Recommended diagnostics for large statistical stock assessment models

Shelton J. Harley and Mark N. Maunder

June 25, 2003

Executive Summary

This report summarizes the discussions and recommendations of the participants of a meeting on diagnostics for large statistical stock assessment models held in La Jolla, U.S.A., on October 2-4, 2002 (see Appendix 1 for a list of participants). The use of such models is becoming more common in the assessment of pelagic species. Because these models are large and complex, there is a need to determine how to best summarize their results so that the quality of the model fit can be investigated. This report is divided into three sections discussing: (1) diagnostics that should be evaluated regularly (*i.e.*, at every assessment); (2) diagnostics that should be evaluated periodically (*i.e.*, every few assessments); and (3) some specific research questions. Included are recommendations about the types of information (*i.e.*, model output) that might be presented at assessment meetings and what should be included in assessment reports, or made available in electronic form to interested parties. Other analyses are possible, and it is expected that in the future other analyses will be found to be appropriate for other data sets. The examples described here are based on data for bigeye tuna in the eastern Pacific Ocean (EPO).

Introduction

A full range of assessment models are used in the assessment of the pelagic fish stocks. Many stocks are assessed with very basic stock production models with only a few parameters and limited data requirements, but increasingly stocks are being assessed using large age-structured statistical stock assessment models, often with over 1000 parameters. These large complex models require considerable data and knowledge of biological parameters.

The majority of the parameters in these models arise because of the relaxing of the critical assumption in traditional cohort analyses – catch-at-age is not assumed to be known without error. Fishing mortality at age is assumed to be separated into selectivity-at-age and an annual fishing mortality scalar. Most parameters are catchability effort deviates – due to not solving the catch equation or using an approximation (*e.g.*, Pope's approximation). This requires the estimation of selectivity parameters and effort deviates to remove the catch. Because of the large number of parameters, a thorough evaluation of convergence and model sensitivity is necessary, but difficult.

We evaluated several types of diagnostic analyses and assigned them to two broad categories:

1. **Regular**: analyses that should be undertaken every time a stock assessment is conducted.
2. **Periodic**: analyses that do not need to be undertaken every assessment, but should be undertaken every so often. These analyses should be undertaken if there have been significant changes in either the data or model structure.

Some specific research questions related to stock assessment that were raised during the evaluation of model diagnostics are also described.

Regular model diagnostics

The group considered that these analyses/activities should be done every time an assessment is conducted. It is recognized that some of these recommendations may reflect the specific data sets used in the analyses, rather than a general issue.

1. Creation of data objects

Given the large number of parameters, it is very difficult to easily evaluate convergence or identify potential problem areas. To help overcome this we developed a suite of S-PLUS functions that read in the output from the assessment model. This data object contains all of the important information about the parameters, *e.g.*, starting values, bounds, values at each phase of estimation, final estimates, and precision of the final estimates.

This was found to be useful to determine which parameters to estimate in each phase and the use of alternative starting values (both described in detail in the section of periodic diagnostics). In our examples, we increased the number of estimation phases from four to nine to allow for a more detailed investigation. We can see that while the starting values were close to the final values, the model assumptions during the first two phases resulted in a large change in the estimates, but that the estimates remained stable over the final six phases (Table 1). Ideally, for important parameters, the final phases of the model should not result in large changes in parameter values.

This summary was also repeated for all components of the objective function. It was very useful to see how the values of various penalties varied over time and how the overall objective function changed with the addition of more model parameters (Table 2). We can see that some penalties were added to the objective function in the early phases, but not in the later phases, *e.g.*, sm.Ftot.penalty. It is clear that the estimation of growth in the final phase greatly reduced the selectivity penalty for the third difference (s.3d.penalty).

2. Likelihood profile for important management quantities

Realistic representation of uncertainty is critical for any stock assessment. For small models, there is a range of options for determining uncertainty *e.g.*, normal approximations, likelihood profiles, bootstrapping, or Bayesian posterior distributions, but few of these can be used with the large statistical models.

Presently, uncertainty is most often presented based on normal approximations via the Hessian matrix and the delta method. It is clear that the uncertainty in many parameters is not symmetrical, so confidence intervals based on normal approximations can be inappropriate. It is strongly recommended that likelihood profiles be calculated for a small range of important management quantities. These may have to be done “manually”, *i.e.*, fix a particularly parameter at a range of values and estimate all the other parameters.

The obvious candidates for likelihood profiling include the standard biological reference points, *e.g.*, B_{MSY} or MSY. Analytical solutions for these quantities seldom exist for these large-scale models so thus they are usually estimated outside of the model with a fishing mortality scalar that will provide the largest long-term yield, given the estimates for all other model parameters. It is not clear whether present computing technologies will allow for profiling of these quantities in large models, as it involves penalizing derived parameters.

3. Sensitivity analyses: selectivity and data weighting

There are many possible sensitivity analyses that can be undertaken, but it is not practical to do them all. Also, many of these result in little change in the results from the basecase model. The group recommended that only a select number of sensitivity analyses be conducted regularly.

Selectivity

Modeling results and yield estimates are often sensitive to different selectivity curves. Often, estimates of selectivity-at-age can be sensitive to smoothness penalties used to ensure that selectivity does not vary greatly from one age class to the next. Estimates of growth (and growth variability) can also lead to uncertainty in selectivity parameters. *It is recommended that a small number of sensitivity analyses to the weighting factors used to control smoothness be presented*.

When different assumptions are made regarding these penalties, the estimated selectivity curves can be very different (Figure 1), resulting in differing interpretations of the patterns and magnitude of recent fishing mortality (Figure 2). While the relative biomass trajectory in our example was quite insensitive to these changes, the sensitivity to the weighting factors is evident in the scaling effect on the absolute biomass trajectories (Figure 3) and important management quantities (*e.g.*, B_{MSY} and MSY).

Data weighting

Most often these models fit to two types of data: (1) catch and effort data; and (2) length-frequency samples of the catch (tagging data are often included in analyses using MULTIFAN-CL). Due to the differences in these data types, different likelihood functions are used for each. As there are often more length-frequency samples than catch and effort observations, the likelihood associated with the length-frequency data often dominates the overall objective function. This can lead to a trade-off between fitting the trend in relative abundance from the catch and effort data and fitting to the length-frequency data. The effects of the trade-off between fitting to each data source can be seen in Figures 4 and 5. The biomass trajectory and impressions about the current status of the stock relative to 1980 is much different when the length-frequency data are given greater weight (ss100) (Figure 4) and recruitment variability is greater as the model attempts to better fit the various modes in the length-frequency data (Figure 5). *It is recommended that a small number of sensitivity analyses to the sample sizes assumed for the length-frequency data be presented*.

4. Comparisons with model estimates/predictions

It is recommended that several comparisons be made between estimates of quantities from the current assessment compared to those made in the previous assessment. Some of these are briefly discussed below, and will be further discussed in the section on retrospective analyses.

Observed versus predicted catches

Each year predictions should be made for catches in the next few time periods. The predictions will be based on assumptions about overall fleet selectivity, catchability, and assumed fishing effort. These predictions should be compared to actual catch estimates when these become available. These comparisons might provide useful insights into model inadequacies, and should be used to help provide a more realistic indication of uncertainty in future catch predictions.

Estimated recruitment in recent years

Estimates of recruitment for the most recent few year classes will generally be uncertain, as there will be few observations of them in the length-frequency data. Previous model estimates should be compared to recent estimates to determine if there is any systematic trend, *e.g.*, initially “strong” year class being estimated to be much weaker than they first appeared or vice versa. Any general information gained from these comparisons can provide useful information in projecting future catches and auxiliary information for management advice.

5. Residual plots for the basecase model

Residual plots from large nonlinear models can be difficult to interpret. Nevertheless, they should be presented for the main model fits, *e.g.*, the length-frequency data and the effort deviates. QQ-plots can be used to determine if the model is over- or under-fitting the data relative to its assumed variance.

The residual plots presented here (Figures 6 and 7) indicate several things. First, the standardized residuals are highly under-dispersed, *i.e.*, there are few residuals more than one standard deviation from the observed value. This indicates that we are fitting the data much better than we would expect, given the assumed sample size. We can see that this effect is greater with some gears (*e.g.*, fishery 9) than others (*e.g.*, fishery 3). Second, most of the larger residuals are negative, indicating that the model is predicting a higher proportion of fish than was observed. This is not surprising, given the often large number of observed zeros with the 1-cm length bins, as assumed for the bigeye analysis.

6. Correlation plots for key management quantities from the basecase model

Often quantities such as recent estimates of recruitment deviates and fishing mortality can be highly correlated. This type of information can be important, as it indicates a flat solution surface that implies that a range of alternative states of nature might be equally likely (given the data). For example, a state of nature indicating high recent recruitment and low fishing mortality might be as likely as one with lower recent recruitment and high fishing mortality. This type of information might be useful for management advice.

Correlation coefficients derived for quantities from large nonlinear models may not always accurately reflect the true relationship between the parameters because of the required assumption of bivariate normality. Scatter plots based on the results of Monte Carlo Markov Chain simulations may be more accurate, but are generally not available for these models because of the enormous computational requirements.

As the surface fisheries in the assessment catch smaller, more recently recruited, individuals, there is generally a strong negative correlation between the effort deviates for those fisheries and the recruitment deviates at the end of the time period (Figure 8). This correlation is lagged from the age at recruitment to the age at which they are selected by the fisheries. The correlation between spawning biomass and recruitment reflects the extent to which spawning biomass in a year is "recruitment-driven", *i.e.*, spawning biomass is made up primarily of first-time spawners (Figure 9). The correlation is positive, and the highest correlations are off the diagonal because of the lag between recruitment to the fishery and the age-at-maturity.

7. Catch-at-age and fishing mortality-at-age matrices.

Many assessment reports of VPA-like analyses routinely include tables of the catch-at-age (input) and fishing mortality-at-age (output). In many statistical models, both may be outputs from the model. These tables should be compared from year to year for consistency. Also, estimated catch-at-age could be used as input for an alternative age-structured assessment. These tables could be made available in an electronic form to interested parties.

When considering alternative models, it will be important to ensure that the assumptions of both models are the same to allow valid comparisons (*e.g.*, biological parameters). Also, most of the current large statistical models assume the fishing mortality can be separated into age- and year-specific components. This assumption will be reflected in the predicted catch-at-age and is different to the assumptions of many VPA-like models.

