A very preliminary attempt to apply Multifan CL to Indian Ocean bigeye tuna

M. Basson
CSIRO Marine Research
Castray Esplanade
Hobart, Australie

Abstract

This working paper presents first attempts to apply the Multifan-CL software to bigeye tuna in the Indian Ocean. This software has the advantage of being able to use size frequency data, which the previously used age-structured production model does not do directly. A simple, single region Multifan-CL model was fitted to data. Catch and size data up to 2001 were used, but the ‘old’ standardised CPUE series from previous assessments, i.e. only up to 1999, was used. The way in which data are required by the software and the decisions which need to be made in this regard are discussed. Some of the settings which define main assumptions about model structure are also considered.

Results, for runs considered here, are most sensitive the relative weighting of size frequency data and assumptions about catchability associated with the standardised CPUE series, but also somewhat sensitive to the length of the time-series and assumptions about natural mortality. The sensitivity analyses are not comprehensive. The goodness of fit to size frequency data is very briefly considered, but in an actual assessment, more attention should obviously be paid to diagnostics than has been the case in these preliminary attempts. Another omission in this paper is scrutiny of covariance matrices of parameter estimates.

Although the preliminary results are promising, further thought needs to be given to the data issues, such as, decisions about the definition of fisheries, length of data series, input values for biological parameters, priors and penalties. These results should not be taken as ‘an assessment’ without further data-checking and evaluation.

1. Introduction

This paper, as the title indicates, is a first attempt to apply the assessment software 'Multifan-CL' (see Fournier et.al. 1990 and 1998; Kleiber et.al. 2002) to Indian Ocean bigeye tuna. The paper does NOT provide an assessment as such, but instead tries to highlight the important decisions that need to be made when starting to develop such an assessment and the tricky issues that need to be borne in mind when using the software. I explore the sensitivities of results to different model or fitting assumptions, but have not examined the diagnostics as carefully as they should be for a 'proper' assessment.

Previous assessments of bigeye tuna have used catch and CPUE in stock production models or in (so-called) age structured production models (e.g. Ricard and Basson, 2002). It has been noted in IOTC reports and in discussion at WPTT meetings, that there are size frequency data for the longline and purse seine fisheries which could in principle be used in an assessment. The age-structured production models do in fact require selectivity at age to be input. These have in the past been derived outside the assessment, by using the size frequency data to devise a catch at age matrix (using e.g. cohort slicing) and then applying cohort analysis or simple VPA to estimate selectivity at age.

The intention was initially to write a relatively simple size-structured model specifically for bigeye, but time and other commitments did not allow this. Instead, I decided to try an existing piece of software. Although we have in-house assessment models (e.g. Kolody et.al. 2004) which could have been tried, there are advantages in using a widely distributed existing piece of software such as Multifan CL. In addition, the method is currently undergoing simulation testing (see e.g. Labelle 2002 and 2003), and there is the potential to subsequently develop a version which incorporates spatial structure and movement. (The capability is already in the software; the structure and data would need to be defined for IO bigeye).

Kolody et. al. (2004) present results from a recently completed large simulation study which tested a wide range of assessment methods. The simulation model generated tuna-like fisheries data. Based on their results, the authors suggest that, although Multifan-CL is at the forefront of single species assessment model development in most respects, they would not yet want to see it universally adopted, if it meant the cessation of development of alternatives. It is interesting to note that the authors were NOT left with a good impression of (at least their implementation of) age-structured production models. These models were prone to numerical problems, and generally required unrealistically good prior knowledge to yield performance comparable with the complicated models. The age-aggregated production models (particularly Fox) yielded results that were better than expected in most cases[1]. The authors also caution against attempting to estimate natural mortality in the complicate models which allow this (e.g. Multifan-CL).

In summary, by using Multifan-CL for bigeye, I am not suggesting that this is the best or only way of proceeding, but I consider it worth trying. The key point is that, as with all off-the-shelf software, it needs to be used carefully.

2. Assessment method: Multifan

This section provides a brief summary of Multifan-CL, particularly in terms of its simplest incarnation as applied to bigeye. Details of input data are given in subsequent sections.

Full documentation can be found in the Multifan-CL user's guide (Kleiber et. al., 2003; the web site www.multifan-cl.org) and associated papers (e.g Fournier et.al. 1990, 1998, and yellowfin assessments in the Pacific). The description here relies heavily on the content of the user's guide. The software implements a statistical length-based, age-structured assessment model. The current version of the software includes spatial structure, fish movement and can use tagging data, but these features have not been used in this application to bigeye tuna. The simplest form of the model fits to time series of catch and size composition data, in terms of length and/or weight frequency distributions. Other inputs include effort data and values, or priors, for biological parameters. The software also requires a file which specifies model and fitting characteristics (e.g. whether selectivity is dome-shaped or not, how to weight data or how to treat penalties in the likelihood). It is in this last input file for running the assessment where most user errors are likely to occur!

The stock dynamics are basically standard fisheries population dynamics with a Beverton-Holt stock recruit function and von Bertalanffy (or modified von B) growth. Selectivity is modelled in terms of size and the shape of the selectivity curve is flexible, but it is assumed to be constant over time. This can lead to relatively poor fit in some years and would not be able to do well where there are systematic changes. Unlike many other stock assessment methods, Multifan-CL allows for, and estimates, changes in catchability over time with a great deal of flexibility as how often changes occur. The method also allows for estimation of natural mortality, but in the absence of tagging data or much information on this in size frequency data, I would use this feature with great caution.

For the bigeye implementation there are essentially two components to the likelihood: a catch component and a size-frequency component. The catch data are assumed to be accurately measured, so the differences between the logged true and observed catches are assumed to be normally distributed with relatively low variance (CV less than about 10%). The size frequency data are fitted with a mixtures of normal distributions with means (i.e. mean length at age) defined by the von Bertalanffy growth curve. Standard deviation of mean length can be modelled as constant over all age classes or an increasing or decreasing function thereof. Here a constant variance was assumed.

3. Data decisions: the '.frq' file

---

[1] There is, of course, a potential complication when applying age-aggregated stock production models (e.g. Fox) to a fishery which started out as longline only, but subsequently developed a purse seine fishery, and where the only index of abundance is from the longline fishery. It is possible to convert the purse seine catch to 'longline equivalent' with the appropriate time lag (e.g. Butterworth... SBT paper).

Defining 'Fisheries'

Unlike some of the previously used assessment models, Multifan-CL ideally requires catch and effort (or a proxy for effort) for each distinct fishery. This means that instead of simply being able to give a CPUE series and total CATCHES for all fleets as in a production model, each 'fishery' should ideally have catch, effort and, if available, size and/or weight frequency. This in turn means that more care needs to be taken when defining the 'fisheries'. Although the effort can be 'missing', the Multifan users' manual suggests that it is better to provide a proxy (even catch, for example) rather than a whole missing effort series.

