and removes the retrospective pattern (ICES 2007). This approach was been demonstrated to work well with simulated data when missing catch is in fact the source of the retrospective pattern, but it will produce incorrect stock abundance and F estimates if something other than catch is the true cause of the retrospective pattern (ICES 2007). Similarly, estimating year effects in catchability within the VPA removes the retrospective pattern in all cases, but will only produce improved results when changes in catchability are the actual source of the retrospective (ICES 2007). An alternative approach to reducing the retrospective pattern is to make the model less flexible in the recent years through a process called shrinkage, thereby enforcing stability on the results (Darby and Flatman 1994). Shrinkage does not remove the retrospective and was not examined by this working group.

Local Influence Surface

As described above, the local influence surface (LIS) approach of Cadigan and Farrell (2002, 2005) using rho as the response variable cannot be used to identify the timing or source of the retrospective pattern. However, one can use it to remove a retrospective pattern from an assessment in a quick and objective manner by changing the input data according to the changes indicated by the LIS. The working group demonstrated this ability by conducting a three by three analysis, whereby the three main sources of retrospective patterns (catch, M, and survey catchability) were fixed by changing each of the three input data matrices according to their LIS (Figures 16a-c). In all cases, the retrospective patterns were removed. However, in no cases were the results closer to the truth. This clearly demonstrates that retrospective patterns can be “fixed” by changing the input data according to the LIS, but that the new assessment will not necessarily provide better estimates of the stock abundance or fishing mortality rate.

Split Series

Another way of modifying the input data to reduce a retrospective pattern is to split the survey time series. This allows estimation of different catchability coefficients for the two time periods of the survey and effectively breaks the link between the early and

recent survey data in the assessment. This approach has been applied in the Georges Bank yellowtail flounder assessment where it greatly reduced the retrospective pattern. However, catchability coefficients increased three to five fold, which could not be explained by any changes in the surveys themselves, and was described as aliasing an unknown mechanism. It is easily demonstrated that if in fact a step change in survey catchability was the source of the retrospective, then splitting the surveys in the appropriate year will remove the retrospective pattern and produce correct stock and F estimates because no model misspecification exists (results not shown).

However, if the true source of the retrospective pattern is a change in catch or natural mortality rate, then splitting the surveys will remove the retrospective pattern (except for years where cohorts span the intervention year) but the stock and F estimates will be biased (Figures 17-18b). Similarly, the natural mortality rate could be split during the time series and by trying different values one could solve for the recent M that removes the retrospective pattern, but does not necessarily produce the correct stock and F estimates if the true source of the retrospective pattern was not M (Figure 19). These “fixed” results all use the correct timing of the intervention that caused the retrospective pattern, and all remove the retrospective pattern, but if the “fix” does not match the true source then biased estimates result. The fixed assessments also have good diagnostics, with no obvious residual patterns in recent years and slight patterning in early years that would most likely be difficult to detect in actual assessments (Figure 20). Thus, in an actual assessment, there would be no obvious reason to reject the fixed assessment, but it could still be as biased as the original assessment, or even further from the truth. Caution is urged when fixing assessments which exhibit retrospective patterns.

Adjust Population Estimates for Projections

When an assessment has exhibited a retrospective pattern for a number of years, one option is to adjust the estimated populations to account for the retrospective pattern before conducting projections to provide management advice. This approach has the advantage of not requiring identification of the correct source of a retrospective pattern. Instead, it acknowledges that a retrospective pattern is present and attempts to adjust the management advice under the assumption that the pattern will continue into the future. One disadvantage of this approach is that exactly how to make this adjustment is still an active area of research. Furthermore, if the retrospective pattern goes away or changes direction in the next assessment, the adjusted advice will be worse than the original advice.

Research on this topic using management strategy evaluations was conducted by the ICES WGMG (2007) and is ongoing at the Northeast Fisheries Science Center. One finding is that if a retrospective pattern is caused by a change in survey catchability, and this change persists into the future, the retrospective pattern does not diminish with increasing years of data under the new survey catchability value (Figure 21). Future work will focus on the effect of different methods of adjusting the population to account for the retrospective pattern within the management strategy evaluation framework.

Alternative States of Nature

Since the source of a retrospective pattern cannot be identified, and research is still needed on how to adjust for a retrospective pattern when making projections for management, different states of nature could be put forward to managers to account for the additional uncertainty in the assessment due to the retrospective pattern. This could be accomplished by using the full time series assessment with the retrospective as one state of nature and then removing the retrospective through any means to produce an alternative state of nature. One can think of this as using the tips of the assessments versus the converged part of the assessment. Alternatively, one can think of this as identifying the two attractors that are causing the retrotension in the model (e.g. the red and blue lines in Figure 11). The alternative states of nature would then correspond to the stock dynamics under these two possibilities. However, it is not possible to ensure that the states of nature selected will correctly bound the truth for all years (e.g. top right panel of Figure 8). More research is needed to determine how to appropriately bound alternative states of nature when an assessment exhibits a retrospective pattern.

Discussion

This working group benefited greatly from some short period of focused research during the past two years. During these week-long periods, only retrospective topics were considered by some of the working group members. Progress was made much faster during these short periods than during the intervening periods when other activities required attention. This model of short, focused research periods should be used in the future for other topics.

Although the working group conducted a large number of analyses and derived insight into the issue of retrospective patterns, it did not find a method to determine the source of a retrospective pattern nor clearly identify how to provide management advice in such a situation. This was not entirely unexpected due to the many previous attempts to “solve” the retrospective problem. For example, the ICES WGMG has been working on the retrospective problem almost continuously since 1991 and has not solved the problem either. Nonetheless, the working group was able to agree to the following conclusions and recommendations during the January 2008 meeting, which we hope will provide guidance for dealing with retrospective patterns.

Conclusions

1. Retrospective pattern is an indication something is inconsistent (data and/or model).
2. Lack of a retrospective pattern does not mean all is well. Based on simulations, data or model inconsistency does not always produce a retrospective pattern. Retrospective patterning is just one diagnostic to be considered when conducting stock assessments.
3. Simulated retrospective patterns can be caused by time trending changes in biological characteristics, catch, survey catchability, or spatial concentration of the population. Multiple sources may occur in assessments.
4. The source(s) of the retrospective pattern can be anywhere in the time series. Some methods were presented to identify when the change took place (moving window, q surface, mean square residual LIS).
5. The true source(s) of a retrospective pattern have not been identified using current methods. Knowledge of events in the fishery or biological information may help identify probable sources.
6. Interventions (correlated errors) are more likely to cause retrospective patterns than random noise.
7. Splitting surveys, changing M , or changing catch may reduce the retrospective pattern, but do not necessarily produce an assessment closer to the truth, although the other diagnostics for the new assessment may be fine.
8. The retrospective statistic, ρ , may be a useful measure of the amount of retrospective pattern. A strong retrospective pattern can be defined by the degree of overlap between confidence intervals from different terminal years.
9. Local influence surface analysis using ρ is not useful for diagnosing the timing or source of retrospective patterns.
10. In many stocks, strong retrospective patterns typically persist.

Recommendations

1. Always check for the presence of a retrospective pattern.
2. If a model shows a retrospective pattern, then consider alternative models or model assumptions.
3. Develop objective and consistent criteria for the acceptance of assessments with retrospective patterns.
4. A strong retrospective pattern is grounds to reject the assessment model as an indication of stock status or the basis for management advice.
5. When a moderate retrospective pattern is encountered: (not an exhaustive list)
 - a. Consider alternative states of nature approach to advice.
 - b. Investigate the performance of alternative methods for retrospective adjustments through management strategy evaluations.
6. Use biological and fishery hypotheses and auxiliary information as a basis for adjustments for retrospective patterns.
7. Consider use of survey swept area numbers instead of mean catch per tow in assessment models.
8. The presence and implications of a retrospective pattern as a source of uncertainty in the assessment should be clearly communicated to managers.

Literature Cited

- 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.
- Cadigan, N.G. and P.J. Farrell. 2002. Generalized local influence with applications to fish stock cohort analysis. *Appl. Stat.* 51: 469-483.
- Cadigan, N.G. and P.J. Farrell. 2005. Local influence diagnostics for the retrospective problem in sequential population analysis. *ICES J. Mar. Sci.* 62: 256-265.
- Darby, C. D., and Flatman, S. 1994. Virtual Population Analysis: version 3.1 (Windows/Dos) user guide. Info. Tech. Ser., MAFF Direct. Fish. Res., Lowestoft, (1): 85pp.
- Evans, G.T. 1996. Using the elementary operations of sequential population analysis to display problems in catch or survey data. *Can. J. Fish. Aquat. Sci.* 53: 239-243.
- ICES. 2007. Report of the Working Group on Methods of Fish Stock Assessment (WGMG). 13-22 March 2007. Woods Hole, USA. 139 p.
- 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.
- 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.
- 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.