8. Model output by phase

The data object created in (1) above will be most useful for evaluating the order of estimation (described later) but may be of general interest to some analysts for evaluating/reviewing an assessment. It includes parameter values, likelihood components, and penalties. This object could be made available in an electronic form (presently a S-PLUS list-object). As mentioned in (1), this object could be examined to investigate model performance in the final phases, *i.e.*, do the values of the important parameters change much.

It may also be useful to provide an electronic version of the data used in the analysis.

Periodic model diagnostics

It is recommended that the following diagnostics and analyses be undertaken every few assessments (in addition to the regular items), particularly when there have been significant changes in either the data (*e.g.*, acquisition of tagging data) or the model structure. The full set of analyses is thought to reflect a thorough review of a stock assessment.

1. Sensitivity analyses: all model and data components

A thorough evaluation of model sensitivity might include the varying the following parameters or assumptions:

- Natural mortality;
- Spawner-recruitment parameters;
- Alternative abundance indices;
- Inclusion/exclusion of environmental factors;
- Alternative starting values for selectivity-at-age;
- Weighting factors;
    - priors on growth parameters

- Inclusion of otolith age-length data;
- Variation in length-at-age.

Some model outputs may be robust to some sensitivity analyses, but not to others. Useful model outputs to compare include:

- Biomass (total and spawning) and recruitment trajectories (both absolute and relative);
- Selectivity curves for the major fisheries;
- Biological reference points, *e.g.*, B_{MSY}, MSY, and F_{MSY};
- Stock status in relation to the reference points;
- Mean recruitment;
- Fishing mortality patterns in the most recent years.

2. Retrospective analyses

Retrospective analyses, which first became widespread in conjunction with VPA-like models, are useful for evaluating large statistical models. The analyses are simple: the assessment model is rerun several times, each time excluding another year's data. There are several things that can be compared. For example, if one has six analyses (basecase plus removing 1-5 years data) you can compare the following quantities up to the first year that data were excluded:

- Biomass (total and spawning) and recruitment trajectories (both absolute and relative);
- Biological reference points, *e.g.*, B_{MSY} or MSY;
- Stock status in relation to the reference points.

These comparisons will illustrate how much recent data have changed our perspective of the past. Also, it can provide insights on which data are best to use for estimating reference points. For example, the most recent estimates of fishing mortality are often quite uncertain, and may provide biased estimates of important reference points when compared to estimates using the some time period (for fishing mortality), but with additional years data (*i.e.*, multiple observation of the cohorts).

If one retains the observed effort data in the retrospective analyses it will be possible to predict catches into the "future" and compare these to the observed catches. This can be used to evaluate prediction error in forecasting catches given known levels of effort, *i.e.*, we can evaluate the prediction variance by comparing the predicted catches from the retrospective analyses to the observed catches.

In our example there is a clear indication that recruitment during 1997-98 was overestimated by the data available at the time the cohorts first entered the fishery, as the estimates are always lower with the addition of new data (Figure 10). This has a magnified effect on the projected biomass trajectories where the models that excluded the last 3-4 years data predicted higher biomass at the end of the data and further increases in the short term. These estimates were also "revised" downward with the addition of further data.

3. Evaluation of phases of estimation

To obtain convergence in large nonlinear statistical models, it is often necessary to estimate parameters in a number of phases, *i.e.*, several sets of parameters are estimated in the first phase, and then more and more parameters are added to the model in subsequent phases. For example, important scaling parameters like mean recruitment and catchability might be estimated in the first phase, recruitment deviates estimated added in the second phase, and selectivity added in the final phase. The aim of estimation-in-phases is to help the model converge at a global solution by keeping the model in a realistic part of the high-dimension parameter space. Essentially, it involves getting good parameter estimates before adding more parameters.

Because the role of estimation in phases is to improve convergence, often the results may not be robust to alternative orders of estimating parameters. For this reason, it is important to carefully evaluate different approaches. This can be best achieved by evaluating the values of parameters at each phase. Ideally, the important model parameters should not change much in the final phase; rather the model should only be "fine-tuned".

In the example presented above, growth was estimated in the final phase after selectivity had been estimated. While the starting values for growth were generally quite good, the model estimated smaller sizes (than the starting values) for the youngest fish. This significantly changed selectivity estimates (Table 3). While the recruitment estimates changed little in phases 6 and 7, some changed considerably in the final phase (Table 4). This type of result is not ideal, but may be unavoidable (without very good information of growth) for highly correlated parameters such as growth, selectivity, and recruitment. Note that in practice, the parameter estimation is not spread out as much; in fact we obtained identical final results using four and nine phases with the bigeye example.

4. Examination of the solution surface using alternative starting values

In large statistical models the solution surface will be very complex. This is often why we use estimation-in-phases described in (11) above. To ensure that the model has converged to a "global" solution, rather than a local minimum, it is important to start the model using alternative starting values for the model parameters.

In models with large numbers of parameters it is not feasible to manually change the starting values of all the parameters. To overcome this, we used an indirect approach in which a strong penalty is imposed on average fishing mortality in the second-to-last phase and then released in the last phase. Multiple runs should be performed with different values of fishing mortality for the penalty. This represents a simple way to examine alternative starting values for all parameters. Imposing the penalty on fishing mortality has the extra benefit as F is related to stock status, *i.e.*, we are examining convergence from a point of qualitatively different stock status. It is important to see how much parameter values differ in the second-to-last phase to evaluate how thorough the analysis was. In particular, it will be important to see how selectivity varies in the analyses as this is a key parameter.

It was found that the penalty did effect a considerable change in some of the variables in the second-to-last phase. Fishing mortality-at-age did differ as expected (Figure 12), and the model was able to converge back to essentially the same place in the final phase (Figure 13).

Special research questions

During the meeting some specific research questions were raised that were believed to be of critical importance to assessments using large statistical models. These were projects that should be evaluated in the interval between assessments, rather than at assessments. The group did not discuss the relative importance of each or any priority.

1. Using retrospective analyses to evaluate model predictions and estimates of key parameters

While it is recommended that retrospective analyses be used periodically, it was also considered important to investigate how well the approach would work for answering some critical assessment questions, including:

- Determining realistic bounds for catch predictions;
- Evaluating the quality of environmental index predictions; including internal versus external estimation;
- Best period to determine the average fishing mortality-at-age for estimating yields and biological reference points;
- Evaluating alternative assumptions about selectivity by comparisons of catch predictions.

2. Evaluation of selectivity functional forms

Selectivity is a key model parameter, and parameter estimates are sensitive to alternative assumptions about how smooth selectivity is from one age to the next. In A-SCALA there are four smoothness penalties that are imposed:

1. The first difference of selectivity-at-age;
2. The second difference of selectivity-at-age;
3. The third difference of selectivity-at-age;
4. Selectivity penalized to be monotonic for longline fisheries.

For each of these penalties there is a weighting factor; when this is set to zero the penalty is ignored, but there is no objective way to determine non-zero scalars for each of the penalties. The critical question is whether there is an objective way to choose between alternative assumptions about the smoothness of the selectivity curves. Presently, different assumptions for the weighting factors can lead to considerable differences in the estimated curves and important management quantities.

Though a separate selectivity parameter is generally estimated for each age class, the use of smoothness penalties effectively reduces the number of degrees of freedom by some amount. If the effective number of degrees of freedom can be determined, then likelihood-based hypothesis testing could be used to choose between models, but there maybe other methods that don’t rely on determining the effective number of degrees of freedom. The research should include:

1. Searching the literature for methods to estimate approximate degrees of freedom in the presence of smoothness penalties;
2. Investigating generalized/nested selectivity functions that can be used for hypothesis testing;
3. Investigating alternative methods to select smoothness penalties, *e.g.*, cross-validation.

3. Numerically solving the catch equation in A-SCALA

A preliminary investigation of incorporating a method to numerically solve the catch equation within A-SCALA was presented at the meeting. Insufficient testing of the method had been undertaken to fully investigate the performance of this approach over the “effort-deviate” approach presently implemented. As, the model almost exactly fits the observed catch; elimination of effort deviates would greatly reduce the number of parameters to be estimated. For example, in the bigeye model 785 out of 970 (80%) of the parameters were effort deviates. This reduction in the number of parameters may reduce the amount of time required to estimate the parameters and the Hessian matrix (required for deriving confidence limits). This method may make the use of Bayesian methods of integration feasible (through increased computing power).

The initial analysis showed that the new method is not necessarily faster, and that there are some stability issues that must be overcome. There are two main issues that need to be examined are:

- Comparison of computing time required for each approach. In particular examining how any differences scale with increased model complexity (*i.e.*, increased number of fishing fleets);
- Technical issues relating to implementation: Often the simple theoretical approach to a problem like the catch equation does not immediately work in a large-nonlinear model. In order to effectively implement the approach there may be several computational “tricks” required.

4. Investigation of changes in fishing power

In the absence of fishery-independent abundance indices, stock assessments of most large pelagic fisheries rely heavily on catch-per-unit-effort (CPUE) indices. For example, in the bigeye assessment, trends in adult biomass were driven almost solely by the CPUE data from the Japanese longline fleet. Assuming that this index is proportional to abundance, the model estimates significant increases in fishing power of the purse-seine fleets. In reality, we do not know if the longline CPUE is proportional to abundance. In fact, there is strong evidence that nominal CPUE is not.

The group recommended that sensitivity analyses be developed that allow for alternative hypotheses about changes in the fishing power in different fleets. Part of this will involve evaluating how CPUE is presently standardized.

5. Simulation studies

Simulation studies play an important role in testing assessment models for both precision and accuracy of parameter estimates and model predictions. Researchers at the South Pacific Community (SPC) have developed a large simulation model that can be used to create “realistic” data sets for testing models such as MULTIFAN-CL and A-SCALA. Results from a simulation pilot study were presented at the Methods Working Group at SCTB 15 (18-19 July, 2002). The Methods Working Group recommended that these studies continue with more simulated data sets. The group strongly supported involvement in this study.

6. CPUE time series for purse seine fisheries

Over the past several decades there have been many attempts to derive abundance indices from purse-seine catch and effort data. Even given this, the catch and effort data from purse seiners is often given very low weight in the assessment of tuna stocks. The recent increased catches of bigeye tuna associated with floating objects (FADs) has made this task of critical importance to assessing these stocks (worldwide). It is a recommendation of this group (also made at the most recent ICCAT bigeye assessment) that considerable efforts be made analyzing and interpreting the catch and effort data from FAD fisheries, with the goal of calculating indices of abundance for juvenile bigeye.

