CPUE Webmeeting 2012 paper #2

Using General Linear Models of SBT CPUE-at-age data to investigate changes in catchability with age and time.

By John G. Pope and Alex. Cutbill
NRC(Europe)Ltd.
popejg@aol.com

Introduction.

CPUE at age data provide an alternative approach to analysing and possibly of verifying fishing effort and catch -at-age data. Shepherd and Nicholson 1991 showed that {Catch Numbers(year,age)} {C(a,y){could be modelled approximately with multiplicative age and yearclass effects.

That is as

$$Ln\{C\ (y,a)\}=Y(y)+YC(y-a)+A(a)+\varepsilon \qquad 1$$

Where Y, YC and A are year, yearclass and age factors.
They also noted that such interpretations are not unique because if a constant h is used to modify each of these factors by adding -h*y, h(y-a) and h(a) respectively then all cell estimates are unchanged since

$$Y(y)\ -h*y\ +YC(y-a)+\ h(y-a)\ +A(a)+\ h(a) = Y(y)+YC(y-a)+A(a)$$

For all y and a for any value of h. Hence the interpretation will give unknown trends in each of YC and A and an opposite trend in Y all with slope h in the respective indices.

The reason that equation 1 works is that for a system with separable fishing mortality rates F(y,a)=f(y)*q(a) catch(a,y) may be written as

$$C(a,y)=f(y)*q(a)*Yearclass(y-a)*exp(-cum(Z)) \qquad 2$$

Where Yearclass is the population at the time fish recruit and cum(Z) is the cumulative total mortality from recruitment to the time the average catch is taken in year y at age a.

Thus

$$Ln(C(a,y))= ln(f(y))+ln(Yearclass(y-a)+ln(q(a))-cum(Z)\ +\eta \qquad 3$$

The first three terms are clearly factors of y, yc and a while –cum(Z) is predominantly an age effect. In fact where F(a,y) is constant cum(Z) is entirely an factor of a but where F(a,y) changes systematically it also contains a small y effect and a very small unresolved y*a effect.

When fishing effort E(y) is proportional to f(y) then

$\mathrm{Ln}\{C(a,y)/E(y)\} = \ln(\mathrm{Yearclass}(y-a)+\ln(q'(a))-\mathrm{cum}(Z) +\eta$ 4

Or as a first approximation

$\mathrm{Ln}(\mathrm{CPUE}(y,a) = YC(y-a)+A(a)+\varepsilon$ 5

Where as a first guess ε might be thought to be normally distributed ($N(0,\sigma^2)$).

Such models are useful in providing a semi-realistic breakdown of CPUE data without the need to perform a full assessment.

Data and methods

The following results came from a set of Japanese LL CPUE(y,a,area) data that Dr Hillary and Professor Butterworth were trying to interpret and of which I received a copy. The data extend from 1991-2010 and ages 4-12 for areas 4-9. They thus cover components of the years-classes 1979-2006 although the extreme years are only covered by one cell.

The additional breakdown by area suggests that at a minimum equation 5 should be extended as

$\mathrm{Ln}(\mathrm{CPUE}(y,a) = YC(y-a)+A(a)*\mathrm{AREA}(\mathrm{area})+\varepsilon.$ 6

Where AREA is an area based factor. This is to assume that q(a) varies between areas, as we in fact know is the case. Equation 6 then forms our basic model. I showed some results from a more limited year range at the 2011 ESC in Bali but was limited by the size of the GLIM package I was then using. R (which I now use thanks to Jim and Trevor) allows the whole data set to be interpreted!

Initial analysis with R suggests the data from area 5 were different from those of the other areas and were excluded from further analysis.

The basic model (equation 6) was then extended to consider

1. possible overall changes in age effects post 2005 by a year factor.
2. possible differences in year-class strength (or more likely distribution) between areas
3. The combination of options 1 and 2
4. Year class, age effect and a post 2005 year effect all crossed with area. I.e. equation 1 applied to each area individually.

This last run might suffer from the h trend problem noted in the introduction so is included for completeness rather than as a viable interpretation.

Additional analyses proposed at the April CPUE Web Meeting

Results of these models were presented at the April CPUE web meeting and the following additional analyses were proposed by the participants.

