Taihoro Nukurangi

Evaluation of the paua stock assessment model with an individual-based operating model

Susan W. Kim, Paul A. Breen
& Neil L. Andrew

National Institute of Water and Atmospheric Research

July 2002

Kim, S.W., Breen, P.A. & Andrew, N.L. 2002. Evaluation of the paua stock assessment model· with an individual-based operating model. Final Research Report.

EXECUTIVE SUMMARY

An individual-based model is described and used to test the bias and precision of the stock assessment model used for paua (*Haliotis iris*). The individual-based model system consists of a data simulator and the estimation model, where the simulator and estimation model use the same method for calculating population dynamics based on model parameters.

A shell program runs the data simulator, which simulates the lives of many individual paua for each of a large number of years, and writes data files with the catch time series, abundance indices and length frequencies. Then it runs the estimation model, which simulates a simple assessment based on the data and the mode of the joint posterior distribution of parameters. The "real" and estimated values for parameters and biomass are written to a file for comparison, the random number seed is changed for the simulator, and this procedure is repeated in a loop for 100 times.

Comparison of the estimated values with simulated values was done with three comparison indices. One compared the mean values to assess bias. A second compared the distributions of values from the estimator and simulator, and a third assessed the mean difference between the two sets of results.

When nearly-perfect data were simulated, with no observation error, no recruitment deviations, and abundant data for every year of the simulated fishery, the estimator recovered the true values with little bias and high precision. This suggests that the estimation model is coded correctly.

When the model was tested with much fewer data, observation error on the abundance indices and recruitment variation, some bias was observed in some estimated parameters, but this was relatively small. Precision was slightly less than in the actual assessment for PAU 7, but the results here are not strictly comparable with the actual assessment, which uses Bayesian procedures and is based on marginal posterior distributions.

We experimented with the model's ability to estimate parameters correctly when the abundance index was not directly proportional to abundance, and when recruitment had serial auto-correlation. Although the model did not do well at estimating the non-linearity and autocorrelation parameters, other parameters were well estimated. Bias and precision both declined slightly when the model was forced to assume linearity in the abundance index, or that auto-correlation was absent.

When we simulated other mis-specifications - density-dependent growth and size-dependent natural mortality, both absent from the estimation model, bias and precision were both affected but the effect was minor.

1. Introduction

Paua supports a valuable fishery in New Zealand, with total annual landings of about 1200 t. Legislation requires that New Zealand fisheries be managed so that stocks are maintained at or above B_{MSY}, the biomass associated with the maximum sustainable yield (*MSY*). However, B_{MSY} is not defined, and Francis (1999) suggested that B_{MSY} varies among different harvest strategies, which are usually undefined. There is current discussion over how to estimate B_{MSY}.

Virgin biomass, *B0*, has also been used as a reference point, but in the paua stock assessment for 2001(Breen et al. 2001) "reference biomasses", B_{ref} and S_{ref}, were introduced to replace B_{MSY} and *B*. These biomass reference points are the average recruited and spawning biomass in some period, or reference years, where the stock was stable. In PAU 7 (Breen et al. 2001), PAU 5B (Breen et al. 2002a) and PAU 5D (Breen et al. 2002b), the years 1985-1987 were used as reference years. The reference biomass concept was also used in the rock lobster stock assessment (Breen et al. 2002c).

Stock assessment of paua (*Haliotis iris*) in New Zealand has been described by a variety of authors (Schiel, 1989; McShane et al., 1994, 1996; Breen et al., 2000a, 2000b, 2001, 2002a, 2002b). A length-based model, implemented in AD Model Builder™, has been used since 1999. It is similar to length-based models developed for abalone by Worthington (1998, see Andrew et al. 1996 for application to the NSW abalone fishery), and for rock lobsters by Punt & Kennedy (1997) and Bentley et al. (2001). This model is driven by reported commercial catches from 1974 through the current year and is fitted to five sets of data: standardised CPUE, a research diver survey index (Andrew et al. 2002), length frequency data from commercial catch sampling and from research diver surveys (Andrew et al. 2002), and a set of growth increment data.

The assessment model has been fitted to data using robust techniques to specify likelihood and to describe prior distribution (see Chen et al. 2001), although the newest version has dropped some of the robust techniques. The model population is initialised and then driven by observed catches. Outputs are the present and projected future states of the stock, estimated using Bayesian methods. The assessment is based on the marginal posterior distributions of the parameters and derived parameters of interest, in turn based on Markov chain - Monte Carlo (McMC) simulations. The model was described by Breen et al (2001) for the 2001 stock assessment for PAU 7 and the current version of the model is described in Breen et al. (2002a, 2002b).

Between the 2001 and 2002 paua assessments, we developed an individual-based model to test the precision and bias of the assessment model (Breen et al. 2001). The individual-based model has two programs run by a shell program in a loop: a data simulator and a version of the estimation model.

The simulator simulates the life of individual paua, from recruitment until death. It uses the same method of calculating the growth curve, selectivity of divers and recruitment as in the estimation model. Parameters estimated by the estimation model, such as mortality, selectivity, growth rate and the asymptotic length, are input to the simulator. Output is in the five datasets required by the estimation model, plus annual catches.

The estimation model estimates parameters of the simulated datasets; these are then compared with the parameter values used for the simulation of the datasets.

The use of a simulator, which is also called an operating model (Punt 1990, 1992), to generate data addresses problems in the use of real data such as bias in observation and missing information. An operating model was used with an estimation model in the southern African hake fishery to compare the model estimation procedures and management procedures (Punt, 1992), and was used alone to construct population dynamics in other fisheries (Rose et al. 1999, Beard & Essington, 2000).

Several tests are done with this individual-based model to investigate the effect of datasets with different amounts of information, explore the effects of serial auto-correlations in recruitment, and

non-linear CPUE, and test the effects of mis-specifications of reality, such as density-dependent growth and size-dependent mortality on the estimated parameters and assessment results.

2. Estimation model

The estimation model used for paua stock assessments is implemented in AD Model Builder™ and was described by Breen et al. (2002a). The estimation model used in this study is a model upgraded from the model used in the 2001 stock assessment (Breen et al. 2001), but was not upgraded as fully as the 2002 stock assessment model. Thus this estimation model has several differences from that described by Breen et al. (2002a).

First, the estimation model does not have the global standard deviation of observation error, $\tilde{\sigma}$, that was added to weight datasets more effectively. In the model used here, the relative weights of datasets were adjusted through the assumed standard deviations for the CPUE and research diver survey indices (both were assumed to be 0.25). For the length frequencies, the relative weight could be adjusted by multiplying each of the effective sample sizes by a constant, which was set to 1 in this study. Thus, the effective sample sizes were the square roots of the numbers of paua in the sample, as in the 2001 assessment.

Second, the estimation model uses different likelihood from the 2002 assessment (Breen et al., 2002) when it fits the data sets. It uses the old likelihood, used in 2001 assessment (Breen et al., 2001). The estimation model calculates likelihoods for the proportion-at-length fits using the robust normal likelihood, which is calculated as:

$$(1) \quad L(\hat{p}_{s,t} \mid \theta) = \prod_t \prod_k \frac{1}{\sqrt{2\pi p_{s,t}(1 - p_{s,t}) + 0.1/\Omega}} \exp\left(\frac{-\tau_t (\hat{p}_{s,t} - p_{s,t})^2}{2(p_{s,t}(1 - p_{s,t}) + 0.1/\Omega)} + 0.01\right),$$

where the τ_t is the effective sample size, $\hat{p}_{s,t}$ and $p_{s,t}$ are predicted and observed proportion at length s in year t, Ω is the number of size bins. For the CPUE and RDSI fits, the estimation model uses fat tail likelihood, which is calculated as:

$$(2) \quad L(\hat{I}_t \mid \theta) = \prod_t \frac{1.329}{\sqrt{4\pi}} \left[1 + \frac{(\ln(I_t) - \ln(\hat{I}_t))^2}{4\sigma^2} \right]^{-2.5},$$

where the I_t and $\hat{I}_t$ are observed and predicted CPUE(or RDSI) index of year t, respectively and σ is the assumed standard deviation, 0.25. A normal likelihood is used to fit the tag data set,

$$(3) \quad L(\Delta \hat{l}_j) = \prod \left[\frac{1}{\sigma_{\Delta l,j} \sqrt{2\pi}} \exp\left(-\frac{(\Delta l_j - \Delta \hat{l}_j)^2}{2\sigma_{\Delta l,j}^2} \right) \right],$$

where the Δl_j and $\Delta \hat{l}_j$ are observed and predicted increment of jth tag-recapture data, respectively and $\sigma_{\Delta l,j}$ is standard deviation of jth tag-recapture data.

Third, the model estimates auto-correlation of the recruitment deviations, a parameter dropped from the 2002 assessment model.

Parameters estimated by the model are:

ln(R0)   the natural logarithm of the base annual recruitment,
M   the instantaneous rate of natural mortality, assumed to be constant over time and size,

K	the instantaneous rate of approach to $L\infty$,
$L\infty$	asymptotic length,
α	CV of the expected growth increments,
σ_{MIN}	the minimum standard deviation of the expected growth increment,
$Rdev$	a vector of recruitment deviations (in arithmetic space) modifying the actual model recruitment in each year from 1984 onwards,
$S50$	the paua size at which research survey divers are 50% effective,
$S95$	the paua size at which research survey divers are 95% effective,
h	the coefficient for the exponent of the relation between CPUE and the recruited biomass, and
ρ	the parameter describing auto-correlation of the recruitment deviations.

The model is fitted to five data sets: catch per unit effort (CPUE), research diver survey index (RDSI), proportions-at-length from commercial catch sampling, proportions-at-length from research diver surveys, and tag-recapture size data.

Paua assessments are based on the marginal posterior distributions obtained from Markov chain Monte Carlo simulations. We did not simulate this procedure; instead we simply ran the estimation model to minimise the objective function, producing the mode of the joint posterior distribution (MPD) for parameters.

3. Data simulator

3.1 Overview

The data simulator is an individual-based model (IBM); it simulates the life of an individual paua from recruitment (to the 70-80 mm size bins) to death. The life of a paua includes its growth in length, the onset of maturity, and death by fishing or natural causes.

The simulator consists of a part that simulates data and a part that writes data. The first part simulates lives of individual paua from specified parameters, such as the initial number of recruited animals for each year, natural mortality rate, exploitation rates, research divers' selectivity, variance of observation errors and process errors. The second part generates and then writes datasets for the estimation model, such as catch, CPUE and RDSI indices, length frequency data from research survey and catch sampling, and the tag-recapture data.

3.2 Input

The simulator has three types of data inputs. The first type of input determines the dimensions of the simulation and provides fixed values. These include the years when the simulation begins and ends (1914 and 2001), the year that fishing starts (1974), minimum length and maximum length (70 and 170 mm), length at 50% and 95% maturity (91 and 105 mm), the bin width for the proportion-at-length datasets (2 mm), and length-weight coefficients (as in Breen et al. 2001), etc. These are data that are also specified for the estimation model, and they remain constant for all trials. The start year, 1914, was chosen to obtain equilibrium before fishing starts in 1974.

The second data type specifies the population parameters (Table 1) that will be estimated by the estimation model. These are the parameters listed in the table in section 2. Different values were used in different trials, and these will be shown in tables discussing the results. The standard deviation of observation errors is specified here. A time series of exploitation rates is also specified in this section.

The third data type specifies how much data will be produced for the estimation model. This includes the specific years for which proportion-at-length data are produced, the number of tag-recapture data and the maximum number of tag-recapture data in each length bin. This data section also includes the initial random number seed (this is changed after every run so that the data generated by the simulator is different for every run).

3.3 Mechanics

In each year the specified number of paua are recruited, one at a time. For each recruited paua, the simulator assigns an initial length, chosen randomly from a uniform distribution between 70 and 80 mm. Then the model simulates life for this paua. For every year, it examines whether the paua has died from natural causes; if the animal is alive the model examines whether the paua is been caught. If the paua is not caught, it grows, and if greater than 90 mm may be tagged in a tag-recapture experiment. The tag-recapture data are recorded here by writing the growth increment of this animal, and its length before it grew. The time-at-liberty is assumed as 365 days. Then the model moves to the paua's next year, adds its length to the population length frequency for that year and does the same step as described above (Figure 1).

If the animal is dead or caught, the simulator moves to the next recruited animal and does the same. When the animal is caught, its length is added to the catch length frequency for that year.

3.4 Dynamics

3.4.1 Recruitment

The recruitment for each year is simulated from a base level modified by annual deviations that act lognormally (Breen et al. 2001):

$$(4) \quad R_y = R0\exp\left(Rdev_y\right)\exp\left(-0.5\sigma_\varepsilon^{\ 2}\right)$$

where

$$(5) \quad Rdev_y = \sqrt{\rho}Rdev_{y-1} + \sqrt{1-\rho}\varepsilon_y,$$

where $R0$ is the base recruitment, ε_y is a normally distributed random number (mean 0) for year y, σ_ε is the standard deviation of the ε_y, and ρ determines the serial auto-correlation. The σ_ε were set to 0.4 except for case 1, which has ε_y set to 0 for all years (Table 2), and ρ was chosen to be 0.0001 except where otherwise stated (see Table 2), to give minimal auto-correlation in recruitment. Hence the $Rdevs$ have the same standard deviation as ε_y. Recruitment in the model is calculated as a vector from the start year to the end year.

3.4.2 Mortality

3.4.2.1 Natural mortality

For each year in the life of each paua, the natural mortality rate (fixed for all lengths and years) is compared with $m_{a,y}$, a uniformly distributed random number between 0 and 1, to determine whether

the paua survives. The individual is dead from natural causes if the $m_{a,y}$ is less than or equal to the probability of death from natural mortality in a year:

(6) Animal is dead if $m_{a,y} \leq 1 - \exp(-M)$

(7) Animal is alive if $m_{a,y} > 1 - \exp(-M)$

where M is the instantaneous rate of annual natural mortality.

The model is able to simulate data with a size-dependent mortality, $M(l)$, specific to size l. To do this, we replace the constant M with a function that calculates M for an animal with size l:

(8) $M(l) = \exp(-0.014l - 0.64)$,

where 0.014 and 0.64 are arbitrary constants chosen to give $M = 0.20$ for small (70 mm) paua and $M = 0.05$ for large (170 mm) paua. This arbitrary M function is shown in Figure 2.

3.4.2.2 Fishing mortality

The simulation for fishing mortality is similar to that for the natural mortality. If the paua is at or above the minimum legal size (MLS, 125 mm), a uniformly distributed random number between 0 and 1, $c_{a,y}$, is compared with the probability of being caught, which is the exploitation rate in year y.

The exploitation rates U_y are specified to the simulator. If an animal is greater than or equal to 125 mm and

(9)   if $c_{a,y} \leq U_y$: the animal is caught,

(10)   if $c_{a,y} > U_y$: the animal is not caught,

3.4.3 Growth

From the von Bertalanffy growth parameters L_∞ and K, the expected annual growth increment for length l is:

(11) $\Delta l = (L_\infty - l)(1 - \exp(-K))$ for $l \leq L_\infty$

and

(12) $\Delta l = 0$ for $l > L_\infty$

Then the increment is simulated as a normally distributed random variable with mean Δl and standard deviation $\sigma_{\Delta l}$:

(13) $\Delta l_{new} = \Delta l + \delta_{inc}\sigma_{\Delta l}$ where $\sigma_{\Delta l} = \max(\sigma_{MIN}, \alpha\Delta l)$

where δ_{inc} is the normally distributed random increment error, $\sigma_{\Delta l}$ is the standard deviation of the growth increment, σ_{MIN} is a minimum standard deviation, 1 is assumed for this, and α is a c.v. of the growth increment.