7. Weighting length-frequency data

We have shown here that the biomass trajectory from the model is sensitive to the assumed sample sizes used for the length-frequency data (Figure 5). Presently, there is no objective method for determining these sample sizes, in particular the relative sample sizes assumed for samples from purse-seine vessels versus longline vessels that are collected in very different ways. We recommend detailed modeling of the sampling process for different fisheries to attempt to derive effective sample sizes for these data.

Acknowledgements

We thank all of the meeting participants (Appendix 1) for their discussions and ideas. In particular we recognize Ramón Conser, who provided an initial list of diagnostics that formed the basis of much of the work presented here. William Bayliff provided many editorial comments to improve an early version of this manuscript.

Appendix 1: List of meeting participants.

Inter-American Tropical Tuna Commission

Richard Deriso

Shelton Harley

Michael Hinton

Mark Maunder

Jenny Suter

National Marine Fisheries Service

Ramón Conser

Paul Crone

Pierre Kleiber

The Secretariat of the Pacific Community

John Hampton

Fideicomiso de Investigación para el Desarrollo del Programa de Aprovechamiento del Atún y Protección de Delfines of Mexico

Michel Dreyfus

Table 1: Summary of information about the mean catchability coefficients for each fleet in the bigeye model. Included in order presented are: phase of estimation, lower and upper bounds, estimates at the end of each phase, and estimated standard error and coefficient of variation. This information is also provided for all other estimated parameters and some derived model quantities and contained in a single S-PLUS list object.

	phz	lb	ub	start	ph1	ph2	ph3	ph4	ph5	ph6	ph7	final	sd	cv
1	1	-500	5	-8.25	-12.38	-12.38	-7.87	-7.93	-7.77	-7.76	-7.76	-7.77	0.1977	0.0255
2	1	-500	5	-6.19	-10.46	-10.46	-5.84	-5.76	-5.92	-5.86	-5.87	-5.73	0.1807	0.0316
3	1	-500	5	-5.49	-10.13	-10.13	-5.47	-5.43	-5.29	-5.25	-3.83	-3.92	0.2777	0.0708
4	1	-500	5	-8.91	-13.04	-13.04	-8.34	-8.30	-8.00	-7.99	-7.99	-8.12	0.2423	0.0298
5	1	-500	5	-8.29	-12.33	-12.33	-7.74	-7.67	-7.75	-7.69	-7.69	-7.57	0.1792	0.0237
6	1	-500	5	-10.89	-15.03	-15.03	-10.56	-10.64	-10.44	-10.43	-10.43	-10.45	0.2021	0.0193
7	1	-500	5	-6.99	-12.25	-12.25	-7.26	-7.30	-6.78	-6.76	-6.76	-6.94	0.2842	0.0409
8	1	-500	5	-6.81	-11.62	-11.62	-6.99	-7.05	-6.59	-6.56	-6.57	-6.63	0.2614	0.0394
9	1	-500	5	-3.03	-8.26	-8.26	-3.44	-3.52	-2.74	-2.70	-2.70	-2.88	0.5288	0.1836
10	1	-500	5	-8.04	-13.40	-13.40	-9.30	-8.98	-8.80	-8.69	-8.70	-8.50	0.1247	0.0147
11	1	-500	5	-7.33	-12.42	-12.42	-8.30	-7.99	-7.82	-7.71	-7.72	-7.53	0.1238	0.0165
12	1	-500	5	-10.23	-14.95	-14.95	-10.83	-10.52	-10.35	-10.24	-10.25	-10.05	0.1239	0.0123
13	1	-500	5	-8.16	-14.25	-14.25	-10.13	-9.82	-9.65	-9.54	-9.55	-9.36	0.1239	0.0132

Table 2: Values for the various penalties and likelihood components in the objective function at the end of each phase of estimation.

	phase 1	phase 2	phase 3	phase 4	phase 5	phase 6	phase 7	final
TotalLike	504.802	504.802	-239032	-239092	-239440	-239448	-239458	-239452
CatLike	6.36603	6.36603	5.9365	5.55654	5.89911	5.83344	5.6916	5.60212
sLike	0	0	0	0	121.193	122.132	121.974	96.9887
qLike	0	0	0	0	0	0	0	0
EffLike	498.404	498.404	435.048	407.28	423.49	421.411	412.446	409.561
steepLike	0	0	0	0	0	0	0	0
RLike	0	0	63.3852	42.318	22.715	17.9488	17.9489	20.7262
RinitLike	0	0	0	2.37745	1.99907	2.05656	2.05011	2.24703
LLike	0	0	-239537	-239550	-240016	-240018	-240018	-240037
GLike	0	0	0	0	0	0	0	50.2313
mean.s.penalty	0	0	0	0	0.221721	0.221903	0.22108	0.123351
monotonic.s.penalty	0	0	0	0	5.0E-05	4.8E-05	4.6E-05	3.2E-04
s.1d.penalty	0	0	0	0	86.9934	87.1147	87.0069	88.2734
s.2d.penalty	0	0	0	0	0	0	0	0
s.3d.penalty	0	0	0	0	34.1991	35.0174	34.9672	8.71533
mean.Effdev.penalty	0	0	0	0	0	0	0	0
sum.Effdev.penalty	0.0117	0.0117	0.0045	0.0042	0.0066	0.0060	0.0051	0.0050
Lsd.penalty	0	0	0	0	0	0	0	0
Rd.penalty	0	0	0	0	0	0	0	0
autocor.logR.penalty	0	0	0	0	0	0	0	0
cv.logR.penalty	0	0	0	0	0	0	0	0
abs.Rdev.penalty	0	0	0	0	0	0	0	0
avg.Ftot.penalty	0	0	0	0	0	0	0	0
qbeta.penalty	0	0	0	0	0	0	0	0
Finit.penalty	0	0	0	0	0	0	0	0
lg.Ftot.penalty	0	0	0.0303	0.0840	0	0	0	0
sm.Ftot.penalty	0.0198	0.0198	0	0	0	0	0	0

Table 3: Estimates of the natural logarithm of selectivity at age by phase for a purse seine fishery.

phz	lb	ub	start	phase 1	phase 2	phase 3	phase 4	phase 5	phase 6	phase 7	final	sd	c.v.
5	-500	5	-13.211	-13.211	-13.211	-13.211	-13.211	-13.211	-13.211	-13.211	-13.211	NA	NA
5	-500	5	-6.281	-6.281	-6.281	-6.281	-6.281	-4.594	-4.487	-4.486	-7.322	0.783	0.107
5	-500	5	-1.461	-1.461	-1.461	-1.461	-1.461	-0.056	0.078	0.078	-3.025	0.876	0.290
5	-500	5	1.219	1.219	1.219	1.219	1.219	1.015	1.113	1.114	-0.387	0.568	1.466
5	-500	5	1.962	1.962	1.962	1.962	1.962	0.804	0.867	0.876	0.715	0.264	0.370
5	-500	5	1.704	1.704	1.704	1.704	1.704	0.636	0.681	0.693	0.980	0.230	0.235
5	-500	5	1.389	1.389	1.389	1.389	1.389	0.659	0.710	0.715	1.043	0.219	0.210
5	-500	5	1.334	1.334	1.334	1.334	1.334	0.660	0.718	0.710	1.166	0.222	0.191
5	-500	5	1.434	1.434	1.434	1.434	1.434	0.671	0.723	0.725	1.220	0.226	0.185
5	-500	5	1.480	1.480	1.480	1.480	1.480	0.714	0.747	0.759	1.223	0.219	0.179
5	-500	5	1.214	1.214	1.214	1.214	1.214	0.689	0.692	0.689	1.245	0.223	0.179
5	-500	5	0.662	0.662	0.662	0.662	0.662	0.561	0.539	0.513	1.131	0.232	0.205
5	-500	5	-0.079	-0.079	-0.079	-0.079	-0.079	0.435	0.388	0.356	0.960	0.257	0.268
5	-500	5	-0.927	-0.927	-0.927	-0.927	-0.927	0.432	0.351	0.327	0.844	0.312	0.370
5	-500	5	-1.842	-1.842	-1.842	-1.842	-1.842	0.594	0.485	0.488	0.826	0.367	0.445
5	-500	5	-2.808	-2.808	-2.808	-2.808	-2.808	0.915	0.804	0.851	0.860	0.359	0.418
5	-500	5	-3.812	-3.812	-3.812	-3.812	-3.812	1.279	1.200	1.276	0.804	0.390	0.485
5	-500	5	-4.851	-4.851	-4.851	-4.851	-4.851	1.444	1.410	1.442	0.481	0.803	1.669
5	-500	5	-5.924	-5.924	-5.924	-5.924	-5.924	1.116	1.116	1.041	-0.272	1.330	4.886
5	-500	5	-6.992	-6.992	-6.992	-6.992	-6.992	0.114	0.136	-0.029	-1.458	1.688	1.158
5	-500	5	-7.951	-7.951	-7.951	-7.951	-7.951	-1.465	-1.437	-1.635	-2.946	1.784	0.606
5	-500	5	-8.659	-8.659	-8.659	-8.659	-8.659	-3.373	-3.348	-3.525	-4.561	1.637	0.359
5	-500	5	-9.000	-9.000	-9.000	-9.000	-9.000	-5.332	-5.315	-5.442	-6.136	1.308	0.213
5	-500	5	-9.000	-9.000	-9.000	-9.000	-9.000	-7.080	-7.071	-7.140	-7.508	0.867	0.115
5	-500	5	-9.000	-9.000	-9.000	-9.000	-9.000	-8.368	-8.365	-8.388	-8.510	0.396	0.047

Table 4: Estimated of the recruitment deviates by phase for the last 20 quarters for which length-frequency data was available.