For the first attempts shown here, three fisheries were defined:
fishery 1 = Japanese longline
fishery 2 = all other longline
fishery 3 = all purse seine

The reason for separating Japanese longline is because we usually have a standardised cpue (or standardised effort) to use for this fleet. Any other longline fleet (e.g. Taiwanese) for which a standardised series may be available can and should probably also be treated separately. All the other 'miscellaneous' catches by longline gear, or catches for which effort has not been standardised can be grouped together. Other gears which may have a selectivity pattern similar to that of longline could also be grouped with this 'other longline' fishery.

The reason for separating longline and purse seine is obvious - the selectivity patterns are substantially different. At this stage all purse seine fleets were grouped together because those that took the bulk of the PS catches seemed to have very similar size frequency distributions.

It is important that the scientist who know the data well (I am not one of those scientists!) consider what the most sensible fishery definitions might be. There are advantages in being parsimonious, while ensuring that the major differences between 'fisheries' are reflected in the definitions. It is also probably not worth treating gear types as separate fisheries if that gear type is only taking a very small percentage of the catch. As usual with assessments, there is also great merit in trying a few different fishery definitions to check whether results are sensitive to the choice.

Catches by other gear types, such as bait boat and gillnet, have not been included here. The catches by the three fisheries included in the input files add up to 99% of the catch over the period 1992-2001 (earlier years were not checked). Given the intent of these trial runs, I consider this to be a very minor issue which can easily be resolved.

Time and spatial scales

For simplicity I chose an annual time scale, so that annual catches were input. The runs assumed that the timing of fishing was in the middle of the year. It would be possible to input quarterly, monthly, or even weekly catches, but these options then raise other questions. For example, how best to deal with CPUE if catches are input quarterly, but the standardised CPUE is annual? Can some fisheries be on an annual scale and others a quarterly scale? I think the answer to that is 'no', unless a fishery only operates in one of the quarters, so that its annual catch = catch in that one quarter.

I chose a single area or region (the whole of the Indian Ocean) for the purposes of these initial trials. The main reasons are (1) simplicity and (2) the fact that there are not yet tagging results which could inform the model about movement or inform us about sensible area definitions.

The length of the time series can (and does) also affect assessment results. I chose to do runs starting with data in 1960 and a few additional trials including data from 1978 only, for all fisheries.

Catch and effort inputs

The catch data for each fishery were extracted from the raised catch at size data files (e.g. betllcyqsf03.txt, betpscyqsf03.txt) since these catches in weight matched well with those in the "Nominal Catch" Tables in the IOTC data summary reports. For fisheries 2 (other LL) and 3 (all PS) catches were input in weight. For the

Japanese LL fishery the catches have been input in terms of numbers to match up with the standardised CPUE which was in numbers.

Effort for the japanese LL fleet is the most important effort component from the model point of view. The standardised CPUE series used in the 2002 assessments (see e.g. Ricard and Basson, 2002) was used to derive a notional 'standardised effort' as:

'standardised' Effort = total catch in numbers / Standardised CPUE index

This index ends in 1999, and the trial assessments were all run only up to 1999. It would be relatively easy to update the input file with results from a new standardised index. Effort was input as 'missing' for 2000 and 2001.

The other two fleets followed the suggestion of the Multifan-CL manual and catches were used as proxy effort. I consider this to be rather unsatisfactory, however, and it may be possible to devise a better proxy at least for some components. It may be better to enter number of hooks, scaled up to total catches for the 'other' longline fishery, but that would require some conversions from catch in weight to catch in numbers (or vice versa) since not all catch data are reported in the same units. Effort for the purse seine fleets is obviously even more tricky since effort is also not recorded in a single unit. The simplest would be to down-weight these fleets in the model, at least as far as their catch-effort contribution to the fit is concerned.

Size frequency inputs

The raised catch numbers at size files (betllcyqsf03.txt and betpscyqsf03.txt) were used to extract annual size frequencies for the three fisheries. These are meant to be (Multifan-CL manual) 'random samples of the catch .. .' The raised size frequencies are not exactly this. The raised size frequency distributions were, however, re-scaled to reflect either the actual sample size (for that fishery in that year), or the average percentage of the catch sampled in that year (e.g. for the purse seine fishery). For years where the sampling level was unavailable, or where a size frequency distribution from another year or fleet was assumed, a very low sample size was used with the hope that this would effectively down-weight those data. They could also have been input as 'missing'.

The sum of the numbers at size determine the year and fishery specific weighting in the model fit, but the run-time flags can also be set to downweight or re-weight the whole series for each fishery. It appears, however, that sampling for the Japanese LL fishery was much higher in the mid 1970-s and 1980's (Figure 1). This would tend to imply high weight for these samples. If the selectivity has been constant over time, then this should not be a problem, but if the selectivity has changed, then a high weight on these samples could lead to a somewhat distorted selectivity for recent years. Note that Multifan-CL does not allow for time-varying selectivity (as far as I can see in the current version, at least).

Two options of year-fishery weights were therefore considered:
I. Base Case
F1 = Jap.LL: actual sample size or 1% of catch in numbers for 1960-1964 (inferred from 1965)
F2 = other LL: 1% of catch in numbers (SF inferred from Japanese samples)
F3 = PS : sample size is % sampled, averaged over SP, FR, NEI-EUR and SYC (see Figure 2)

II. Alternative (down-weight mid 70's and 80's SF - runs called ‘SF2’)
F1 = Jap.LL: actual sample size where that is <10% or 10% of the catch in numbers (1% for 1960-1964 as in I)
F2 = other LL: 1% of catch in numbers, up to a maximum of 1000
F3 = PS : as in I; sample size is % sampled, averaged over SP, FR, NEI-EUR and SYC

Note also that the assumption for F2 in the Base Case could lead to artificially high sample weights when the catches are high eventhough the percentage sampled is assumed to be only 1%. The alternative weighting scheme attempts to avoid that situation.

The data discussed in this section are all essentially in the '.frq' input file. (An electronic version of this file has been sent to the secretariat. I do not guarantee the file to be bug free!)

4. Other Inputs: the '.ini' file

There are a few other inputs required by Multifan-CL. These define the population dynamics assumptions, such as the number of age classes, proportion mature, starting values for some parameters (e.g. mortality and growth) or values to use if these parameters are not estimated.

One important point to note is that the AGE of the first age class is not explicitly entered. Although the first age class can therefore be any age, the default time-lag between recruitment and spawning (or adult) stock is 1, and I don't know whether a value of 0 can be entered (larger values can be entered). The growth parameters are estimated in terms of the mean size of the first age class (L1), the mean size of the last age class (LA) and the growth rate, K. It is therefore important to make sure that the proportion mature (which is entered) is compatible with the mean sizes at age.