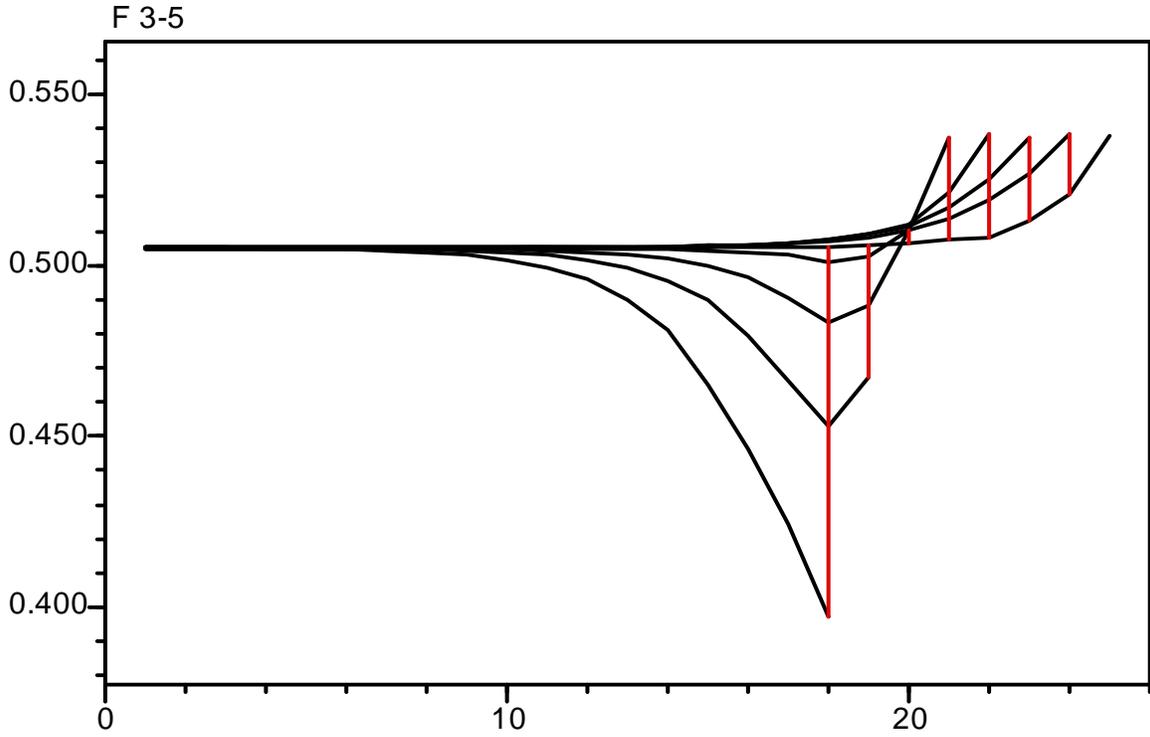


Figure 1. Example of the comparisons made when calculating Mohn rho.

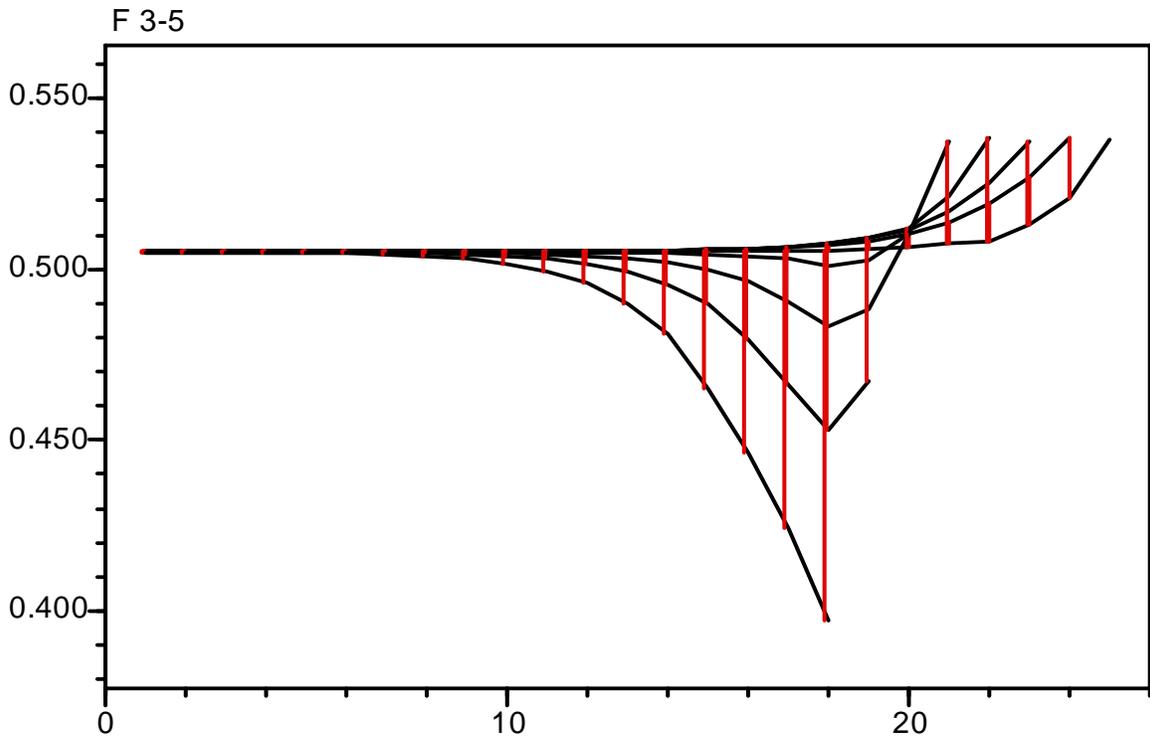


Figure 2. Example of the comparisons made when calculating Woods Hole rho.

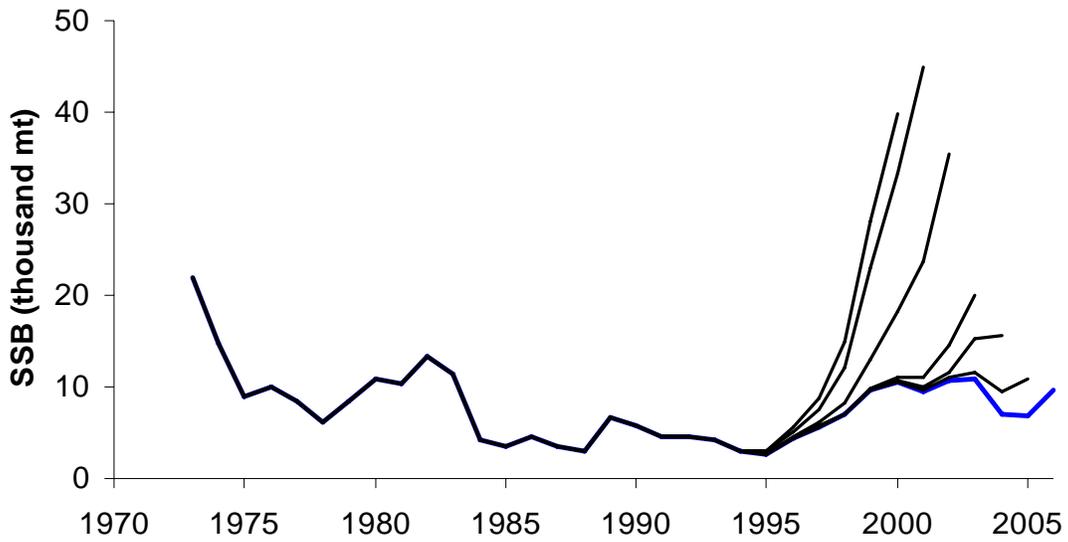


Figure 3. A strong retrospective pattern observed in the 2007 Georges Bank yellowtail flounder Base Case assessment.

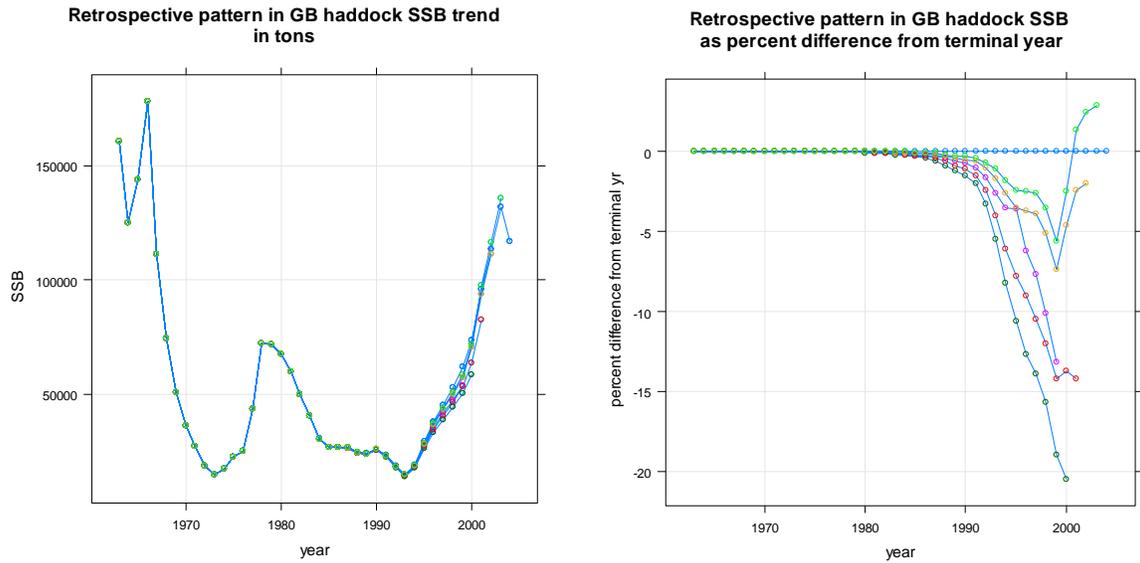


Figure 4. Retrospective plots for Georges Bank haddock, standard plot on left and relative plot on right.