1. Investigation of the retrospective patterns of recruitment to see if the high recruitment to the most recently observed (2005 and 2006) yearclassses were an artifact of the method.
2. Investigation of the error distribution with age and to check if residuals were consistent with the assumptions of the model.
3. To consider if the seeming increase in catchability seen since 2006 might alternatively be explained by a reduction in total mortality rate.
4. To attempt a reconciliation between the analyses presented here and the standard CPUE ANOVAs.

Item 1 of these additional tasks was addressed by forming retrospective analyses of the data by removing successive years from the analysis from 2010 to 2007.Retrospective patterns were constructed for the basic model (equation 6) and for the basic model extended to consider possible overall changes in age effects post 2005 by a year factor.

Item 2 was addressed by plotting mean sum of squares by age and by considering q q plots and other diagnostics.

Item 3 was addressed by fitting the basic model extended to consider possible overall changes in age effects post 2005 by including a “year step” variable with a value of 0 for the years up to 2005 and a value of (year-2005.5) for years 2006 onwards (i.e. the sequence, 0, 0.5, 1.5,2.5…..4.5). The rationale for this was to consider the effect of a step reduction in total mortality rate of β in 2006, i.e. one possible effect of post 2006 management changes. If post 2005 the total mortality was constant on all ages then equation 4 would indicate that the Ln(CPUE) would increase as β*(“year step” variable) (n.b. CumZ is composed of the full total mortality in preceding years and approximately 0.5 the total mortality in the final year. Hence in 2006, 2007 etc. the 0.5, 1.5, 2.5 represent the amount that the changed values Z would modify the year effect). The coefficient of the “year” variable is thus a simple minded estimator of the scale of a step reduction in Z post 2005. The question of whether catchability has increased is then a question of whether the reduction in Z implied by the estimate of β is plausible given the post 2005 reduction in the total allowable catch and also we would be happier if we saw similar reductions in all areas.

Item 4 is an interesting question which is pertinent to how we develop CPUE series in the future. However it is not possible to address it solely in the context of this CPUE at age data set. We have therefore also drawn on some results of the 2010 base run tuning to illuminate this question.

Results and Discussion
The models, their degrees of Freedom and AIC values are shown in Table 1.

Table 1

Text Name	R #	Model	df	AIC	Table 2 Name
Basic	1	**Basic** = Yearclass+age:area	73	1345	Model 1
Bullet 1	10e	Basic + year(post 2005)	78	1328	Model 2
Bullet 2	10c	Basic +YC*area	181	1255	Model 3
Bullet 3	10d	Basic +YC*area+year(post 2005)	186	1226	Model 4
Bullet 4	10b	Basic +{YC +year(post 2005)}*area	206	1154	Model 5

The progressively extended models (bullet 1-4) all show significant improvements on their precursors and improvements in AIC despite reductions in the residual DF (See Table 2).

Table 2

```
Model 1: CPUE ~ yc + a:area
Model 2: CPUE ~ yc + a:area + rr
Model 3: CPUE ~ yc:area + a:area
Model 4: CPUE ~ yc:area + a:area + rr
Model 5: CPUE ~ yc:area + a:area + area:rr
  Res.Df    RSS  Df Sum of Sq      F    Pr(>F)
1    801 201.88
2    796 195.81   5     6.068 6.7768 3.552e-06 ***
3    693 142.32 103    53.492 2.8999 3.397e-16 ***
4    688 136.05   5     6.266 6.9981 2.191e-06 ***
5    668 119.63  20    16.425 4.5858 2.386e-10 ***
```

The root mean squares of the residuals suggests coefficients of variation of the CPUE(y,a) data of between 50% with the basic model (model 1)and about 40% with the bullet 4 model (model 5). These seem rather higher than for stocks with which I am more familiar (these have CV's of perhaps about 20-30%) so there might be more sum of squares to explain with an appropriate model ~ but it might be that the SBT is less well sampled or more variable!

Figure 1 show the age effect (ln(selection)-Cum(Z)) obtained for areas 4, and 6-9 with the model 4 (bullet 3). The shape differs between areas. In particular areas 4, 6 and 7 show declines after age 5 while area 8 declines from age 4 but has a secondary peak at age 8. Area 9 declines from age 4. If catchability were constant after these ages the declines would suggest annual increments in Z of about 0.33. However, increasing cum(z) might be confounded with increases in selection. Area 4 shows a somewhat steeper decline than in other areas possibly suggesting that it is an area from which older fish migrate. Differences in catchability by area could result in changes in overall stock catchability when or if fishing effort changes its areal distribution.