Inputs were as follows: (value, # comment - what is input?)
Number of age classes 15
von Bertalanffy parameters (mean length 1, mean length nage, K)

	Initial value[2]	Lower bound	Upper bound
mean L, first age class	42	10	50
mean L, oldest age class	220	210	250
von Bertalanffy K value	0.1	0.05	0.4
Length-weight parameters	3.66e-05	2.90	

Variance parameters (Average SD of length by age class, SD dependency on mean length)

	Initial value	Lower bound	Upper bound
Average SD of length at age	7.0	3.0	10.0

Proportion mature by age class (if the first is age 1) 0 0 0.5 1 1 1 1 1 1 1 1 1 1 1 1

Although all the runs presented here used the assumption of constant SD of mean length at all ages, it is possible to impose other models, for example, SD of mean lenght increasing or decreasing with mean length (i.e. age). Runs were done with natural mortality fixed at M=0.4 or M=0.6 and also for M estimated.

The selection above is not the full .ini file required for running the model. (An example of the full file has been sent to the Secretariat)

5. Running the model

The details of how the model is structured, which parameters are estimated and how penalties are dealt with are handled with 'flags' which are set during run time. All results presented here were run with a script file which defines the flags in several phases (similar to the phases used in AD model Builder estimation). This is potentially confusing, because flags can be turned 'off' in an early phase and turned 'on' again at a later phase. For example, one might want to estimate the von Bertalanffy growth parameters right at the end rather than in each phase. As noted above, many inadvertant errors can occur here, so care needs to be taken. The following is a summary of some of the assumptions in the scripts used for the base case runs.

Base Case
Catchability: Assume time-varying catchability for fleets 2 and 3, but estimate constant catchability for fleet 1 under the assumption that the effort has been standardised.
Selectivity: Selectivity the same for ages 10 and older in fisheries 1 and 2; no other shape constraints.
Dome shaped for fishery 3, with selectivity at 0 for age classes 7 and above.
Mortality: Input a value (0.4 and 0.6). The value of 0.4 was assumed for age 1 and older in previous assessments and 0.6 is reasonable as an average value for a vector of (0.8, 0.4, 0.4 ...). A few runs were also done where a constant average mortality was estimated.
Weights and Penalties

---

[2] If the parameter is being estimated, the initial value serves as the starting value, otherwise (i.e. if the parameter is not being estimated) it is fixed at that value.

The penalties for effort deviations (fishery flag 13) was set to imply a CV of around 0.22 for fishery 1 and 0.7 for fisheries 2 and 3 (flag values: 10,1,1).
Size frequency weights were 10,10,200 for fisheries 1,2,3. This weighting implies roughly similar numbers in samples.

6. Preliminary results and sensitivity trials

Recall that all results are based on catches and size frequency data up to 2001, but Japanese CPUE (in terms of standardised effort) only up to 1999. The last two years have missing effort values, but the model still estimates expected catches for those years. 'Long' runs start with data in 1960 for the longline fisheries, 1978 for purse seine. 'Short' runs start with data in 1978 for all fisheries (these runs have a truncated dataset, NOT just 'missing' values). Some runs were done with data starting in 1952 (F1) or 1954 (F2) for interest sake. It is interesting that these runs tended to estimate a higher mean length for the oldest age class, but lower von Bertalanffy k. Results for these runs are not presented.

Results are taken from three files: plot.rep, the last .par file (16.par for these runs) and the length.fit file. As noted, due to time constraints the diagnostics were not as carefully examined as they should be.

It is hard to know how best to present results, because the different assumptions and different parameters appear not to be independent. This means that a change in, for example, M from 0.4 to 0.6, has a different effect on results depending on, for example, whether a short or long data series is used. This clearly has implications for running a 'proper' or 'thorough' sensitivity analysis.

6.1 'Long' runs - from 1960

Natural mortality
Estimates of biomass for three values of natural mortality, m, are shown in Figure 3 (the value of 0.48 was estimated). The slight increase in biomass between about 1972 and 1978 is interesting, and appears to be driven (at least to some extent) by the relative weights of the size frequencies during that period, and the penalties on the CVs of effort deviations for the 3 fleets. For example, the increase is more dramatic when the CVs are forced to be unrealistically small. Also, the biomass trajectories are not parallell for the different m-values; this is often the case with simpler assessment models.

Figure 4 shows that the selectivities for fishery 1 are subtly different. This is partly because estimates of growth are slightly different (Table 1)

Table 1. Estimates of growth parameters for base case 'Long' runs.

Run	vonB. K	Oldest age FL (~age 15)
M=0.4	0.103	210 cm
M=0.48 (estimated)	0.088	223 cm
M=0.6	0.088	224 cm

Selectivities for fishery 2 are similar to those for fishery 1 (F1), and should probably be assumed to be the same given that most of the size frequency data for F2 are derived from F1. Selectivities for F3 are very similar for the 3 values of m with highest selectivity at the first age class. Given that the size of that age class is estimated at ~43cm, it coincides probably with age 1 (see below for runs with first age class mean size ~20cm).

An average constant catchability was estimated for F1 (forced) under the assumption that the effort has been standardised to remove most of any catchability change. Fleets 2 and 3 had catchability changes every 2 years and estimated trends are shown in Figure 5 for M=0.4. There is little difference in patterns over time for other m-values. It may make sense to keep q fixed for slightly longer periods of time to reduce the number of parameters being estimated.

Downweighted size frequency (SF2 run)

Results of the ‘SF2’ run which down-weights the size data (as illustrated in Figure 2) for fleet1 are shown in Figure 6 for m=0.4. The down-weighted version estimates a much lower level of biomass. The selectivities and growth parameters are quite similar. The main difference appears to lie in the estimates of catchability. The levels are higher in the SF2 case, and the increase over time, particularly for fleet 2 is much greater. The base case run estimates q2001/q1960 at 1.7 whereas the SF2 run estimates this ratio at 2.9 for F2. For fleet 3, the ratios are more similar: q2001/q1978 is 1.1 for the base case and 1.2 for the SF2 run.

Note that other relative fishery weightings were not tried. The relatively poor fit of fishery 3 size frequency data (see below) could be due to the choice of these weights.

Change in q for F1 (2step q run)
Given the apparent ‘jump’ in standardised CPUE in 1977, a run was done where q is allowed to change for F1 every 17 years. This estimates an increase in 1977 and also in 1994 given the way the software is set up. The q-estimates relative to the value in 1960 are 1.34 in 1977 and 1.52 in 1994. This leads to more of a decline in the biomass trajectory as expected (Figure 7). The change in fit is also reflected in the effort deviations for fishery 1 (Figure 8).

Lower length for first age class (L20 run)
All the above runs had a relatively wide prior on the mean length of the first age class. The range was 10cm to 50cm, and the starting value was 42cm. Recall that the actual age is not specified, but estimated size for all runs were 42-43cm, suggesting an age of around 6 months according to Tankevich’s growth curve for females (Indian Ocean) and Lehodey’s growth curve for both sexes combined (Pacific Ocean) (see e.g.). In the alternative run the range was set to 5 - 35 cm with starting value of 20. The vector of maturity at age was changed to reflect this (0, 0, 0, 0.5, 1,1 ... 1) instead of (0, 0, 0.5, 1, ... 1).

