ICES CIEM
International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

ICES Advisory Committee on Fishery Management
ICES CM 2004/ACFM:14, Ref. AFWG, WGNPBW, WGMHSA

Report of the
Study Group on Assessment Methods Applicable to Assessment of Norwegian Spring-Spawning Herring and Blue Whiting Stocks (SGAMHBW)

19-22 February, 2004
IPIMAR, Lisboa, Portugal

International Council for the Exploration of the Sea

Conseil International pour l'Exploration de la Mer

H. C. Andersens Boulevard 44–46 · DK-1553 Copenhagen V · Denmark
Telephone + 45 33 38 67 00 · Telefax +45 33 93 42 15
www.ices.dk · info@ices.dk

Contents

1 INTRODUCTION

1.1 Participants

Azevedo	Manuela	Portugal (part time)
Belikov	Sergei	Russia
Björnsson	Höskuldur	Iceland
De Oliveira FM	Jose	UK (England)
Efimov	Yuir	Russia
Gudmundsdottir	Asta	Iceland
Heino	Mikko	Norway
Holst	Jens Christian	Norway
Jacobsen	Jan Arge	Faroe Islands
Murawski	Steve (chair)	USA
Silva	Alexandra	Portugal
Simmonds	John	UK (Scotland)
Skagen	Dankert	Norway
Tjelmeland	Sigurd	Norway
Toresen	Reidar	Norway
van Beek	Frans	Netherlands
Vasilyev	Dimitri	Russia

1.2 Terms of Reference

The **Study Group on Assessment Methods Applicable to Assessment of Norwegian Spring-Spawning Herring and Blue Whiting Stocks** [SGAMHBW] will be established (chair: S. Murawski, USA) and will meet in Lisbon, Portugal, from 19-22 February, 2004 to:

a) analyse and evaluate the assessment methods that are considered in assessing Norwegian spring-spawning herring and blue whiting;
b) identify for each method the types of population dynamics and data availability for which the method is applicable and relate this to the dynamics observed for the Norwegian spring-spawning herring and blue whiting;
c) devise one method that includes the strong points of all the proposed methods.

1.3 Scientific Justification for this Meeting

In 2001 and in 2002 the Northern Pelagic and Blue Whiting Fisheries Working Group (WGNPBW) reviewed and attempted to apply a number of methods for the assessment of Norwegian spring spawning herring and for blue whiting. The different methods make different assumptions of the error structure in the observed data and apply different subsets of the available data. The estimates produced by these different methods are in some cases widely different. This Study Group was established to clarify the conditions under which each method is applicable and to relate these conditions to the population dynamics and data available for Norwegian spring-spawning herring and blue whiting.

It appears that each method has certain strong points and focuses on particular features of the data. It appears to the WGNPBW and to ACFM that it should be possible to devise a method that would pick up the strong points of each method and construct a combined method that would be preferable to each of the existing methods. The Study Group is asked to consider this possibility.

The Study Group was also asked to report to the Working Group on Mackerel, Horse Mackerel, Sardine and Anchovy (WGMHSA) as these assessment method analyses may prove useful also in connection with assessment of sardine.

1.4 Background and Overview of Study Group Report

1.4.1 Background

The Northern Pelagic and Blue Whiting Fisheries Working Group (WGNPBW; ICES 2003a) is responsible for producing stock assessments and related advice on a number of important fishery resources, including Norwegian Spring-Spawning herring (NSSH), and blue whiting (BW). In aggregate, these two stocks produced nearly 2.4 million tonnes of landings in 2002, with a combined estimated spawning stock size in excess of 10 million tonnes (ICES 2003a). Assessments of these resources are complicated by the vast areas of ocean habitat in which these stocks range, the competing set of coastal state and international fisheries, and complex life cycles and migrations undertaken by the species. Because of the complicated migrations and life cycles of the two stocks, and the large spatial scales they occupy, providing precise and unequivocal indices of stock abundance for all the relevant age groups (pre-recruits and age groups contributing to the bulk of catches) has proved difficult. As a result, the Working Group has used a large number of abundance indices from both surveys directed to the species of interest, and catches of the species taken in surveys for which NSS herring and blue whiting were not the primary target species. These indices include trawling surveys, acoustic surveys, and larval sampling. Some of the survey time series are recent and rather short, while some others occurred for some time and were discontinued. Additionally, for NSS herring, there has been an ongoing set of tagging studies that provide some information in the recapture rate of the stock by age group and year.

The WGNPBW has, in the past, provided analytical stock assessments using two models each for NSS herring and blue whiting (ICES 2003a). For NSS herring, the models used are ISVPA (the catch-controlled version) (Kizner and Vasilyev 1997; Vasilyev 2003) and SeaStar (Tjelmeland 2003). ISVPA is a multipurpose modelling tool for providing assessments, assuming a separable pattern of fishing mortality at age and by year (Pope and Shepherd 1982), and is programmed in FORTRAN 77. SeaStar is a tool specifically designed to model the NSS herring resource. It is implemented in the *Mathematica* environment, and is a standard backward-calculating VPA utilizing estimated fishing mortality rates for the terminal year. The most recent assessments of NSS herring using these two approaches (Figure 1.4.1) show broad concurrence in the time series of estimated fishing mortality on fully-recruited age groups, and in the estimated spawning stock biomasses. However, the two models produced divergent perceptions of recent trends and quantities of recruitment, which, when taken into medium-term projections, produce markedly different results. Also, while the total SSBs estimated by the two models are similar, the relative strengths of the year classes contributing to spawning are different (ICES 2003a).

For blue whiting, the two assessment models that have been applied by the WGNPBW are ISVPA (the effort-controlled version) and AMCI (Skagen 2003). AMCI is a forward projection model that utilizes a separable (age group, year) model for fishing mortality rate, but allows the selection-at-age to change slowly through the use of a "gain factor" (Skagen 2003). When run in 2003, the two assessment models produced results that showed rough concurrence in time trends of fishing mortality, SSB and recruitment, but diverging trends in recent years and, resultantly, different medium-term catch forecasts (Figure 1.4.2; ICES 2003a).

A number of efforts have been made in the past to understand the features of the models and supporting data that contribute to the differences seen in the estimates of fishing mortality, abundance and biomasses for the two stocks. The WGNPBW has investigated the properties of the models since 2001. In its 2003 report, the WGNPBW concluded for NSSH herring:

"The Working Group concluded that both models are relevant and applicable to assess the state of the NSS herring. The main difference of the models is the estimation of the year classes entering into the spawning stock. This is a period of dynamic changes, the herring migrates from the nursery area in the Barents Sea to the Norwegian Sea, ….The amount of herring migrating to the northern Norwegian Sea varies from year to year according to year-class strength. This year-to-year variation may be important to take into account when discussing the separability issue. SeaStar lacks information from this phase, while ISVPA relies on a constant selection pattern".

Similarly, sets of model runs conducted in the 2003 meeting of WGNPBW for blue whiting used differing assumptions in ISVPA and AMCI. These test runs highlighted the sensitivity of model results to a variety of assumptions regarding data series used, and weighting of parameters in the models.