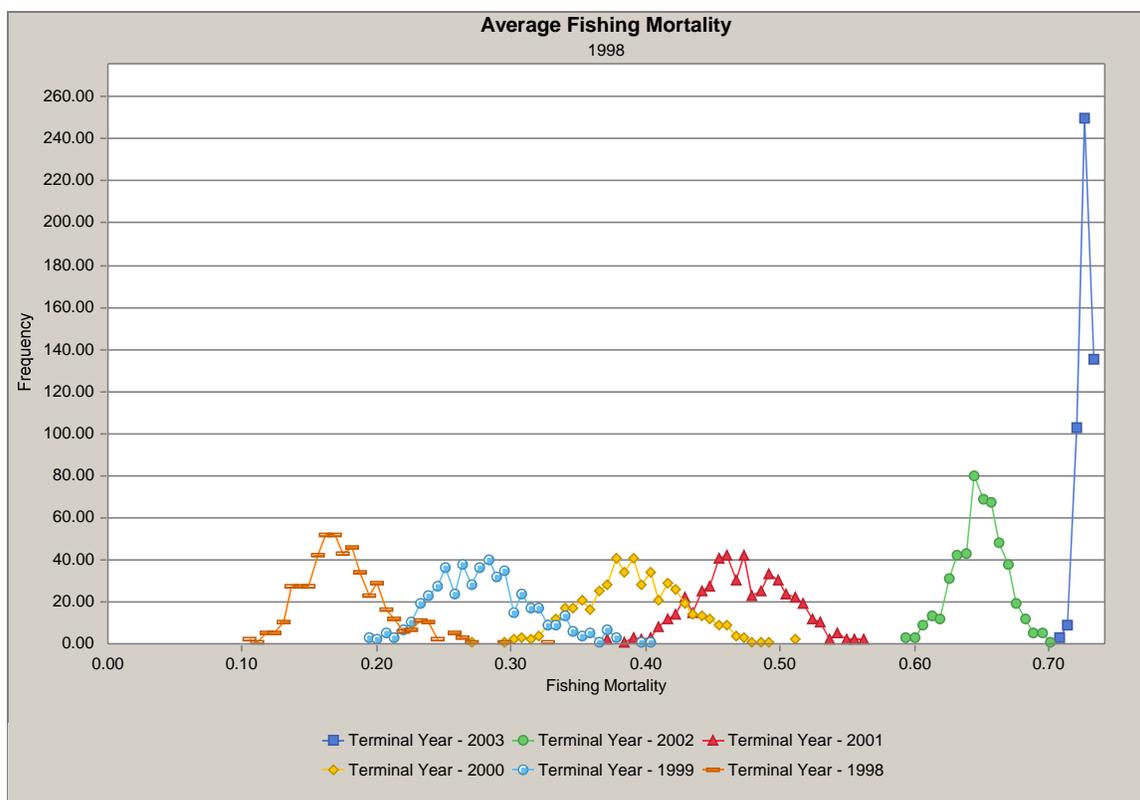


Figure 5. Bootstrap distributions of fishing mortality rate in 1998 from peeled assessments of Georges Bank yellowtail flounder with different terminal years. The lack of overlap between distributions indicates the presence of a strong retrospective pattern.

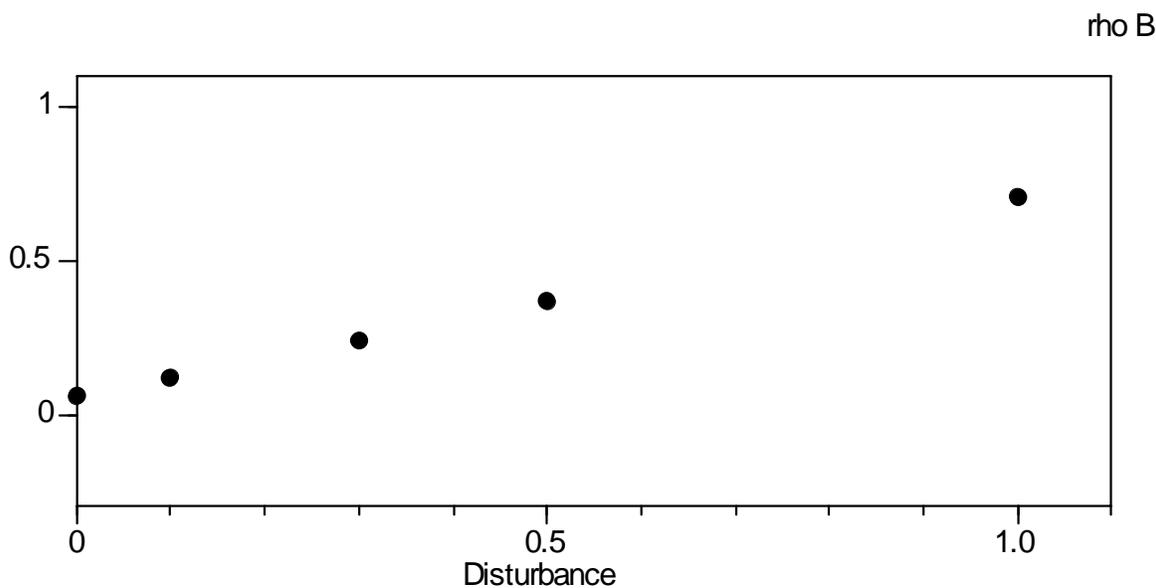


Figure 6. Relationship between rho (y-axis) and the level of bias introduced in the data (x-axis). The monotonically increasing relationship indicates that larger biases in the data do produce larger rho values.

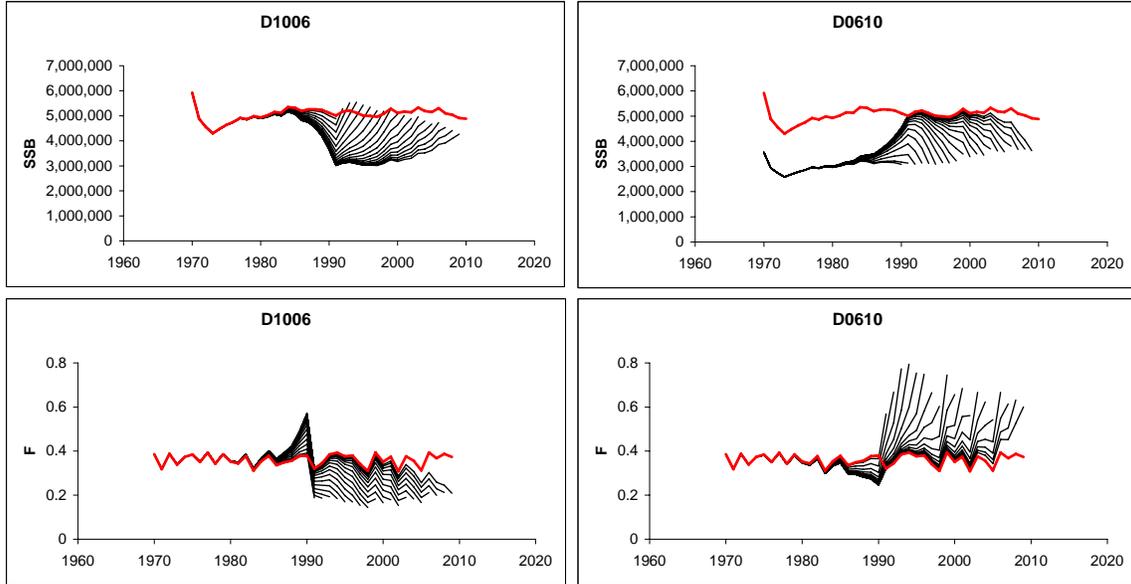


Figure 7. Retrospective patterns (black lines) created by reducing the catch input to the VPA by 60% relative to the actual catch either in recent years (since 1990, left panels) or in the early years (through 1989, right panels). In all plots, the thick red line denotes the true value.