The model is able to simulate density-dependent growth by replacing the constant K with K_y, the growth parameter K for animals recruited in year y, which is relative to the recruitment deviation, $Rdev_y$. A function relating K_y to $Rdev_y$ is:

$$(14) \qquad K_y = -0.085\exp\left(Rdev_y\right)\exp\left(-0.5\sigma_\varepsilon^2\right) + 0.28 .$$

The gradient (-0.085) and the intercept (0.28) were chosen to produce a K value of 0.25 when $Rdev_y$ is one standard deviation less than the mean and 0.12 when $Rdev_y$ is one standard deviation greater than the mean. This K function gives less growth when there are more recruited animals (Figure 3).

3.5 Output

The simulated data is written for the estimation model. There are three data files: for abundance data, length frequency data and the tag-recapture data. The abundance data include catch, CPUE and RDSI data; the length frequency dataset includes the research survey and the catch sample length frequencies. The derived parameters from the simulator, such as biomass (Table 1), are written for later comparison with results from the estimation model.

3.5.1 Length frequency data

The population length frequency data is calculated at the beginning of the year before mortality occurs. After every calculation of length (including the first length calculation), the length l of the individual is rounded down and added to the appropriate 2-mm bin for the length frequency for that year. The population length frequency data matrix has columns of length bin from 70 to 168 and rows of year from 1914 to 2001.

At the end of the simulation, the research diver survey length frequency data are sampled from the population length frequency data using the research divers' selectivity:

$$(15) \qquad f_{y,k}^s = f_{y,k}^p \times SIS_k$$

where $f_{y,k}^s$ is the frequency of paua in bin k for the research diver survey for year y, $f_{y,k}^p$ is the frequency of paua in bin k of the population length frequency data for year y, and SIS_k is the research diver selectivity for paua in bin k. The last is a logistic curve defined by two selectivity parameters - lengths with at 50% and 95% diver effectiveness (Breen et al. 2002).

The effective sample size of research diver survey length frequency data for each year is the square root of the sum of the length frequencies greater than 90 mm (the estimation model ignores paua less than 90 mm length).

The catch length frequency data are recorded if the individual is caught. Again, the length l of the individual is rounded down and added to the appropriate 2-mm bin for the catch length frequency for that year. The effective sample size of the catch length frequency data for each year is the square root of the sum of the length frequencies greater than the minimum legal size, 125 mm.

For both the research diver survey and catch length frequency data, the sample size (i.e. total number of fish in the length frequency data within a year) was not specified for the simulation. With the recruitment values specified, the length frequency sample sizes were comparable to the actual dataset sizes, without the need for a re-sampling process.

3.5.2 Biomass

The biomass for each year is calculated as the sum of the product of the length frequency and weight at each length. The weight, w_k of an individual in length class l_k can be calculated as in Schiel & Breen (1991):

$$(16) \qquad w_k = 2.59 \times 10^{-8} l_k^{3.322}.$$

Then the recruited biomass is the sum of these weights of legal-sized paua:

$$(17) \qquad B_y = \sum_k f_{y,k}^p w_k$$

and similarly the spawning biomass is the sum of the weight of mature paua

$$(18) \qquad S_y = \sum_k f_{y,k}^p d_k w_k.$$

The maturity, d_k is calculated from the constants $Mat50$ and $Mat95$, which the lengths at 50% and 95% maturity respectively:

$$(19) \qquad d_k = \frac{1}{1 + \exp\left(-\ln(19) \times \dfrac{l_k - Mat50}{Mat95 - Mat50}\right)}.$$

3.5.3 Population catch series, CPUE and RDSI data

The input to the simulator specifies the series of annual exploitation rates, U_y. The simulator calculates biomass as described above, so catch is calculated as:

$$(20) \qquad C_y = U_y B_y.$$

This is an estimated catch, because the biomass is estimated in the simulation with some sampling error. In the real world, catch is also estimated with observation error, so this was thought to be realistic.

The CPUE index is based on the abundance of paua greater than the MLS and is calculated from the population length frequency data matrix and some observation error. CPUE in year y is calculated as

$$(21) \qquad CPUE_y = (B_y)^h \exp\left(\delta_{y,CPUE}\sigma_{CPUE} - 0.5(\sigma_{CPUE})^2\right)$$

where $\delta_{y,CPUE}$ is a normally distributed random number (mean 0, standard deviation of 1) and σ_{CPUE} is the specified standard deviation of the observation error. The power h is 1 for most runs (see below) which specifies a linear relationship between CPUE and recruited biomass.

The RDSI is based on the number of paua sampled from the population by research survey divers:

$$(22) \qquad RDSI_y = \sum_k f^s_{y,k} \exp\!\left(\delta_{y,RDSI} \sigma_{RDSI} - 0.5 (\sigma_{RDSI})^2 \right)$$

where $\delta_{y,RDSI}$ is a normally distributed random number (mean 0, standard deviation of 1) and σ_{RDSI} is the specified standard deviation of the observation error.

3.5.4 Tag-recapture data

The tagging simulation is done for animals greater than 90mm of length. The model checks to see whether the paua has been recaptured in earlier years, whether the paua's length-class has been filled with enough tagged animals and whether the specified total number of recaptures has been filled. The increment of the recaptured animal is recorded by subtracting the length in a year before from the length at the time of recapturing. The days at liberty are recorded as 365 assuming there is no leap year.

In the data input to the simulator, the total number of tagged animals and the number of tagged animals in each length class is given. The smaller length-classes are filled first.

4. Study designs

A simple shell program runs the simulator and then the estimation model, and repeats this loop 100 times for each run in this study. At the end of every loop, the simulator writes an output file of seeds that will be used by the simulator in the next loop. Estimates from the estimation model are collated for comparison with the parameter file from the simulator.

Most runs of the estimation model converged to an apparently reasonable parameter set. Occasionally, the estimation would not converge - this was apparent from the estimated M hitting the upper bound of 0.50 or the lower bound of 0.01 - and such runs were discarded. Their frequency ranged among sets of runs from 0-4%.

4.1 Cases

The general data and parameter of inputs for the simulator were changed so that we could compare two extremes. The first extreme (Table 2, case 1) used exploitation rates that ascend for the first half of the series from 1974 then descend for the last half. This case has neither recruitment deviations nor errors for the CPUE and RDSI datasets, and the simulator produced 10 000 tag-recaptures, with a maximum of 250 data from each initial length greater than 90mm. Both abundance indices and length frequency data were produced for all years from 1974.

The second extreme case (case 2 in Table 2) used continuously descending exploitation rate from 1974. This case used recruitment deviations with a CV of 0.4 in log space and observation errors on both CPUE and RDSI dataset with CVs of 0.25. The tag-recapture dataset size was more realistic and similar to the PAU7 stock assessment in 2001: 713 data with maximum of 13 data from each initial length greater than 90mm. The commercial catch sampling and research survey diver length frequency datasets were similar to those in the 2001 assessment, with eight years of catch sampling data, (from 1990-94 and 1999-2001) and five years of research diver survey data (from 1990, 1993, 1996, 1999 and 2001). We used indices from all years from 1974 for both indices.

More experiments were done on case 2 with some other options. First, we examined the effect of having h different from 1. By fixing h = 0.5 in the simulator, we compared the result from the estimation model: in case 3, we estimated h in the estimation and in case 4 we fixed h to 1 in the

estimation model. When we simulated data for cases 3 and 4 we used other inputs as for case 2 except for the h.

Second, in case 2, the auto-correlation parameter ρ in the simulator and the estimation model was fixed as 0.0001 (no auto-correlation). In cases 5 and 6 we examined estimation of ρ when there was serial auto-correlation. For cases 5 and 6, we fixed $\rho = 0.5$ in the simulator to give autocorrelation in the recruitment deviations. In case 5, we estimated ρ in the estimation model and in case 6 we fixed ρ as 0.0001. The same parameter values and data inputs were used as in case 2 except for the ρ.

Third, density-dependent growth was simulated in case 7, and size-dependent mortality was simulated in case 8. The estimation model was not changed; it assumes no density- or size-dependence.

Because there were big differences between case 1 and case 2, we increased the number of tag-recapture data to 2500 from 713 in case 9 to examine whether the number of tagging data has any influence on the degrees of variability and precision of the estimates. We also examined the effect of increasing the years of length frequency data in case 10 by adding 8 sample years to the commercial catch sampling data: 1974, 1976, 1978, 1980, 1982, 1984 and 1988, and 10 years to the research diver survey length frequency data: 1970, 1972, 1974, 1976, 1978, 1980, 1982, 1984 and 1988.

4.2 Comparison method

Two performance measures were used to compare the results from the simulator and estimation model. These are based on the average (PB) and the standard deviation (CI2) of the ratio of difference between the estimated valued and the simulated values to the simulated values of each run. They are calculated as:

$$(23) \qquad PB(\beta) = 100 \times \frac{\sum_i E_i}{n}, \text{ and}$$

$$(24) \qquad CI2(\beta) = 100 \times \sqrt{\frac{\sum_i (E_i)^2 - \frac{\left(\sum_i E_i\right)^2}{n}}{n}},$$

$$\text{where} \qquad E_i = \frac{\beta_i^{EST} - \beta_i^{SIM}}{\beta_i^{SIM}}.$$

The β_i^{EST} is the ith value of parameter β from the estimation model, β_i^{SIM} is the ith value of parameter β from the simulator, and n is the number of runs used in each case (usually 100).

Then, the standard errors for PB and CI2 can be calculated to show the significance of the performance measures (Sukhatme & Sukhatme 1970):

$$(25) \qquad s.e.(PB) \approx \frac{CI2}{\sqrt{n}}$$

$$(26) \qquad s.e.(CI2) \approx \frac{CI2}{\sqrt{2n}},$$

where n is the number of runs used in each case (usually 100).

5 Results

5.1 Case 1

Most parameters were estimated very well (Table 4). The minimum observed standard deviation, σ_{MIN}, was slightly underestimated (mean of 0.89 vs the real value of 1.0, giving a bias of -10%). Most biomass estimates were very slight over-estimates; the extreme was S_{ref}, where the bias was 2.31%. S_{01} was overestimated by 1.87%; all the other parameter estimates had bias of less than 1%. The distributions of biomass (Figure 4) are very similar between 100 runs of simulator and estimation model.

The CI2 indicator reflects the distribution of the differences between individual runs from the simulator and the estimation model. The largest was σ_{MIN}, which showed an average deviation of 7.4% from the true value. The next largest were U_{01}, B_{01} and B_{01}/B_{ref}, all near 5%; all others were 3% or less.

Figure 5 shows the fit to the CPUE index, RDSI, and growth increment data, and Figure 6 shows the fit to the proportions-at-length from catch sampling and research diver surveys. These fits are almost perfect for the CPUE and RDSI, and very good for the growth increment data. Standardised residuals from both CPUE (Figure 5B) and RDSI (Figure 5D) were small and the standardised residuals for tag-recapture data (Figure 5F) lie between −4.0 and 4.0. There are no observed data greater than 140 mm length for the tag-recapture data (Figure 5E), but the fit is good. Figure 7 shows that the standardised residuals from proportions-at-length from the commercial catch sample and the research diver survey are small (between −0.06 and 0.08) and there are no signs of pattern. Figure 8 shows the distributions of standardised residuals from proportions-at-length and the tag-recapture data. The distributions for proportions-at-length are very narrow, but all residuals are distributed with a mean near zero.

Figure 9A shows the exploitation rate from the simulator and the estimation model. Both exploitation rates are similar in shape, increasing for the first decade then decreasing for the next decade. The estimation model over-estimated exploitation rate at the end of the simulated period. Figure 9B shows the recruitment from the simulator and the estimation model. The simulated recruitment was constant, and estimated recruitments were constant at very close to the simulated level.

5.2 Case 2

One extreme run from the estimation model was taken out because it hit the upper bounds for M and $S95IS$ and had very high $ln(R0)$ (11.3), giving very high biomass. This run was probably not converged properly. This run had a large influence on calculation of the PB, which uses the average value of percentage difference from 100 runs. The comparison indicators (Table 5) were calculated after this extreme run was taken out.

Most parameters were estimated well (Table 5) although some are slightly overestimated. M was over-estimated by about 10% (from 0.13 to 0.14) and this was accompanied by an over-estimate in recruitment. Most biomasses and biomass indicators were slightly overestimated except for the virgin recruited biomass ($B0$), but bias was small (less than 4% except for S_{ref}).

Distributions of biomasses are similar between the simulator and estimation models (Figure 10).

The comparison indices CI2 show more variability than in case 1. Most have CI2 about 3 to 10 times bigger than in case 1; most lie in the range 3-17%.

Figure 11 shows the fit to the CPUE index, RDS index and growth increment data; Figure 12 shows fits to the proportions-at-length. Generally, the fits are good despite the high observation error on the abundance indices. Standardised residuals for CPUE (Figure 11B) and RDS indices (Figure 11D) are between -3 and 3, and the standardised residuals for tagging data (Figure 11F) lie between -4 and 4. There are no observed data greater than 150 mm length for the tagging data (Figure 11E), but the fit is good. The residuals of proportions-at-length (Figure 13) show no signs of pattern. Figure 14 shows the distributions of residuals, all of which have means around zero.

Figure 15A shows the exploitation rate from the simulator and the estimation model from one run. Both exploitation rates have roughly similar shape, declining over time, but in this run the estimated trajectory is under-estimated for part of the period. Figure 15B shows the recruitment from the simulator and the estimation model. Estimated recruitment is smoother than the simulated recruitment, which has lots of spikes, but the estimation model follows the trends of the recruitment from the simulator.

5.3 Experiments based on Case 2

5.3.1 Cases 3 and 4: non-linear CPUE index

The same simulated data, with $h = 0.5$, were used for Cases 3 and 4. In Case 3, the estimation model estimated h and in Case 4 the estimation model had h fixed to 1.0. Two runs from Case 3 and one run from Case 4 showed high M, $ln(R0)$, and $S50IS$ from the estimation model, and were eliminated because of probable convergence failure.

Most parameters in case 3 were estimated well (Table 6) although some were slightly overestimated. The $S95IS$ hit the upper bound for some runs and slightly overestimated. The $S50IS$ was also slightly overestimated but never hit the upper bound in any of 100 runs. The median of h was 0.56, which is close to the set value, 0.5. But the 90% range of the 100 runs showed that estimated h is highly variable, ranging from 0.34 to 0.78 (Table 6).

Bias was present for recruitment M, L_∞, α, h, $S50IS$, $S95IS$, S_{0I}, $S0$, S_{ref}, and B_{ref}, and higher bias was present for M, h, S_{0I}, S_{ref}, and B_{ref}, but all bias were less than 11% (Table 6). Spawning biomass was less well estimated than recruited biomass.

The CI2 indicators show that Case 3 has more variability, especially for biomass ratios, than Case 2.

In case 4, the parameters $ln(R0)$, M, $S50IS$ and $S95IS$ were slightly overestimated, but the growth parameters L_∞ and K were estimated with little bias.. The $S95IS$ hit the upper bound for some runs as it did for Case 3 (Table 7).

Biomasses were slightly overestimated except for the virgin recruited biomass ($B0$), but medians of the simulated value are all in the 90% range of the estimated value. The $B0$ were underestimated in both case 2 and case 3. The ratios of biomass were larger in the estimation model if they used the virgin biomass ($B0$ or $S0$) and smaller if they used the average biomass (B_{ref} or S_{ref}) (Table 7).

The PB was high for M (12.18%) but the real difference was not large (0.13 vs 0.14) (Table 7). For other parameters, the PB was generally less than 15%. The PB for h is not shown because the simulator used 0.5 but the estimation model was fixed at 1. The PB for the biomasses in the current year and reference year was higher than the parameters and the virgin biomass. The PB of biomass indicators in case 4 was always higher than in case 3 except for the $S0$.

The CI2 shows that Case 4 has similar variability of distribution of differences for most parameters and biomass indicators to Case 3.