Total biomass and SSB are quite similar inspite of the fact that they represent a different first age class. Although selectivity at age CLASS (not age itself) is different, the selectivities at size are very similar for F1 and F2, and reasonably similar for F3. Figure 9 shows that selectivity of 1 remains at the size of ~ 42cm, and the smaller class (estimated at 20.5cm in L20 run) which does not appear in the base run, has selectivity of only 0.05.

The interaction between the size of the first age class, the maturity vector and the relative timings of the fishery and of recruitment obviously need some care in the input files and in interpretation of results (and I do not guarantee these preliminary analyses error free in this regard!).

Model fit to length frequencies
Only results for the base case run (m=0.4) are shown, but these general features also appeared in other runs. Figures 10 and 11 show that although the size frequencies are reasonably well fitted in some years, they are very poorly fitted in some other years. For F1 the fit is generally poorer in recent years when sample sizes have been lower and size frequencies therefore more ‘noisy’. Fishery 3 size frequencies are particularly poorly fitted at the higher sizes (~ 95cm). Selectivity for Fishery 3 was set to 0 for age classes 7 and above, but this should be above those lengths. Nonetheless, alternative runs with a higher age class for 0 selectivity could be tried. Also note the comment above about relative weightings for the three fisheries.

6.2 ‘Short’ runs - from 1978 - and comparison with ‘Long’ runs.

Although the magnitude of biomass does not differ greatly for the short vs long series, the estimated depletion and time-trajectories are different. Also, the change in length of time series has a different effect on results depending on the assumption about natural mortality (compare figures 12a and b). It is interesting that the short run with m=0.6 estimated the mean size of the first age class at 20cm unlike all the other runs which estimated it at 42cm. Selectivities at size for F1 are, however, most different for older larger fish.

Since Multifan CL makes a separability assumption for fishing mortality and assumes that selectivity is constant over time, it is quite possible that there have really been changes in selectivity and that the length of series therefore affects the estimates.

6.3 Steepness

As observed in previous assessments, there is very little information about steepness (h in the stock-recruit relationship) in the data. Multifan-CL allows specification of a prior for steepness and the results below (Table 2) show the effect of different priors on results. As one might expect, the actual estimates of biomass, selectivity etc. are not much affected, but estimates of MSY, Bmsy and related quantities ARE strongly affected.

Table 2. Results for a range of priors on steepness. Other assumptions as for base case runs and M=0.4 (Runs in this table were inadvertantly done with high penalties (unrealistically low CV) on effort deviations. Since they illustrate the main points the runs were not repeated with the correct penalties)

	no prior	wide prior u-shaped	wide prior mode ~0.9	narrow prior mode ~0.75	narrow prior mode ~0.5
Steepness	0.961	0.961	0.901	0.751	0.501
objective Fn	7875.21	7878.17	7871.95	7851.34	7737.84
No. of parameters	197	198	198	198	198
Biomass 1978	2288	2277	2286	2290	2307
Biomass 2001	1285	1280	1286	1288	1298
Ratio 2001/1978	0.56	0.56	0.56	0.56	0.56
SSB 1978	1946	1940	1947	1950	1965
SSB 2001	1001	997	1002	1004	1013
Ratio 2001/1978	0.51	0.51	0.51	0.51	0.52
Bmsy	728	724	760	828	1048
SSBmsy	510	507	542	613	826
MSY	118	118	113	100	74
Fmsy	0.162	0.163	0.149	0.121	0.071
age flag 153 value[3]	na	1	100	328	1000
age flag 154 value	na	1	20	116	1000

Concluding remarks

These preliminary results are promising, but further though needs to be given to the data issues raised above. For example, decisions on the definition of fisheries, length of data series, input values, priors and penalties need to be made with care. Ideally, sensitivity runs should be done. Unsurprisingly, results are sensitive to relative weightings of data pieces and some other assumptions.

From a practical point of view, it is very easy to make mistakes when running the software because there potentially many ‘flags’ to set and they can be switched on and off in different phases of fitting. There are many output files and results so that it is not always easy to see at a glance which flags were set and to what values. Experience supports the Multifan-CL manual’s guidance to run different versions in different directories to minimise confusion.

The more complicated assessment models make it more difficult to compare diagnostics between runs, because there are several components to the likelihood and some, like the size frequencies are time-consuming to evaluate. Diagnostics are obviously important for detecting model failure or weakness and when comparing different runs. Harley and Maunder (2002) recommend approaches to evaluating diagnostics of large statistical assessment models. The other component of outputs which was neglected here is the estimated covariance matrix of parameters. (As far as I know, this output is generated by setting the appropriate flags).

---

[3] These flage are set in the .script file. Also see table 5.2 in the Multifan-CL manual.

Acknowledgements

Thanks to Jason Hartog and Dale Kolody for their assistance with Multifan-CL. This work was supported by the Australian Fisheries Management Authority and CSIRO.

References

Fournier, D.A., J.R. Sibert, J. Majkowski, and J. Hampton. 1990. MULTIFAN: a likelihood based method for estimating growth parameters and age composition from multiple length frequency data sets illustrated using data from southern bluefin tuna (*Thunnus maccoyii*). Can. J. Fish. Aquat. Sci. 47: 301-17.

Fournier, D.A., Hampton, J. and Sibert, J.R. 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*. Can. J. Fish. Aquat. Sci. 57: 1002-1010.

Harley, SJ, Maunder, MN. Draft Document (2002). Recommended diagnostics for large statistical stock assessment models. (from SPC and ICCAT websites)

Kleiber P, Hampton J and D.A. Fournier. 2003. Multifan-CL User's Guide. Obtainable from www. multifan-cl.org.

Kolody, D.S., Jumppanen P.C., Ricard, D.G., Hartog, J.R., Preece, A.L., and Polacheck, T. 2004.
SESAME: a simulation-estimation stock assessment model evaluation project focused on large pelagic species. CSIRO Marine Laboratories Report 241.

Labelle, M. 2002. Testing the accuracy of MULTIFAN-CL assessments of the WCPO yellowfin tuna fishery conditions. Meeting of the Standing Committee on Tuna and Billfish 15, Working Paper MWG-1.

Labelle, M. 2003. Testing the accuracy of MULTIFAN-CL assessments using an operational model of yellowfin tuna (*Thunnus albacares*) fisheries in the western and central Pacific Ocean. Meeting of the Standing Committee on Tuna and Billfish 16, Working Paper MWG-1.

Ricard, D and M. Basson. 2002. Application of an age-structured production model (ASPM) to the Indian Ocean bigeye tuna (Thunnus obesus) resources. Working paper to IOTC WPTT-02-07.