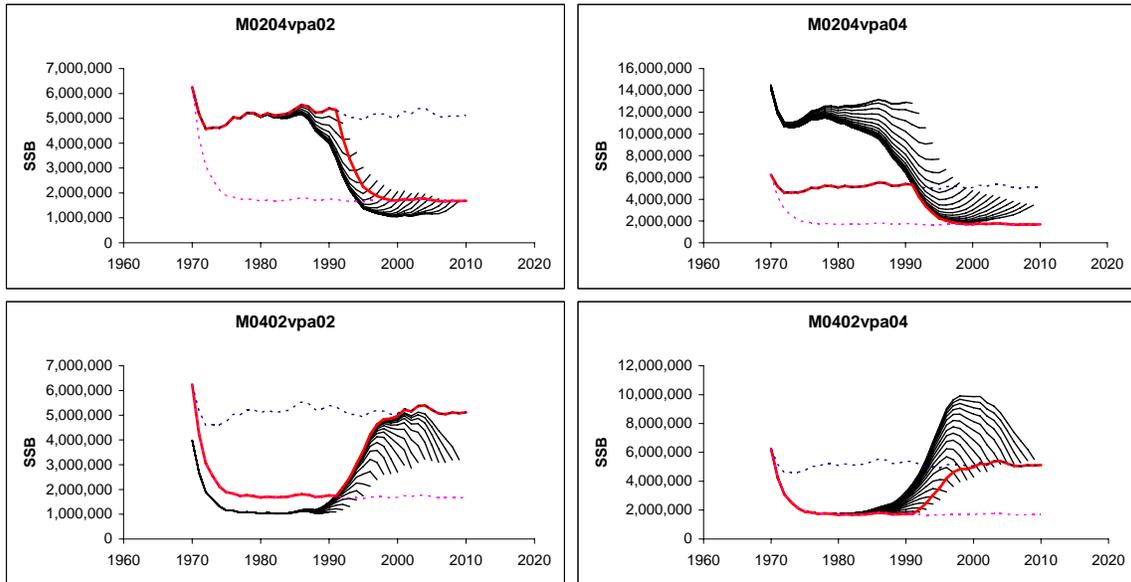


Figure 8a. Retrospective patterns (black lines) created by changing the true natural mortality rate (from 0.2 to 0.4 in 1990 in top panels, from 0.4 to 0.2 in 1990 in bottom panels) with the VPA M set to 0.2 in all years (left panels) or set to 0.4 in all years (right panels). In all plots, the thick red line denotes the true value, the pink dashed line denotes what the SSB would have been if M was 0.4 for all years, and the blue dashed line denotes what the SSB would have been if M was 0.2 for all years (the “attractors”).

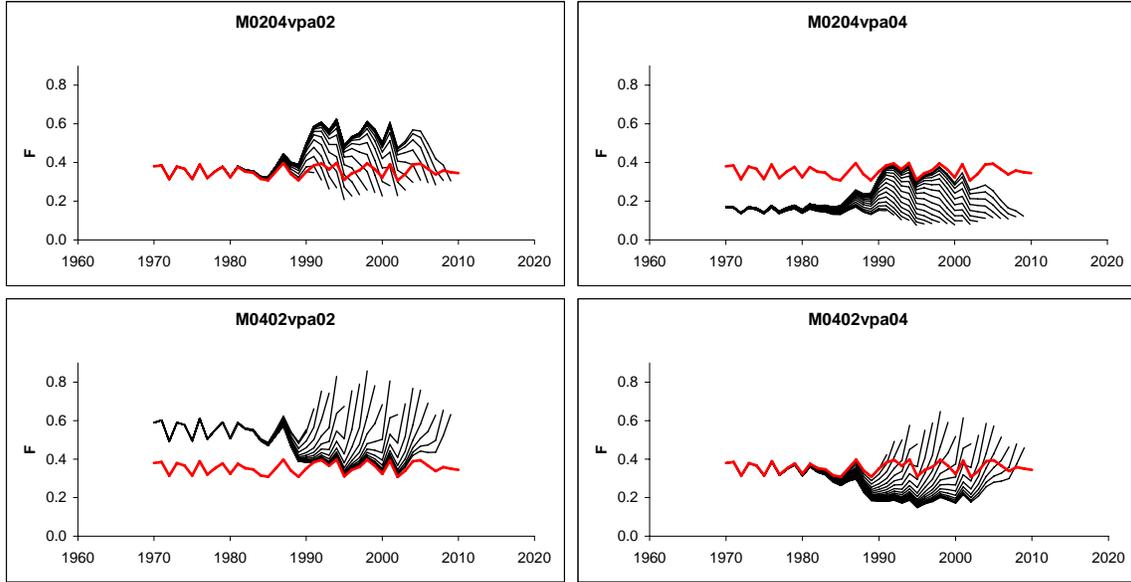


Figure 8b. Retrospective patterns (black lines) created by changing the true natural mortality rate (from 0.2 to 0.4 in 1990 in top panels, from 0.4 to 0.2 in 1990 in bottom panels) with the VPA M set to 0.2 in all years (left panels) or set to 0.4 in all years (right panels). In all plots, the thick red line denotes the true value.

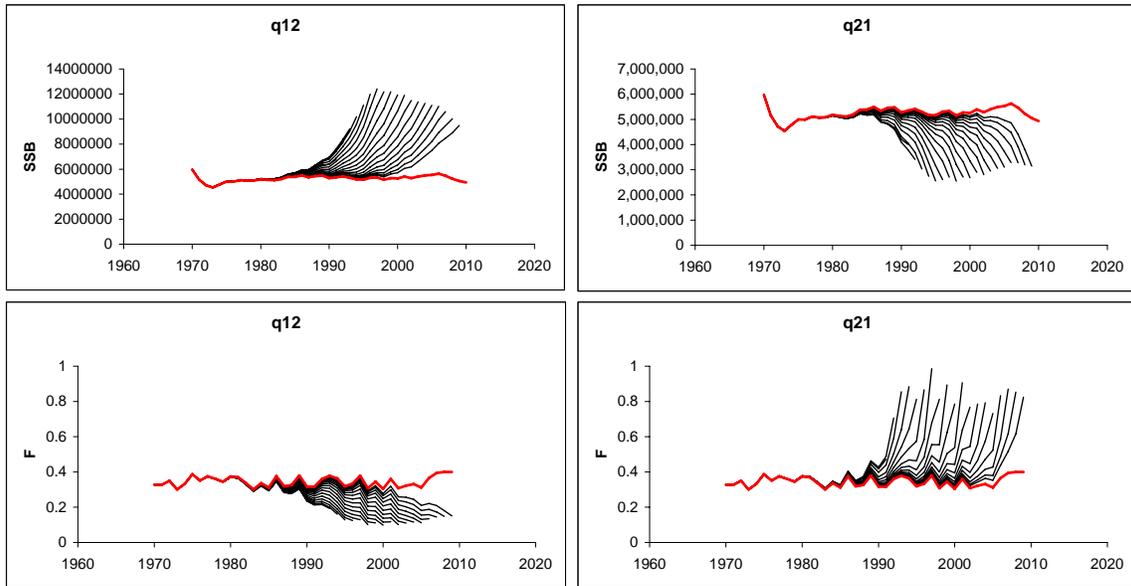


Figure 9. Retrospective patterns (black lines) created by changing the true survey catchabilities for all tuning indices (doubled for years after 1990 in left panels, halved for years after 1990 in right panels). In all plots, the thick red line denotes the true value.

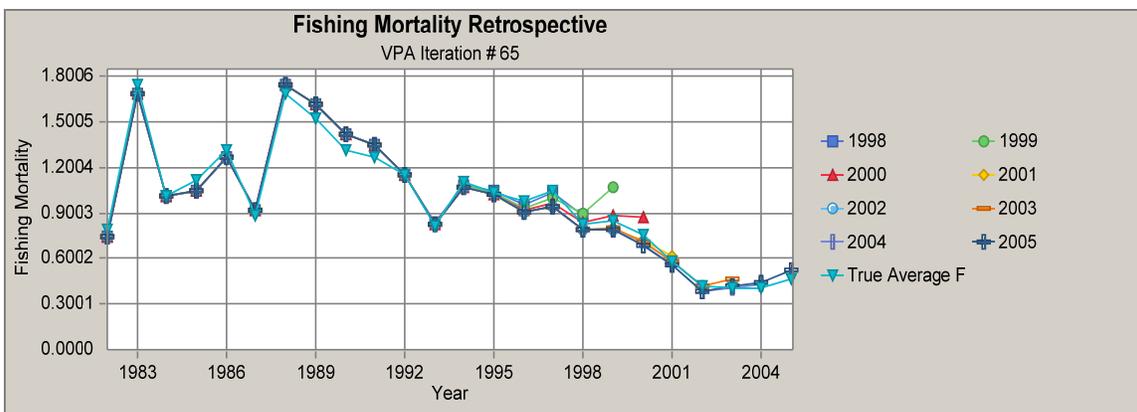


Figure 10. The retrospective pattern corresponding to the 5th percentile of rho statistic from the fluke-like simulation.