Figure 2 shows the ratio of catchability on all ages in the most recent 5 years compared to the joint estimate for the period 1991-2005. The results suggest that catchability was somewhat higher in 2006 and in 2008 than in the 1991-2005 period. However, in 2007 it was 20% lower while in 2009 and 2009 it is about 35% higher than in the 1991-2005 period. Hence could it be that catchability has increased in the most recent years compared to the 1991-2005 period. This might result from changes in the management regime or it might also be that catchability was low in the earlier period, for example if catch but not effort were affected by the market anomaly problems.

Figure 3 shows the relative year-class sizes 1979-2006 obtained with model 2 (bullet 1) (i.e. the exp(yearclass factors).

Including Yearclass*area interactions improves fits significantly (see the bullets 2, 3 and 4 model results in tables 1 and 2. Figure 4 shows the relative year-class sizes 1979-2006) for each area obtained from model 4 (bullet 3). Broadly similar trends in year-class sizes 1979-2006 are seen to those of figure 3. However, the separate areas do show some different trends particularly in the first and last years. It is important to recall that the earliest and most recent year-class estimates are based on only one cpue(y,a) estimate and thus will be very variable.

Since as far as we are aware the stock has a single spawning population if there really are differences in year-class strength between areas this would suggest differential initial distribution by area which tend to persist over the life cycle of cohorts. I.e our model of a freely mixing stock might not be entirely true.

These results (tables 1, 2 and figures 1-4) were shown at the April Web Meeting. Suggestions were made at the meeting about how this model might be extended to consider questions relevant to our understanding of the SBT catch and effort data sets.

These questions are discussed below..

<u>Web meeting questions</u>

1) Investigation of the retrospective patterns of recruitment to see if the high recruitment to the most recently observed (2005 and 2006) yearclassses were artefacts of the method.

The year-class strengths were calculated retrospectively using data terminated at 2006, 2007, … ,2010 under bullet 1 (i.e. model 2), and the results are shown in Figure 5. Hence the 2010 line on Figure 5 is the line shown in Figure 3. Figure 6 shows the retrospective year-class strengths estimated under the basic model. These retrospective plots do not suggest any tendency for other “last years” to be overestimated. Indeed the year-class strengths are in general lower in the earlier years models than in the 2010 model which uses all of the data available. Therefore it is believed that the high 2005 year-class strength is not a consequence of systematic error. In the basic model (Figure 6, '2010'), the 2005 year-class strength is 3.5 times that of the average from 1991 to 2004. Under model 2 (Figure 5) fitting year factors for years subsequent to 2005 gives a lower 2005-class strength (1.76 compared to 2.53), still by far the largest but only 2.6 times the average year-class strength seen between 1991 and 2004. This suggests that the increase in strength could result either from increased catchability or an increase in stock size.

2).Investigations of residuals.
Using a bubble plot of the sum of squares of residuals by age and area for model 2 (Figure 7) it is clear that a greater than expected proportion of the residual sum of squares comes from area 6, and for age 4. This is confirmed by the pie charts of squared residuals by age and by area (Figure 8)

In the light of these results the retrospective runs were repeated for the basic model both with area 6 omitted and with it included. A further run was made for area 6 alone. Results showing the ratio of year-class strengths are shown in figure 9. It is clear that the large 2005 year class effect is amplified in area 6 though increases also occur in other areas

It is worth noting that the catch was on average much lower in area 6. (149 for area 6 compared to 1915 as an average for the other areas) so it is not surprising that it is more variable and a more thorough analysis might weight by sample sizes.

3).To consider if the seeming increase in catchability seen since 2006 might alternatively be explained by a reduction in total mortality rate.