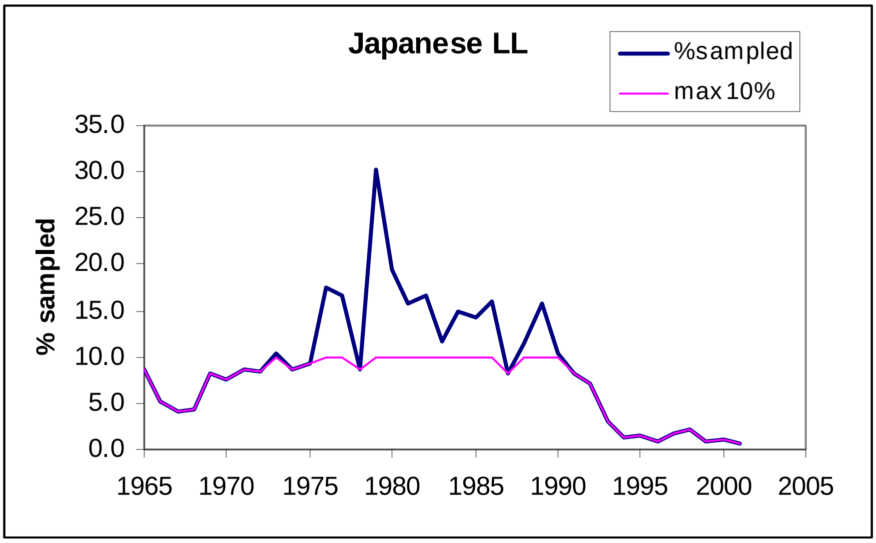


Figure 1. Sample size of Japanese LL bigeye size frequencies as a percentage of the catch in numbers (from the raised CL files). Also shown is an alternative scheme which downweights samples from the mid70's to the early '90s to a maximum of 10% of the catch (runs using this weighting is called ‘SF2’).

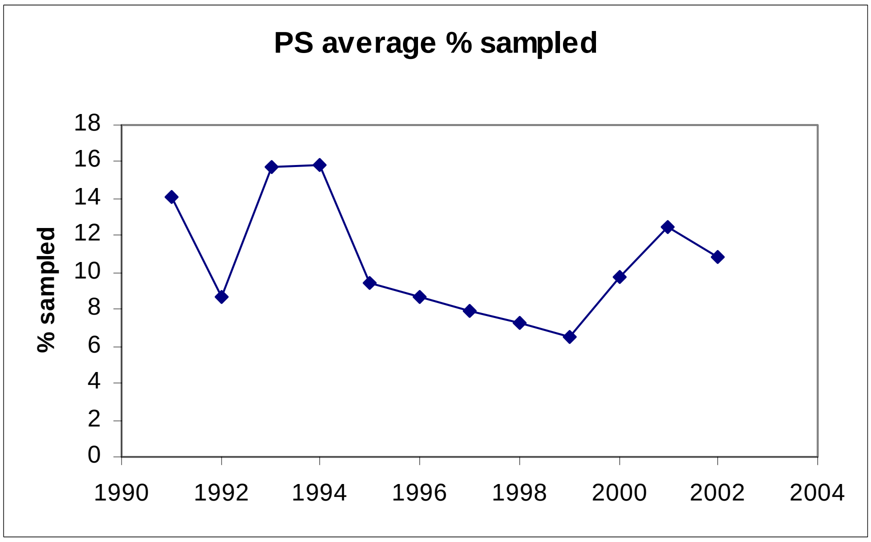


Figure 2. Average percentage of catch sampled in PS fisheries: SPN, FRA, NEI-EUR and SYC. These averages were used to define sample sizes of length frequency data for fishery 3.

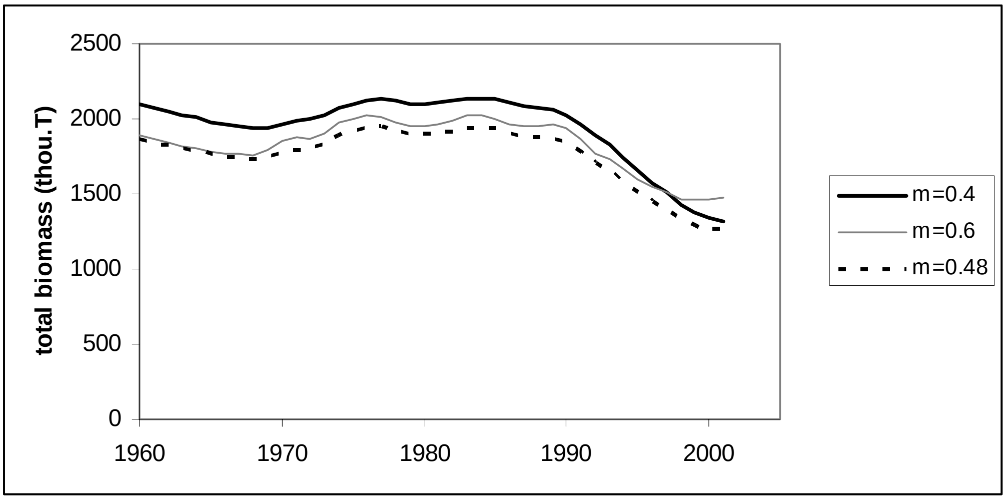


Figure 3. Estimates of total biomass for three values of natural mortality: 0.4, 0.6 and m estimated at 0.48.

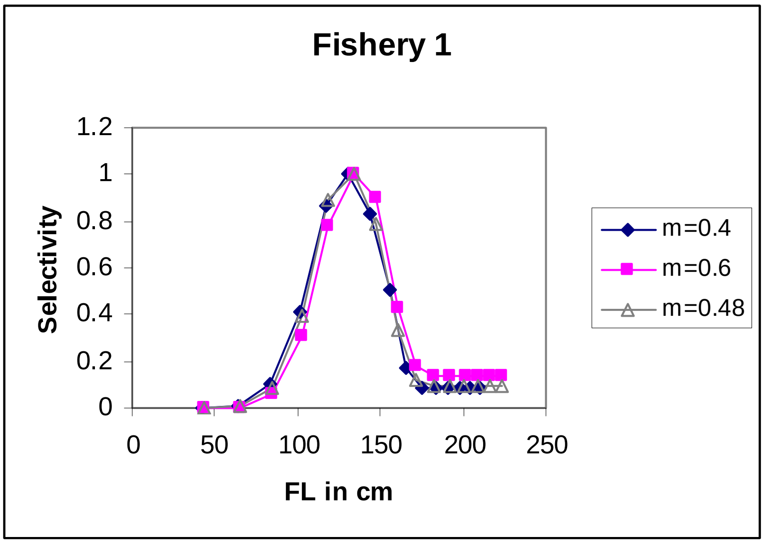


Figure 4. Estimated selectivity at size for fishery 1 (Jap. LL) at different values of natural mortality.