phz	lb	ub	start	phase 1	phase 2	phase 3	phase 4	phase 5	phase 6	phase 7	final	sd	cv
3	-15	15	0	0	0	0.656	0.713	0.648	0.104	0.091	-0.210	0.400	1.908
3	-15	15	0	0	0	0.181	0.225	0.046	-0.549	-0.544	0.254	0.334	1.314
3	-15	15	0	0	0	-0.120	-0.068	0.415	-0.243	-0.216	0.071	0.363	5.097
3	-15	15	0	0	0	0.636	0.737	0.937	0.717	0.744	0.985	0.352	0.357
3	-15	15	0	0	0	0.931	1.033	0.491	0.234	0.244	0.816	0.281	0.345
3	-15	15	0	0	0	1.682	1.772	1.107	0.618	0.638	-0.098	0.367	3.748
3	-15	15	0	0	0	0.447	0.492	0.439	-0.036	-0.027	-0.955	0.368	0.385
3	-15	15	0	0	0	0.156	0.245	-0.045	-0.161	-0.158	-0.970	0.344	0.355
3	-15	15	0	0	0	-0.757	-0.655	-0.672	-0.514	-0.510	-0.696	0.363	0.522
3	-15	15	0	0	0	-1.529	-1.437	-1.008	-0.902	-0.899	-0.420	0.316	0.752
3	-15	15	0	0	0	-1.314	-1.199	-0.835	-0.646	-0.647	-0.185	0.306	1.656
3	-15	15	0	0	0	-0.594	-0.487	-0.448	-0.206	-0.205	-0.460	0.385	0.837
3	-15	15	0	0	0	0.271	0.260	-0.339	-0.173	-0.167	-0.748	0.358	0.478
3	-15	15	0	0	0	-0.357	-0.429	-0.562	-0.457	-0.453	-0.356	0.389	1.093
3	-15	15	0	0	0	-0.890	-0.995	-0.751	-0.304	-0.306	0.210	0.429	2.041
3	-15	15	0	0	0	0.222	-0.159	-0.248	0.397	0.397	-0.374	0.457	1.221
3	-15	15	0	0	0	0.819	-0.656	-0.115	0.481	0.493	-0.101	0.547	5.406
3	-15	15	0	0	0	-0.005	-1.050	-0.812	-0.549	-0.535	0.237	0.555	2.341
3	-15	15	0	0	0	4.108	0.716	0.193	0.287	0.306	-0.165	0.554	3.359
3	-15	15	0	0	0	0.150	-0.069	-0.064	-0.019	-0.020	-0.060	0.569	9.428

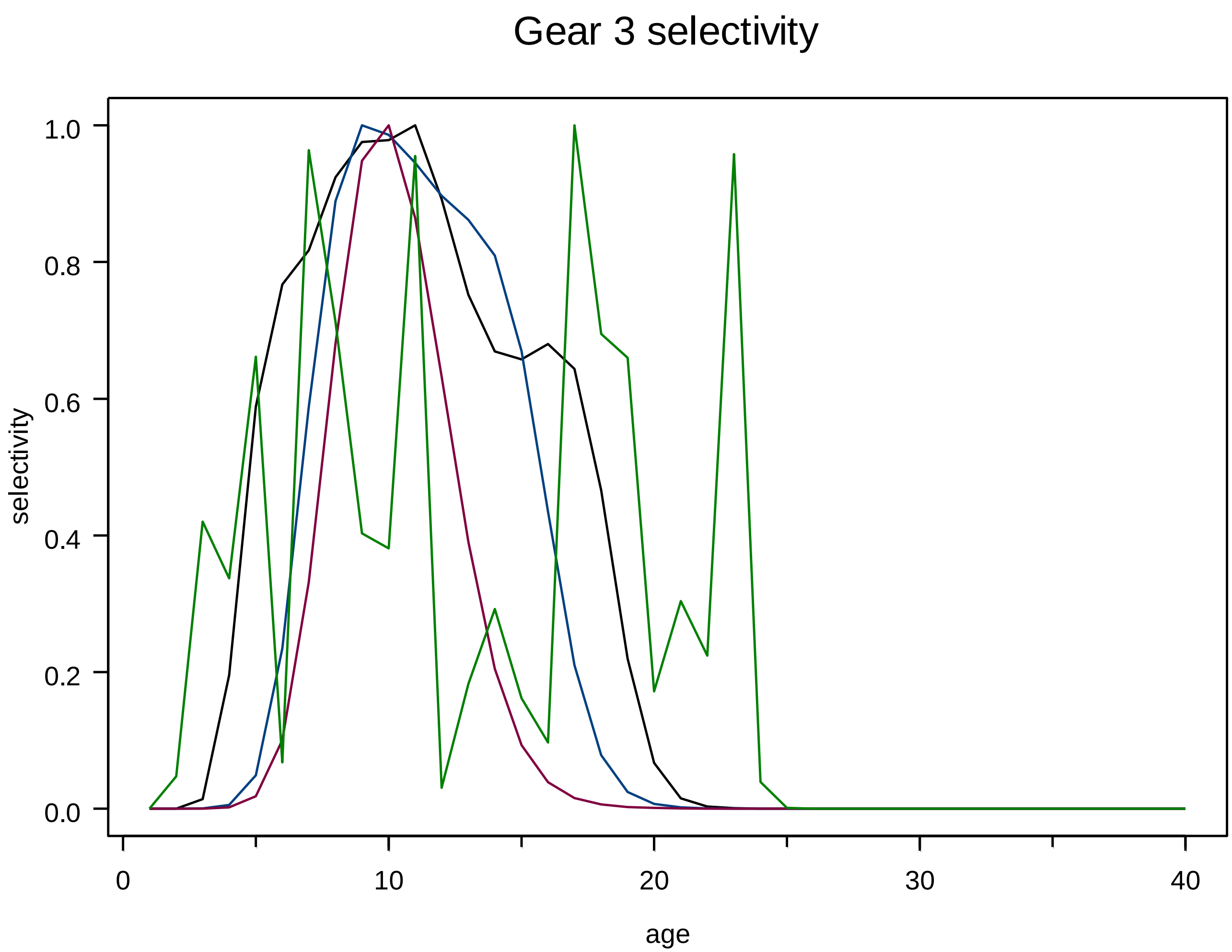

Figure 1: Estimated selectivity curves for a purse seine fishery for bigeye tuna based on different assumptions for the weightings applied to the smoothness parameters.

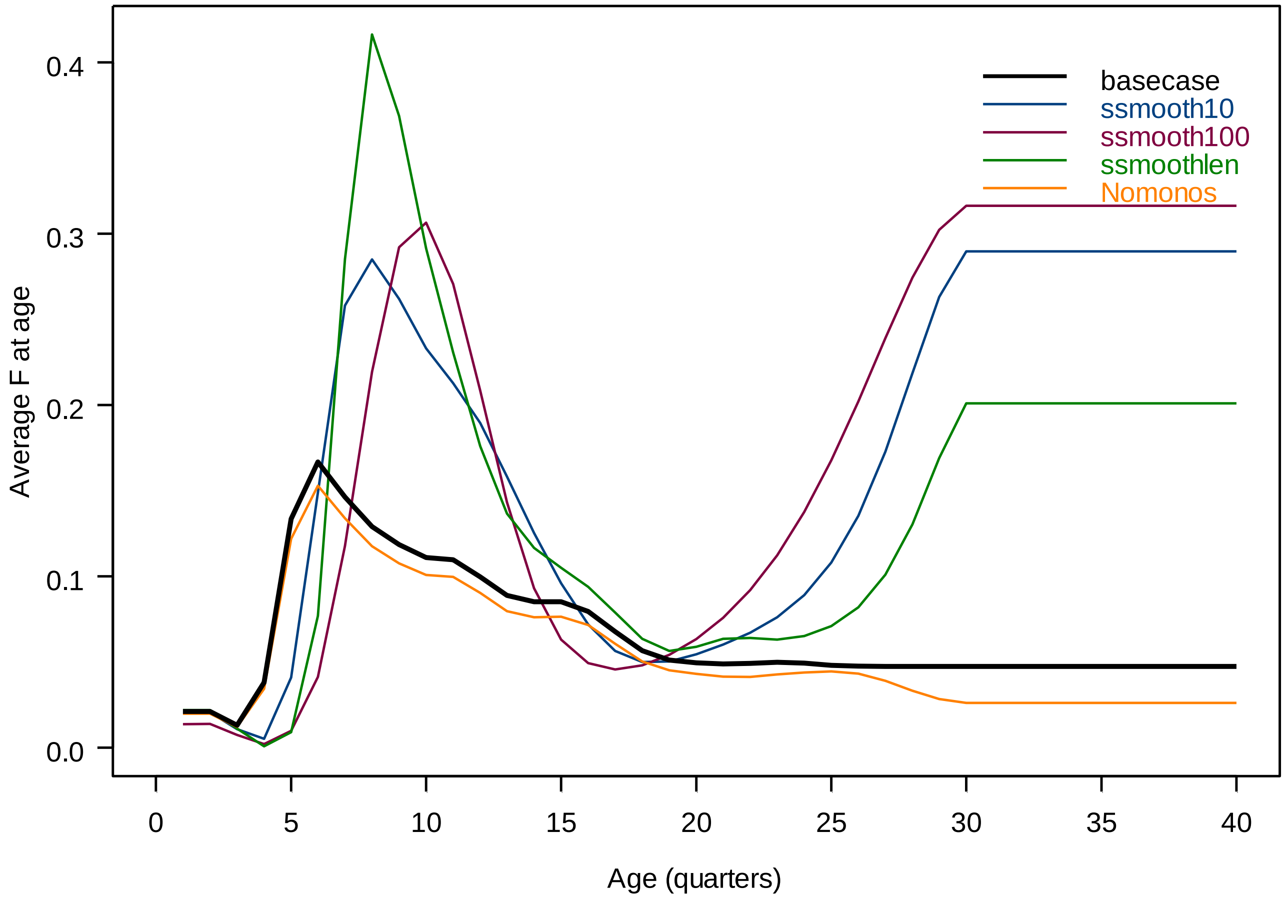

Figure 2: Estimated average fishing mortality-at-age for the most recent two years using different assumptions for the weightings applied to the smoothness parameters.