Results of fitting the "year-step" variable suggests a step reduction in total mortality rate of 0.074 in 2006. This made a small reduction but a significant reduction in residual sum of squares from 202 to 200 and a small reduction in AIC from 1345 to 1337. The year effect seemed to be partly confounded with the age:area effect but although small was significant (1DF, P<.002). The slope of the "year step" variable at .074 seems a reasonably plausible figure for a reduction in total mortality rate. Crossing this variable with areas suggests that mortality decreased in all areas except area 4.
The area slopes were (area-4=-.081, area-6=.17, area-7=.19 and area-8=.06. Thus apparent reductions in mortality rate seem particularly strong in areas 6 and area 7. The added reduction in residual sum of squares from 200 with the single year-step" variable to 185 with this variable crossed with area was greater than for including the variable uncrossed, as was the reduction in AIC (from 1337 to 1281). The reduction in residual sum of squares was significant (DF 4, P<5E-12). Thus the area variations in the change seem much stronger than the main "year-step" effect. Yet again it seems area year effects are prominent in the changing CPUE of SBT even when age differences are taken away as in these analyses. The apparent increase in catchability might be interpreted as being due to a reduction in total mortality rate presumable as a result of quota reductions after 2005 and more stringent management. However, these effects are rather uneven across areas!

4. To attempt a reconciliation between the analyses presented here and the standard CPUE ANOVAs.
The present analyses do not allow a very close comparison with the standard CPUE model. However, they do perhaps provide a few guidelines. Firstly strong regional changes in catchability seem to occur and these cannot be explained purely in terms of age structure changes. Thus the inclusion of area:year effects (and perhaps non-targeting effects) in the main model seem justified.

This age based data set does not allow of consideration of area:month effects. However a study of results from trial fittings of the base model (circulated by Dr Itou in 2010 when we were chosing models) suggests that there are regular patterns of change in CPUE with season. Figure 10 shows these results standardized to the maximum CPUE of each area. These seem consistent with a migration away from area 7as the year progresses. Area 7 peaks in April, while areas 8 peaks in June and area 9 in August. In the opposite direction area 4 peaks in July and combined areas 5and 6 in September. It is tempting to think that these might be modelled by annual or twice annual cycles which might also be used to interpret future data in months not fished in the past. By contrast the pattern of the latitudinal:year factors (see Figure 11) are less easy to interpret. Perhaps they relate to changing hydrographic conditions?

Perhaps the point to consider under this heading is whether we might be able to base a future CPUE standardization on a realistic model of how SBT (and perhaps how long line fishers) behave. Such a model might well be informed by the ongoing results of the archival tagging project (Basson *et al*. 2011).

Reference

Shepherd, J.G. and Nicholson, M.D., 1991.
Multiplicative modelling of catch-at-age data, and its application to catch forecasts
J. Cons. int. Explor. Mer (1991) 47(3): 284-294.doi: 10.1093/icesjms/47.3.284

CCSBT ESC Paper 20. (Australia) Update on the global spatial dynamics archival tagging project – 2011
(Basson, Eveson, Hobday, Patterson, Lansdell)