At its 2003 meeting, the Working Group on Methods of Fish Stock Assessment (ICES 2003b) undertook a number of tests of AMCI and ISVPA (as well as other assessment techniques including XSA, CADAPT, and ICA). This testing involved both simulated data with noise and the actual data sets used by the WGNPBW in its 2002 assessment of blue whiting (ICES 2002). The SeaStar model was not applied to the simulated data at that time; those comparisons are now included in this report (see section 4.5).

Based on the application of ISVPA and AMCI to "noisy" simulated data, the Methods Working Group found that both AMCI and ISVPA performed reasonably well (as compared with these and other assessment methods) in recovering the "true" spawning stock biomasses, fishing mortality rates and recruitment series. Thus, there is nothing inherent in these two models themselves that would, *a priori*, favour the selection of one method over the other for conducting assessment calculations. Rather, it is the specifics of how individual data sets are fitted and weighted, and conditions such as changes in selectivity-at-age over time that result in divergent views of the stocks. This was confirmed in tests performed by the Methods Group when applying a variety of methods to data for blue whiting.

With respect to the comparison of results from various techniques as applied to the assessment of blue whiting, the Methods Working Group (ICES 2003b, p.85-86) provided three main findings:

that:

1) "Conflicting sources of information appear to present the main problem in the blue whiting assessment. The conflict in the data sources is handled differently by the different methods that have been applied to this stock (e.g., AMCI and ISVPA)"
2) "There are indications of changes in selection of the most recent (strong) year classes which appear to have a higher exploitation on the younger ages compared to older ages. Although this may be a relative change only, it could seriously affect models that assumed fixed selection patterns over a longer period of time."
3) "The constraint of zero row- and column sums of the residual matrix in ISVPA seems to be a contributing factor to the difference between ISVPA and other separable models. Further work is necessary in order to fully understand the causes and implications of these constraints"

These previous findings are the point of departure for the current Study Group. A brief overview of its work, findings and conclusions is given below.

1.4.2 Overview

The main new work undertaken at this Study Group meeting included: (1) an objective examination of the signal-noise characteristics of the tuning data used in the models, with an objective of recommending which data series to use in future assessments, (2) testing of the three assessment approaches (ISVPA, AMCI, SeaStar) using simulated data with trends in the dynamics of the stock, sampling variation, and changes in stock characteristics, and (3) calculation of the influence of various data input series on management parameters of interest coming from the models (plenty of work for a 4-day meeting!). Additionally, the Study Group considered how the Northern Pelagic and Blue Whiting Group might proceed, in light of analyses produced by the Methods Working Group and this Study Group, in updating assessments for these two stocks.

Given the Methods Working Group conclusion that conflicting data sources were a major issue in these assessments (especially for blue whiting), the Study Group undertook a significant effort to evaluate the quality and consistency of data inputs into the assessments for both Norwegian spring-spawning herring and blue whiting (section 2 of the report). We reasoned that if data provided little signal with regards to the abundance of a particular age group, then little information was likely being added to inform the assessment models. The data quality analyses provided in the report included, for each survey series, correlations between adjacent age groups within-series, as well as correlations among particular ages between surveys. These correlations utilize linear-scale data and are combined with corresponding scatter plots. Based on these correlations and corresponding scattergrams, the Study Group recommended a subset of ages within each survey series that likely have sufficient tracking ability for cohorts to be useful in assessment modelling (e.g., where correlation coefficients are at least 0.9 and correlations were based on sufficiently long series with adequate contrast and minimum numbers of zero values). In some cases there is an indication that combinations of survey data may be warranted. The Study Group recommends to the Working Group that these revised data ranges be considered in updated assessments.

Section 0 of the Study Group report updates and expands model descriptions using a standard format proposed by the Methods Working Group. Several of the models have been updated to allow additional functionality and for bug fixes, and these changes are documented.

Sensitivity analyses were previously undertaken by the Working Group examining various properties of ISVPA, AMCI and SeaStar. Additionally, the Methods Working Group (ICES 2003b) conducted simulation testing of ISVPA and AMCI. The Study Group undertook extensive simulation testing using three standard data sets, each having specific properties intended to exercise the models relative to properties thought to be handled differently by them. Section 4 of our report describes the simulated data sets as well as application of the three assessment models to the data. These analyses revealed that all three methods can recover the essential details of the stock dynamics, even with complicated and time-varying selection patterns, changes in fishing mortality rates and exploitation patterns. Differences between models are due to the details on how conflicts in data are handled, and, to a certain extent, the subjective decisions of the analysts.

The study group outlines a number of conclusions and recommendations for handling assessment data and for model development.

1.5 Acknowledgements

The Study Group expresses its appreciation for the hospitality of the directorate and staff of the Instituto de Investigação das Pescas e do Mar (IPIMAR), in Lisboa, Portugal, for hosting the meeting and providing excellent services and facilities.

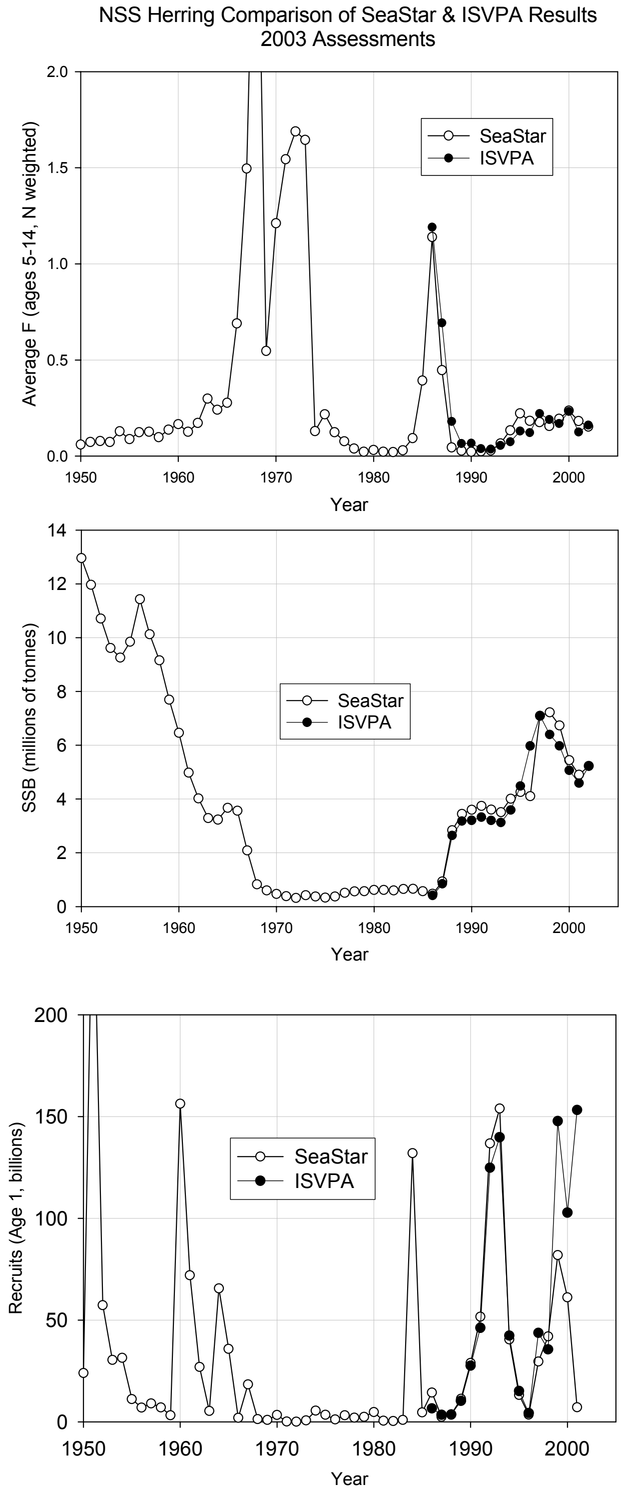


Figure 1.4.1. Comparison of stock assessment results for NSS herring from 2003, for two assessment methods (SeaStar and ISVPA).