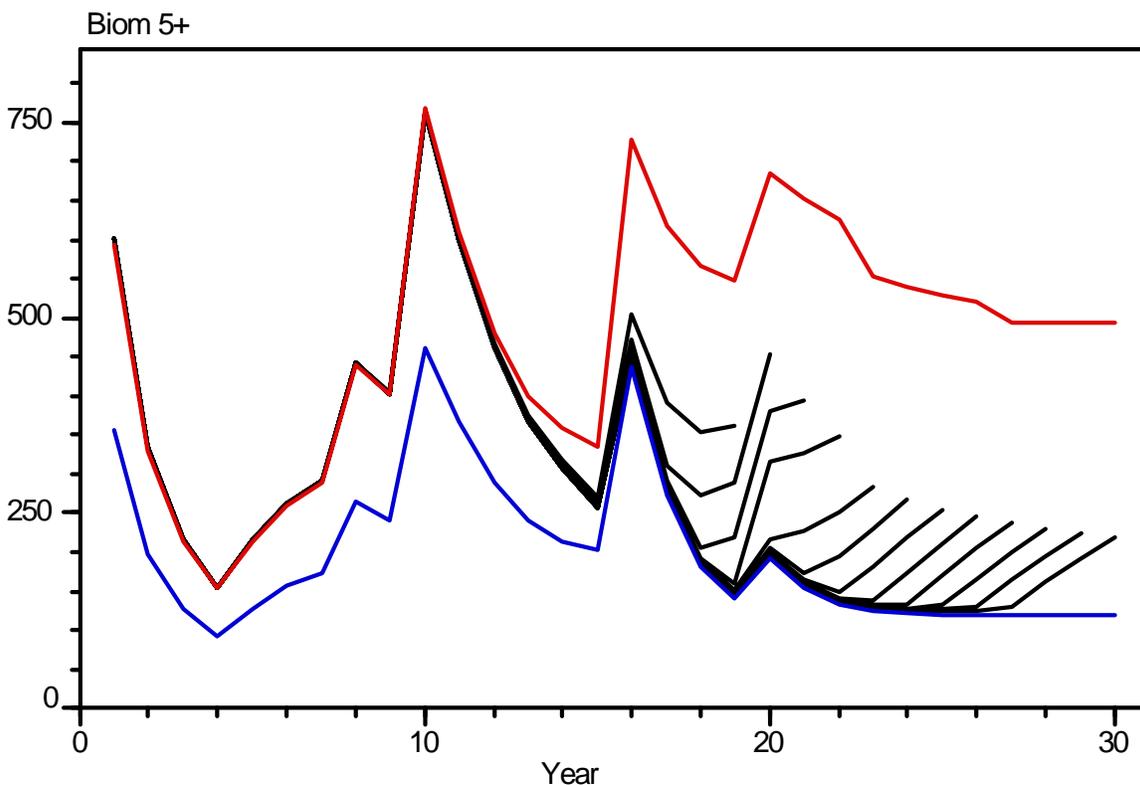


Figure 11. Retrospective pattern (black lines) generated by implementing a closed area half way through the time series for a sessile species. The red line denotes the abundance in the closed area in recent years and the total area in the early years, while the blue line denotes the abundance in the open area in recent years and can be ignored in the early years (the “attractors”).

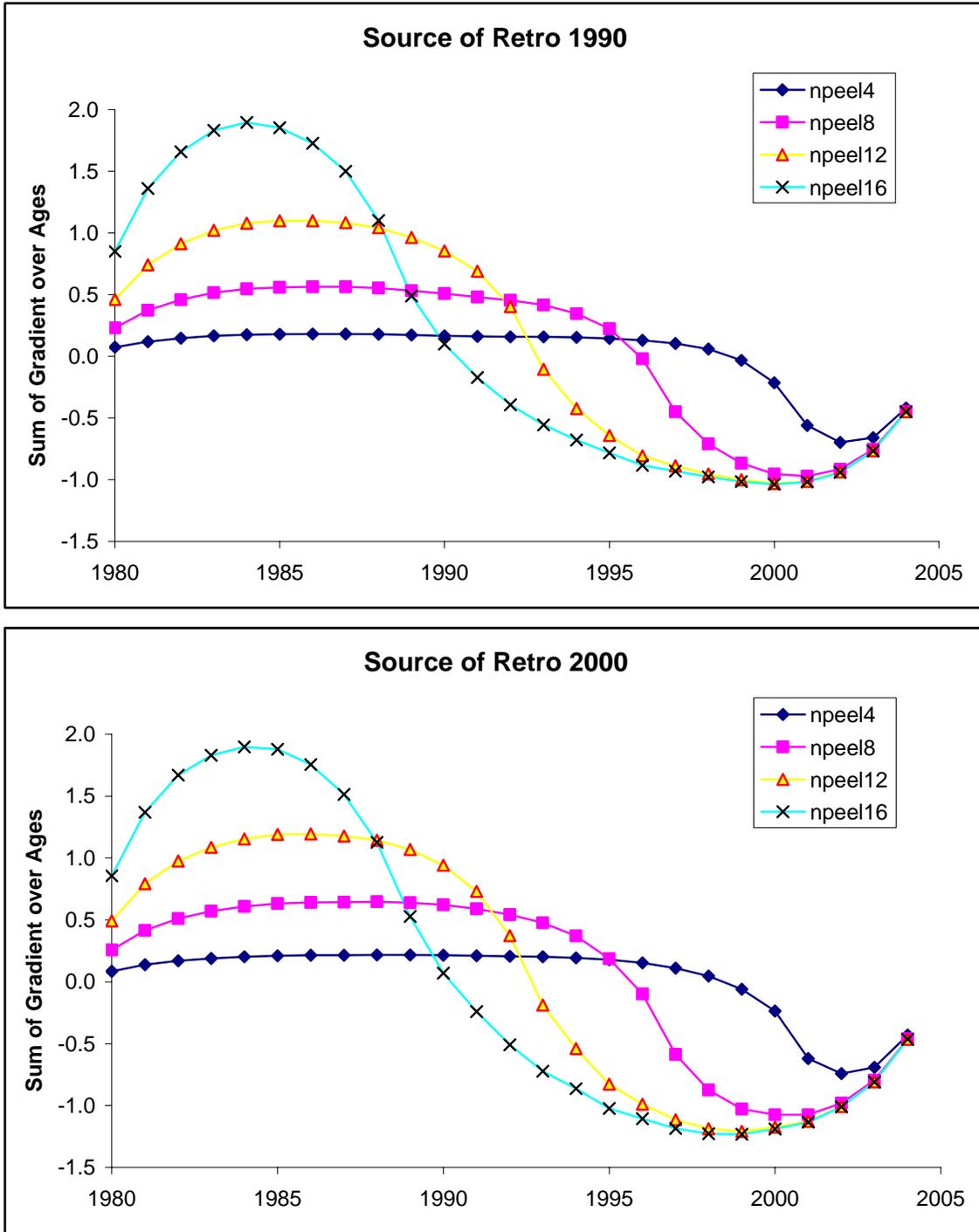


Figure 12. Comparison of local influence surfaces summed over ages for four different number of peels used when computing rho when the true source of the retrospective occurs in 1990 (top panel) and when it occurs in 2000 (bottom panel).

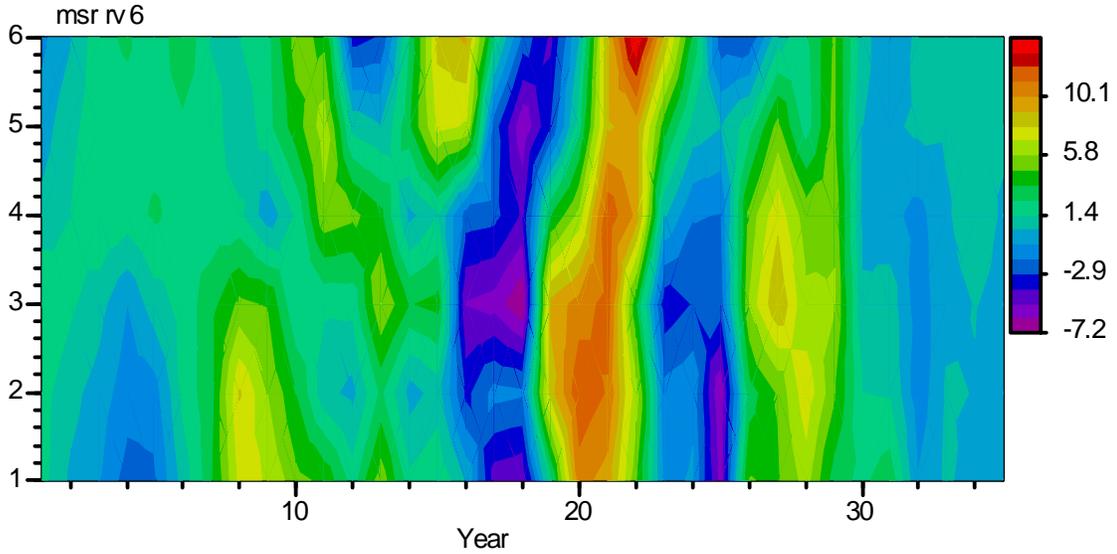


Figure 13. Local influence surface with mean square residual as the response variable and timing of intervention in year 20.