Figure 1 Age effects of model 4 by area

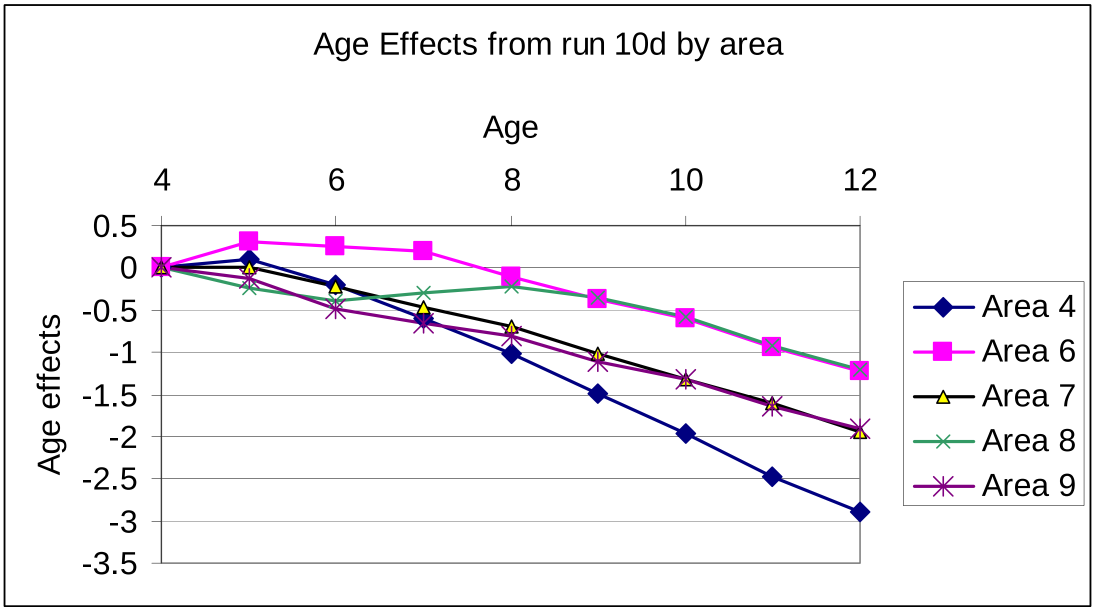


Figure 2

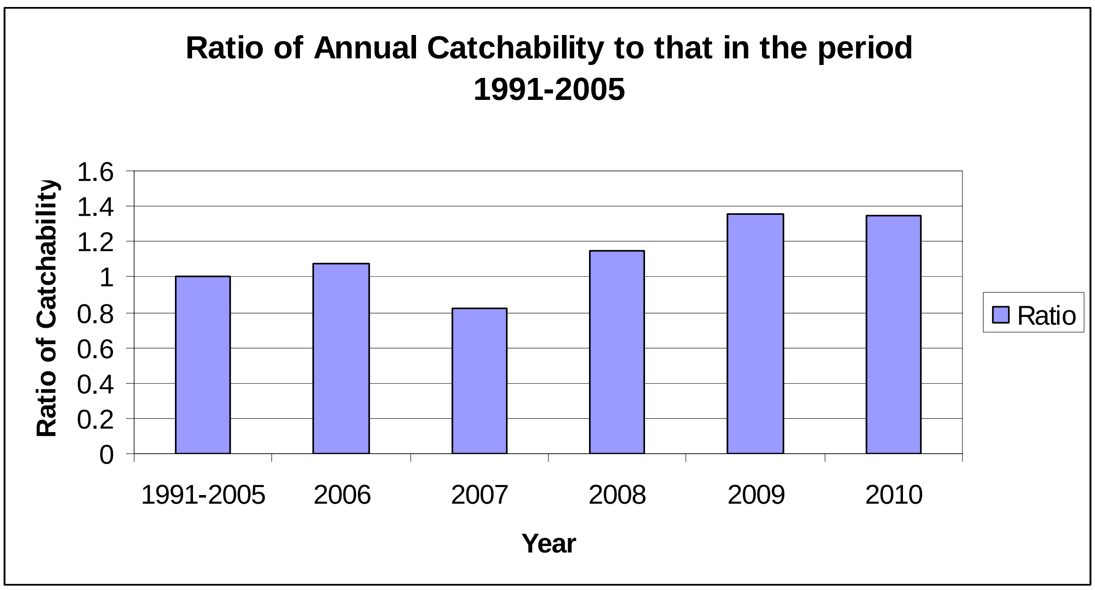


Figure 3

Yearclass multipliers from model
2 (bullet 1) fitted with
Overall yearclass effects
and overall year effects

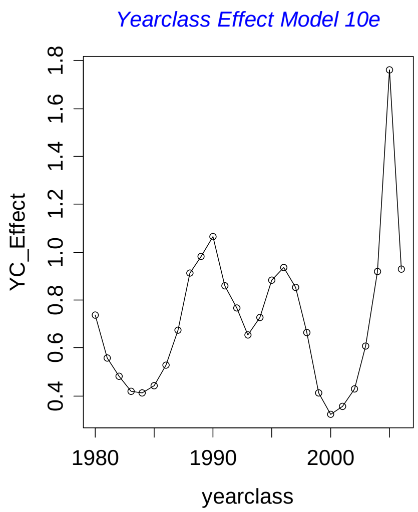


Figure 4

Yearclass multipliers from model
4 (bullet 3) fitted with
area based yearclass effects and
and overall year effects

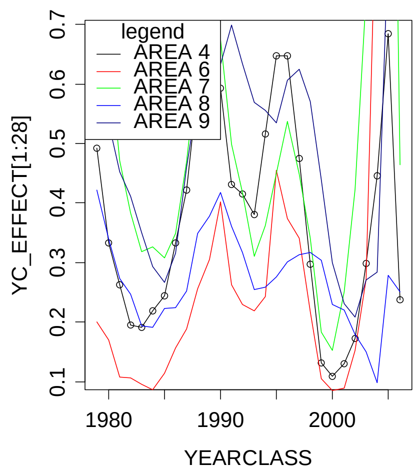


Figure 5 Retrospective Year-class effects as shown by model 2 (Bullet 1)