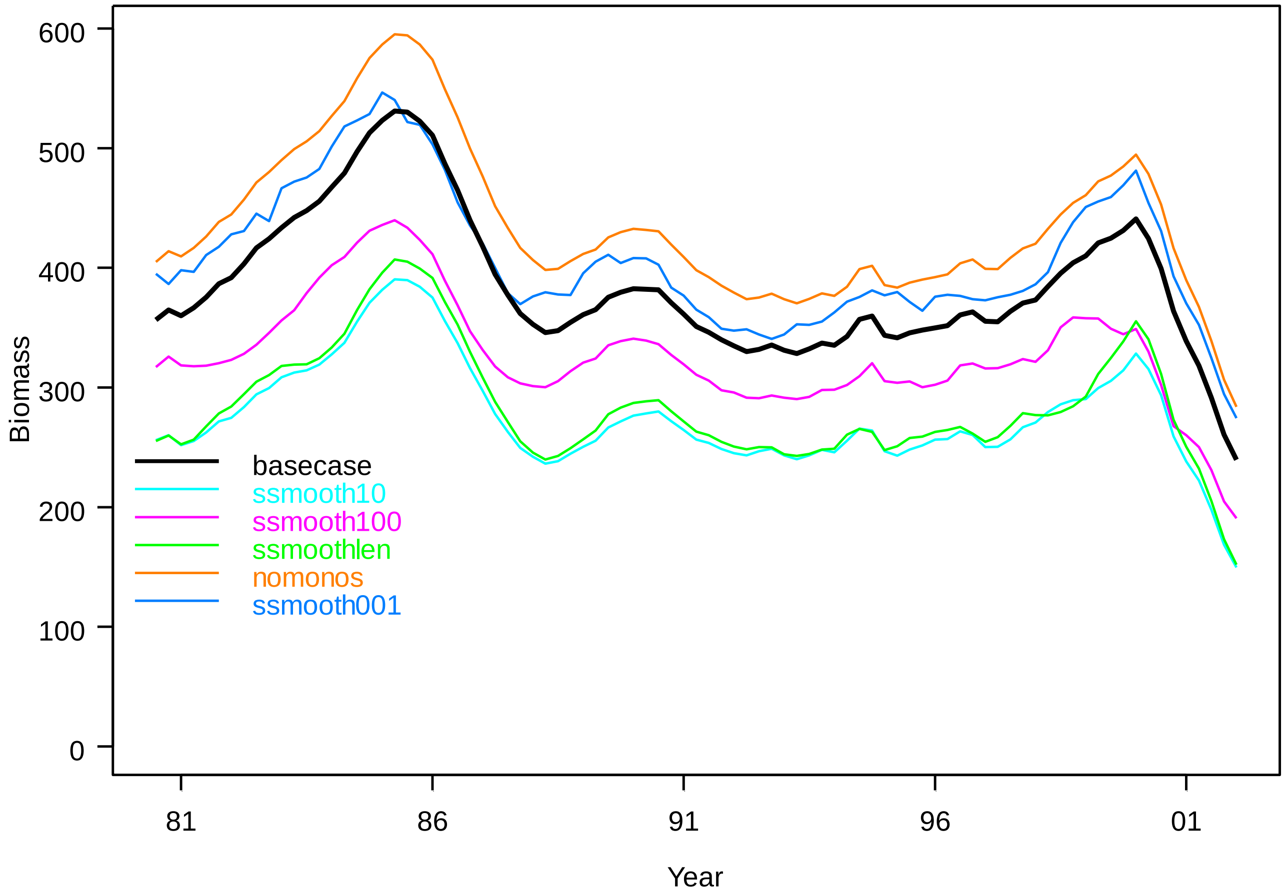

Figure 3: Estimated biomass trajectories using different assumptions for the weightings applied to the smoothness parameters.

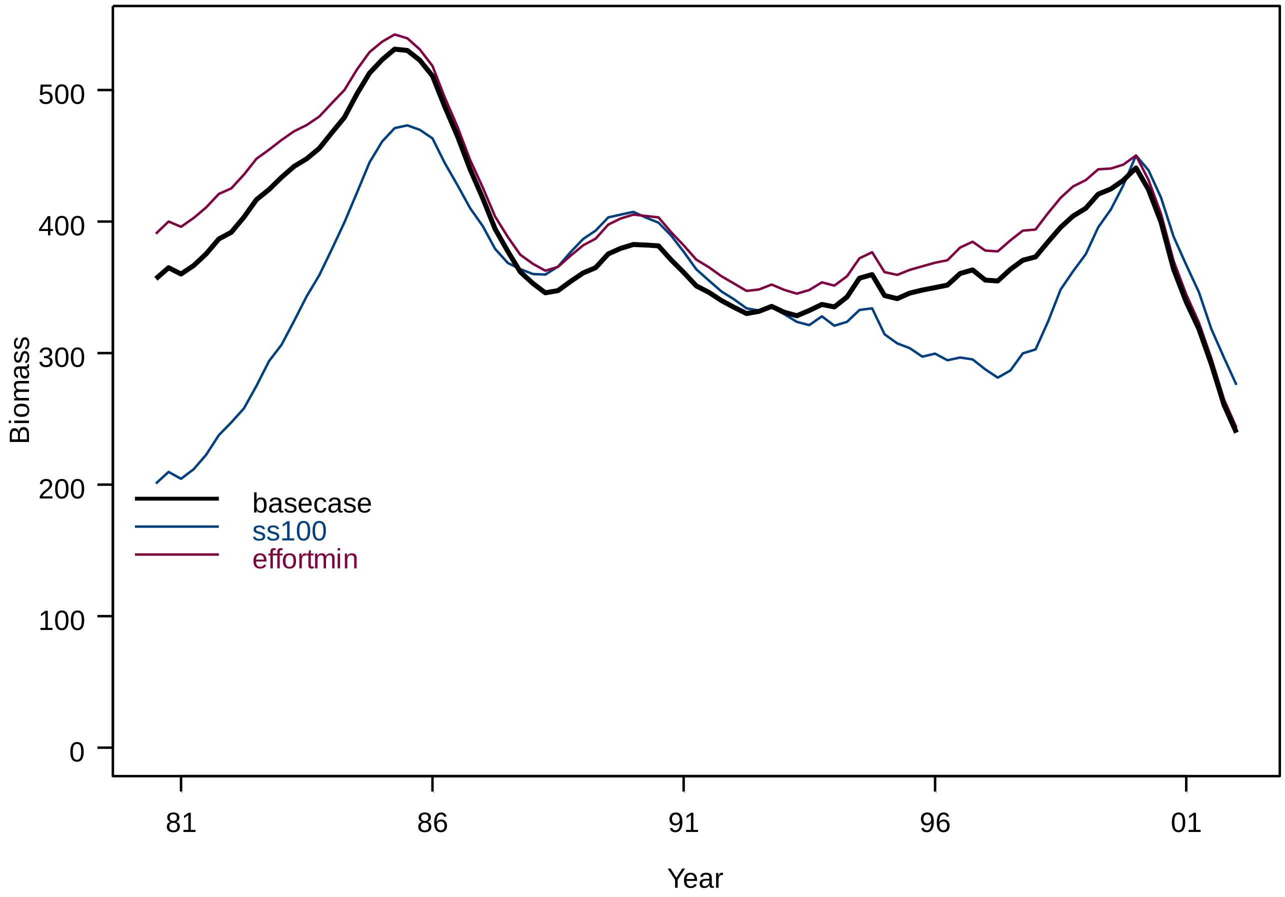

Figure 4: Estimated biomass trajectories for the basecase model and two sensitivity analyses to the weighting applied by the model to the catch and effort data and the length-frequency data.

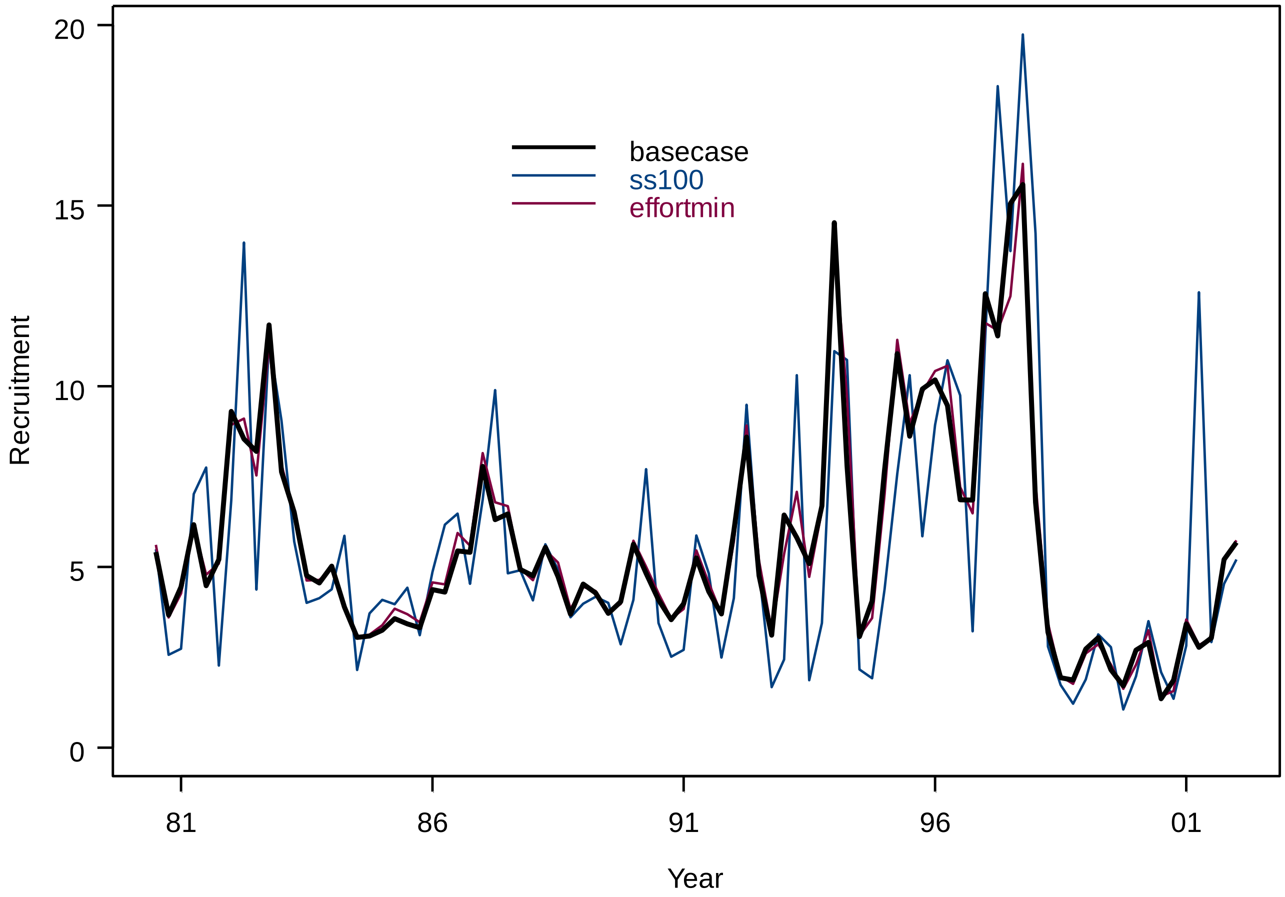

Figure 5: Estimated recruitment time series for the basecase model and two sensitivity analyses to the weighting applied by the model to the catch and effort data and the length-frequency data.