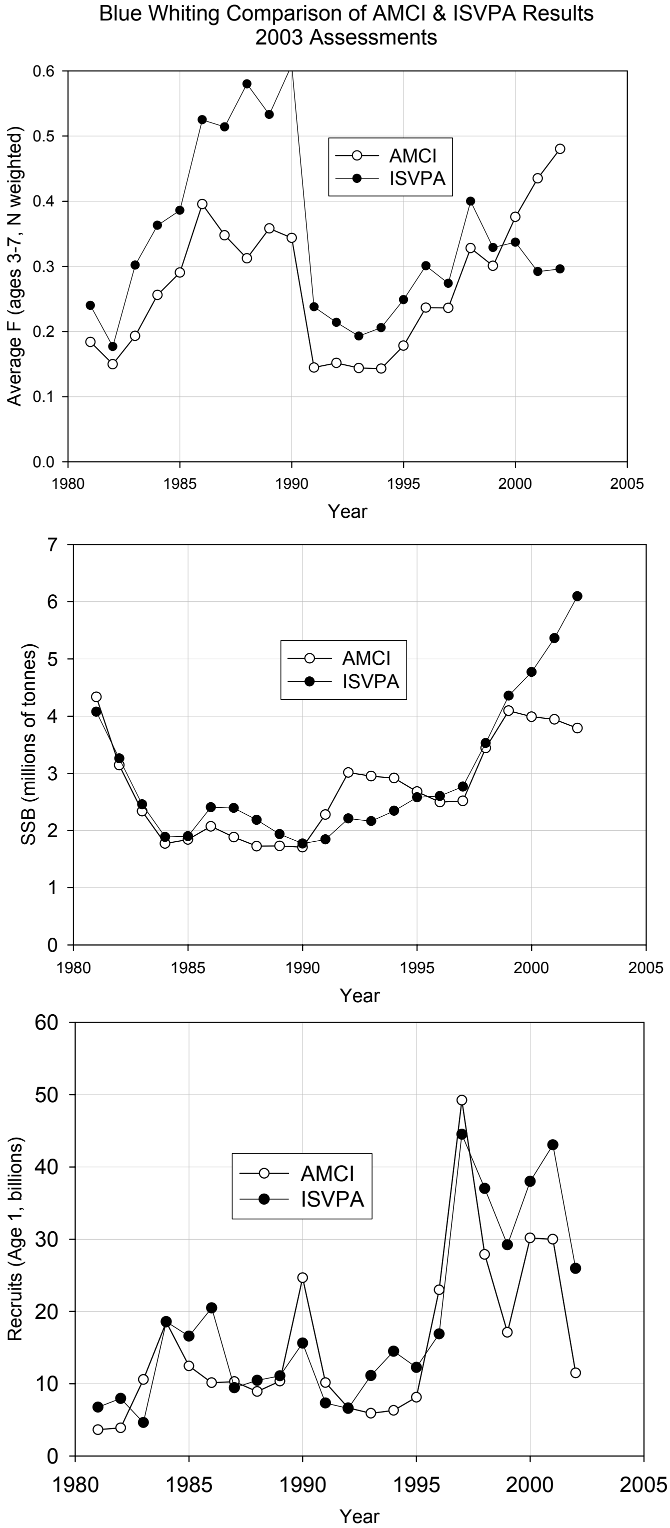


Figure 1.4.2. Comparison of stock assessment results for blue whiting from 2003, for two assessment methods (AMCI and ISVPA).

2 EVALUATION OF DATA INPUTS INTO ASSESSMENT MODELS FOR NORWEGIAN SPRING-SPAWNING HERRING AND BLUE WHITING

2.1 Description of Survey Data Series

This section provides a brief description of the different surveys used in the assessment of the Norwegian spring-spawning herring and the blue whiting stocks. Map showing typical area and distribution of the stocks for the different surveys are also shown. In some of the surveys both stocks are covered and the survey is subsequently referred to twice.

2.1.1 Norwegian Spring-Spawning Herring

All trawl-acoustic estimates are all carried out after the same principles but under varying conditions with regard to relative density in acoustic sampling and also with regard to species mix in the acoustic registrations. In some areas (in particular in the western areas during the May survey) herring is found in a mix with blue whiting, which is a challenge in allocating Sa values to species. In the wintering area the herring is in pure concentrations.

The survey areas, cruise tracks and distribution of herring from these surveys are illustrated in Figures 2.1.1.1-2.1.1.11. The acoustic backscattering data are allocated to species categories using the Bergen Echo Integrator (BEI) system, and are assembled as average Sa values per nautical mile for each 5 miles ('5-mile values'). These values are allocated to squares of 0.5° latitude and 1° longitude, so that the average Sa value per square is recorded.

Samples of the fish are obtained by pelagic trawling on registrations, using most often the 'Åkratrål', which is a 30x30 m midwater trawl. In some cases a capelin trawl is used in the wintering areas.

2.1.1.1 Norwegian January survey for Norwegian spring spawning herring

This survey is carried out by Norway in the fjords in the Vestfjord area. The survey is carried out within the Ofotfjord, Tysfjord and inner parts of the Vestfjord. The area and survey design is adapted to suit the shape of area and the distribution of the stock which can change from year to year and throughout the winter. In most cases each of the three fjords is covered successively 2 times, thus giving a very high aerial coverage as compared to the oceanic surveys.

This survey series was ended in 1998.

2.1.1.2 February March survey for Norwegian spring spawning herring

This survey covers the area of spawning grounds and is limited by 69 ° N in the north and about 62° N in the south. The survey period used to be about three weeks. The procedures were similar to those used for the other surveys.

This survey was ended in 2000.

2.1.1.3 May survey for Norwegian spring spawning herring

The survey is carried out in the Norwegian Sea north of 62° N covering the area to the west of the Norwegian shelf and east of Iceland. The extent of the distribution depends on the movement of the herring off the shelf as they follow sources of food to the North and West. Participation in the survey is by Norwegian, Faroes, Icelandic and EU vessels, with varying participation from the EU. Transects are widely spaced over the extensive area (Figures 2.1.1.2-2.1.1.3) and are carried out alternately by different vessels and single and combined estimates have been made. The survey design and analysis is documented in report of the Planning Group on Surveys on Pelagic Fish in the Norwegian Sea (ICES PGSPFN). At the PFSPFN homepage reports for every year during the period 1995-2003 are found (www.imr.no/PGSPFN).