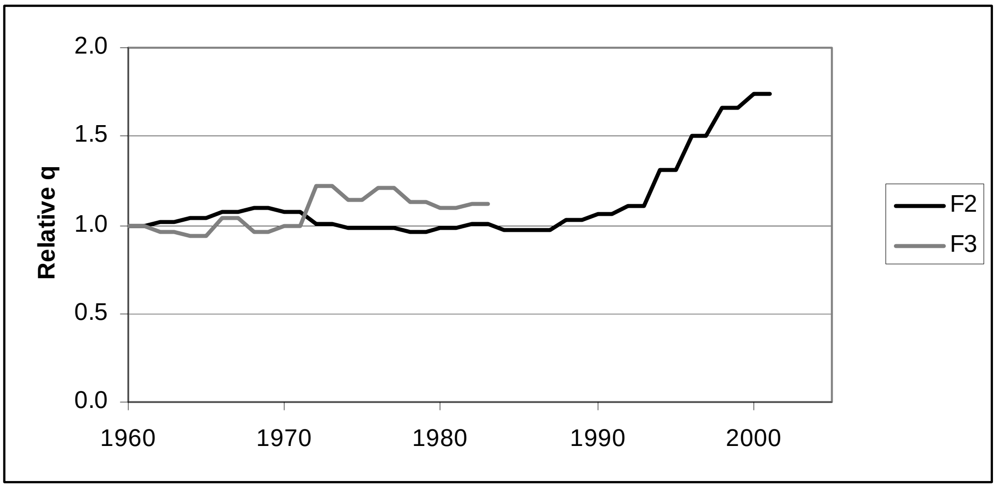


Figure 5. Relative catchability (scaled to value in 1960) for fisheries 2 and 3, M=0.4.

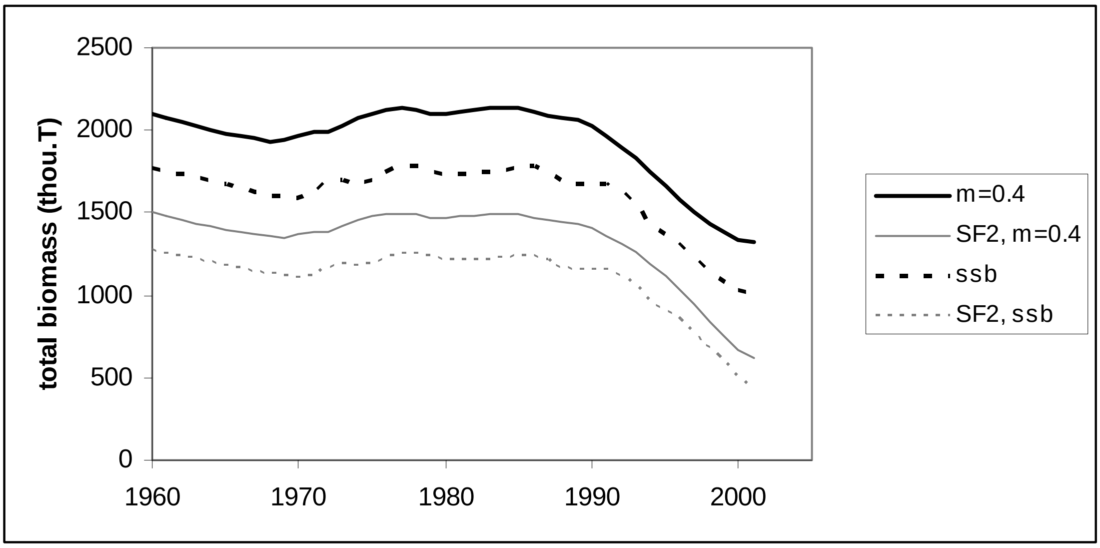


Figure 6. Total biomass (solid line) and SSB (dashed line) for the 'base case' and SF2 weightings for fishery 1. M=0.4 in both cases.

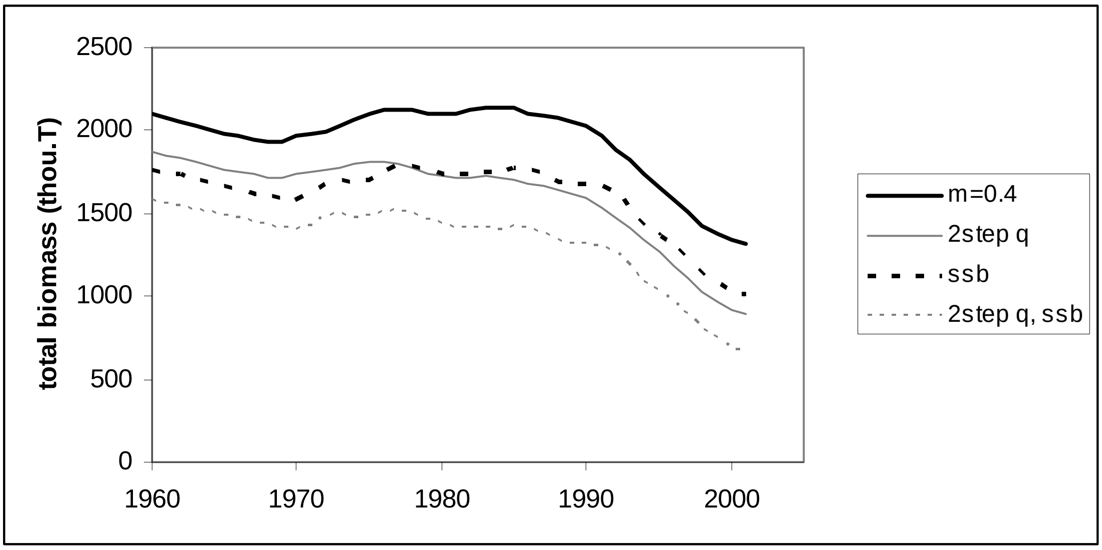


Figure 7. Total biomass (solid line) and SSB (dashed line) for the 'base case' and 2 changes in catchability for F1 (changes in 1977 and 1994). M=0.4 in both cases.

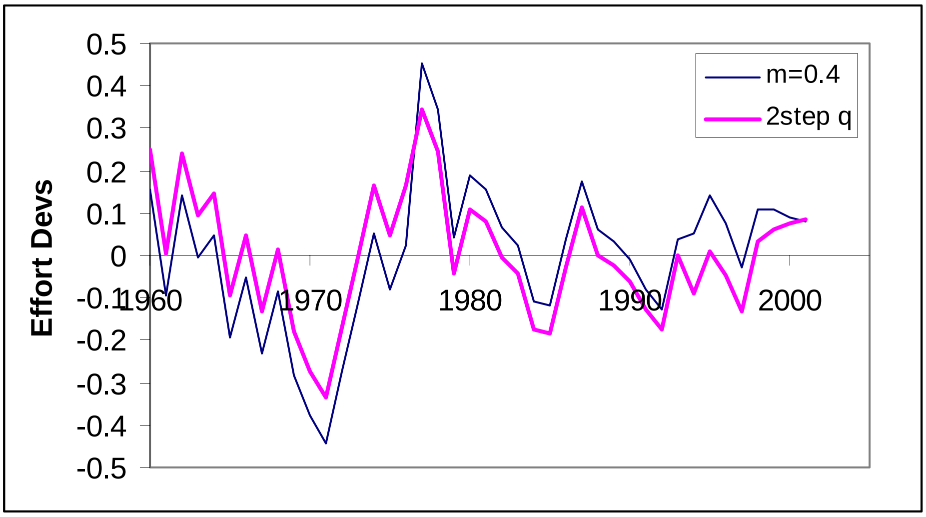


Figure 8. Effort deviations for the base case and 2 changes in catchability for F1.