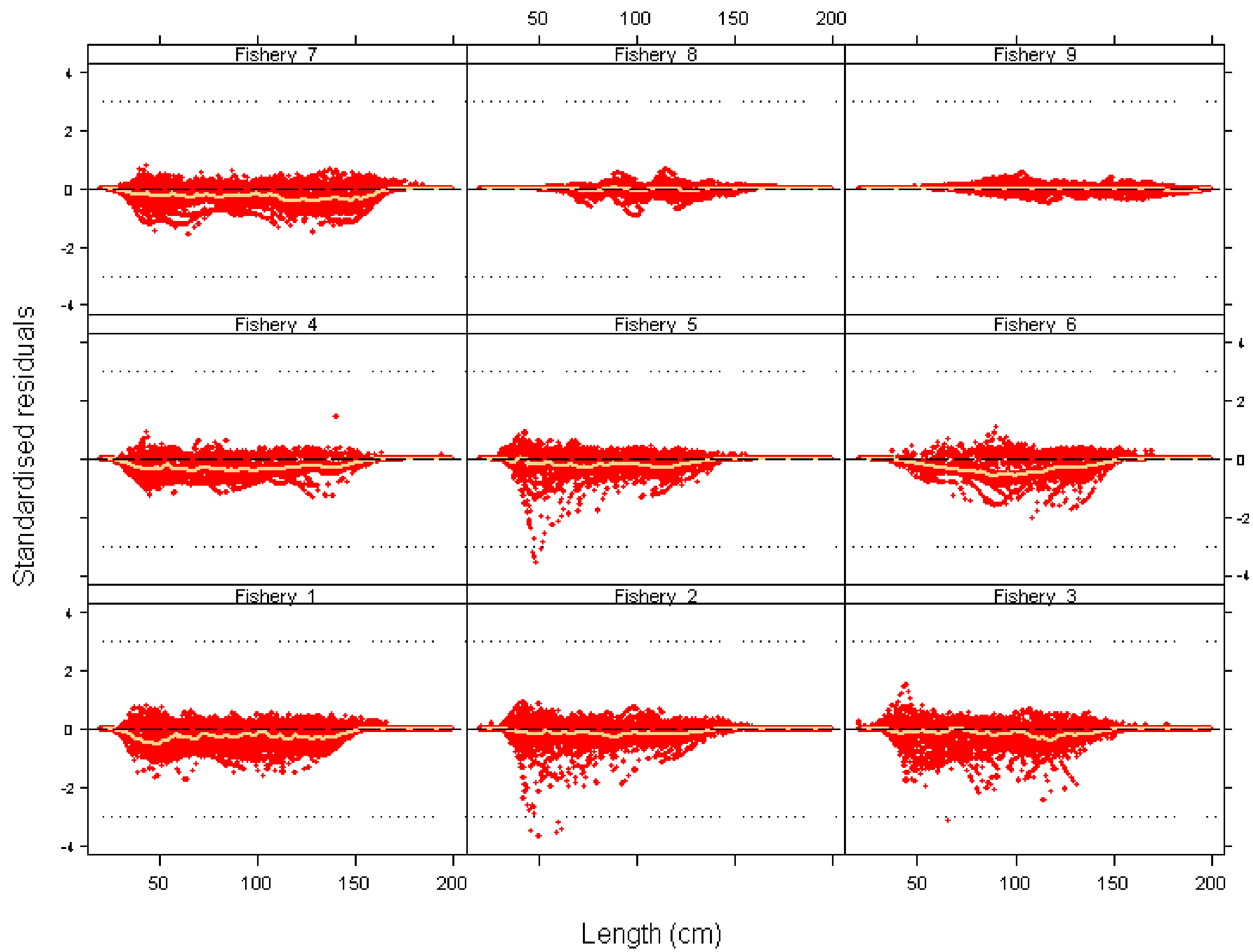

Figure 6: Standardized residuals by length class and fishery for the fit to the length-frequency data for the basecase model for bigeye. A weighted mean trend line is fit through the data as a summary.

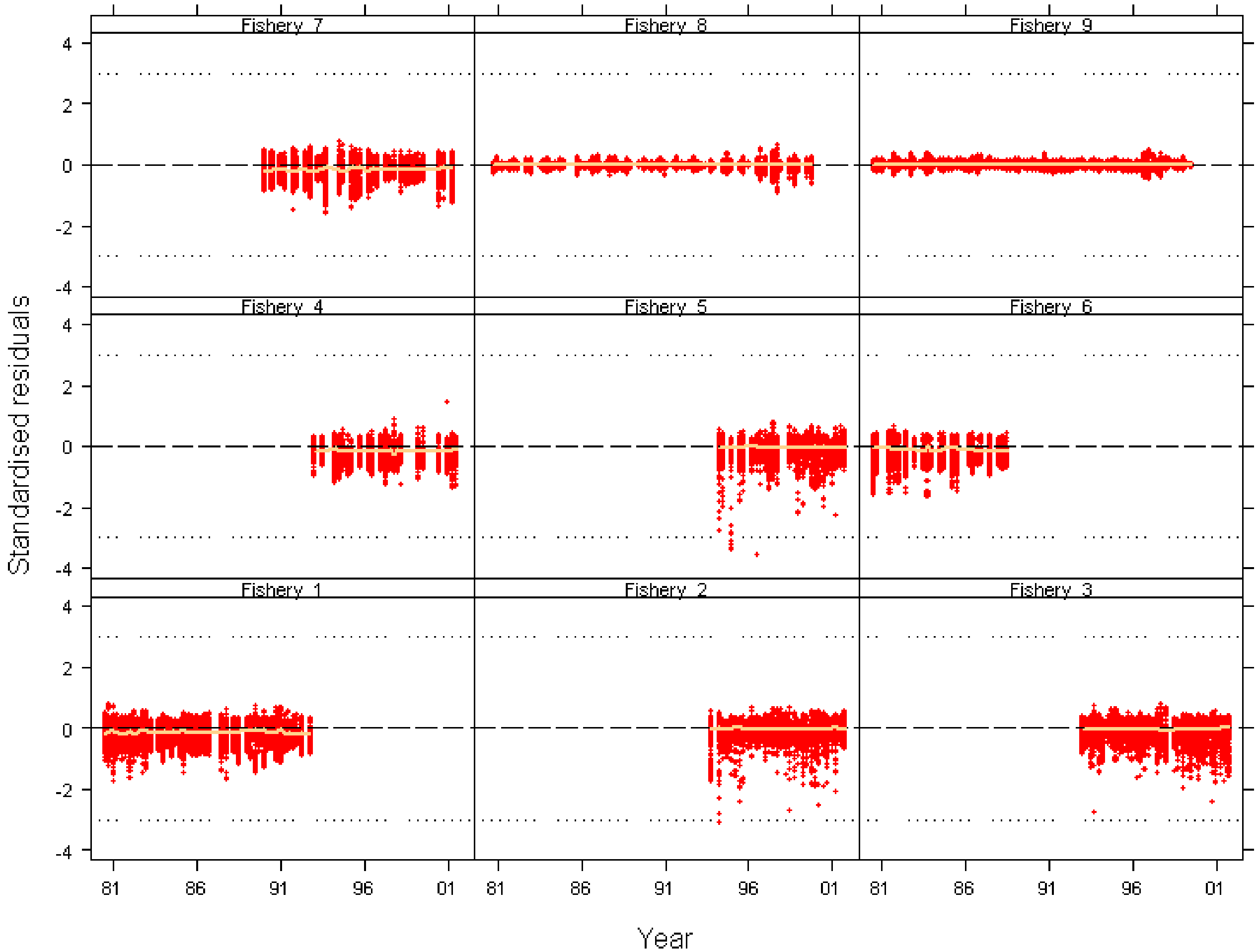

Figure 7: Standardized residuals by year and fishery for the fit to the length-frequency data for the basecase model for bigeye. A weighted mean trend line is fit through the data as a summary.