5.3.2 Cases 5 and 6: serial auto-correlation factor in recruitment

The same simulated data, with $\rho = 0.5$, were used for Cases 5 and 6. In Case 5, the estimation model estimated ρ and in Case 6 the estimation model had ρ fixed to 0.0001, a value giving no autocorrelation. Four runs from Case 6, including two from Case 5, showed high M, $ln(R0)$, and $S50IS$ from the estimation model, and were eliminated because of probable convergence failure.

Most parameters in Case 5 were estimated well (Table 8) although some were slightly overestimated. The $S95IS$ hit the upper bound for some runs and was slightly overestimated. The median of ρ was 0.38, which is lower than the fixed value, 0.5, and the 90% range of the 100 runs showed that ρ is mostly underestimated, ranging from 0.23 to 0.57.

Biomass as slightly overestimated except for the virgin recruited biomass ($B0$) and the current recruited biomass (B_{0i}). The ratios of biomass are similar for those using the virgin biomass (Table 8) but those using the average biomass were under-estimated in the estimation model. The PB was less than 7% except for ρ (22.8%). Although PB of M is high (6.4%), the value is not significantly different from zero. S_{ref} had the highest PB, over 5%.

Biomass was similar between 100 runs of simulator and estimation model. The indicator CI2 shows that Case 5 has a similar degree of precision for most parameters and biomass indicators to those in Case 2. The parameters $ln(R0)$, M, σ_{MIN} and ratios of current to virgin biomass have twice-higher CI2 than in Case 2.

In Case 6, parameters were estimated well (Table 9) although selectivity parameters for the research diving survey were slightly overestimated. The $S95IS$ hits the upper bound for some runs. The virgin and current spawning biomasses were overestimated while the virgin and current recruited biomasses were underestimated. The reference biomasses were estimated well. The ratios of biomass are similar for those using the virgin biomass (Table 9) but those using the average biomass were lower in the estimation model as for Case 5. The precision was slightly higher than in Case 5.

The PB was always less than 10% for the parameters, biomass and the ratios. M and S_{ref} had the highest PB, over 5%. Bias tended to be slightly higher in Case 6.

5.3.3 Case 7: density-dependent growth

All 100 runs appeared to have converged. Most parameters in Case 7 (Table 10) were estimated very well. $S95IS$ hit the upper bound for some runs but was estimated reasonably well. The median of K was 0.2, which is similar to the average value set in the simulator. Although the estimation model used constant K over the time, other parameters were estimated reasonably well.

Biomasses were slightly underestimated, but were similar to the values from the simulator. The ratios of biomass from the estimation model are similar to those from the simulator (Table 10).

The PB shows that there is almost no bias, except for the M, α, σ_{MIN}, and S95IS. The PB is generally under 10% for the parameters, biomass and the ratios. M and σ_{MIN} had higher bias than other parameters (more than 6%). The large PB of σ_{MIN} means that the growth parameters are not estimated very well.

The CI2 show that Case 7 has less precision than in Case 2 except for the ratio between the current spawning biomass and the spawning biomass in reference years. The CI2 is especially higher for the biomasses in the reference years and in the current year and it is 5 times higher for L_∞.

5.3.4. Case 8: size-dependent mortality

Thirteen runs from Case 8 showed high M, $ln(R0)$, and $S50IS$ from the estimation model, and were eliminated because of probable convergence failure.

Most parameters in case 8 were estimated well (Table 11). $S95IS$ hit the upper bound for some runs but was estimated reasonably well. The median of M was 0.07 from the estimation model; the value in the simulator depends on the size structure. Although the estimation model used constant M for all sizes, other parameters were estimated reasonably well.

All biomasses except for virgin biomasses were slightly underestimated but were similar to the values from the simulator. Both spawning and recruited virgin biomasses were slightly overestimated. The ratios of biomass from the estimation model are very similar to those from the simulator (Table 11).

The PB is generally under 10% for the parameters, biomass and the ratios. All parameters were slightly biased. The minimum standard deviation of the growth increment error, σ_{MIN}, had very high PB (about 20%), but the median σ_{MIN} was similar to the real value (1.0 vs 0.99). The large PB of σ_{MIN} was influenced by the extremely low estimates (0.15 for the 5% tail, Table 11).

The CI2 show that Case 8 has less precision than Case 2 except for L_∞. The CI2 is especially higher for the recruited biomass in the reference years and it is 5 times as high as in Case 2.

5.3.5. Case 9: more tag-recapture data

Two runs from Case 9 showed high M, $ln(R0)$, and $S50IS$ from the estimation model, and were eliminated because of probable convergence failure.

Results from this case were generally similar to those from Case 2 (Table 12). The precision (CI2) of estimated growth parameters improved except for σ_{MIN}, where they deteriorated to some extent. But the PB of K shows that K is underestimated with 0.42% of bias despite the fact the K was not a biased estimate in Case 2. Precision but not bias of some biomass estimates and ratios became slightly worse with more tag-recapture data. Overall, there was relatively little difference.

5.3.6. Case 10: more length frequency data

One run from Case 10 showed low M from the estimation model, and was eliminated because of probable convergence failure. Results from this case were generally similar to those from Case 2. Bias generally improved slightly over Case 2, but increased for current exploitation rate and S_{0I} biomass ratios. The CI2 indicator improved for M, σ_{MIN} and the selectivity parameters but got worse for $S_{0I}/S0$ and $B_{0I}/B0$.

5.4. Comparison of PB and CI2 of ratios by cases

Figure 16 and Figure 17 shows the PB and CI2 comparison indicators for four biomass ratios: S_{01}/S_{ref}, $S_{01}/S0$, B_{01}/B_{ref}, and $B_{01}/B0$. Generally case 1 has least PB and CI2 for all four ratios, and other cases have less bias and precision compared to Case 1. For ratios of current biomass to virgin biomass, Case 2 has lower CI2 than cases 3 to 8. CI2 is higher in cases with mis-specified parameters (Cases 4 and 6) than in cases with estimated parameters (Cases 3 and 5). Less PB is shown in ratios with the reference biomass for most cases, but for Cases 5 and 6, the bias of S_{01}/S_{ref} is worse than the bias of $S_{01}/S0$. Ratios involving recruited biomass have less bias than those with spawning biomass. When parameters h and ρ were mis-specified (Cases 4 and 6), bias was sometimes, not always, higher than when they were estimated (Cases 3 and case 5). When the growth is density dependent (Case 7), bias was small for all ratios compared to other cases, but CI2 was large.

6. Discussion

The individual-based model for paua examined whether the paua stock assessment model is biased or not, and examined the expected degree of precision of estimation. Case 1 was a deliberately extreme case, without observation error or recruitment deviations and with large amounts of tagging data. It was designed to reveal coding errors, because the estimation model should have been able to make accurate and precise estimates from this dataset. High and accuracy and precision (Table 4) suggest (but do not prove) that the estimation model is correctly coded.

The case with much fewer data, Case 2, suggests that the model's accuracy and precision are good, even when observation error is substantial and data are limited.

The tag-recapture data appear to be very important. Because the estimation model depends heavily on tagging data to estimate the growth parameters - L_∞, K, α, and σ_{MIN}, it is important to have observed increments from the whole range of length in the tagging data, and especially lengths close to L_∞. When too few large paua are tagged, as in Case 1, σ_{MIN} is biased. Increasing the number of tagging data, as in Case 9, has almost no effect on the precision or bias in the estimation.

Spawning and recruited biomass were slightly overestimated in most cases. Recruited and spawning biomass indicators appeared to be generally similar to each other in quality. The reference biomass indicators tended to be slightly more precise than the virgin biomass indicators.

Precision of the estimates was slightly lower in these trials than the Bayesian results from the 2001 assessment (Breen et al. 2001) suggested. For instance, the 90% range of B_{01}/B_{ref} was 90-208% in Case 2 (Table 5), but only 37-60% in the PAU 7 assessment. The assumed observation error used for the simulator may have been higher than the real observation error in the assessment data. Further, in this study we compared the results from 100 point-estimate fits to 100 different data sets; whereas in the real assessment we use millions of Markov chain - Monte Carlo simulations based on one data set.

When simulated recruitments varied each year, the estimation model appeared to follow the trends but could not follow the fine-scale of recruitment variation (Figure 15B). This arises from the stochastic growth model - a strong year-class is "mushed out" as it grows, so that it appears as a series of several good year-classes when the paua appear in the proportions-at-length data. The estimation model will probably always under-estimate recruitment variability.

When recruitment variability contained auto-correlation (Cases 5 and 6), the estimation model did poorly at estimating ρ (Table 8). When the estimation was mis-specified with respect to this

parameter, bias and precision deteriorated slightly, but there was no substantial change in the assessment results.

The standardised residuals (Figures 11 and 14) show different scales for different datasets. The proportion-at-age residuals are very small, while those from the tag-recapture data are between -4 and 4. This reflects the different arbitrary weights given to the different datasets in this study. The new procedure of Breen et al. (2002a) uses iterative re-weighting to balance the residuals from all datasets, which avoids this problem and probably reflects true uncertainty more reasonably.

When we simulated non-linear CPUE, using the exponent h, this constant was estimated without much bias, but with little precision: the range was 0.34 to 0.78 when the true value was 0.50. When the model is mis-specified with respect to this parameter (we fixed h to 1 in the estimation model when the simulated value was 0.50; Case 4), there was little real deterioration of the assessment results.

We also mis-specified the model by simulating data with different assumptions: we used density-dependent K and the size-dependent M. In these runs, precision and bias were both somewhat worse than compared with Case 2, but the differences were not dramatic.

In many cases in this study, some runs showed probable convergence problems, whose symptoms were very high M values. In a real assessment, these would probably disappear when the model was given different starting values for parameters, allowing the model to converge.

Generally the stock assessment model performed well and the stock assessment indicators were reliable. The model performs best when there is no mis-specification, but performs adequately even when mis-specified. In most trials, precision was the major concern, and bias was a relatively minor problem.

The estimation model used in this study is not the current paua assessment model. The assessment model used in 2002 (Breen et al. 2002a) stock assessment could show different results, principally because the relative dataset weighting has been made more defensible; we might have less variation and bias if we used the new model.

7. Acknowledgements

8. References

Andrew, N.L., Kim S.W., Naylor, J.R., Gerring, P. & Notman, P.R. 2000. Fishery independent surveys of paua (*Haliotis iris*) in PAU 5B and PAU 5D. *New Zealand Fisheries Assessment Report 2002/xx:* (in the system)

Andrew, N.L., Worthington, D.B., Chick, R.C. & Brett, P.A. 1996. An assessment of the NSW abalone fishery in 1996. Report to the TAC Committee. Unpublished report NSW Fisheries Research Institute, Cronulla, NSW, Australia.

Beard, T.D. Jr. & Essington, T.E. 2000. Effect of angling and life history processes on bluegill size structure: insights from an individual-based-model. *Transactions of the American Fisheries Society 129(2):* 561-568.

Bentley, N., Breen, P.A., Starr. P.J. & Kendrick, T.H. 2001. Assessment of the CRA3 and NSS stock of red rock lobster (*Jasus edwardsii*) for 2000. *New Zealand Fisheries Assessment Report 2001/69.* 27 p.

Breen, P.A., Andrew, N.L. & Kendrick, T.H. 2000a. Stock assessment of paua (*Haliotis iris*) in PAU 5B and PAU 5D using a new length-based model. *New Zealand Fisheries Assessment Report 2000/33*. 37 p.

Breen, P.A., Andrew, N.L. & Kendrick, T.H. 2000b. The 2000 stock assessment of paua (*Haliotis iris*) in PAU 5B using an improved Bayesian length-based model. *New Zealand Fisheries Assessment Report 2000/48*. 36 p.

Breen, P.A., Andrew, N.L. & Kim, S.W. 2001 The 2001 stock assessment of paua (*Haliotis iris*) in PAU 7. *New Zealand Fisheries Assessment Report 2001/55*. 53 p

Breen. P.A., Kim S.W. & Andrew, N.L. 2002a. The 2002 stock assessment of paua (*Haliotis iris*) in PAU 5B. *New Zealand Fisheries Assessment Report 2002/xx*. (in the system)

Breen. P.A., Kim S.W. & Andrew, N.L. 2002b. The 2002 stock assessment of paua (*Haliotis iris*) in PAU 5D. *New Zealand Fisheries Assessment Report 2002/xx*. (in the system)

Breen. P.A., Kim, S.W., Starr, P.J. & Bentley, N. 2002c. Assessment of the red rock lobsters (*Jasus edwardsii*) in area CRA3 in 2001. *New Zealand Fisheries Assessment Report 2002/xx*. (in the system)

Chen, Y., Breen, P.A. & Andrew, N.L. 2001 Impacts of outliers and mis-specification of priors on Bayesian fisheries stock assessment. *Canadian Journal of Fisheries and Aquatic Sciences 57*: 2293–2305.

Francis, R.I.C.C. 1999. Moving towards B_{MSY}. New Zealand Fisheries Assessment Research Document 99/42: 30 p. Unpublished report held in NIWA library, Wellington, New Zealand.

McShane, P.E., Mercer, S.F. & Naylor, J.R. 1994. Spatial variation and commercial fishing of the New Zealand abalone (*Haliotis iris* and *H. australis*). *New Zealand Journal of Marine and Freshwater Research 28*: 345-355.

McShane, P.E., Mercer, S.F., Naylor, J.R. & Notman, P.R. 1996. Paua (*Haliotis iris*) fishery assessment in PAU 5, 6, and 7. New Zealand Fisheries Assessment Research Document 96/11. 35 p. Unpublished report held in NIWA library, Wellington, New Zealand.

Punt, A.E. 1990. Is B_1=K an appropriate assumption when applying an observation error production-model estimation model to catch-effort data? *South African Journal of Marine Science 9:* 249-259.

Punt, A.E. 1992 Selecting management methodologies for marine resources, with an illustration for southern African hake. *Benguela Trophic Functioning: South African Journal of Marine Science 12*: 943-958.

Punt, A.E. & Kennedy, R.B. 1997. Population modelling of Tasmanian rock lobster resources. *Marine and Freshwater Research 48*: 967–980.

Rose, K.E., Cowan J.H. Jr., Clark, M.E., Houde, E.D. & Wang, S. 1999. An individual-based model of bay anchovy population dynamics in the mesohaline region of Chesapeake Bay. *Marine Ecology Progress Series 185*: 113-132.

Schiel, D.R. & Breen, P.A. 1991. Population structure, ageing, and fishing mortality of the New Zealand abalone *Haliotis iris*. *Fishery Bulletin 89(4):* 681-691.

Schiel, D.R. 1989. Paua fishery assessment 1989. New Zealand Fisheries Assessment Research Document 89/9. 20 p. Unpublished report held in NIWA library, Wellington, New Zealand.

Schiel, D.R. 1991. The paua (abalone) fishery of New Zealand. pp 427-437 *In* Shepherd, S.A., Tegner, M.J. & Guzman del Proo, S. (Eds.) Abalone of the world: Biology fisheries and culture. Blackwell Scientific, Oxford.

Worthington, D.G., Chick, R.C., Blount, C., Brett, P.A. & Gibson, P.T.A. 1998. Final assessment of the NSW abalone fishery in 1997. Unpub. Report, Fisheries Research Institute, New South Wales Fisheries, Cronulla, NSW, Australia. 67 p.

Table 1. Model parameters specified for the simulator (above the line) and derived parameters (below the line).