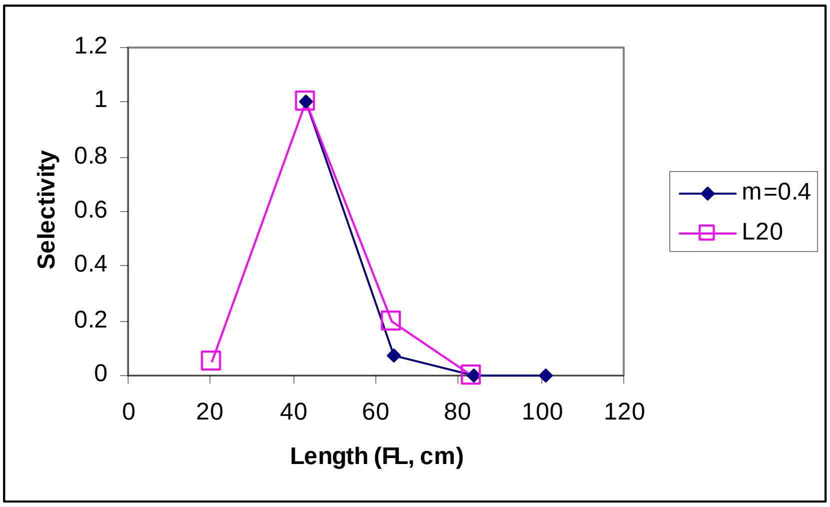


Figure 9. Selectivity at length for fishery 3 base case run and low length for the first age class, L20 run.

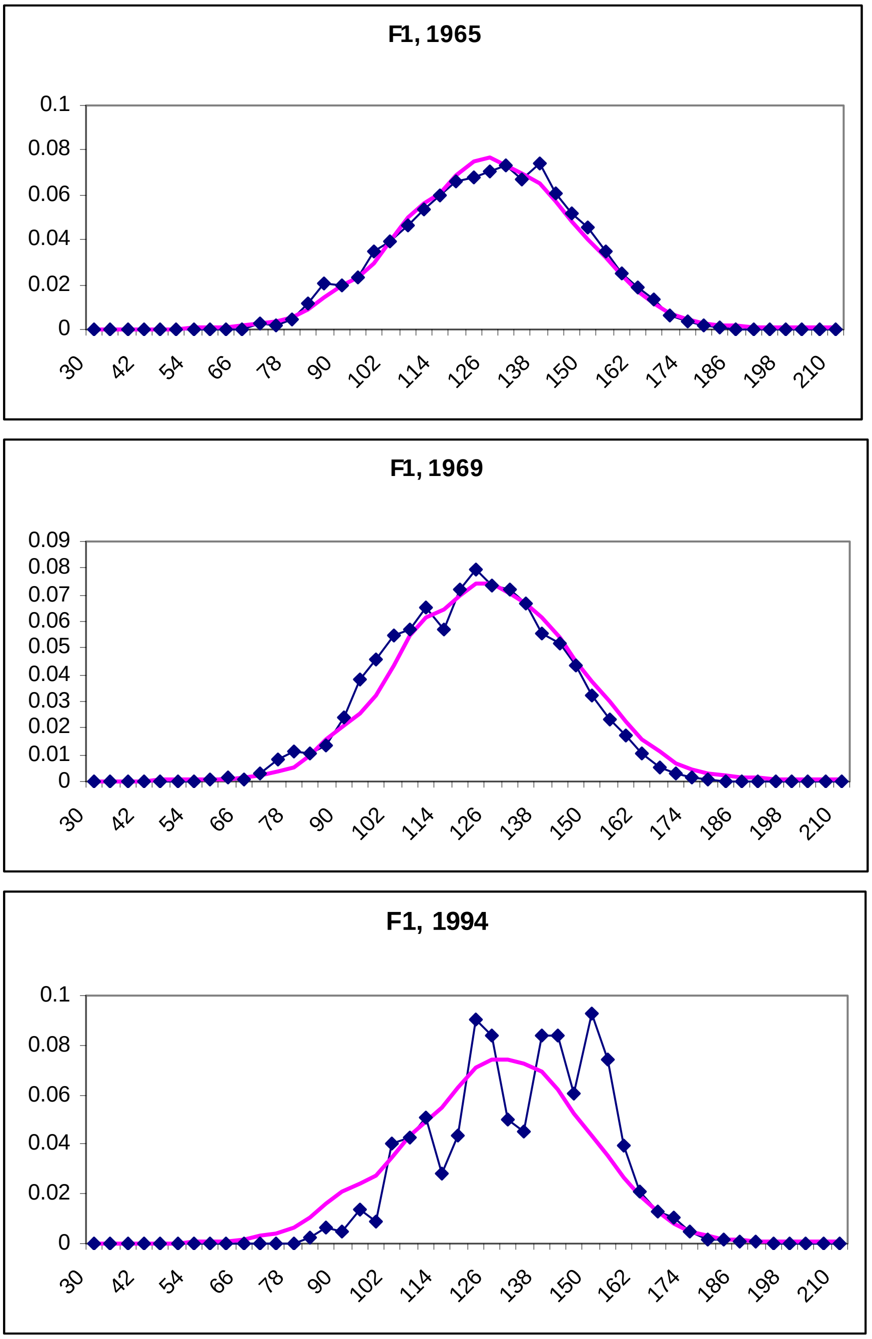
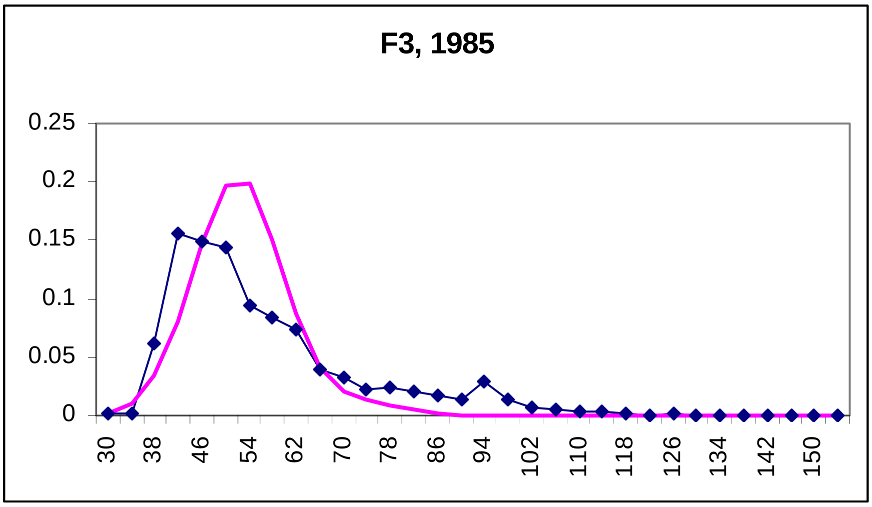
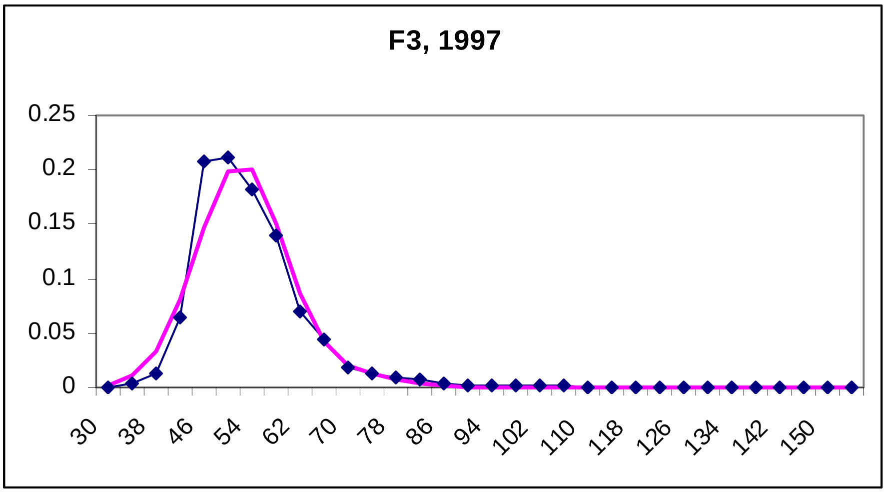