2.1.1.4 Norwegian July-August survey for Norwegian spring spawning herring

The survey has normally been carried out as a one-boat survey to map the distribution of the NSSH during the late feeding period. In some years it has been carried out by two boats over an area stretching further to the south in order to map a larger portion of the blue whiting stock but in recent years it has been concentrated on the NSSH in the areas north of 70ºN (Figures 2.1.1.4-2.1.1.5).

The survey will not be carried out in 2004 due to the MARECO (www.mar-eco.no) survey on the mid Atlantic ridge but will again be run in 2005

2.1.1.5 November-December survey for Norwegian spring spawning herring

This survey is carried out by Norway in the fjords in the Lofoten area (Figures 2.1.1.6-2.1.1.9) and is today regarded the most important survey in this stock together with the May survey.

In December 2002 the recruiting 1998 and partly the 1999 year classes were expected to enter the fjord system. According to the survey results this appeared not to be the case. As a consequence, based on observations by fishermen, an additional coverage was carried out in the areas off Vesterålen, from about 68º to 69º N just outside the shelf edge. Herring was observed in the area and a trawl haul confirmed that the herring in this area was the missing 1998 and 1999 year classes. Based on these observations a survey carried out by the RV Johan Hjort and the RV G.O.Sars in the area during the 2003 December coverage of the herring stock (Figure 2.1.1.10).

2.1.1.6 Juvenile survey for Norwegian spring spawning herring

This survey is covering the young herring in the Barents Seas. It is carried out by Norwegian or Russian vessels depending on availability (Figure 2.1.1.11).

2.1.2 Blue Whiting

Examples of survey areas cruise tracks and distribution patterns from the various fishery-independent surveys used to inform stock assessments of blue whiting are provided in Figures 2.1.2.1-2.1.2.3.

2.1.2.1 Norwegian acoustic survey for blue whiting on the spawning grounds.

This survey covers the spawning grounds for blue whiting in March April, and is aiming specifically at estimating the abundance of spawning blue whiting (Figure 2.1.2.1). It covers the shelf break to the West of Ireland and Scotland with a zig-zag cruise track, the area is surveyed during a period of 4-5 weeks. Traditionally, the abundance estimate is several times higher than indicated by the analytical assessment. One likely reason is that the target strength (TS = 21.8logL – 72.7) is too low. The weather conditions can be quite rough, and it is not unlikely that this may cause a year effect in the survey results. The survey results may also be influenced by migrations of the fish. The data used in the assessment extend back to 1981. A shift in catchabilities is assumed in 1991, due to change both in vessel and in acoustic equipment

2.1.2.2 Russian acoustic survey for blue whiting on the spawning grounds.

Since 1983 during the spring-time one research vessel conducted target strength surveys TAS of blue whiting west off the British Islands on the spawning grounds. For several years this survey was joint together with the Norwegian vessel and sometimes a common estimate was done. This survey was conducted in the same area and time as the Norwegian one, although sometimes in the opposite direction. The series was stopped in 1996 and was continued from 2001.

2.1.2.3 May survey for blue whiting

Besides Norwegian spring spawning herring, blue whiting is measured during the international survey in May. An example of relative densities of blue whiting is given in Figure 2.1.2.2.

2.1.2.4 Norwegian survey for blue whiting in the Norwegian Sea in July – August.

This survey is primarily conducted to follow the migration of Norwegian spring spawning herring and to relate the migration to hydrographic conditions. The cruise track is not ideal for an acoustic survey, with tracks usually 60 nm apart (Figure 2.1.2.3). This survey is the first indication of the incoming year class, however, which is measured at age 1.

2.2 Correlations Among Data Series

Data investigation methods

Two methods of examining survey consistency were used for both Norwegian Spring Spawning herring and for blue whiting, within-survey consistency and between-survey consistency. These investigations were similar to the methods used within the EVARES project (Anon. 2003)

Within-survey consistency

$N_{a,y,s}$ is the abundance index for age a, year y, and survey s. Within survey consistency may be expressed as correlation coefficients calculated over years between the $N_{a,y,s}$ and $N_{a+1,y+1,s}$ offer an indication of the ability of survey s to track year class strength effects. This has been done in the linear domain to allow for zeros as these are often present in the Norwegian Spring Spawning herring data, if correlation of log(N) was preferred the log of (N+k) would need to be used where *k* is a small constant depending on the scaling of N. A value of k of half of the min{N} might be preferred. In addition to the correlation coefficients, bi-variate plots were examined to check for linearity and the absence of a spuriously high correlation resulting from one or two outliers.

There are limits to the interpretation of such correlation coefficients. If for a stock the variability of the true year class strength is low within the observed period, this leads to lower correlations and conversely high variability in recruitment leads to potentially high correlation. Also, when we calculate a correlation coefficient between the two variables X1(y) and X2(y) with $X1(y) = N_{a,y,s}$ and $X2(y) = N_{a+1,y+1,s}$ we are measuring the adequacy of a linear relation of the form $X2(y) = \alpha X1(y) + \beta$. We accept or assume that the corresponding value for α may not be equal to one due to mortality or survey catchability. But this also implies that we may need to accept that the catchability coefficients associated to age a and/or a+1 may vary with year class strength. In most cases, in assessments this is not allowed. However, for the sake of simplicity it was decided to use basic correlation coefficients, as they prove a useful indicator. They may highlight specific difficulties, including phenomena that would deserve further biological interpretation, for instance when it appears that a survey can efficiently track year class strength effects within an age range, but not necessarily the same age range as another survey. This implies even for adult it may be preferable to limit the upper ages used for tuning for some surveys.

To visualize the correlation in the surveys plots were made where the numbers at age a are plotted versus the numbers at age a+1 in the same survey. The points are marked as the year class so it is possible to follow the year classes through the survey. A linear regression was made where the line is forced through the origin. The fitted line is shown.

Between-survey consistency

Correlations for a given age between abundance indices provided by two surveys, s_1 and s_2, the corresponding two time series being:

$X1_y = U_{a,y,s1}$ and $X2_y = U_{a,y,s2.}$

A review of the corresponding correlation coefficients makes it possible to assess the consistency between surveys for each age. Identification of a strong correlation pattern between independently conducted surveys could pave the way for tuning techniques that would recognize them. A comparison of within survey consistency and between-survey consistency may be used as a first stage to identifying ages that may be unsuitable for tuning.

To see if there are correlations between surveys, plots were made where the numbers at the same age in the surveys were plotted against each other. A linear regression was made where the line was forced through the origin. The fitted line is shown in the plots.

2.2.1 Norwegian Spring-Spawning Herring

Within-survey consistency is illustrated with scatter plots of $N_{a,y,f}$ against $N_{a+1,y+1,f}$ in Figures 2.2.1.1-2.2.1.6 and correlations within surveys in Table 2.2.1. The surveys for Norwegian spring spawning herring are described by the following mnemonics.