$\ln(R0)$	log of the base number of recruited animals
$Rdev$	a vector of recruitment deviations modifying the actual model recruitment in each year
ρ	a parameter determining serial auto-correlation used in $Rdev$ calculation
h	the exponent of the relation between CPUE and the recruited biomass
M	the instantaneous rate of natural mortality, assumed to be constant over time and paua sizes
L_∞	the asymptotic length
K	the instantaneous rate of approach to L_∞ over a year
α	the CV of the growth increment error
σ_{MIN}	the minimum standard deviation of the growth increment error
σ_ε	standard deviation of the recruitment residuals in log space
$S50IS$	the paua size at which research divers are 50% effective
$S95IS$	the paua size at which research divers are 95% effective
U_y	the exploitation rate for year y

$B0$	the virgin recruited biomass, B_{1974}
$S0$	the virgin spawning biomass, S_{1974}
B_{01}	the current recruited biomass, B_{2001}
S_{01}	the current spawning biomass, S_{2001}
B_{ref}	the average recruited biomass over B_{1985}-B_{1987}
S_{ref}	the average spawning biomass over S_{1985}-S_{1987}

Table 2. Description of cases.

	Simulator			Estimation model
Cases	observation error	data	others	
1	No error	As many as possible		$\varepsilon = 0$
2	With relevant CVs	far fewer data		ε estimated
3	As in case 2	As in case 2	$h = 0.5$	estimated h
4	As in case 2	As in case 2	$h = 0.5$	$h = 1$
5	As in case 2	As in case 2	$\rho = 0.5$	Estimated ρ
6	As in case 2	As in case 2	$\rho = 0.5$	$\rho = 0.0001$
7	As in case 2	As in case 2	density-dependent growth	
8	As in case 2	As in case 2	size-dependent mortality	
9	As in case 2	As in case 2, but more tag-recaptured data		As in case 2
10	As in case 2	As in case 2, but more length frequency data		· As in case 2

Table 3. Parameters estimated in the model, their bounds and prior distributions. LB: lower bound; UB: upper bound; prior types: U, uniform; N, normal; L, lognormal. For definitions of parameters see Table 1. Initial values in bold indicate a parameter held fixed in Case 2, also indicated by the negative value for estimation phase. Dash indicates not applicable.

Parameters	Phase	LB	UB	Prior	Mean	CV	Initial values
$R0$	1	5	50	U	–	–	7
M	1	0.01	0.5	U	–	–	0.13
L_∞	2	100	250	U	–	–	155.95
K	2	0.01	0.8	U	–	–	0.2
$S50IS$	2	70	125	U	–	–	99.1
$S95IS$	2	70	125	U	–	–	116.35
ε_y^+	-1	-2.3	2.3	N	1	0.4	0
ρ	-3	0.00001	0.99	U	–	–	**0.0001**
α	.1	0	1	U	–	–	0.4
σ_{MIN}	1	0	10	U	–	–	1
h	-1	0	2	U	–	–	**1**

+ ε_y is fixed at 0 for case 1.

Table 4. Summary of parameter values from 100 runs of the simulator and the estimation model for Case 1, the comparison indices (CI2) and percentage of bias (PB). Negative PB indicates parameter has been underestimated in the estimation model.

	Simulator			Estimation model			Tests	
	0.05	median	0.95	0.05	median	0.95	PB (%)	CI2 (%)
$ln(R0)$		7.0		6.96	7.00	7.03	0.01	0.3
M		0.13		0.13	0.13	0.13	0.13	1.9
K		0.2		0.19	0.20	0.20	-0.96*	1.0
L_∞		155.95		155.25	155.78	156.20	-0.12*	0.2
α		0.40		0.40	0.40	0.41	1.04*	0.8
σ_{MIN}		1		0.78	0.89	1.00	-10.83*	7.4
$S50IS$		99.1		99.41	99.74	100.08	0.65*	0.2
$S95IS$		116.35		116.53	117.39	118.25	0.85*	0.4
U_{or}		0.47		0.43	0.47	0.51	-0.21	5.1
S_{or}	694.2	711.5	728.4	698.9	724.4	749.4	1.87*	2.6
B_{or}	272.7	286.5	299.3	268.9	285.9	305.0	0.47	5.1
$S0$	2117.4	2148.7	2180.9	2119.9	2166.0	2207.9	0.70*	1.0
$B0$	1689.8	1725.0	1758.5	1676.5	1725.8	1770.1	0.05	1.3
S_{ref}	653.9	668.1	682.1	665.1	683.0	702.2	2.31*	1.7
B_{ref}	234.1	243.1	251.0	234.6	244.2	256.4	1.00*	3.1
$S_{or}/S0$	0.32	0.33	0.34	0.32	0.33	0.35	1.16*	2.5
$B_{or}/B0$	0.16	0.17	0.18	0.16	0.17	0.18	0.41	4.8
S_{or}/S_{ref}	1.03	1.06	1.10	1.04.	1.06	1.09	-0.41	2.9
B_{or}/B_{ref}	1.11	1.18	1.26	1.10	1.16	1.26	-0.41	6.3

Table 5. Summary of parameter values from 99 runs of the simulator and the estimation model for Case 2, the comparison indices (CI2) and percentage of bias (PB).

	Simulator			Estimation model			Tests	
	0.05	median	0.95	0.05	median	0.95	PB (%)	CI2 (%)
$ln(R0)$		7.0		6.79	7.16	7.52	2.15*	2.9
M		0.13		0.11	0.14	0.18	10.94*	16.7
K		0.2		0.19	0.20	0.21	-0.01	3.4
L_∞		155.95		154.17	155.42	156.45	-0.35*	0.5
α		0.40		0.38	0.41	0.43	1.32*	3.2
σ_{MIN}		1		0.60	1.03	1.21	-0.28	19.9
$S50IS$		99.1		98.05	100.42	103.13	1.30*	1.5
$S95IS$		116.35		112.69	118.66	125.00	1.69*	3.3
U_{or}		0.35		0.27	0.35	0.46	0.76	16.0
S_{or}	664.9	801.6	1049.1	647.0	831.2	1128.6	3.46*	16.3
B_{or}	281.0	370.7	478.8	266.0	379.7	522.7	1.74	16.2
$S0$	1870.3	2146.9	2452.4	1891.0	2168.0	2515.8	1.53*	3.2
$B0$	1437.8	1729.2	1978.0	1408.8	1683.9	2074.1	-1.75*	6.9
S_{ref}	557.2	690.8	915.0	562.4	734.1	1044.2	5.78*	6.2
B_{ref}	214.8	264.7	377.5	195.2	271.9	399.2	0.31	11.6
$S_{or}/S0$	0.28	0.38	0.51	0.27	0.38	0.52	1.83	15.0
$B_{or}/B0$	0.15	0.22	0.31	0.14	0.22	0.33	3.92*	17.1
S_{or}/S_{ref}	0.79	1.20	1.60	0.83	1.14	1.54	-2.22	14.0
B_{or}/B_{ref}	0.90	1.36	2.02	0.90	1.37	2.08	2.44	18.4

Table 6. Summary of parameter values from 97 runs of the simulator and the estimation model for Case 3, non linear CPUE, h=0.5 in the simulator, estimating h in the estimation model, the comparison indices (CI2) and percentage of bias (PB).

	Simulator			Estimation model			Tests	
	0.05	median	0.95	0.05	median	0.95	PB (%)	CI2 (%)
$ln(R0)$		7.0		6.76	7.12	7.51	1.87*	3.4
M		0.13		0.10	0.14	0.19	8.78*	19.5
K		0.2		0.19	0.20	0.21	-0.45	2.8
L_∞		155.95		154.29	155.51	156.34	-0.31*	0.4
α		0.40		0.38	0.41	0.43	1.56*	3.4
σ_{MIN}		1		0.49	1.07	1.21	-0.05	24.0
$S50IS$		99.1		98.16	100.00	103.32	1.24*	1.6
$S95IS$		116.35		112.19	118.20	125.00	1.57*	3.3
U_{0I}		0.35		0.26	0.34	0.46	-0.40	18.0
S_{0I}	611.5	753.4	1026.9	598.6	806.9	1118.9	6.03*	18.8
B_{0I}	273.2	356.5	476.4	264.4	368.7	498.3	3.74	19.2
$S0$	1909.4	2142.5	2529.1	1965.0	2186.1	2503.4	2.71*	5.2
$B0$	1523.3	1736.4	2011.9	1423.2	1695.3	2047.4	-1.10	9.1
S_{ref}	543.8	692.9	897.1	546.8	740.2	990.2	6.28*	7.5
B_{ref}	206.2	270.2	371.0	192.0	276.8	401.5	3.89*	13.0
$S_{0I}/S0$	0.28	0.35	0.50	0.27	0.37	0.54	3.44	18.4
$B_{0I}/B0$	0.15	0.20	0.28	0.14	0.21	0.34	6.02*	22.8
S_{0I}/S_{ref}	0.79	1.10	1.50	0.84	1.09	1.47	-0.42	14.9
B_{0I}/B_{ref}	0.92	1.31	1.85	0.93	1.27	1.82	1.10	21.4
h		0.50		0.34	0.56	0.78	10.83*	26.7

Table 7. Summary of parameter values from 97 runs of the simulator and the estimation model for Case 4, the same data as in Case 3 but fixed h (= 1) in the estimation model, the comparison indices (CI2) and percentage of bias (PB).

	Simulator			Estimation model			Tests	
	0.05	median	0.95	0.05	median	0.95	PB (%)	CI2 (%)
$ln(R0)$		7.0		6.81	7.19	7.55	2.57*	3.3
M		0.13		0.11	0.14	0.19	12.18*	18.5
K		0.2		0.19	0.20	0.21	-0.50	2.9
L_∞		155.95		154.26	155.48	156.33	-0.33*	0.4
α		0.40		0.38	0.41	0.43	1.61*	3.5
σ_{MIN}		1		0.54	1.06	1.24	0.23	23.6
$S50IS$		99.1		98.16	100.40	103.24	1.40*	1.5
$S95IS$		116.35		112.51	118.61	125.00	1.77*	3.1
U_{0I}		0.35		0.24	0.32	0.45	-4.58*	17.3
S_{0I}	611.5	753.4	1026.9	635.1	857.1	1162.5	11.43*	19.3
B_{0I}	273.2	356.5	476.4	282.7	391.0	525.3	8.26*	19.9
$S0$	1909.4	2142.5	2529.1	2010.6	2183.4	2463.8	2.44*	3.8
$B0$	1523.3	1736.4	2011.9	1444.0	1686.8	2047.2	-2.54*	5.7
S_{ref}	543.8	692.9	897.1	580.3	758.9	1018.9	10.33*	7.6
B_{ref}	206.2	270.2	371.0	232.4	300.7	417.3	13.89*	13.0
$S_{0I}/S0$	0.28	0.35	0.50	0.28	0.39	0.55	8.72*	18.0
$B_{0I}/B0$	0.15	0.20	0.28	0.15	0.23	0.33	11.63*	22.3
S_{0I}/S_{ref}	0.79	1.10	1.50	0.86	1.12	1.46	0.82	14.8
B_{0I}/B_{ref}	0.92	1.31	1.85	0.95	1.23	1.63	-4.07*	19.0
h		0.50			1.00		-	-

Table 8. Summary of parameter values from 96 runs of the simulator and the estimation model for Case 5, auto-correlated recruitment deviations ($\rho = 0.5$) in the simulator and estimated ρ in the estimation model, the comparison index (CI2) and percentage of bias (PB).

	Simulator			Estimation model			Tests	
	0.05	median	0.95	0.05	median	0.95	PB (%)	CI2 (%)
$ln(R0)$		7.0		6.47	7.00	7.73	0.69	5.7
M		0.13		0.09	0.14	0.22	6.39	32.8
K		0.2		0.19	0.20	0.21	-0.48	3.0
L_∞		155.95		154.33	155.50	156.85	-0.27*	0.5
α		0.40		0.39	0.41	0.43	1.46*	3.3
σ_{MIN}		1		0.20	1.05	1.28	-1.69	27.7
$S50IS$		99.1		97.40	99.68	103.89	1.03*	2.1
$S95IS$		116.35		111.59	117.44	125.00	·1.44*	3.7
U_{0I}		0.35		0.28	0.36	0.49	4.82*	18.7
S_{0I}	484.9	720.1	1196.9	489.2	722.4	1143.7	0.32	19.6
B_{0I}	219.0	342.7	553.0	206.5	341.3	540.4	-1.53	17.8
$S0$	1657.5	2115.4	2954.6	1745.1	2173.0	2888.5	2.43*	5.5
$B0$	1308.4	1710.6	2440.3	1241.8	1673.1	2505.7	-1.03	10.1
S_{ref}	427.0	661.9	1140.7	396.3	660.3	1374.4	5.04*	11.5
B_{ref}	150.6	259.4	469.3	149.7	260.6	531.0	0.84	10.8
$S_{0I}/S0$	0.21	0.36	0.63	0.20	0.33	0.56	-2.22	17.2
$B_{0I}/B0$	0.12	0.20	0.35	0.10	0.19	0.40	0.58	21.3
S_{0I}/S_{ref}	0.59	1.12	2.21	0.55	1.06	1.81	-4.42*	15.6
B_{0I}/B_{ref}	0.68	1.35	2.64	0.67	1.30	2.59	-2.11	15.8
ρ		0.50		0.23	0.38	0.57	-22.77*	23.6

Table 9. Summary of parameter values from 96 runs of the simulator and the estimation model for Case 6, the same data as Case 5, but fixed ρ (= 0.0001) in the estimation model, the comparison indices (CI2) and percentage of bias (PB).

	Simulator			Estimation model			Tests	
	0.05	median	0.95	0.05	median	0.95	PB (%)	CI2 (%)
$ln(R0)$		7.0		6.47	7.02	7.87	1.29*	6.3
M		0.13		0.08	0.14	0.23	9.50*	36.4
K		0.2		0.19	0.20	0.21	-0.45	3.0
L_∞		155.95		154.34	155.48	156.86	-0.26*	0.5
α		0.40		0.39	0.41	0.43	1.43*	3.3
σ_{MIN}		1		0.20	1.05	1.29	-2.40	28.0
$S50IS$		99.1		96.87	99.84	104.39	1.27*	2.3
$S95IS$		116.35		110.50	118.16	125.00	1.72*	4.0
U_{0I}		0.35		0.27	0.35	0.50	4.01	20.2
S_{0I}	484.9	720.1	1196.9	496.0	755.0	1192.6	3.21	23.7
B_{0I}	219.0	342.7	553.0	214.6	334.7	522.9	-0.10	20.2
$S0$	1657.5	2115.4	2954.6	1777.1	2176.2	2906.4	3.26*	6.6
$B0$	1308.4	1710.6	2440.3	1231.0	1650.0	2505.6	-1.17	10.7
S_{ref}	427.0	661.9	1140.7	395.8	667.7	1413.5	6.35*	13.3
B_{ref}	150.6	259.4	469.3	152.8	259.8	519.8	1.61	12.3
$S_{0I}/S0$	0.21	0.36	0.63	0.21	0.35	0.57	-0.39	19.9
$B_{0I}/B0$	0.12	0.20	0.35	0.11	0.20	0.38	2.28	24.1
S_{0I}/S_{ref}	0.59	1.12	2.21	0.62	1.07	1.72	-3.03	17.6
B_{0I}/B_{ref}	0.68	1.35	2.64	0.74	1.29	2.40	-1.35	17.8
ρ		0.50			0.0001		-	-

Table 10. Summary of parameter values from 100 runs of the simulator and the estimation model for Case 7 (K from the simulator is fixed and density dependent, hence cannot be directly compared with the K from the estimation model), the comparison indices (CI2) and percentage of bias (PB).