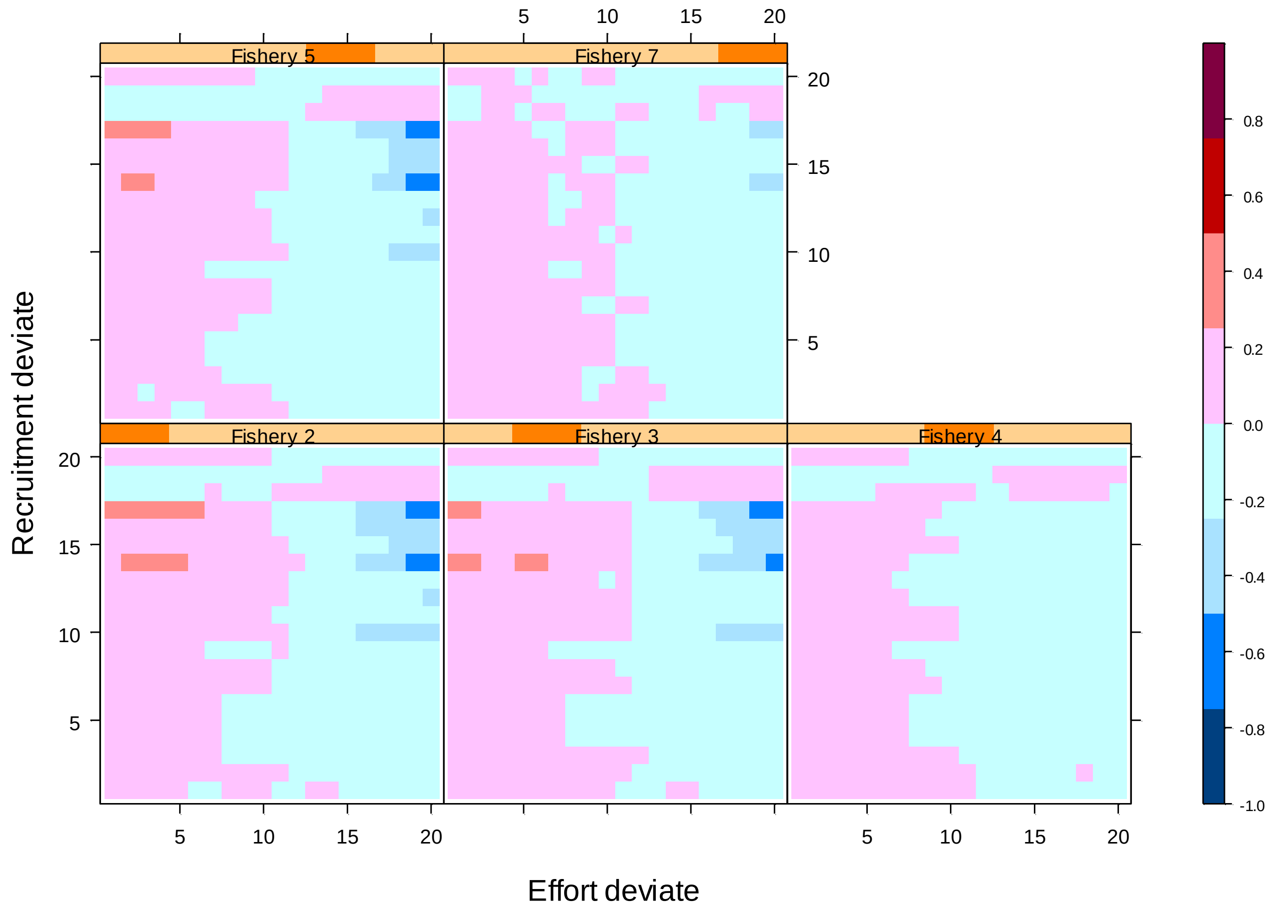

Figure 8: Correlations between estimates of recruitment deviates and effort deviates for five surface fisheries from the basecase model for bigeye for the last twenty quarters. The top right corner of each panel represents the most recent period.

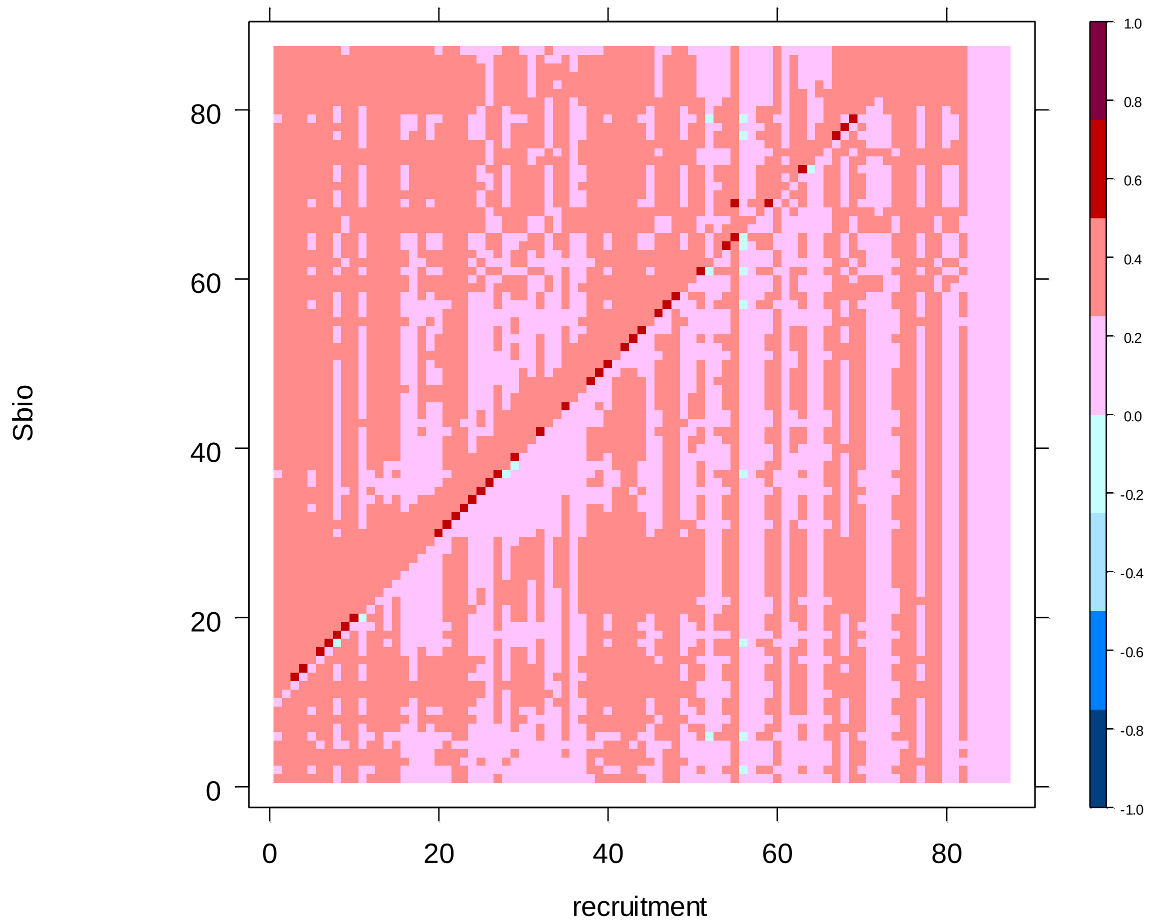

Figure 9: Correlations between spawner biomass and recruitment for the basecase model for bigeye.

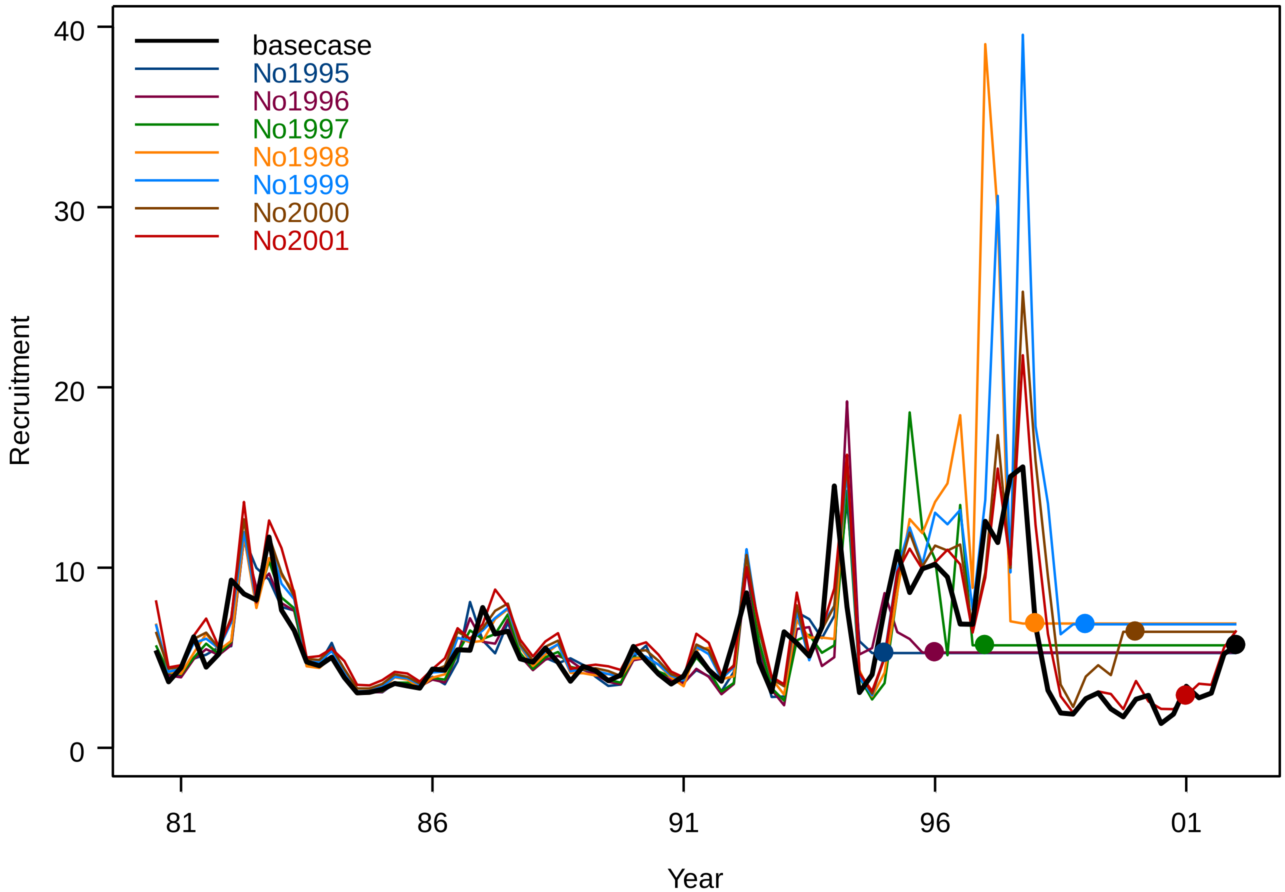

Figure 10: Estimated recruitment trajectories from the retrospective analyses. The solid points represent the last estimate for which data were available. Lines after the point represent average recruitment for each model.