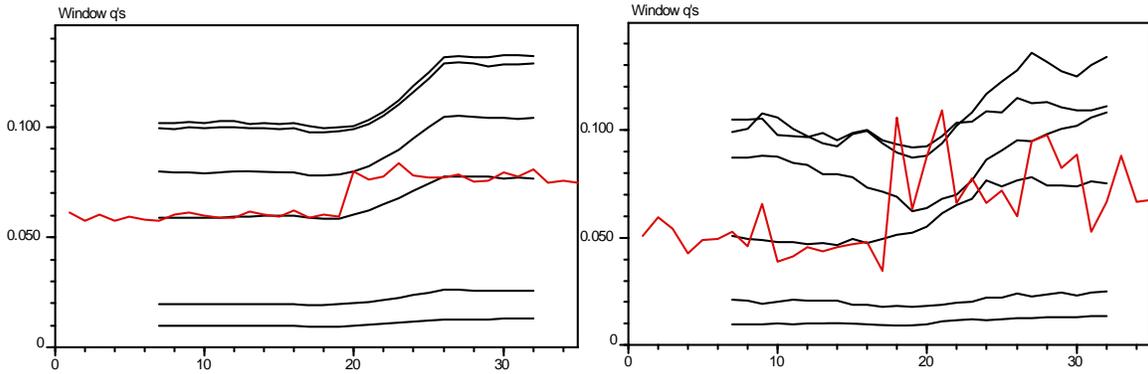


Figure 14. Average catchability (red line) by year from moving window analysis when true timing of intervention is year 20, along with trends in age specific catchabilities (black lines), for 2% noise in data (left panel) and 20% noise in data (right panel).

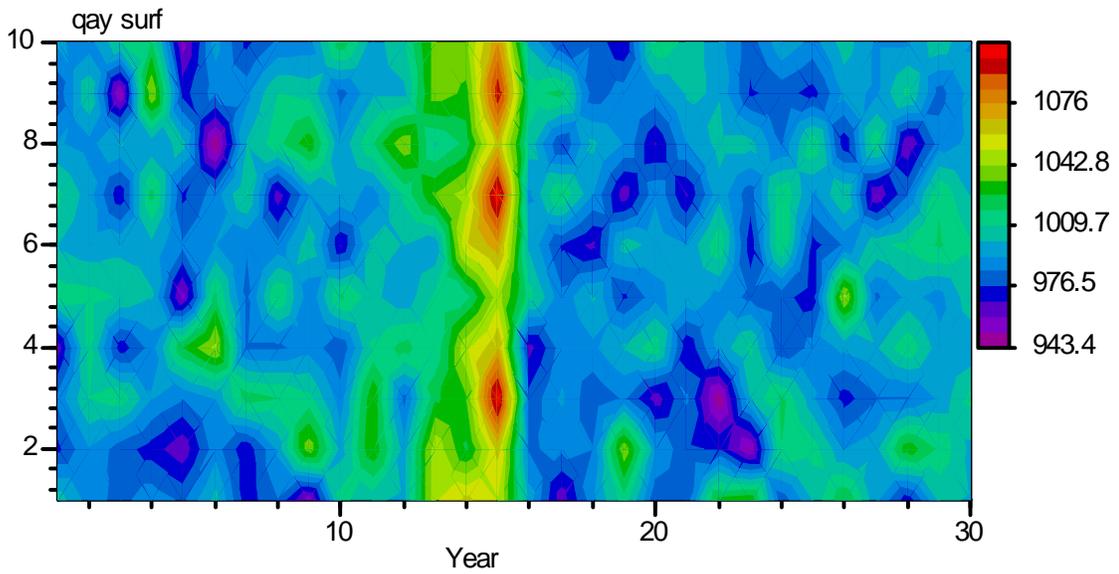


Figure 15. Standardized derived q values from an assessment with a pulse intervention in year 15, called a “ q surface.”

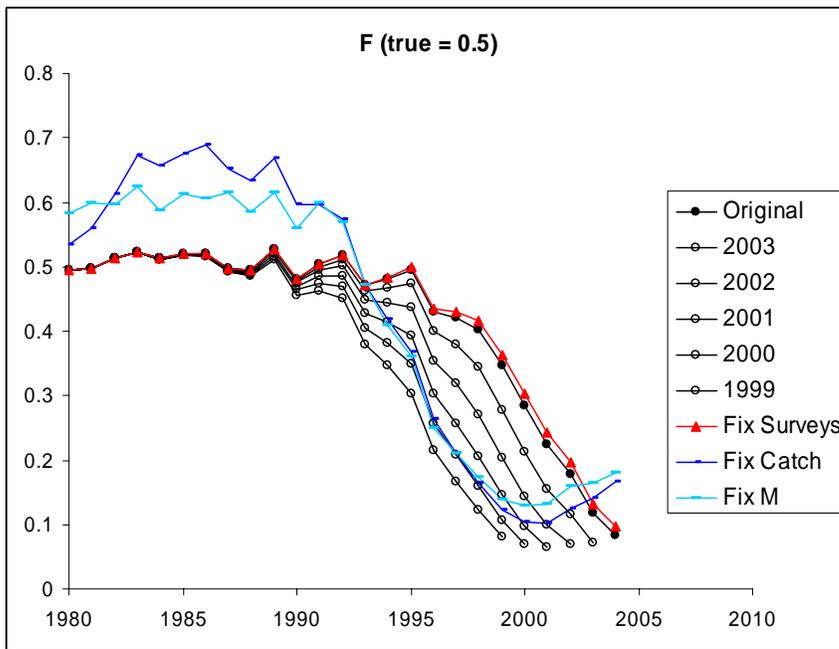


Figure 16a. Comparison of F from initial assessment with retrospective pattern caused by change in survey catchability (black lines) with the “fixed” assessments determined by changes in survey catchability (red line), catch (dark blue line), or natural mortality (light blue line) according to their linear influence surface. The true F for all years was 0.5.

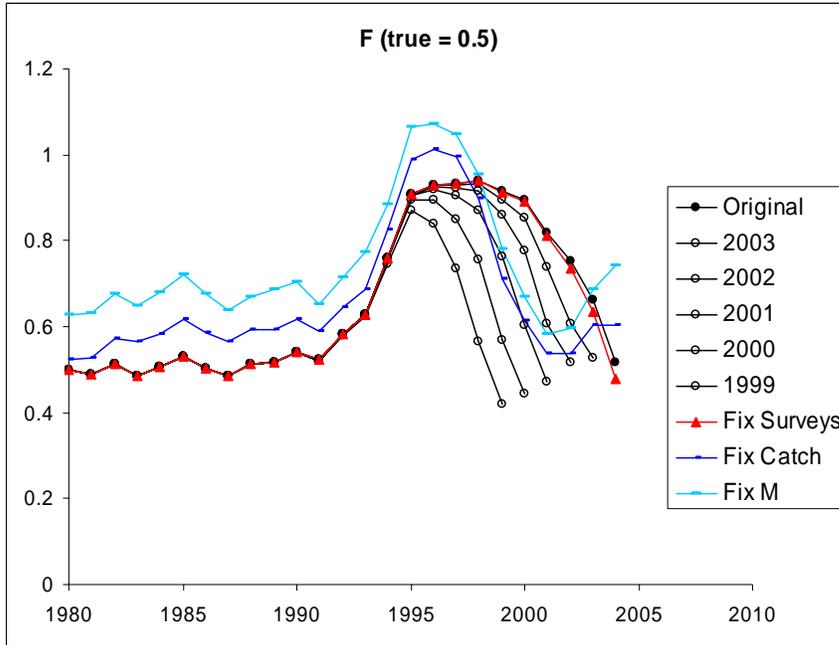


Figure 16b. Same as Figure 16a, except source of retrospective pattern was change in natural mortality rate.

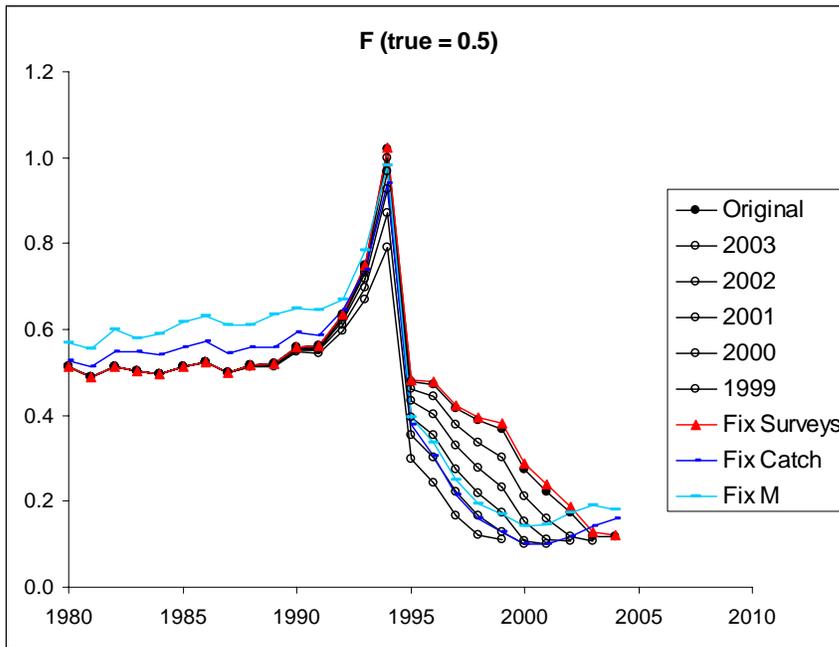


Figure 16c. Same as Figure 16a, except source of retrospective pattern was missing catch.