	Simulator			Estimation model			Tests	
	0.05	median	0.95	0.05	median	0.95	PB (%)	CI2 (%)
$ln(R0)$		7.0		6.71	7.01	7.49	0.56	3.5
M		0.13		0.11	0.14	0.17	6.81*	14.2
K		-		0.12	0.20	0.24	-	-
L_∞		155.95		152.26	155.66	164.63	0.34	2.6
α		0.40		0.38	0.41	0.43	1.43*	3.7
σ_{MIN}		1		0.65	1.10	1.39	6.79*	25.5
$S50IS$		99.1		96.18	99.45	104.62	0.65*	2.7
$S95IS$		116.35		107.09	115.90	125.00	0.22	5.5
U_o		0.35		0.24	0.35	0.50	2.83	23.8
S_o	636.2	763.6	921.0	557.4	751.4	1053.0	1.95	25.5
B_o	223.8	328.5	422.2	243.8	314.5	444.6	2.67	25.3
$S0$	1853.3	1991.7	2184.4	1769.1	1976.2	2298.2	0.10	8.7
$B0$	1334.4	1563.3	1739.4	1344.3	1523.8	1784.0	-0.89	8.5
S_{ref}	529.9	660.9	835.6	454.4	654.8	1019.9	3.54	27.3
B_{ref}	189.4	245.2	320.3	164.9	243.0	345.4	-0.46	17.8
$S_o/S0$	0.31	0.39	0.46	0.29	0.38	0.49	1.04	17.8
$B_o/B0$	0.15	0.21	0.28	0.16	0.21	0.30	3.79	24.5
S_o/S_{ref}	0.87	1.16	1.53	0.84	1.13	1.49	-0.38	14.6
B_o/B_{ref}	0.87	1.39	1.81	0.90	1.32	1.98	4.25	22.6

Table 11. Summary of parameter values from 87 runs of the simulator and the estimation model for Case 8 (M in simulator is size dependent and fixed, hence cannot be compared directly with M from the estimation model), the comparison indices (CI2) and percentage of bias (PB).

	Simulator			Estimation model			Tests	
	0.05	median	0.95	0.05	median	0.95	PB (%)	CI2 (%)
$ln(R0)$		7.0		6.01	6.68	6.89	-4.92*	3.0
M		-		0.03	0.07	0.09	--	-
K		0.2		0.19	0.20	0.21	-0.74*	3.5
L_∞		155.95		154.24	155.63	156.67	-0.24*	0.5
α		0.40		0.39	0.41	0.44	2.31*	3.9
σ_{MIN}		1		0.15	0.99	1.21	-19.08*	39.2
$S50IS$		99.1		95.43	97.69	100.82	-1.37*	1.5
$S95IS$		116.35		105.49	113.33	120.63	-2.53*	3.6
U_o		0.35		0.28	0.36	0.51	5.50*	18.6
S_o	605.3	762.0	1020.7	489.0	732.8	1005.4	-5.64*	15.7
B_o	284.3	383.3	493.9	245.9	369.9	504.0	-2.50	16.0
$S0$	3534.2	3826.2	4379.4	3484.0	3869.2	4624.8	0.29	4.6
$B0$	3170.4	3445.4	3963.6	3050.0	3464.8	4234.6	0.33	7.1
S_{ref}	538.7	676.8	872.0	513.1	655.9	858.0	-3.57*	3.5
B_{ref}	211.0	278.4	376.3	204.8	279.0	399.8	-1.02	9.7
$S_o/S0$	0.15	0.20	0.27	0.12	0.19	0.26	-5.55*	16.8
$B_o/B0$	0.08	0.11	0.15	0.07	0.10	0.15	-2.13	18.3
S_o/S_{ref}	0.83	1.12	1.53	0.78	1.10	1.55	-2.32	14.9
B_o/B_{ref}	0.97	1.35	1.87	0.85	1.32	1.90	-0.75	18.1

Table 12. Summary of parameter values from 98 runs of the simulator and the estimation model for Case 9 (upgrade version of case 2 with more tag recapture data), the comparison indices (CI2) and percentage of bias (PB).

	Simulator 0.05	median	0.95	Estimation model 0.05	median	0.95	Tests PB (%)	CI2 (%)
$ln(R0)$		7.0		6.77	7.17	7.50	2.16*	2.9
M		0.13		0.11	0.14	0.18	11.02*	16.8
K		0.2		0.19	0.20	0.21	-0.42*	1.7
L_∞		155.95		154.76	155.49	156.14	-0.29*	0.3
α		0.40		0.39	0.41	0.42	1.54*	1.6
σ_{MIN}		1		0.20	1.08	1.24	-1.83	27.6
$S50IS$		99.1		98.00	100.35	103.19	1.36*	1.5
$S95IS$		116.35		112.55	118.32	125.00	1.82*	3.3
U_{0l}		0.35		0.27	0.35	0.46	0.92	16.0
S_{0l}	664.6	800.4	1049.7	626.0	834.5	1114.4	3.39*	16.1
B_{0l}	280.9	367.9	482.3	263.6	374.7	519.6	1.57	16.2
$S0$	1870.0	2143.2	2438.7	1907.3	2171.4	2490.8	1.52*	3.2
$B0$	1437.8	1725.7	1964.0	1403.1	1677.4	2061.2	-1.79*	6.9
S_{ref}	556.2	690.2	911.7	555.0	735.5	996.2	5.86*	6.4
B_{ref}	214.6	262.6	376.1	195.9	270.5	397.6	0.10	11.5
$S_{0l}/S0$	0.28	0.38	0.51	0.27	0.38	0.51	1.79	15.0
$B_{0l}/B0$	0.15	0.22	0.31	0.14	0.22	0.32	3.81*	17.2
S_{0l}/S_{ref}	0.79	1.20	1.60	0.83	1.14	1.54	-2.32	14.1
B_{0l}/B_{ref}	0.90	1.39	2.02	0.90	1.38	2.07	2.48	18.4

Table 13. Summary of parameter values from 100 runs of the simulator and the estimation model for Case 10 (upgrade version of case 2 with more length frequency data), the comparison indices (CI2) and percentage of bias (PB).

|	Simulator 0.05	median	0.95	Estimation model 0.05	median	0.95	Tests PB (%)	CI2 (%)
$ln(R0)$		7.0		6.84	7.09	7.35	1.18*	2.1
M		0.13		0.11	0.14	0.16	5.87*	11.8
K		0.2		0.19	0.20	0.21	-0.79*	2.4
L_∞		155.95		154.99	155.66	156.47	-0.17*	0.3
α		0.40		0.38	0.41	0.43	1.10*	3.2
σ_{MIN}		1		0.72	1.03	1.22	0.88	14.3
$S50IS$		99.1		98.79	99.94	101.38	0.93*	0.8
$S95IS$		116.35		114.83	117.69	121.33	1.37*	1.9
U_{0l}		0.35		0.27	0.36	0.47	4.99*	17.7
S_{0l}	664.7	802.6	1049.4	587.8	778.5	1033.4	-2.50	15.6
B_{0l}	280.9	372.7	480.6	258.4	360.3	484.6	-2.10	16.3
$S0$	1870.2	2144.1	2455.8	1905.3	2170.1	2475.9	1.33*	2.1
$B0$	1437.8	1727.3	1979.0	1460.5	1712.5	2024.0	-0.54	4.3
S_{ref}	556.7	690.9	915.9	571.0	721.9	945.9	3.93*	4.3
B_{ref}	214.7	265.1	377.7	214.0	272.1	390.6	0.61	4.3
$S_{0l}/S0$	0.28	0.38	0.51	0.28	0.36	0.46	-3.81*	15.1
$B_{0l}/B0$	0.15	0.22	0.31	0.15	0.21	0.29	-1.66	15.1
S_{0l}/S_{ref}	0.79	1.20	1.60	0.77	1.08	1.50	-6.12*	15.0
B_{0l}/B_{ref}	0.90	1.34	2.02	0.86	1.27	1.91	-2.49	17.0
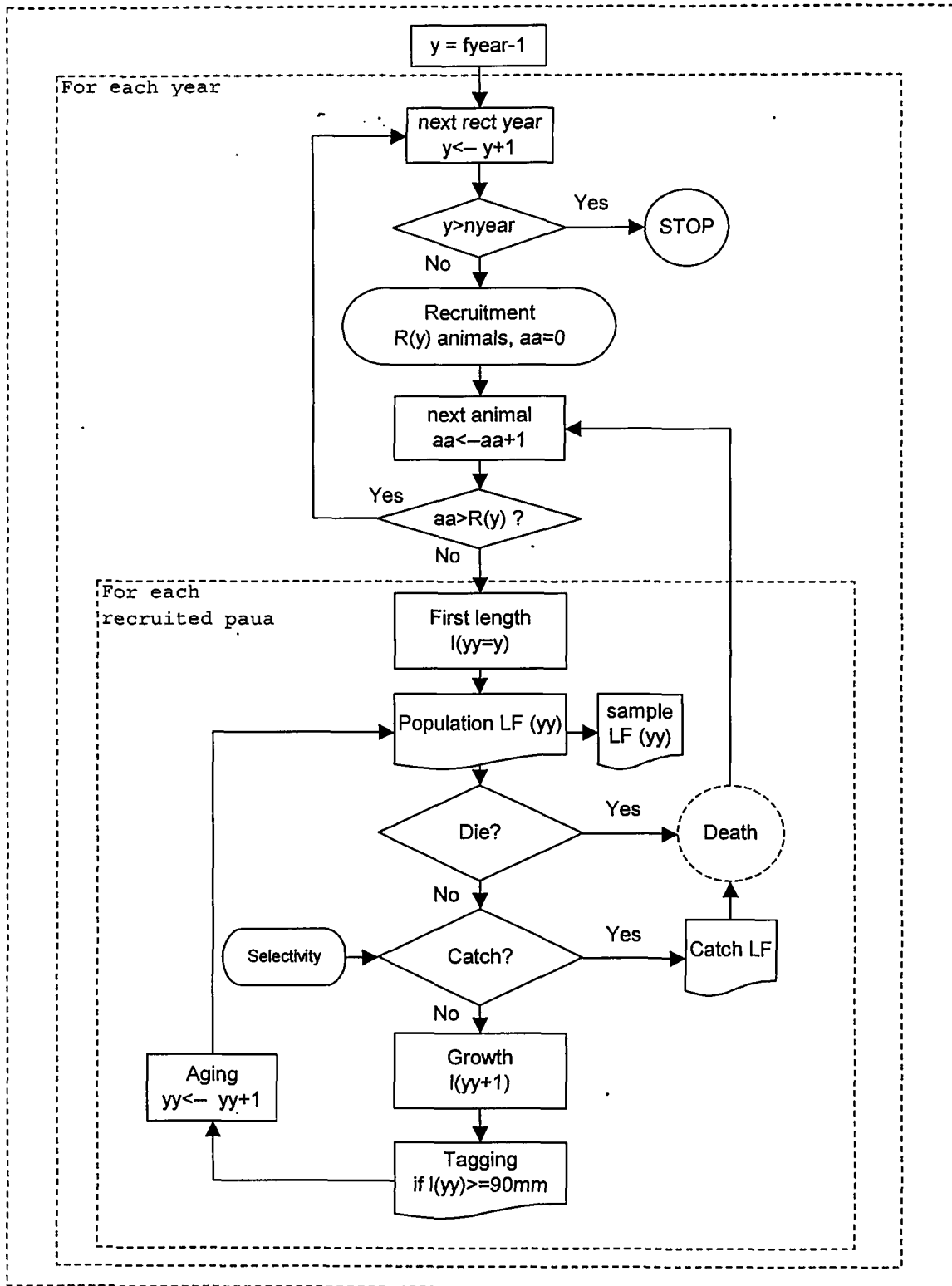
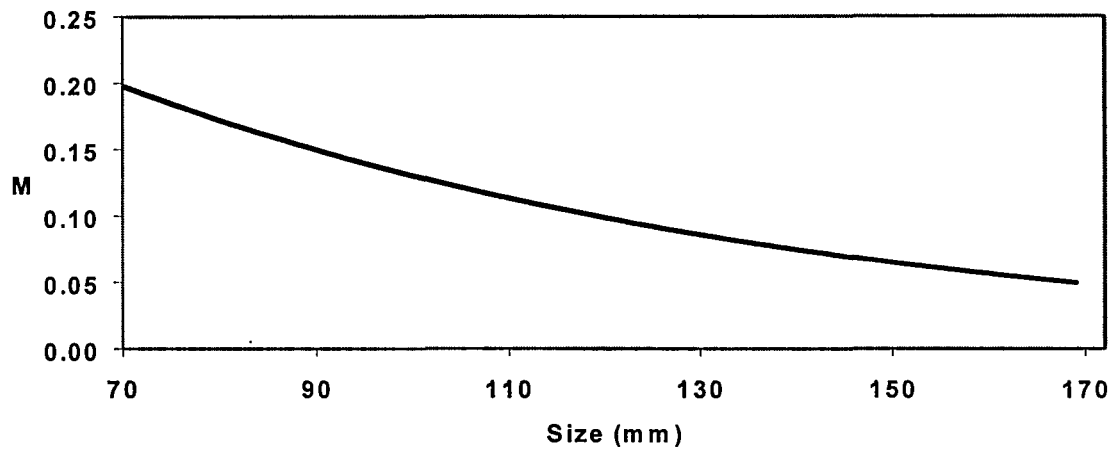**Figure 1: Flow diagram of the data simulator.**

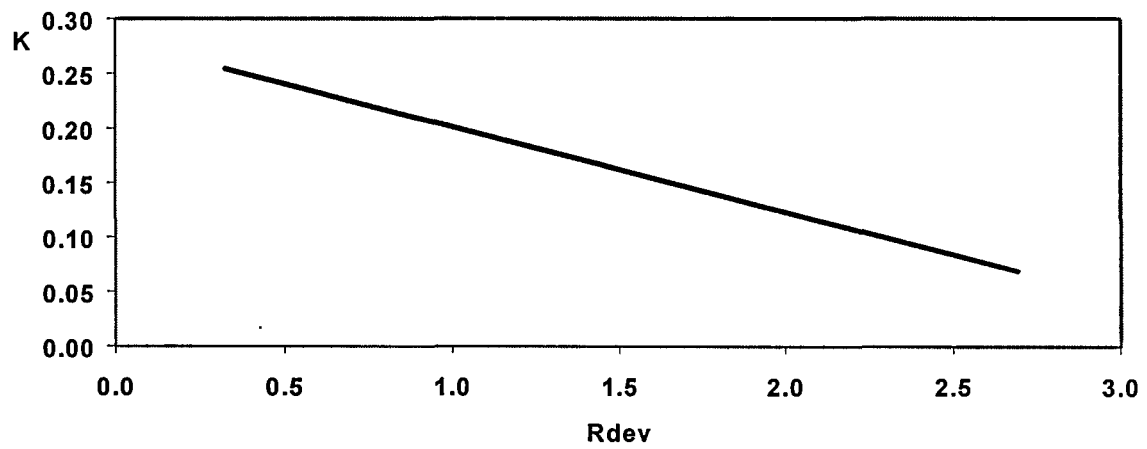Figure 2. Assumed relation between *M* and paua length in Case 8, with size-dependent *M*.