Jan-OW	January survey for Norwegian spring spawning herring in the overwintering grounds.
Mar-Sp	February and March survey for Norwegian spring spawning herring in the spawning areas
May-FD	May survey for Norwegian spring spawning herring in the feeding grounds in the Norwegian Sea
May- BS	May survey for juvenile Norwegian spring spawning herring in the Barents Sea
Aug-BS	Juvenile Norwegian Acoustic Surveys in assessments of Norwegian Spring Spawning herring
Nov –OW	November-December survey for Norwegian spring spawning herring

Table 2.2.1 Within-survey consistency for Norwegian spring spawning herring tuning indices, correlation of N at age a in year y with N at age a+1 in year y+1 over all years of the survey. (n is the number of pairs of years for the correlation; bold type indicate correlation of more than 0.9, small type indicates spurious correlation due to coincidence of zeros; the surveys are described by the same mnemonics given above)

a1	a2	Mar-Sp	Nov-OW	Jan-OW	May-FD	May-BS	Aug-BS
		n = 7	10	6	6	6	3
1	2		0.11		-0.33	0.70	0.78
2	3	-0.39	-0.11	0.74	0.31	0.64	0.57
3	4	0.51	0.70	-0.36	**0.97**	0.44	
4	5	**0.96**	0.78	0.77	**1.00**	0.39	
5	6	**0.92**	0.82	0.87	**0.99**		
6	7	**0.99**	0.80	**0.99**	**0.99**		
7	8	**0.95**	**0.96**	**0.99**	**0.96**		
8	9	**0.96**	**0.97**	**0.99**	**0.99**		
9	10	0.78	**0.90**	**0.99**	0.77		
10	11	0.33	**0.97**	**1.00**	0.43		
11	12	1.00	**0.99**	**1.00**	0.43		
12	13	1.00	**0.99**	**1.00**	0.19		
13	14	1.00	**1.00**	**1.00**	1.00		
14	15	1.00	**1.00**		1.00		
15	16	1.00	**1.00**	1.00			

Table 2.2.2 Between-survey consistency for Norwegian spring spawning herring tuning indices, expressed as correlation coefficients of N at age. (The surveys are described by the same mnemonics given above; n is the number of pairs of years for the calculations; bold type indicates correlation at greater than 0.9

Surveys	Mar-Sp	Mar-Sp	Mar-Sp	Mar-Sp	Mar-Sp	Nov-OW	Nov-OW	Nov-OW	Nov-OW	Jan-OW	Jan-OW	May-FD	May-FD	May-BS
Age	Nov-OW	Jan-OW	May-FD	May-BS	Aug-BS	Jan-OW	May-FD	May-BS	Aug-BS	May-FD	May-BS	May-BS	Aug-BS	Aug-BS
	n=6	6	4	8	4	7	7	11	5	3	8	7	4	6
1							-0.05	-0.19	-0.44			-0.16		0.51
2	-0.02	-0.46	-0.48	0.18	0.99	0.21	0.20	-0.16	0.86	0.99	0.00	-0.05	0.29	0.75
3	0.55	0.22	0.83	0.05	0.74	-0.12	0.38	0.66	0.82	0.82	0.02	0.72	0.74	0.73
4	0.78	0.78	**0.99**	0.40	**0.97**	**0.99**	0.86	0.38	0.60	**0.99**	0.43	0.81		1.00
5	0.81	**0.95**	**0.98**	**0.94**		**0.92**	**0.91**	0.81		**1.00**		**0.98**	0.00	
6	**0.98**	**0.97**	**0.99**			**0.91**	**0.90**			**0.99**				
7	**0.88**	**1.00**	**0.94**			**0.88**	**0.96**			**0.92**				
8	**0.97**	**0.98**	**0.99**			**0.96**	**0.95**			**0.99**				
9	**0.96**	**0.98**	**0.94**			**0.98**	**0.97**			0.79				
10	0.80	0.89	0.64			**1.00**	0.76			0.68				
11	**1.00**	**1.00**	0.06			**1.00**	0.87			**1.00**				
12	**1.00**	**1.00**	-0.33			**1.00**	0.60			-0.27				
13	**1.00**	**1.00**	**1.00**			**1.00**	**0.99**			**1.00**				
14	**1.00**	**1.00**	**1.00**			**1.00**	**1.00**			**1.00**				
15	**1.00**	**1.00**	**0.99**			**1.00**	**0.99**			**1.00**				
16	**1.00**	**1.00**	**1.00**			**1.00**	**1.00**			**1.00**				

Table 2.2.2 shows the between-survey consistency expressed as correlation coefficients at age and the scatterplots in Figures 2.2.1.7-2.2.1.12 show the consistency. The surveys are described by the same mnemonics given above.

The surveys on adult spring spawning herring were investigated to see if they could be used as biomass indices. New indices were calculated from the age disaggregated survey data using the mean weights and fractions mature in the stock from the assessment data (ICES 2003). These are then compared with two indices calculated by different methods from a larvae survey. The time series are illustrated in Figure 2.2.1.13 and the correlation coefficients are given in Table 2.2.3. For this illustration the November December overwintering grounds survey (Nov-OW) has been allocated to the following year. This is because the incoming year class is expected to join the rest of the stock for the first time in this period so that the biomass observed in Nov-Dec is more similar to the Biomass in the first few months of the following year than any of the surveys earlier in the same year.

Table 2.2.3 Correlation between biomass indices derived from surveys of adult Norwegian spring spawning herring using stock weight and fractions mature and indices derived from larvae survey. Correlation coefficient (r) and numbers of pairs (n).

Surveys	Mar-Sp		Nov-OW		Jan-OW		May-FD		Lav Ind 1		Lav Ind 2	
	r	n	r	n	r	n	R	n	r	n	r	n
Mar-Sp	1.00	10										
Nov-OW	0.79	6	1.00	11								
Jan-OW	0.97	6	0.88	6	1.00	8						
May-FD	0.09	4	-0.24	7	-0.49	3	1.00	7				
Lav Ind 1	0.57	10	-0.03	11	0.52	8	0.61	7	1.00	17		
Lav Ind 2	0.71	10	0.06	9	0.57	8	0.49	5	0.85	14	1.00	14

Discussion

The scatter plots and correlation tables suggest that there are a number of ages that are tracked well by the surveys. Comparison with the correlations given in the biomass index table suggests that biomass indices derived from all the surveys may not perform as well as the age disaggregated indices. While the reason for this has not been full established there are indications that year classes do not fully recruit to these surveys as they mature and therefore each biomass index has different amounts of the same ages. On this basis it is suggested that these indices are not selected initially but could be explored later to see if they provide stability. If the age disaggregated survey indices are all used separately the list of ages is given in Table 2.2.4

Table 2.2.4 The best performing ages for Norwegian spring spawning herring tuning indices that should be selected initially as age disaggregated indices for use in an assessment.

Ages Surveys	Minimum	Maximum	Possible
Mar-Sp	3	9	
Nov-OW	4	16	
Jan-OW	5	16	
May-FD	3	9	
May-BS	1	2	3
Aug-BS	1	2	3

The juvenile surveys both provide coherent information on recruits at 1 and 2 years (Table 2.2.1) with some evidence of useful information at age 3 (Table 2.2.3). Age 3 data and older can be provided from both the March spawning ground survey and the May feeding area survey. However, both these surveys do not seem to provide good data at ages older than 9. Both the over wintering surveys provide data for the same cohorts at 4 in Nov/Dec and age 5 in Jan/Feb. These surveys also provide useful data over the full age range to 16 years.