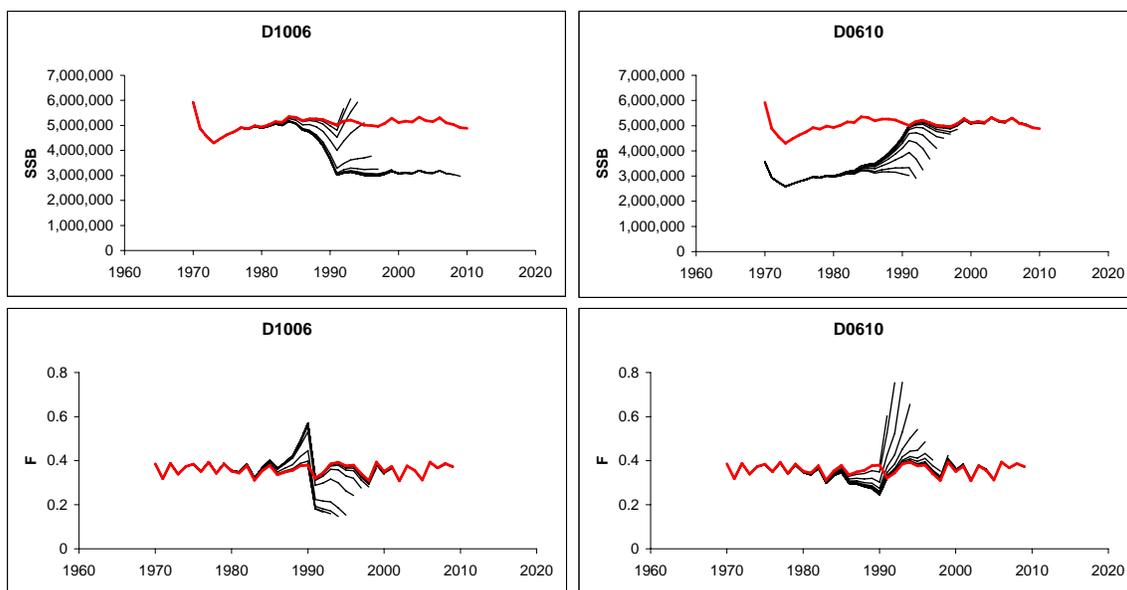


Figure 17. Same simulated data as in Figure 7, except surveys are split in 1990, the year catch was changed. Note the lack of a retrospective pattern in the most recent decade in all cases.

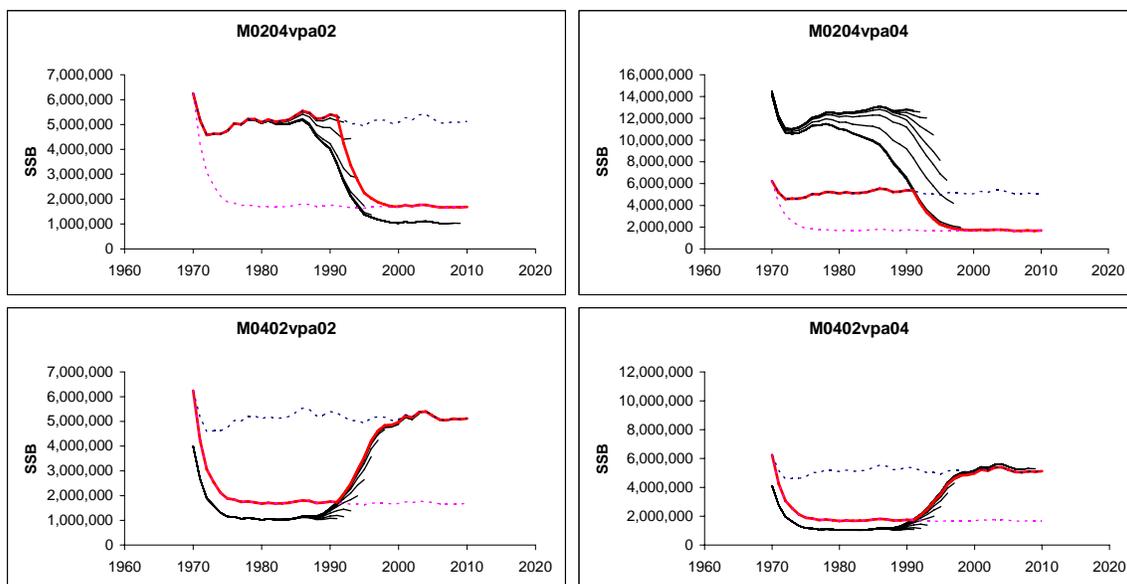


Figure 18a. Same simulated data as in Figure 8a, except surveys have been split in 1990, the year M changed.

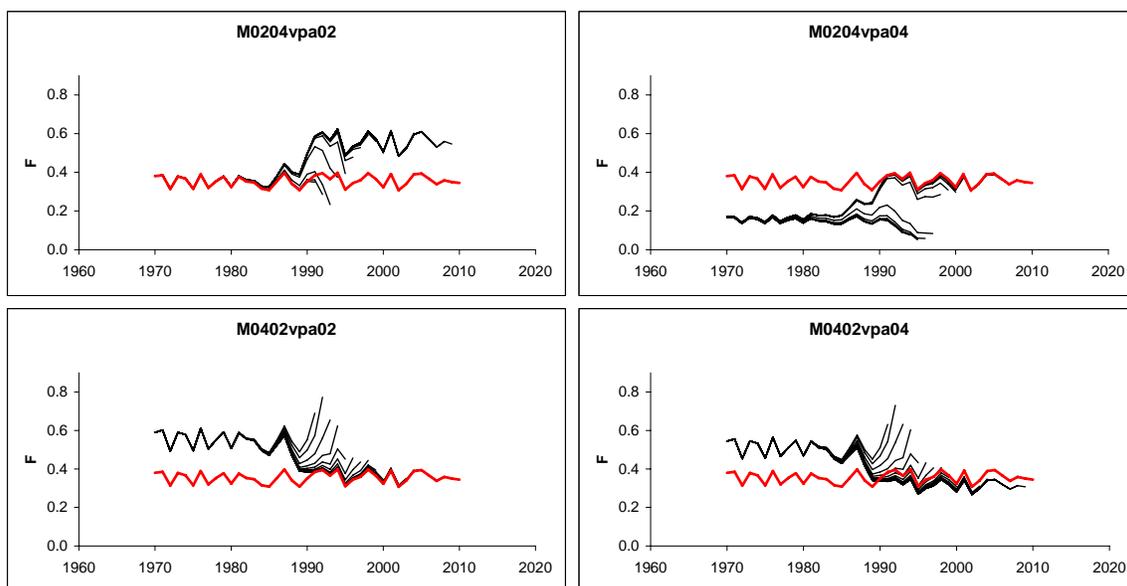


Figure 8b. Same simulated data as in Figure 8b, except surveys have been split in 1990, the year M changed.

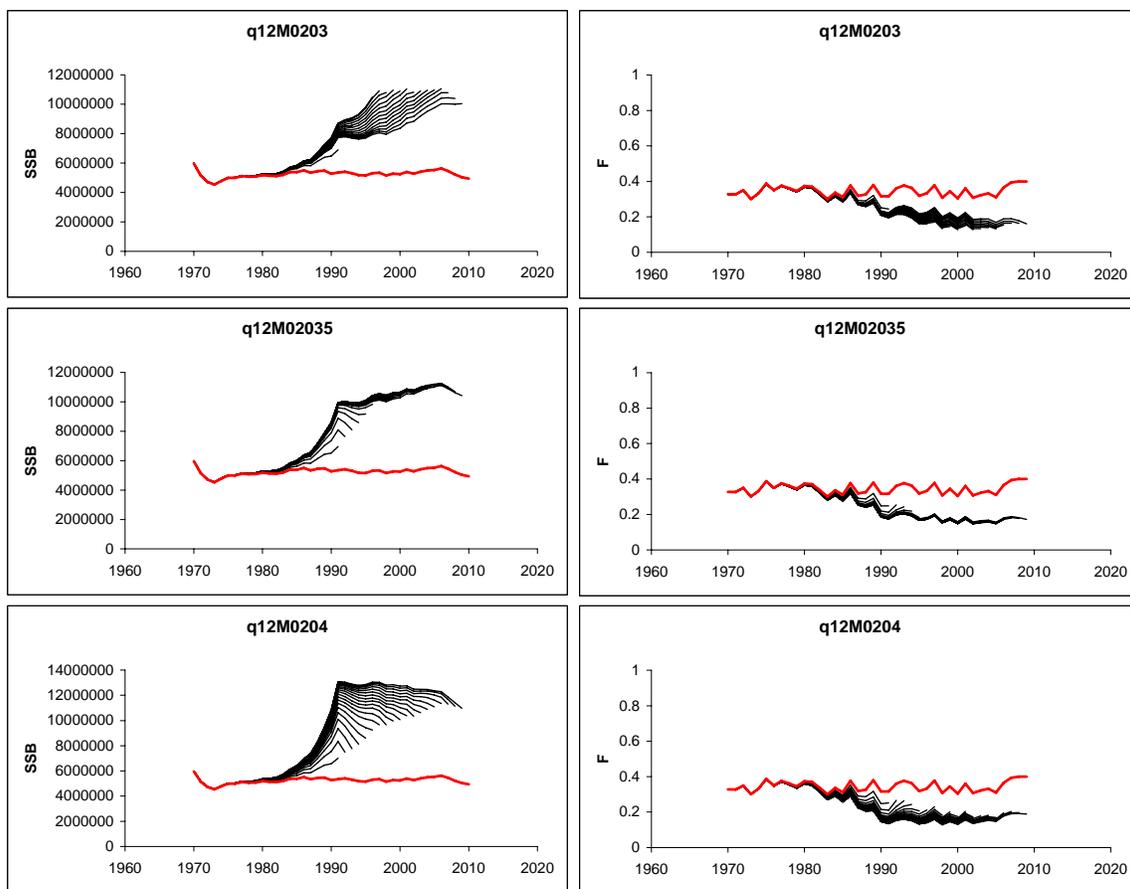


Figure 19. Same data as in Figure 9, except M is increased from 0.2 beginning in 1990 to 0.3 (top panels), 0.35 (middle panels), or 0.4 (bottom panels).