Figure 3. Assumed relation between *K* and the recruitment deviation (*Rdev*) in Case 7, with density-dependent *K*.
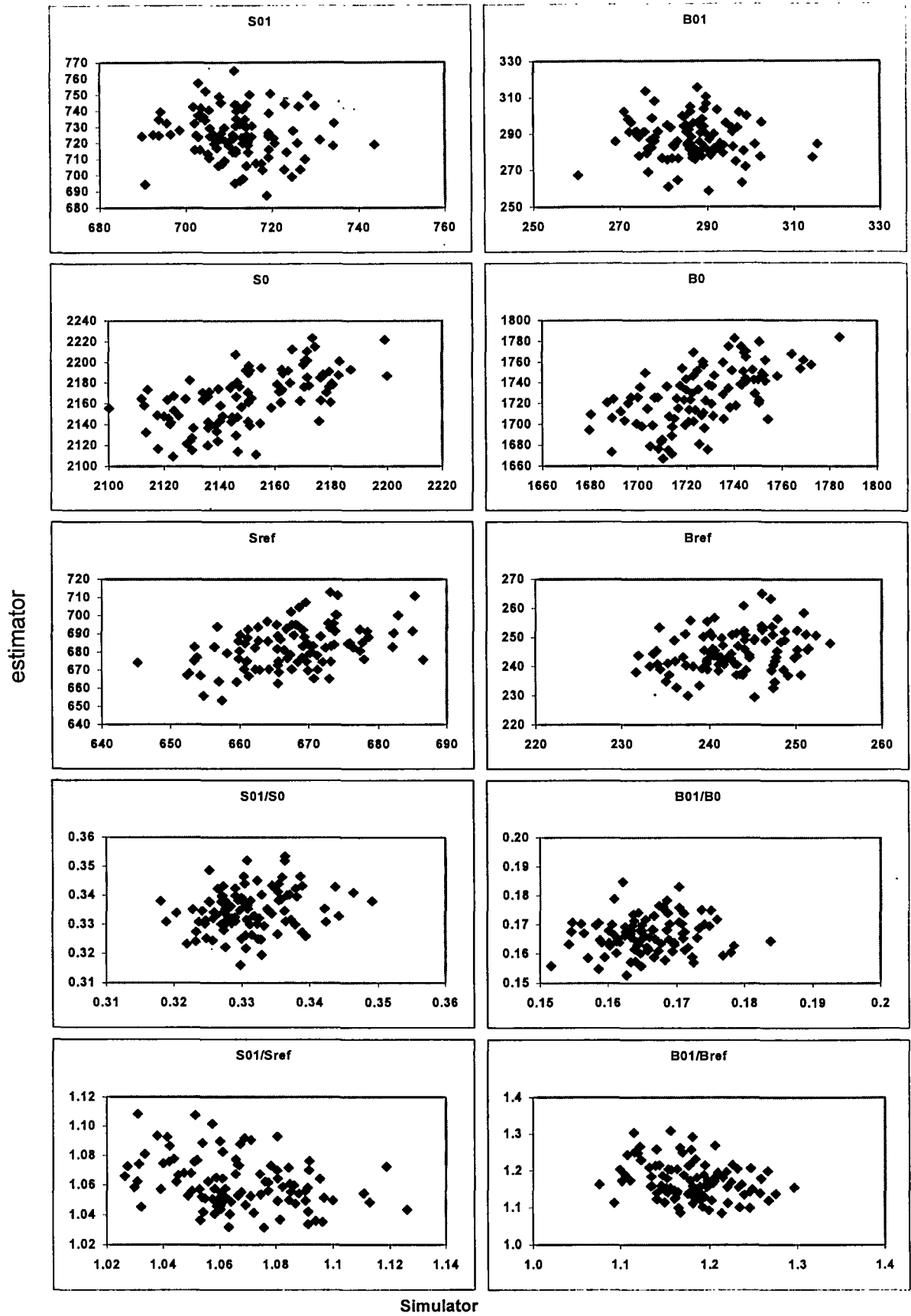
Figure 4. Scatter plot of biomass and biomass ratios from the estimator against the values from the simulator of Case 1.
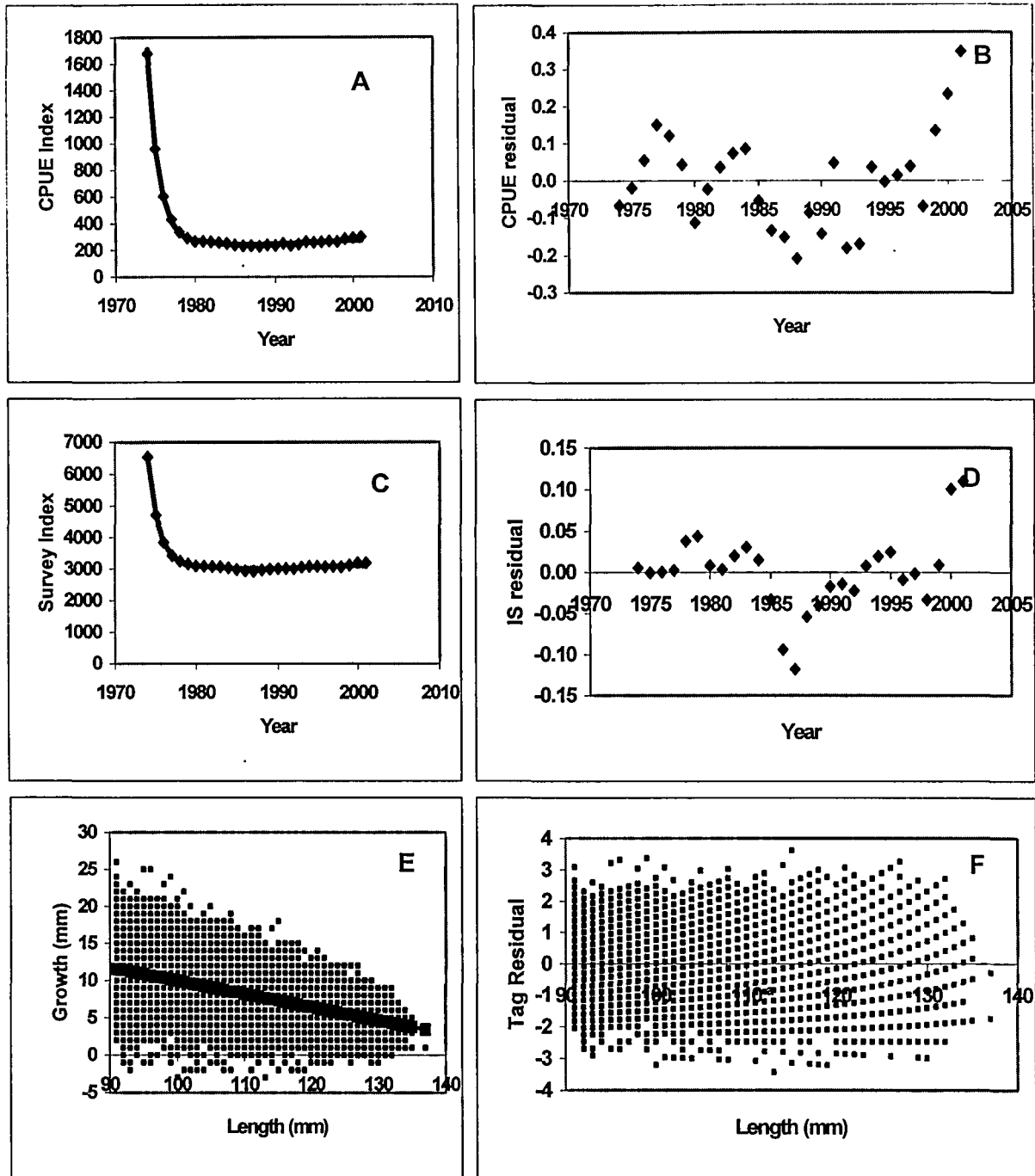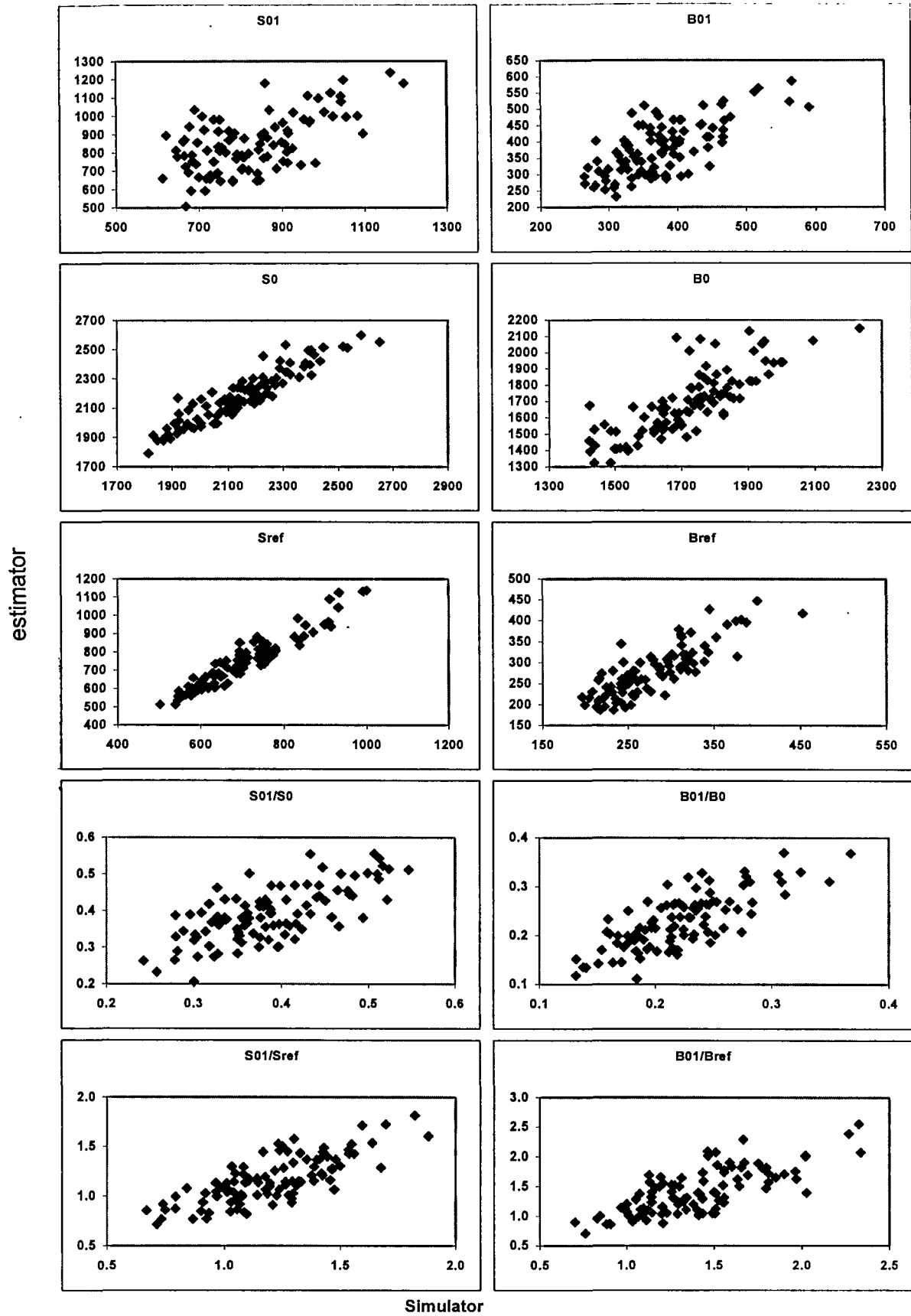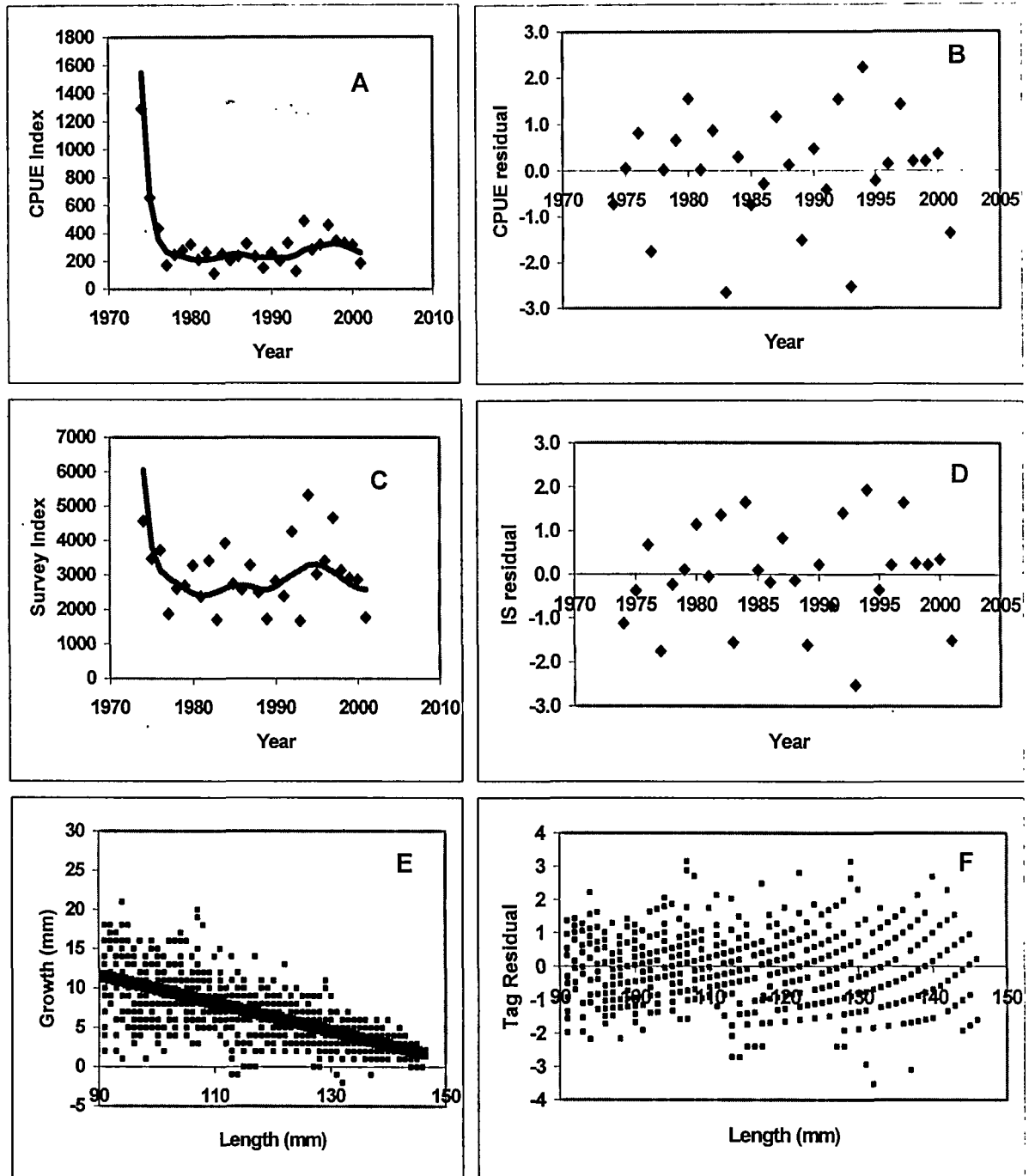
Figure 5: Results from the fit for case 1. A: Observed (squares) and predicted (solid line) CPUE; B: standardised CPUE residuals; C: observed (squares) and predicted (solid line) research diver survey index; D: standardised research diver survey index residuals; E: observed (small squares) and predicted (larger squares) length increments; F: standardised growth increment residuals.
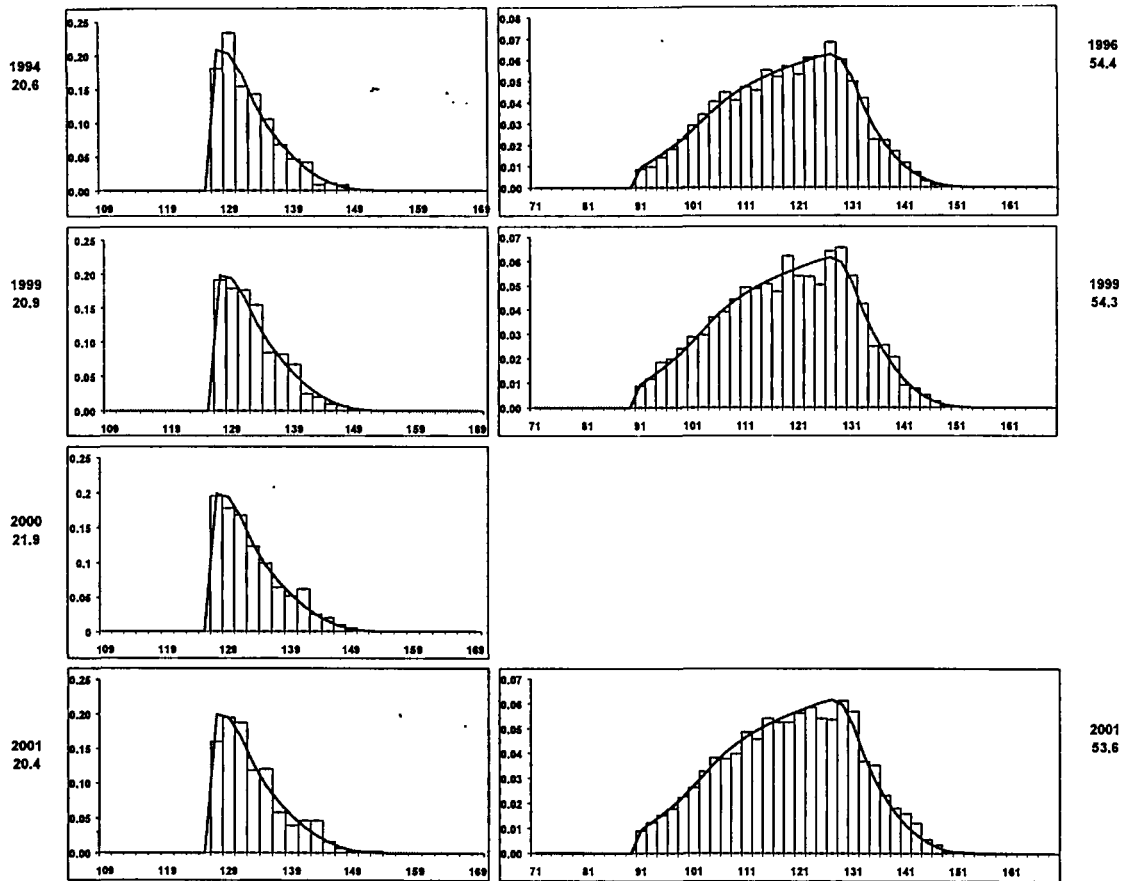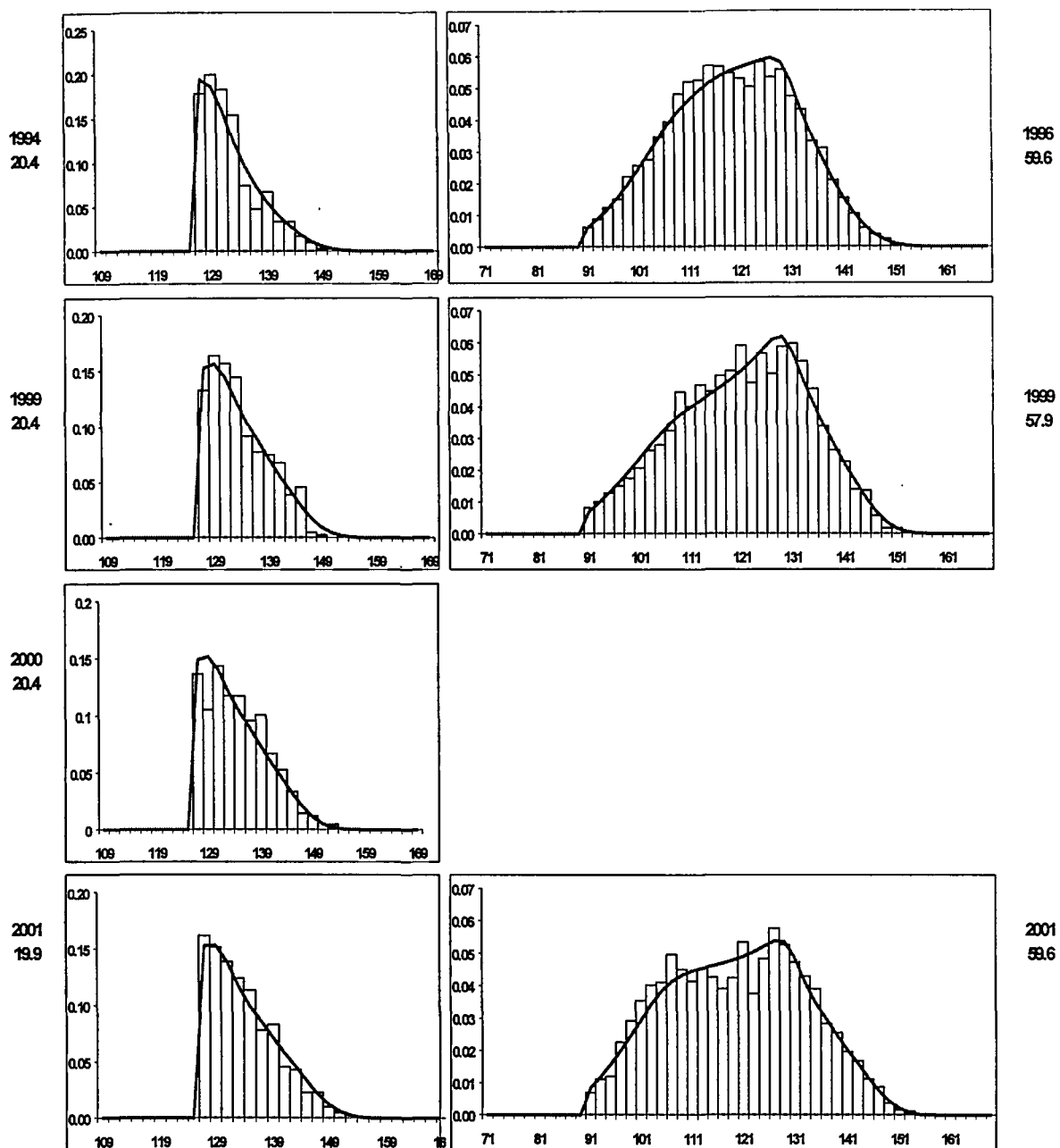