There are some indications that combining the data from the feeding grounds survey with the juvenile survey by adding the observed abundance with equal weighting (see Table 2.2.5) may provide an improvement over using the surveys independently. If this is done it is possible to obtain a short time series for age 2 that performs better than the age 2 from either survey separately. The improved correlation at age 2 to 3 suggests that an age 2 index from a combined survey may give better data for the assessment than the juvenile survey alone with no loss of years of data. Some of this apparent improvement comes from truncating the time series but nevertheless the combined survey does seem to perform better for ages 2 and 3 and these ages are not well described by other surveys.

While this analysis provides a good preliminary indication of data that is suitable for taking into an assessment model, no information is given on the relative merits of the selected data. Estimates of the variance at age of each series would be required if weighting of the data were also to be included.

Table 2.2.5 Comparison of within-survey consistency for Norwegian spring spawning herring tuning indices for a combined survey using the May Barents Sea Juvenile survey and the May Feeding area survey.

age1	age2	May-BS	May-FD	BS+FD
1	2	0.70	-0.33	0.45
2	3	0.64	0.31	0.98
3	4	0.44	0.97	0.98

2.2.2 Blue Whiting

Within-survey consistency is illustrated with scatter plots of $N_{a,y,f}$ against $N_{a+1,y+1,f}$ in Figures 2.2.2.1-2.2.2.7 and correlations within surveys in Table 2.2.6. The surveys for blue whiting are described by the following mnemonics.

Barents	Norwegian winter survey in the Barents Sea
Icelandic	Icelandic blue whiting survey in summer
Norwsea	Norwegian survey in the Norwegian Sea in July – August.
Pgspnf	International survey in the Norwegian Sea in May
Spanish	Spanish pair trawl series (cpue)
Spawnnor	Norwegian acoustic survey for blue whiting on the spawning grounds.
Spawnrus	Russian acoustic survey for blue whiting on the spawning grounds.

Table 2.2.6 Within-survey consistency for blue whiting tuning indices, correlation of N at age a in year y with N at age a+1 in year y+1 over all years of the survey. (n is the number of pairs of years for the correlation; bold type indicate correlation of more than 0.8, small type indicates spurious correlation due to coincidence of zeros; the surveys are described by the same mnemonics given above)

age1	age2	Barents	Icelandic	Norwsea	Pgspnf	Spanish	Spawnnor	Spawnrus
		n=22	4	15	3	19	16	12
0	1		0.59					
1	2	0.45	0.98	0.85	0.93	0.38		
2	3	0.64	0.19	0.91	0.96	0.25	0.86	
3	4		0.03	0.64	0.77	0.11	0.86	0.79
4	5		-0.45	-0.01	-0.83	0.38	0.56	0.46
5	6		0.09	0.41	1.00	0.59	0.70	0.69
6	7		0.87	0.71			0.73	0.31
7	8		0.99				0.38	-0.20

Table 2.2.7 Between-survey consistency for blue whiting tuning indices expressed as correlation coefficients of N at age. (The surveys are described by the same mnemonics given above; n is the number of pairs of years for the calculations; bold type indicates correlation at greater than 0.8.

Survey		n	0	1	2	3	4	5	6	7	8
Icelandic	Northsea	5	0.56								
Barents	Icelandic	5		-0.02	-0.28	-0.11					
Barents	Norwsea	18		0.86	0.14	0.47					
Barents	Pgspnf	4		0.71	0.23	0.30					
Barents	Spanish	20		0.21	0.12	0.12					
Barents	Spawnnor	20			0.06	0.17					
Barents	Spawnrus	14				0.54					
Icelandic	Norwsea	3		0.02	0.97	0.75	1.00	0.40	0.99	0.91	
Icelandic	Pgspnf	4		-0.99	0.33	0.97	0.66	-0.44	0.74		
Icelandic	Spanish	4		0.19	0.53	0.01	0.86	0.39	-0.27		
Icelandic	Spawnnor	5			0.32	0.61	0.09	-0.72	0.65	0.84	0.33
Norwsea	Spanish	15		0.11	-0.01	0.09	0.23	0.11	0.17		
Norwsea	Spawnnor	14			0.61	0.82	0.51	-0.06	0.18	0.76	
Norwsea	Spawnrus	12				0.52	0.68	-0.03	-0.25	-0.01	
Pgspnf	Spanish	3		0.59	-0.64	-0.37	-0.53	0.25	-1.00		
Pgspnf	Spawnnor	4			0.97	0.85	-0.62	0.76	0.96		
Spanish	Spawnnor	18			-0.05	0.12	0.01	-0.15	-0.24		
Spanish	Spawnrus	13				-0.12	-0.10	0.48	0.73		
Spawnnor	Spawnrus	12				0.58	0.40	0.67	0.53	0.13	0.08

Table 2.2.7 shows the between survey consistency expressed as correlation coefficients at age. The surveys are described by the same mnemonics given above.

Discussion

The scatter plots and correlation tables suggest that there are very few ages that are tracked well by the surveys. Judgement requires careful examination of the scatter plots. If the age disaggregated survey indices are all used separately, the appropriate list of ages to consider in assessment calculations is provided in Table 2.2.8, Barents, Icelandic and Norwsea surveys all show some information on recruitment at 1 and 2. The Barents Sea survey provides useful age based indices for ages up to 3, and the NORwsea survey to age 4 and perhaps age 5. Even though Pgspnf seems to perform well in the correlation matrix the series of 3 years is too short and the correlation is driven by the variable recruitment. While this survey may provide useful data it is too early to tell how well it is performing. The Spanish tuning fleet does not seem to contain useful data and is not currently used. This analysis supports its removal in the tuning of the assessment by WGNPBW. Information on adults is available from the Icelandic survey but the results look variable and should not be used without more detailed examination. The best information seems to come from the two surveys on the spawning grounds, Spawnnor and Spawnrus. The ages indicated for these are 2-7 and 3-6 respectively. There are problems around age 5. These are caused particularly, but not exclusively, by low numbers in the 1995 yearclass at age 5 in 2000. The abundance has been checked and is based on only a very small number of otoliths. However, the low value may not be in error and may be a sign of rapid depletion of that cohort. Nevertheless it seems likely that, if exploitation continues at the current rate, there will be a need to ensure that older year classes are represented appropriately and age estimation through sampling stratified by length may be helpful. Though it may be necessary to take increasing numbers of otoliths at the greater lengths.

Table 2.2.8 The best performing ages for blue whiting tuning indices that should be selected initially as age disaggregated indices for use in an assessment.

Ages Surveys	Minimum	Maximum	Possible
Barents	1	3	
Icelandic	1	2	
Norwsea	1	4	5+?
Pgspnf	None		
Spanish	None		
Spawnnor	2	7	8
Spawnrus	3	6	

While this analysis provides a preliminary indication of suitable data, no information is given on the relative merits of the selected data. Estimates of the variance at age of each series would be required if weighting of the data were also to be included. There are signs of some problems in the data and care will be required to ensure that noisy or parts of the data with apparently different catchabilities are not taken at face value. In particular the Norwsea survey shows some signs of a change in catchability from the early period to the later one following changes of EK400 to EK500 sounders. The earlier part could be fitted with a different catchability but if this is done it may add little useful information and could be excluded.