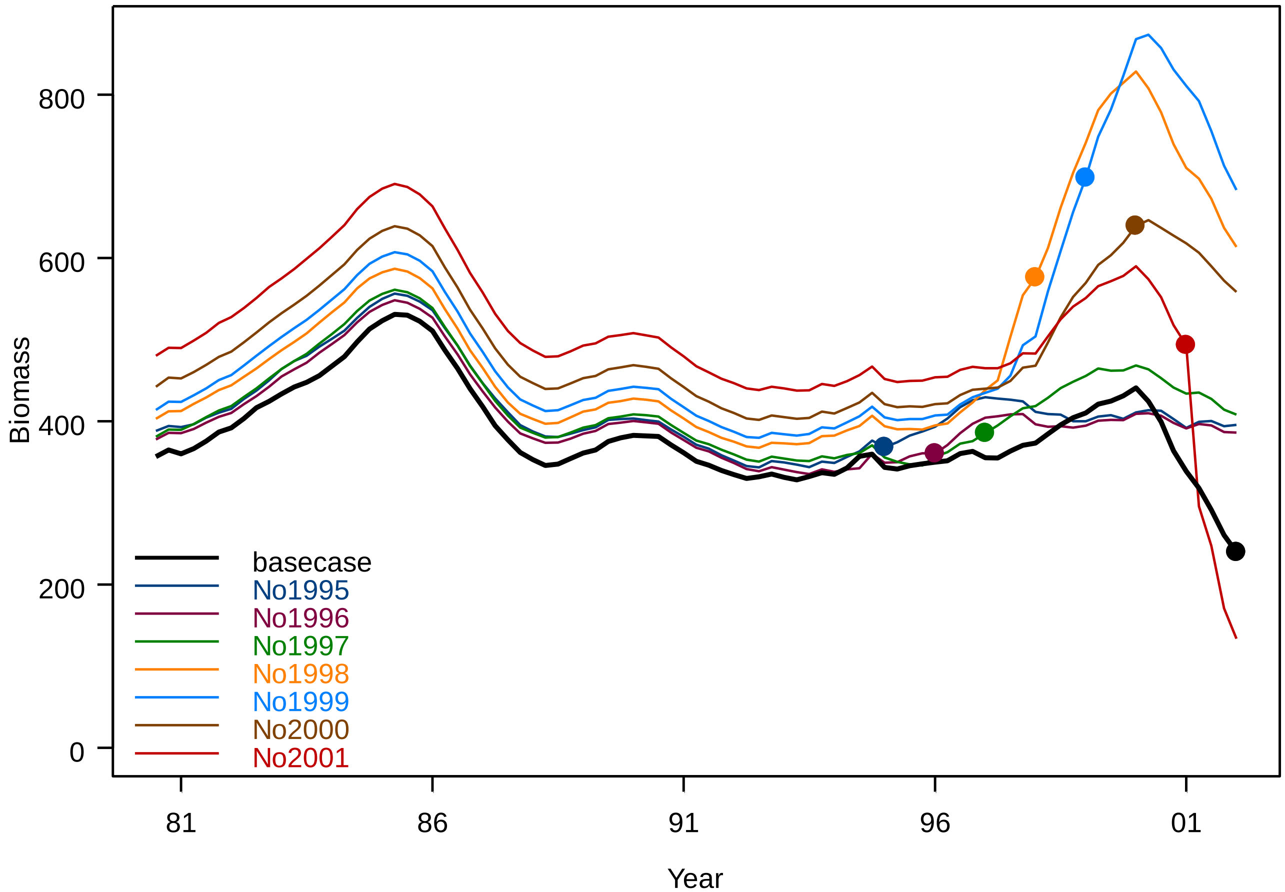

Figure 11: Estimated biomass trajectories from the retrospective analyses. The solid points represent the last estimate for which data were available. Lines after the points represent predictions based on observed effort.

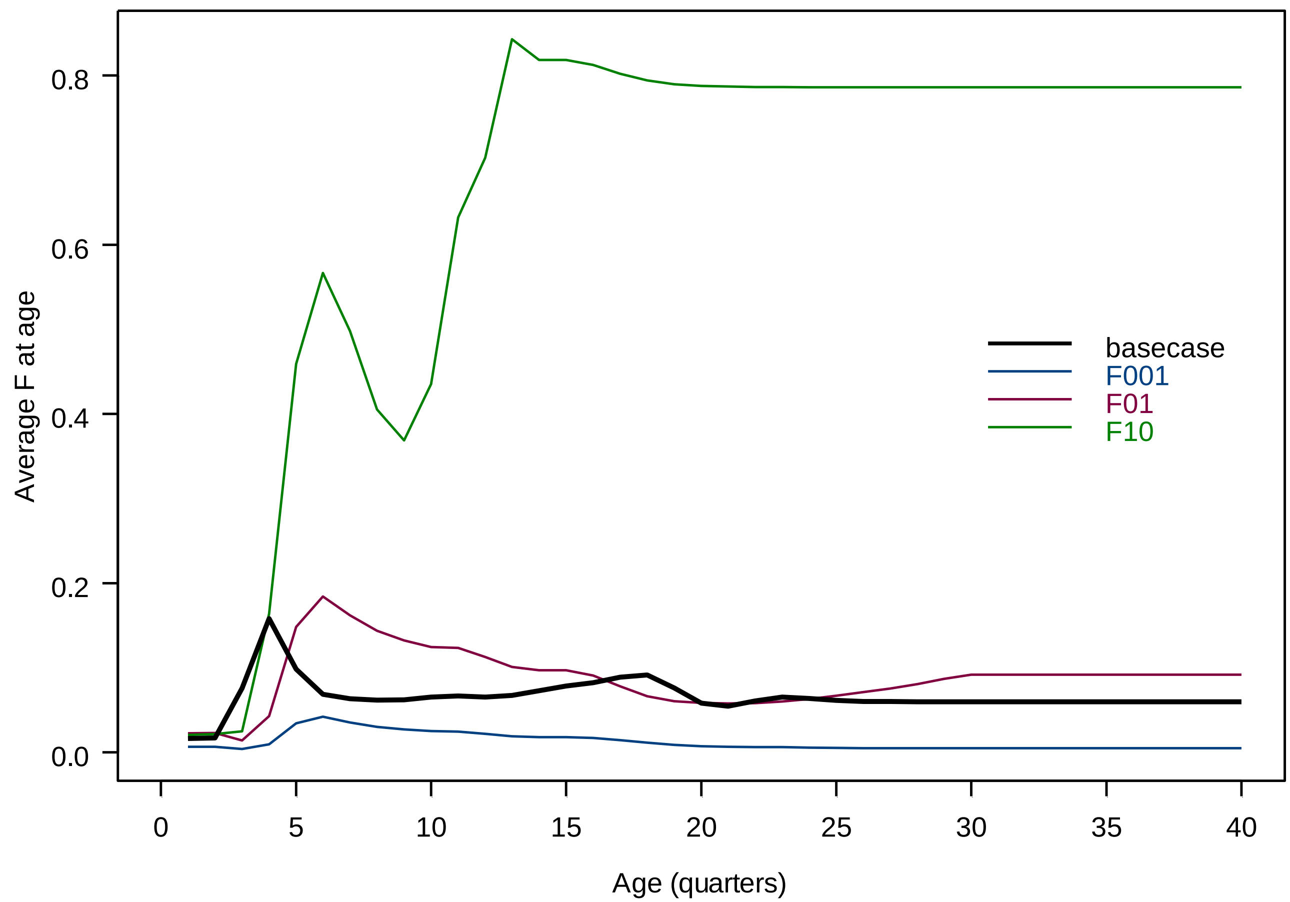

Figure 12: Estimated average fishing mortality-at-age for the last two years at the end of the second to last phase of estimation for different penalties.

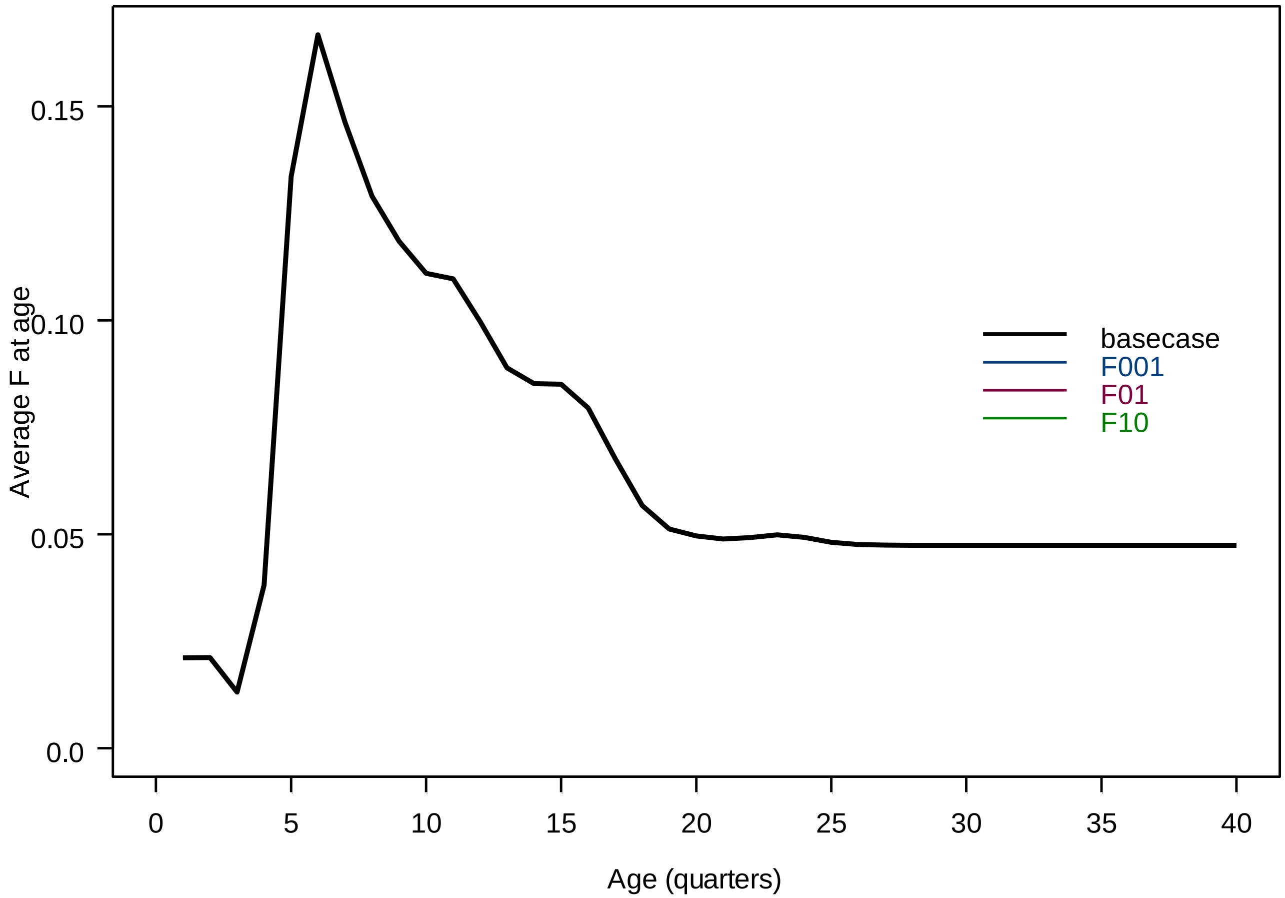

Figure 13: Estimated average fishing mortality-at-age for the last two years at the end of the final phase of estimation for different penalties. All runs converged to the same solution.