Figure 6: Observed (bars) and predicted (lines) proportions-at-length from commercial catch sampling (left) and research diver surveys (right) from case 1. The number under the year is the effective sample size.
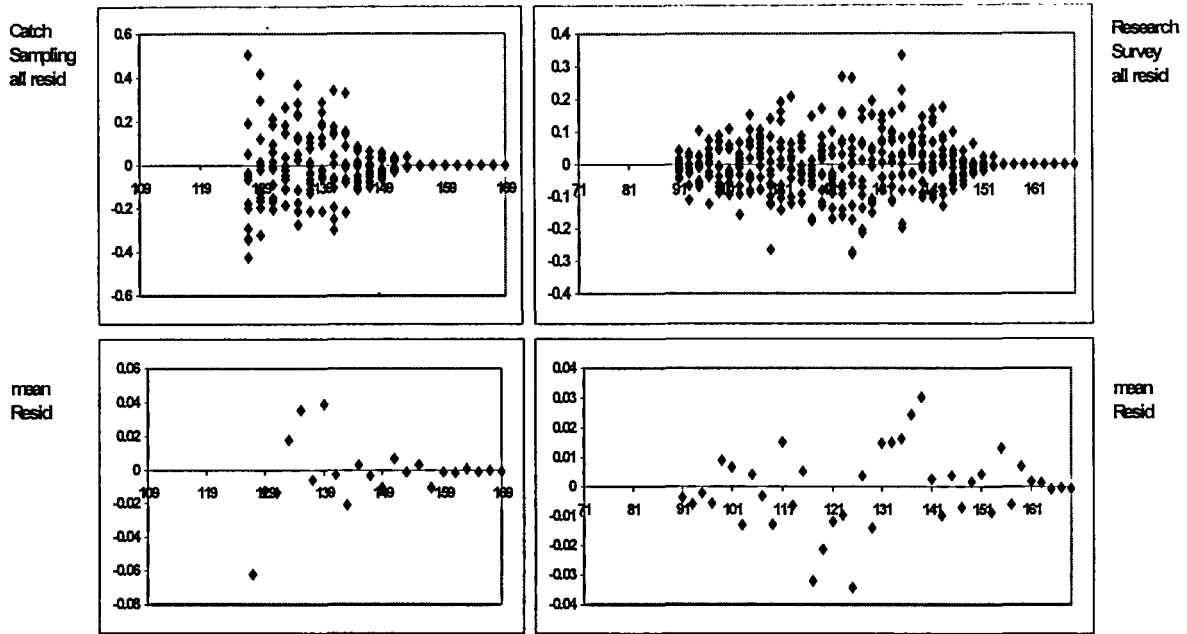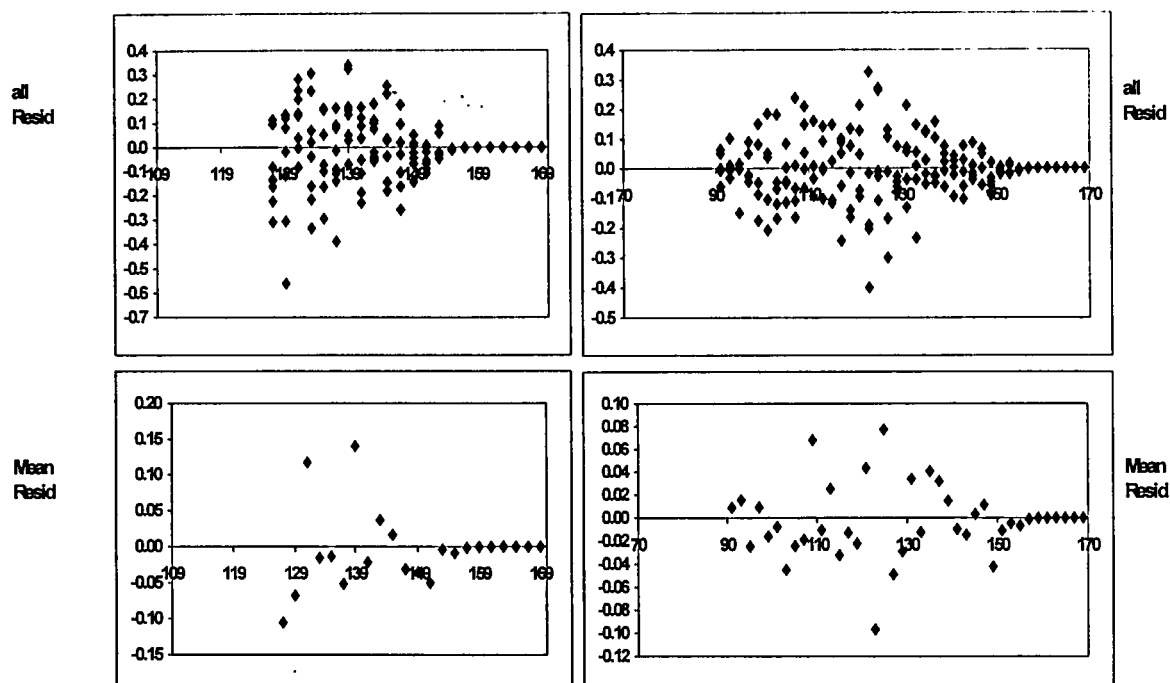
Figure 7: Standardised residuals from proportions-at-length (top) and means of all standardised proportion-at-length residuals (bottom) *vs* observed length from commercial catch sampling (left) and research diver surveys (right) from case 1.
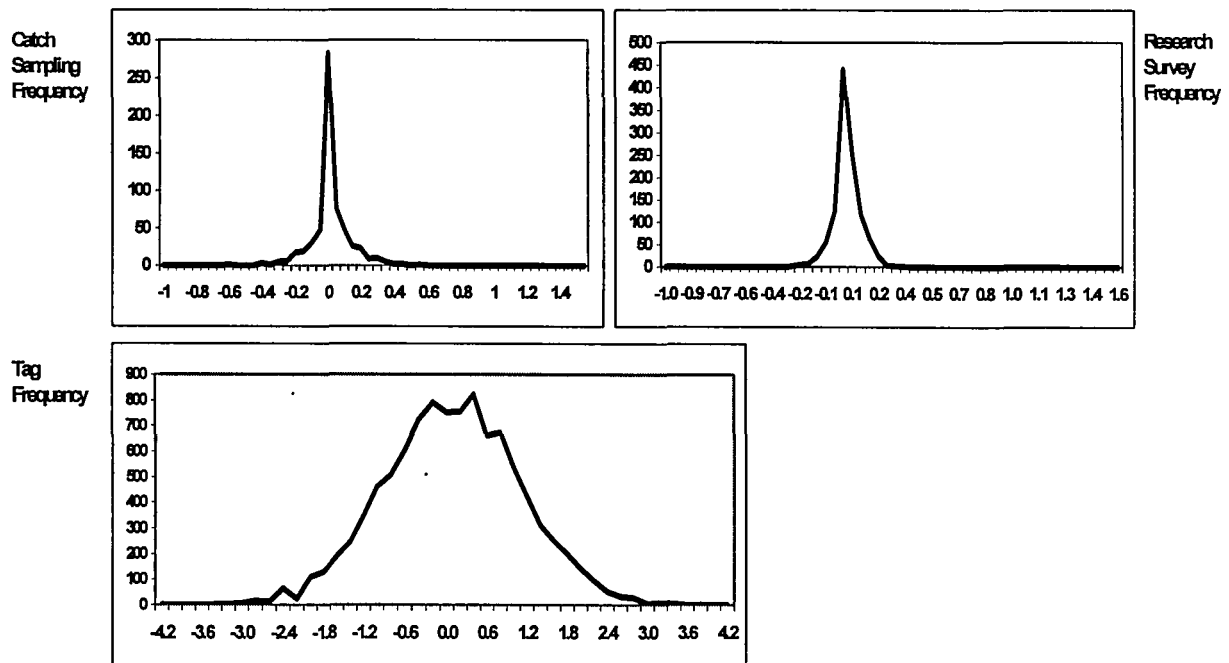
Figure 8: Frequency distributions of standardised residuals from proportion-at-length from commercial catch sampling (upper left) and research diver surveys (upper right), and from growth increment data (lower left) from case 1.
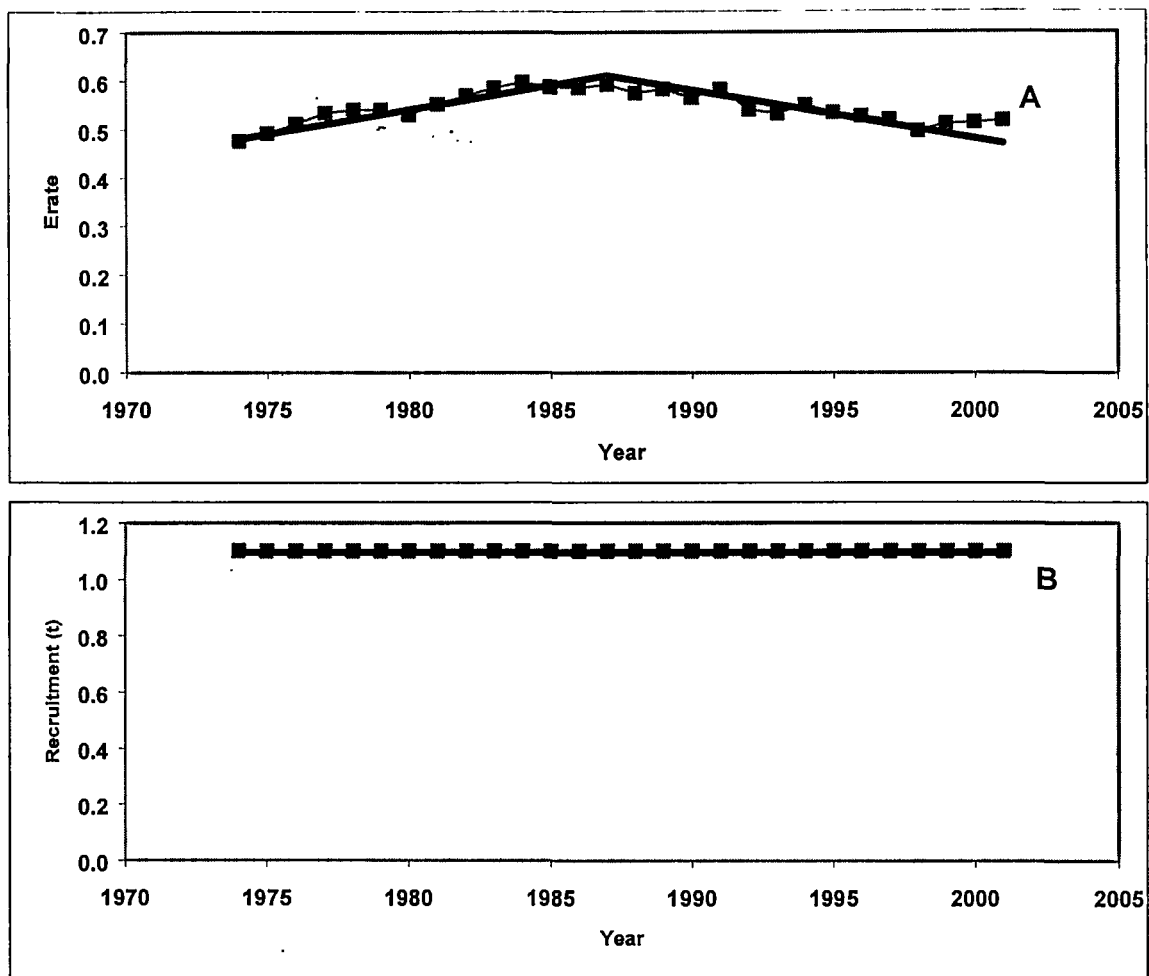
Figure 9. Exploitation rate (panel A) and recruitment (panel B) from the simulator (thick line) and the estimation model (thin line with squares) from case 1.