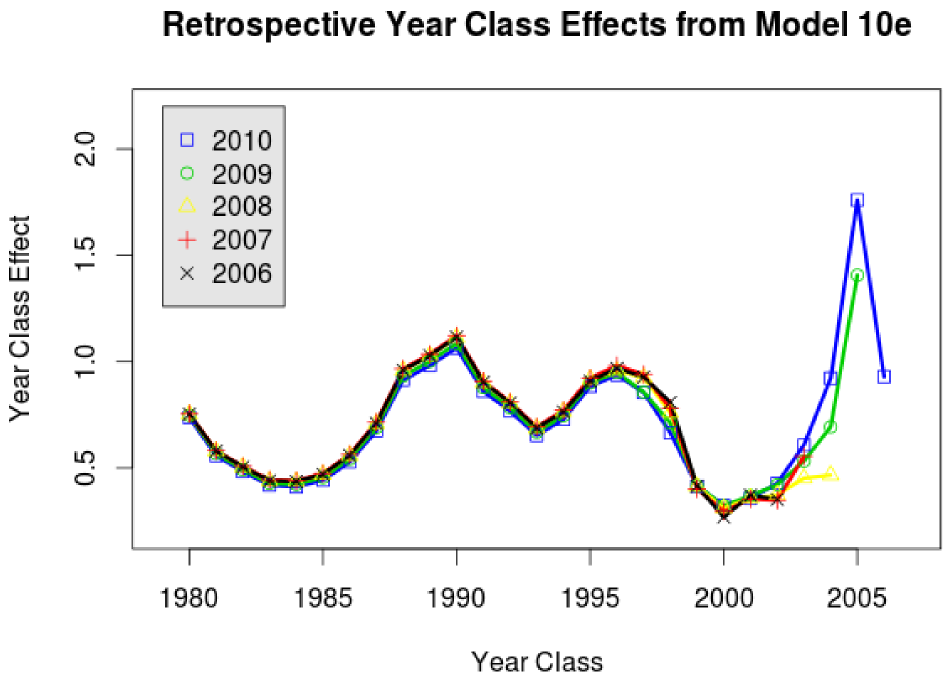


Figure 6 Retrospective Year-class effects as shown by model 1 (Basic model)