2.3 Evaluation of the Consistency of Commercial and Survey Catches-at-Age using Catch Curves

These analyses were undertaken to examine the consistency of commercial and survey catch-at-age data used to track cohorts, using catch curves plotted by year and year class. These analyses plotted the catches by age group on logarithmic scales that allow rapid assessment of the consistency of the catches with the presumed model that such catches (in numbers) should decline consistently with age, as influenced by natural and fishing mortality and appropriate catchabilities at age for commercial catches and survey catches. Obviously, if cohorts are poorly tracked due to fluctuating distribution patterns, poor sampling or other factors influencing seasonal or annual catchability, then catch curves should not demonstrate consistent descending right-hand limbs.

2.3.1 Norwegian Spring-Spawning Herring

Catch curves of different year classes in commercial catches of herring are shown in Figures 2.3.1.1 and 2.3.1.2 with lines corresponding to Z=0.4 and Figure 2.3.1.3 shows log of the ratio of the catch in numbers of an year class and the catch of the same year class the year after. The Figures show few blocks.

1) Year classes 1983-1988 that are caught in large number already at age 0 and for many of the year classes the number caught is relatively similar for most age groups between 0 and 10. Year class 1988 is a little different from these year classes, possibly closer to block 2.
2) Eyeballing 1989 and 1990 where the number caught is low at age 0 - 2 but, peaking at ages 5 to 7 then declining relatively fast (Z around 0.5) after that
3) Year classes 1991 - 1994. Very little is caught of those year classes at ages 0 to 2 but the catch in number peaks at age 5 to 6 declining very slowly after that.
4) Year classes 1995 and later. Very little is caught at age 0 and 1, the catch in number peaks at ages 3 to 5 and declines slowly after that. Some of these year classes are though short way through the catches so there is not much to say about how the catches develop.

The catch curves indicate quite well that total mortality is not very high on this stock but what do the curves tell us about the selection pattern? To understand the data it must be recalled that 3 of the year classes (1983, 1991 and 1992) were nearly an order of magnitude larger than the other year classes and therefore dominate the catches. If their spatial distribution is different from the other year classes one might assume that the fishery would target those year classes. There are major differences in spatial distribution for young fish (less than 5) as they grow to maturity and the feeding ground surveys indicate that there are also differences in spatial distribution at older ages. Also, the catch decreased between 1986 and 1990 but increased dramatically between 1994 and 1996, mostly due to international fishery outside the Norwegian economic zone but reduced considerably between 2000 and 2002 following agreement on the splitting of the catches.

It is clear that fishery for small herring reduced dramatically around 1990 both due to a reduction in effort and that all the fishery was directed towards the 1983 year class which was nearly all the stock in those years. The following years are characterized by low but increasing fishing effort targeting the 1983 year class. This is shown by the increasing catch in numbers for this year class from age 7 onwards and peaking at age 12 and 13. The expansion of the catches between 1995 and 1997 outside the Norwegian economic zone is caused by the recruiting of the 1991 and 1992 year classes in the fishery, possibly reducing the pressure on the 1983 year class which by that time was possibly old enough to be reliably aged. The figures also indicate that year classes 1991 and 1992 had reduced the fishing pressure on year classes 1989 and 1990. The reduction in catches (and effort) 1999 to 2000 can clearly be seen in the catch curves for age classes 1988 - 1991 and little less clearly for the 1992 year class.

What are the implications for model selection?

1) There is clearly a major change in 1990 and again around 1995 - 1996. The latter change is be caused by the international fleet which is fishing in the Norwegian Sea.
2) Modelling the proportion of an age group in the "harvestable biomass" should be considered. It might have to be considered if adjacent age groups (1991-1992) should be treated as one in this context. Modelling correlation in the catch residuals each year might help here.
3) Using a fleet disaggregated model using separable model for each fleet.
4) Using flexible selection the extreme being VPA with perfect flexibility i.e., with all the error modelled as process error.
5) Using a relatively short separable period for the period after 1996.

It is difficult to say what is the correct way forward is and there is probably not one correct solution. Developing a model containing a mixture of elements 2 and 3 (or 1-3) in the list would be an helpful exercise to understand the nature of the herring fishery.

Catch curves of different year classes in the surveys are shown in Figures 2.3.1.4-2.3.1.15 and log of the ratio of the numbers of a year class in the survey and the numbers of the same year class the year after in Figures 2.3.1.16-2.3.1.19. Some of these series are still short and provide little information. By inspecting Figure 2.3.1.7 it can been seen that there is an apparent year effect that occurs in the 1998 survey, suggesting that the availability of fish in all age groups was different in that year. It can also been seen from this same picture that there is a drop between 2001 and 2002, which can mean that there was less measured.

2.3.2 Blue Whiting

The catch curves and log of the ratios of the catch in number of a year class to the catch in numbers of the same year class the following year are shown in Figures 2.3.2.1, 2.3.2.2 and 2.3.2.3.

What characterizes those catches is that commonly, high numbers of age 0 blue whiting are caught and sometimes the catch in numbers peaks at age 1. Year classes 1994-1998 show somewhat different pattern with catch in numbers peaking at ages 3 to 4 while the catch in numbers for the 1999 year class seems to have peaked at age 2.

Landings of blue whiting increased dramatically between 1997 and 1998 and have been at high level since then. The catch-at-age data indicate that most of year classes 1996 to 2000 have been relatively large possibly explaining the high catch.

Interpreting the catch curves is rather difficult but they seem to indicate that a change in selection might have occurred in 1994 - 1995. They also indicate increasing fishing mortality in the period 1997 - 2001 possibly dropping between 2001 and 2002. The catch curves indicate that the age groups have been exploited at a rate of about Z=0.4.

Investigating fleet-disaggregated catch in numbers by age and fleet (or area) would be an important step forward, especially if it turns out that one could assume constant selection pattern for each fleet.

Use of a separable period that is longer than the period back to 1996 (when the large change in the fishery occurs) seems unlikely to be correct, and would need careful checking if applied.

Catch curves based on the surveys conducted for blue whiting are plotted in Figures 2.3.2.4-2.3.2.19 and log of the ratio of the numbers of an year class in the survey and the numbers of the same year class the year after in Figures 2.3.2.20-2.3.2.26. It is clear that in some surveys the full age spectrum is available (e.g., Icelandic summer survey. Norwegian survey in the spawning grounds, Russian survey on the spawning grounds and Norwegian summer survey in the Norwegian Sea), while fewer age groups are represented elsewhere (e.g., Spanish CPUE, International survey on the feeding grounds in May, Norwegian winter survey in the Barents Sea). The length of these time series and their age selection characteristics have a major influence on assessment results and the variation in catch curves and log catch ratios supports the relatively poor levels of between- and within-survey correlation (Tables 2.2.6 and 2.2.7).