Figure 10. Scatter plot of biomass and biomass ratios from the estimator against the values from the simulator of Case 2.

Figure 11: Results from the fit for case 2. A: Observed (squares) and predicted (solid line) CPUE; B: standardised CPUE residuals; C: observed (squares) and predicted (solid line) research diver survey index; D: standardised research diver survey index residuals;E: observed (small squares) and predicted (larger squares) length increments; F: standardised growth increment residuals.

Figure 12: Observed (bars) and predicted (lines) proportions-at-length from commercial catch sampling (left) and research diver surveys (right) from case 2. The number under the year is the effective sample size.

Figure 13: Standardised residuals from proportions-at-length (top) and means of all proportion-at-length residuals (bottom) vs observed length from commercial catch sampling (left) and research diver surveys (right).
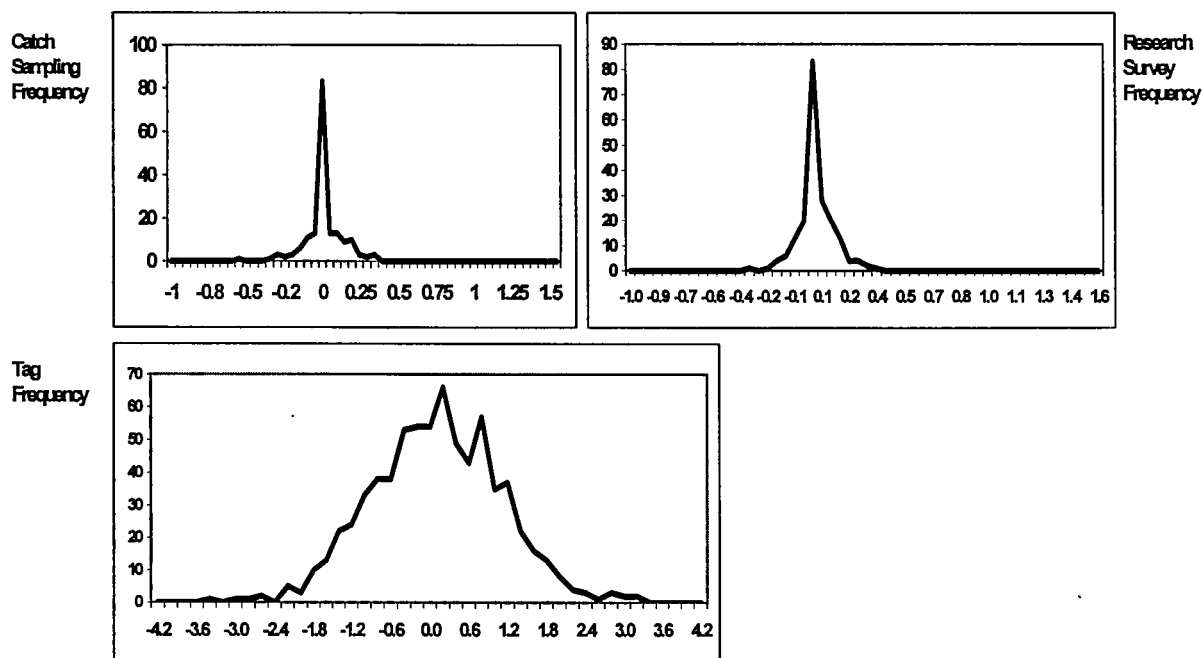

Figure 14: Frequency distributions of standardised residuals from proportion-at-length from commercial catch sampling (upper left) and research diver surveys (upper right), and from growth increment data (lower left).
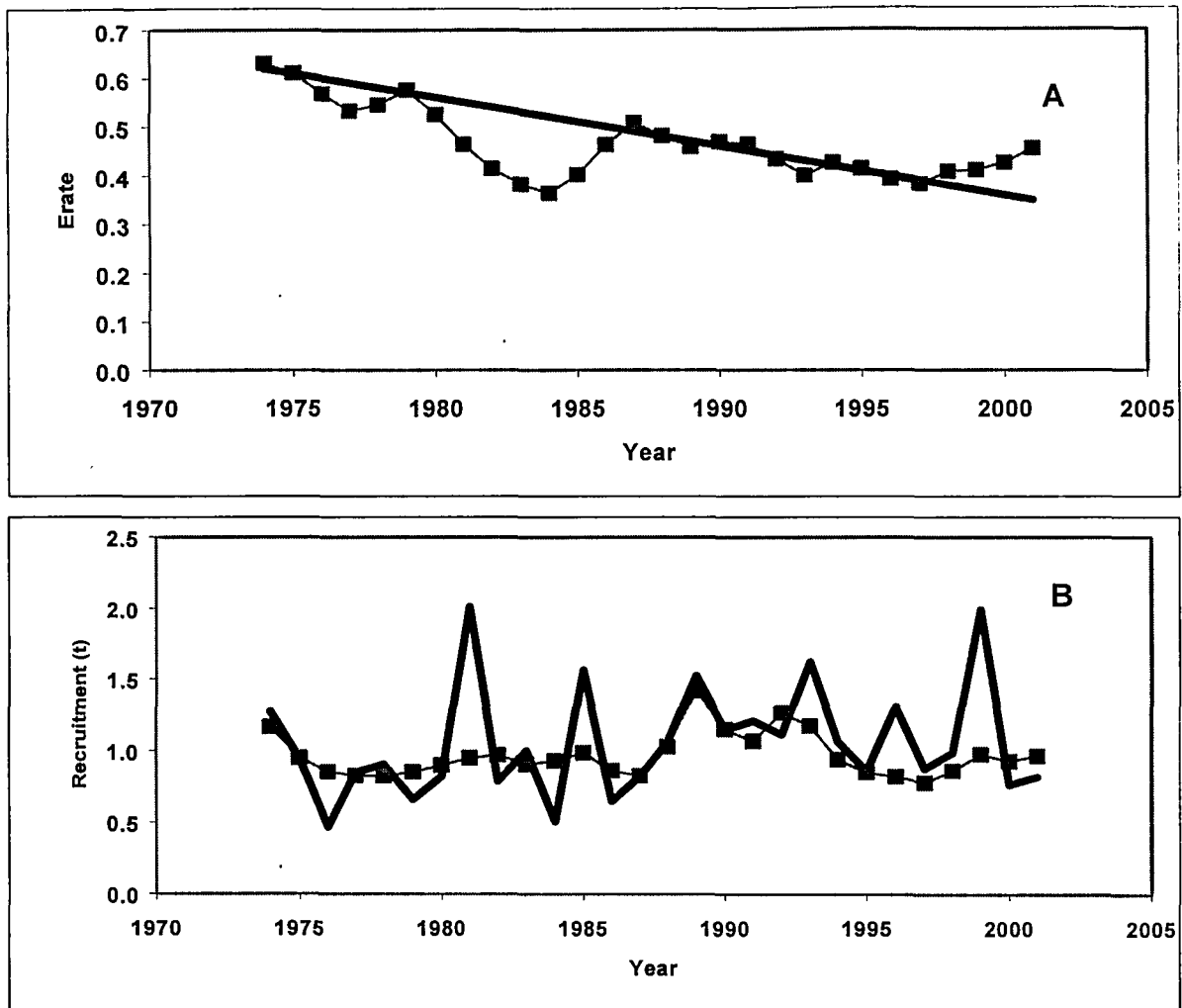
Figure 15. Exploitation rate (panel A) and recruitment (panel B) from the simulator (thick line) and the estimation model (thin line with squares) from case 2.
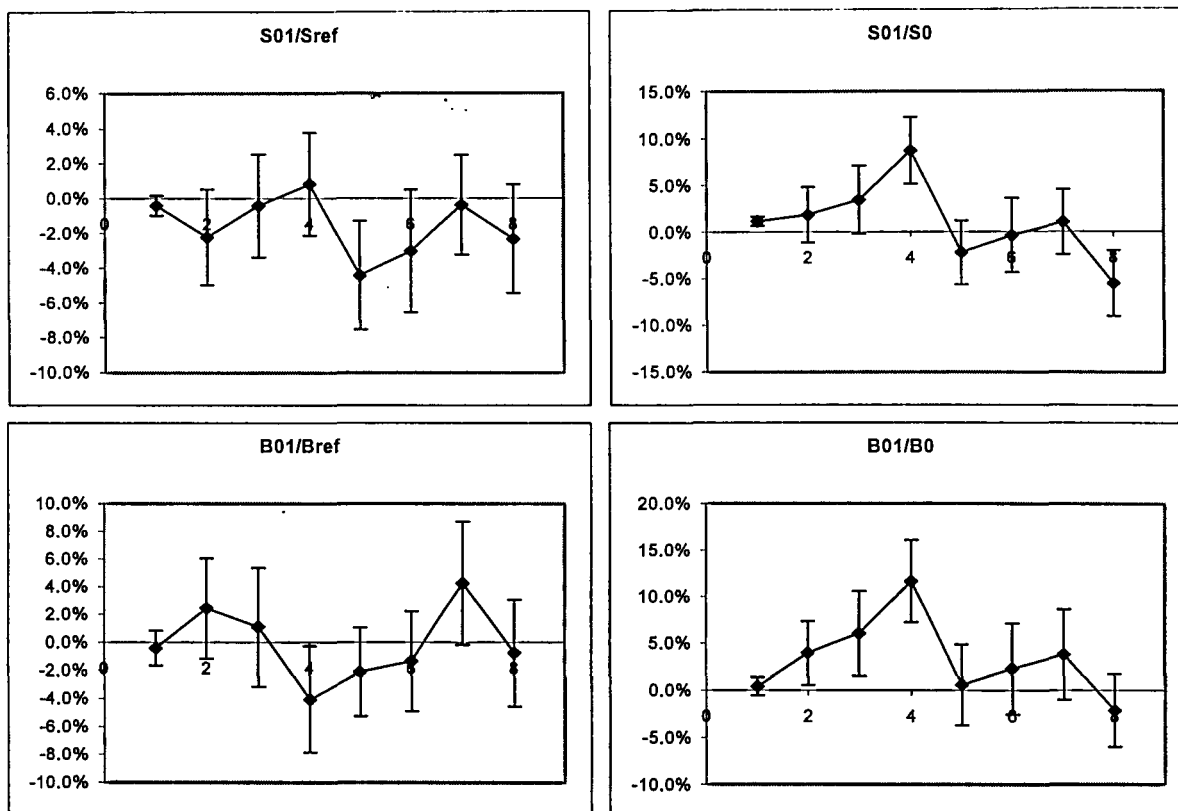
Figure 16. Percentage of bias (PB) and 95% confidence error bars of four biomass ratios for each of cases 1 to 8.
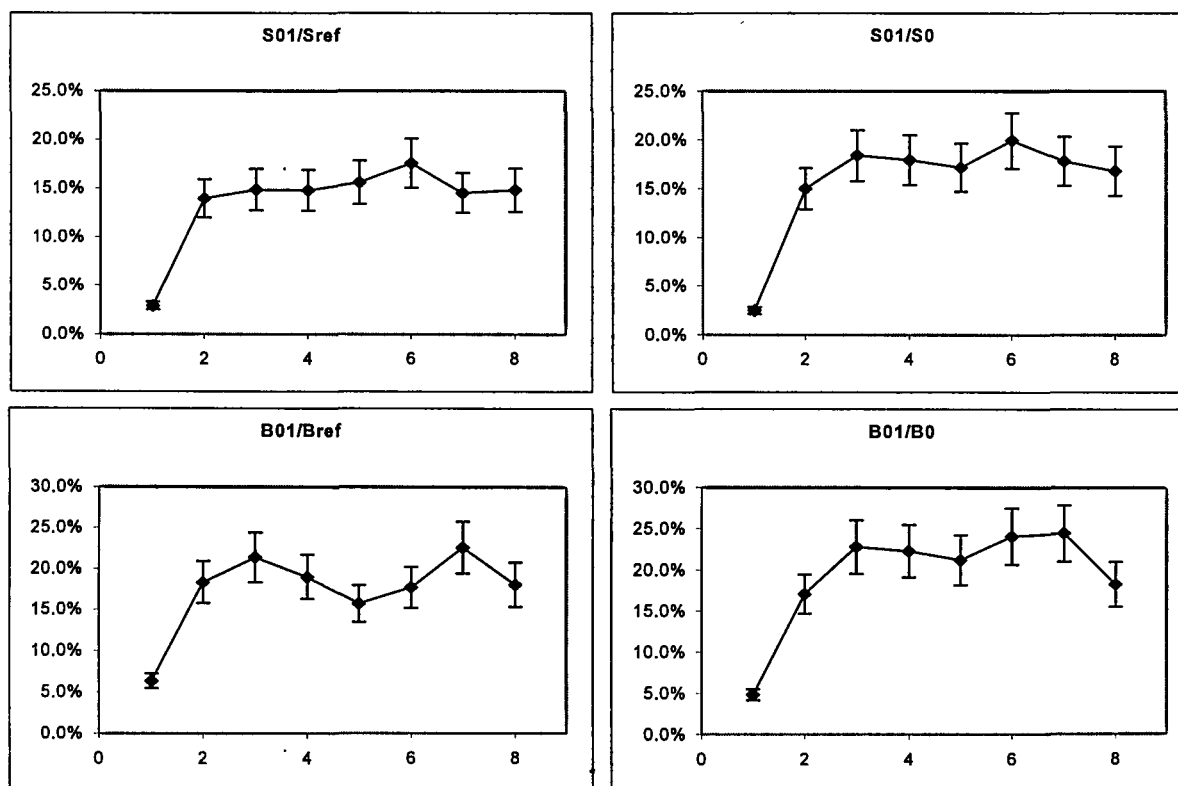

Figure 17. Comparison index (CI2) and 95% confidence error bars of four biomass ratios for each of cases 1 to 8.