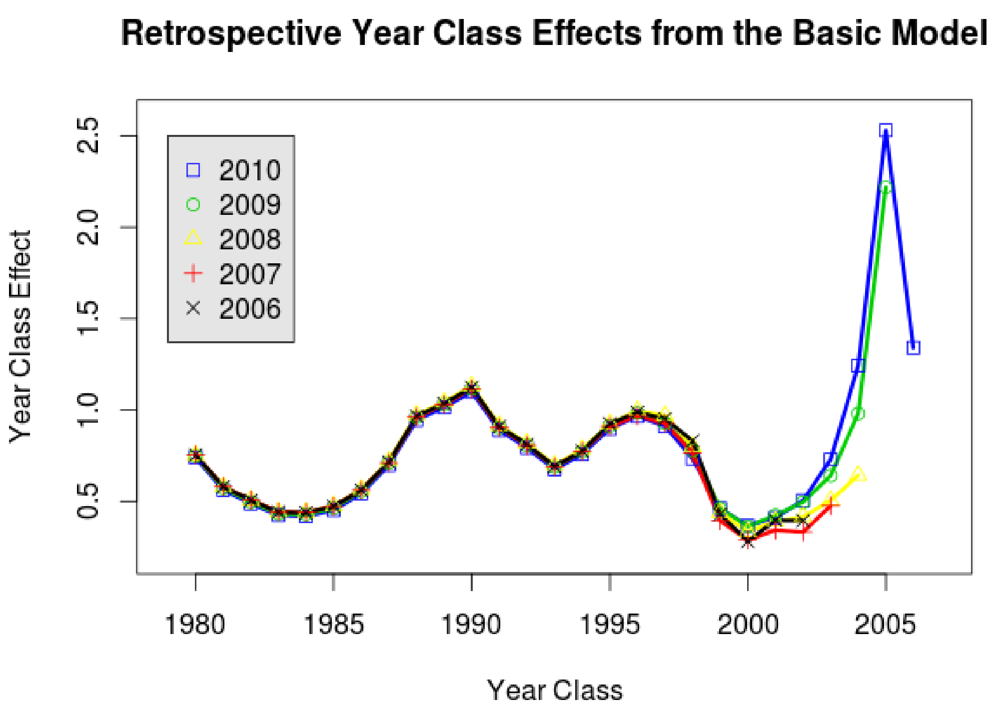


Figure 7

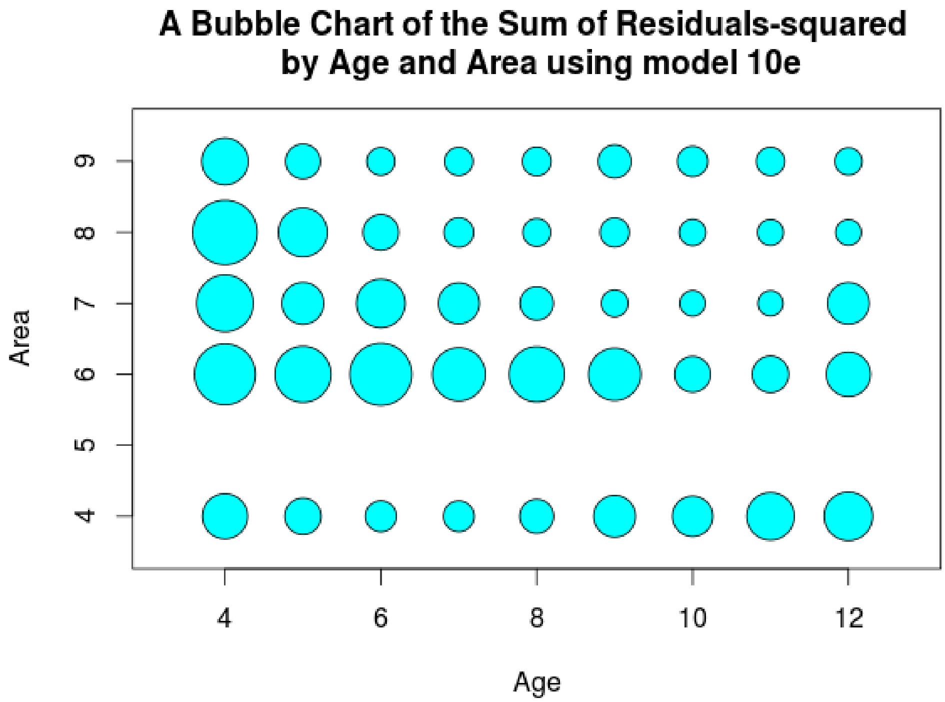
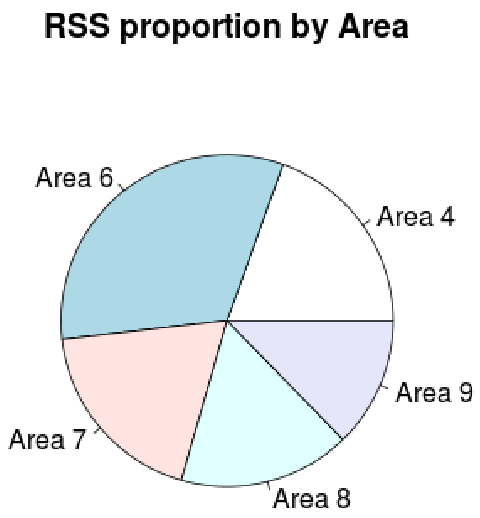

A Bubble Chart of the Sum of Residuals-squared
by Age and Area using model 10e
Area
9
8
7
6
5
4
4
6
8
10
12
Age

Figure 8 Residual Sums of Squares by area and by age.



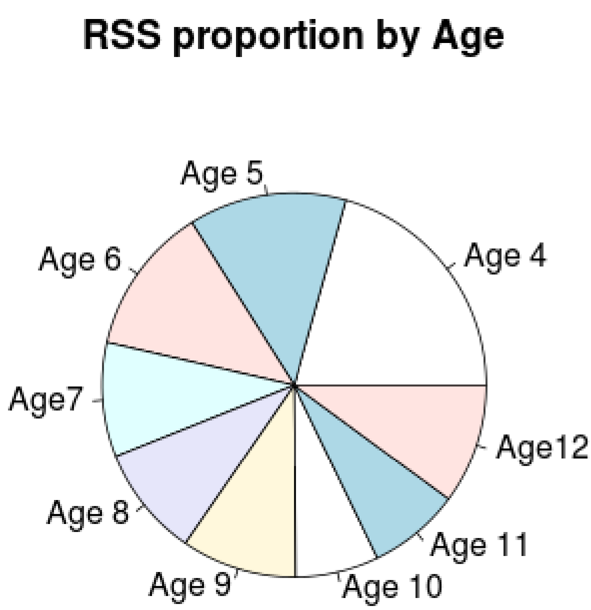


Figure 9 Investigation of how area 6 influences the estimate of the 2005 yearclass.