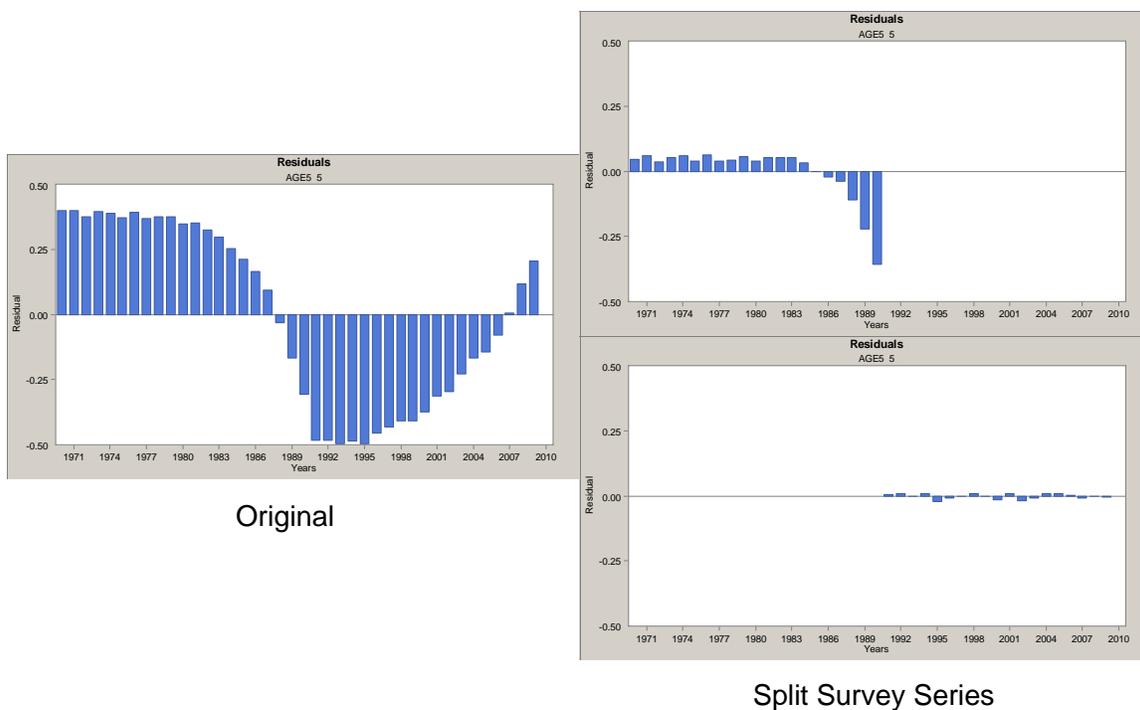


Figure 20. Comparison of age 5 residuals from an assessment with a strong retrospective pattern due to a change in natural mortality rate (left panel) with the age 5 residuals from the same assessment when the survey time series is split in 1990 (right panels). Note that all ages showed a similar pattern and very low noise was included in the dataset.

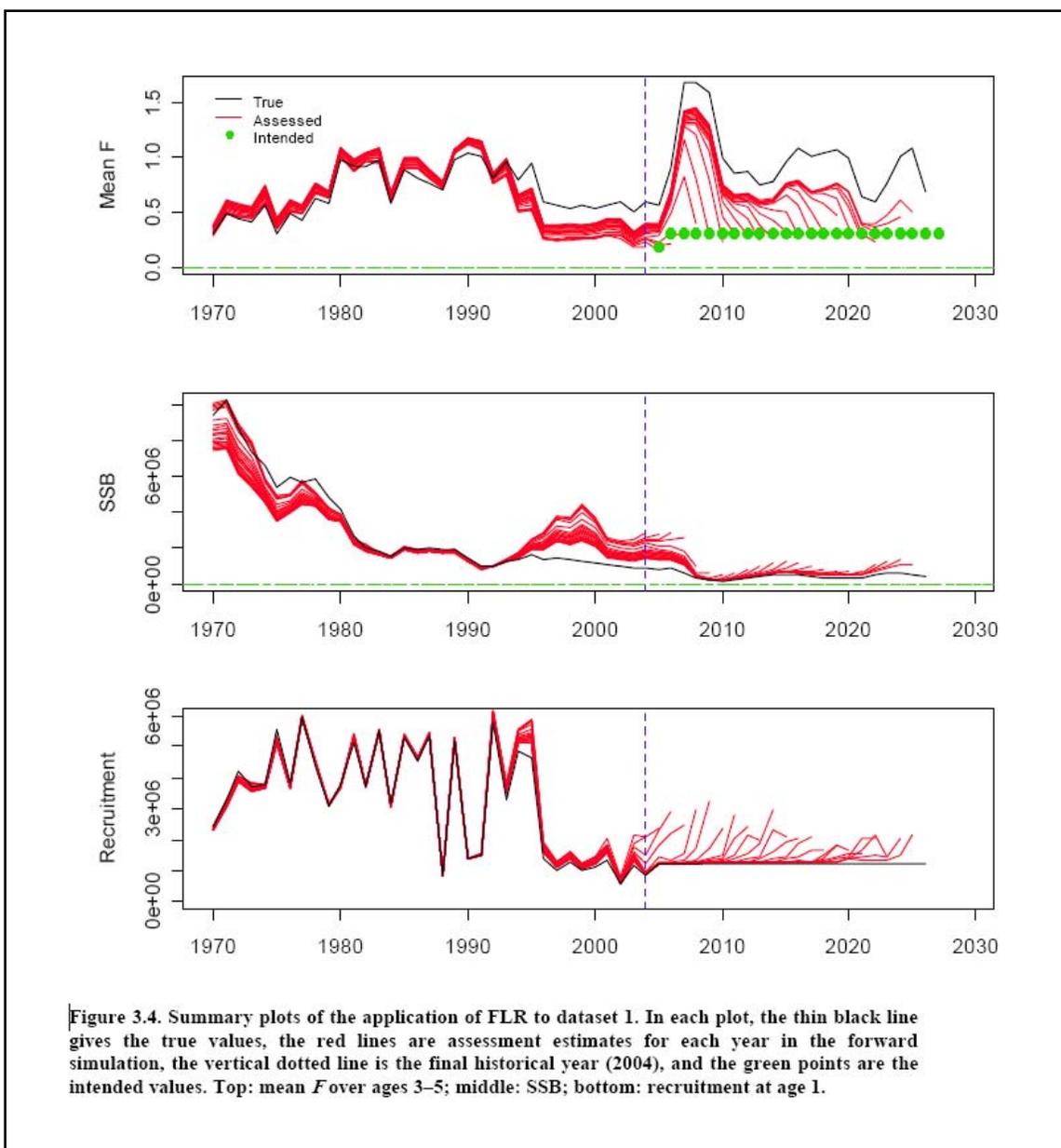


Figure 21. Example of management strategy evaluation in the presence of a retrospective pattern. (From ICES 2007).

Appendix: List of Working Group Participants

<u>Name</u>	<u>Affiliation</u>
Larry Alade	NEFSC
Andy Applegate	NEFMC
Jessica Blaylock	NEFSC
Liz Brooks	NEFSC
Steve Cadrin	NEFSC
Laurel Col	NEFSC
Steve Corriea	MADMF
Deborah Hart	NEFSC
Lisa Hendrickson	NEFSC
Joe Idoine	NEFSC
Larry Jacobson	NEFSC
Chad Keith	NEFSC
Chris Legault (chair)	NEFSC
Jennifer Martin	NEFSC
Ralph Mayo	NEFSC
Tim Miller	NEFSC
Bob Mohn	DFO
Joshua Moser	NEFSC
Paul Nitschke	NEFSC
Loretta O'Brien	NEFSC
Bill Overholtz	NEFSC
Michael Palmer	NEFSC
Paul Rago	NEFSC
Anne Richards	NEFSC
Alan Seaver	NEFSC
Fred Serchuk	NEFSC
Gary Shepherd	NEFSC
Kathy Sosebee	NEFSC
Jiashen Tang	NEFSC
Mark Terceiro	NEFSC
Michele Traver	NEFSC
Jim Weinberg	NEFSC
Susan Wigley	NEFSC

Additionally, a special word of thanks goes to the ICES Working Group on Stock Assessment Methods whose members have been examining the retrospective pattern issue for many years.