Figure 10. Three examples of the fit to size frequency data for fishery 1 (solid line = fitted; diamonds = observed).





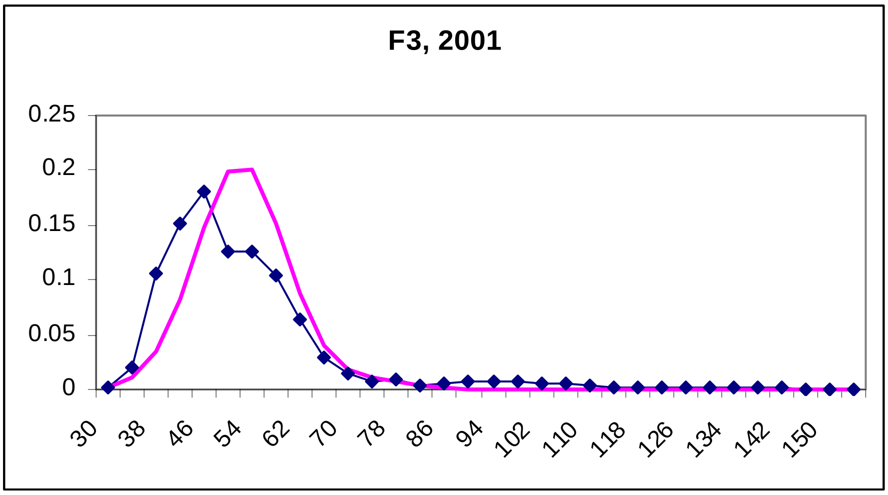


Figure 11. Examples of the fit to size frequency data for fishery 3 (solid line = fitted; diamonds = observed).

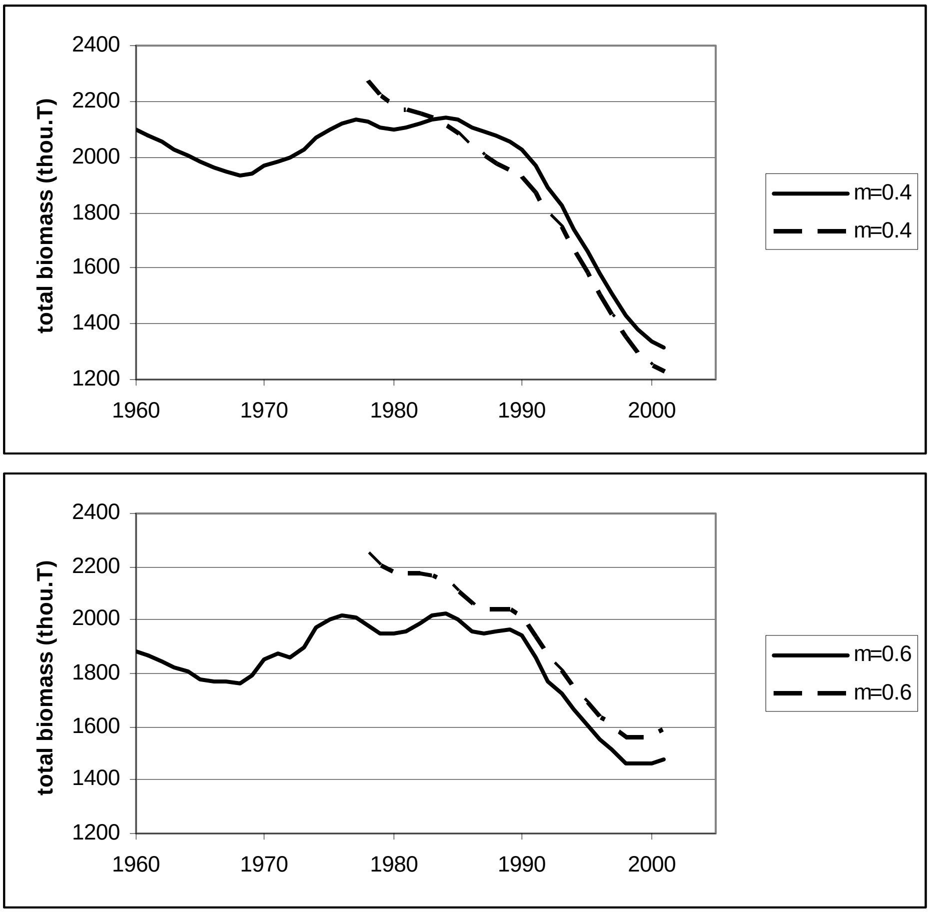


Figure 12. Total biomass for ‘Long’ (solid lines, 1960+) and ‘short’ (dashed lines, 1978+) runs and for m=0.4 (top panel) and m=0.6 (lower panel).