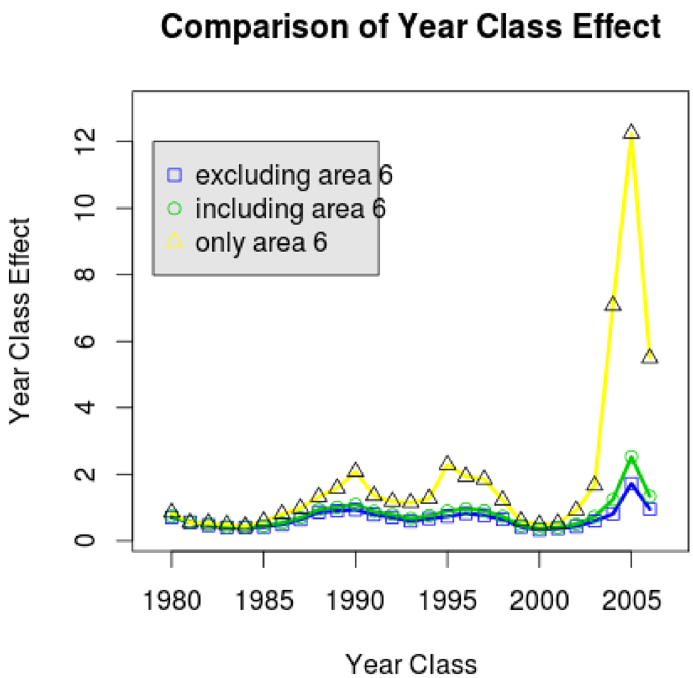


Figure 10 Relative CPUE by area and month from the 2010 base model

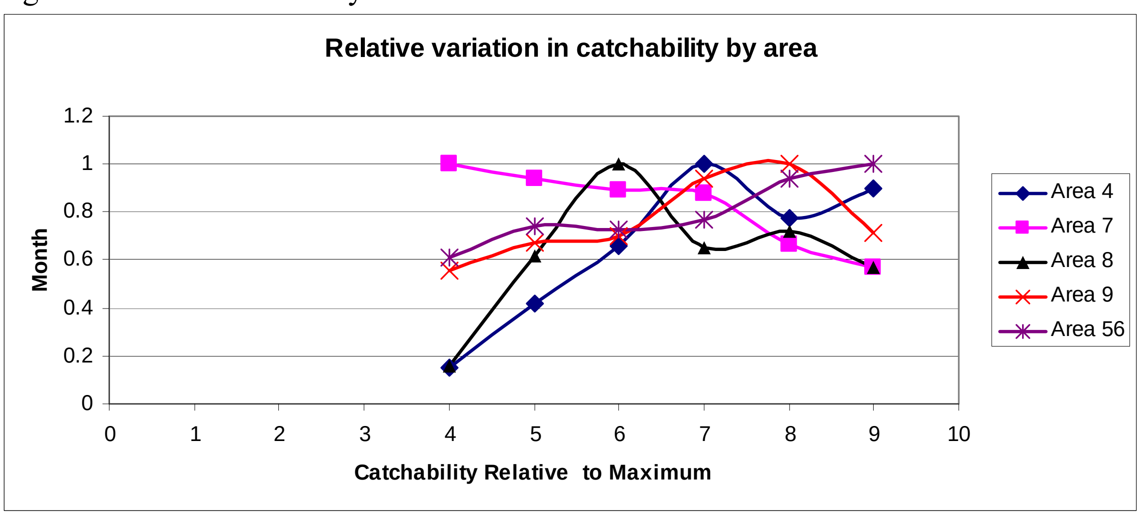


Figure 11 Relative CPUE by year and latitude as estimated by the 2010 base model