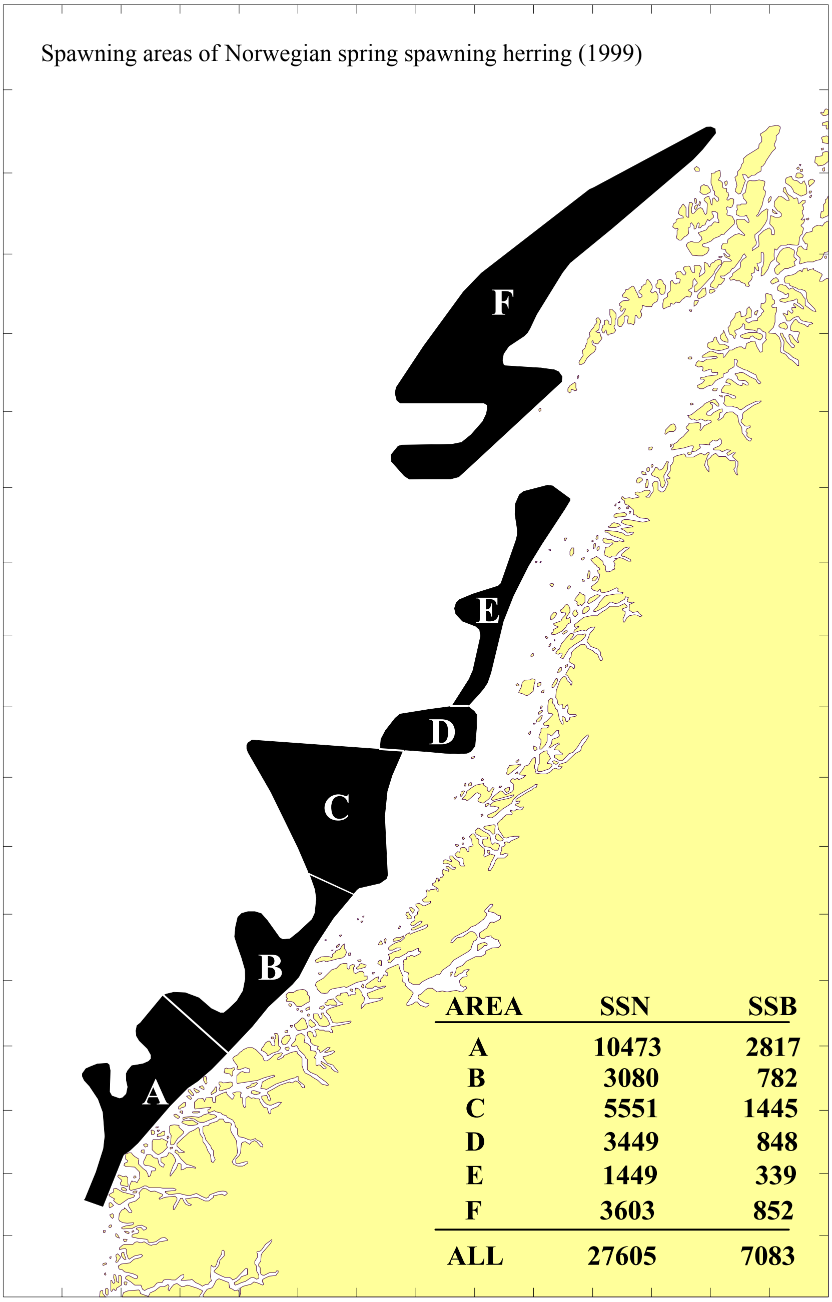


Figure 2.1.1.1. Distribution and abundance (SSN=spawning stock number in millions, and SSB=spawning stock weight in 1000 tonnes) of Norwegian spring spawning herring during the spawning season 1999.

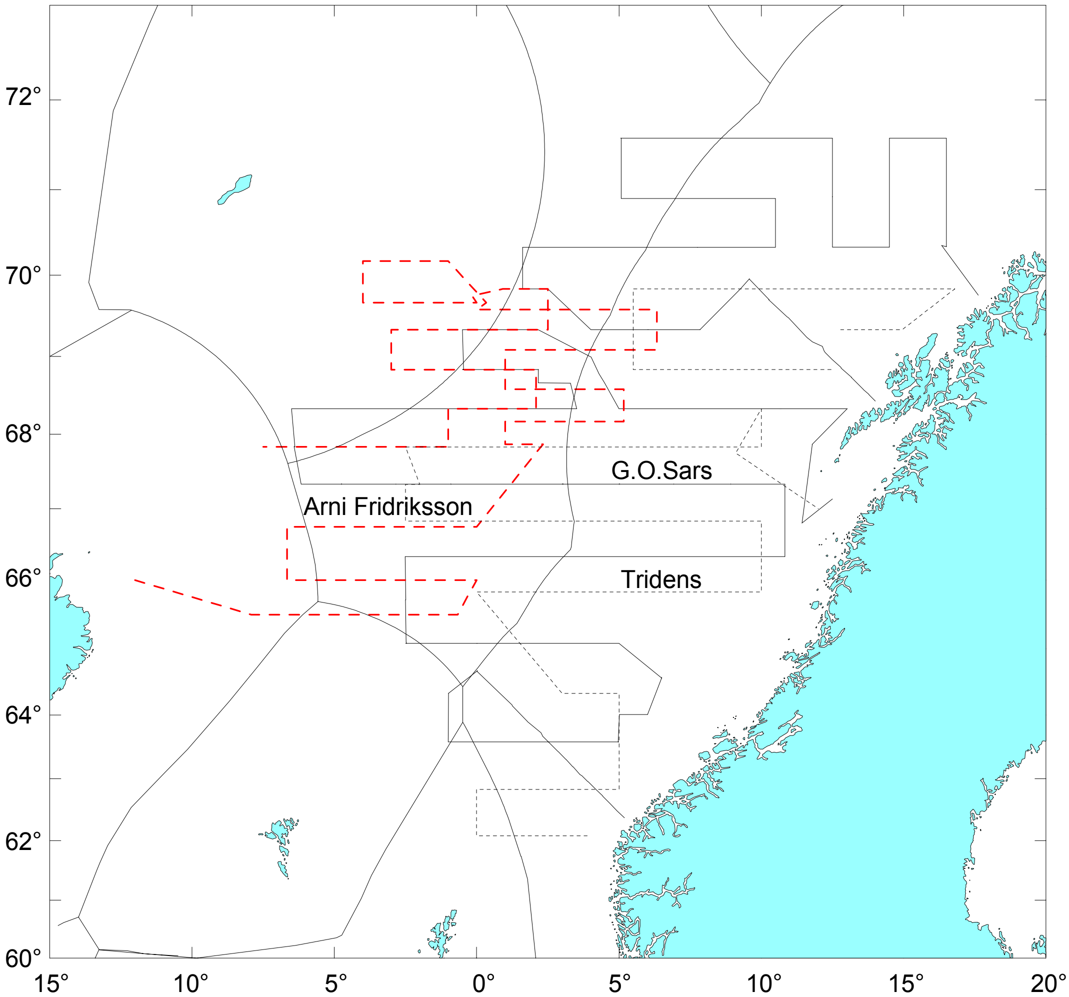


Figure 2.1.1.2. Survey transects of the R/V "Tridens", R/V "Arni Fridriksson" and R/V "G.O.Sars", May 2000.

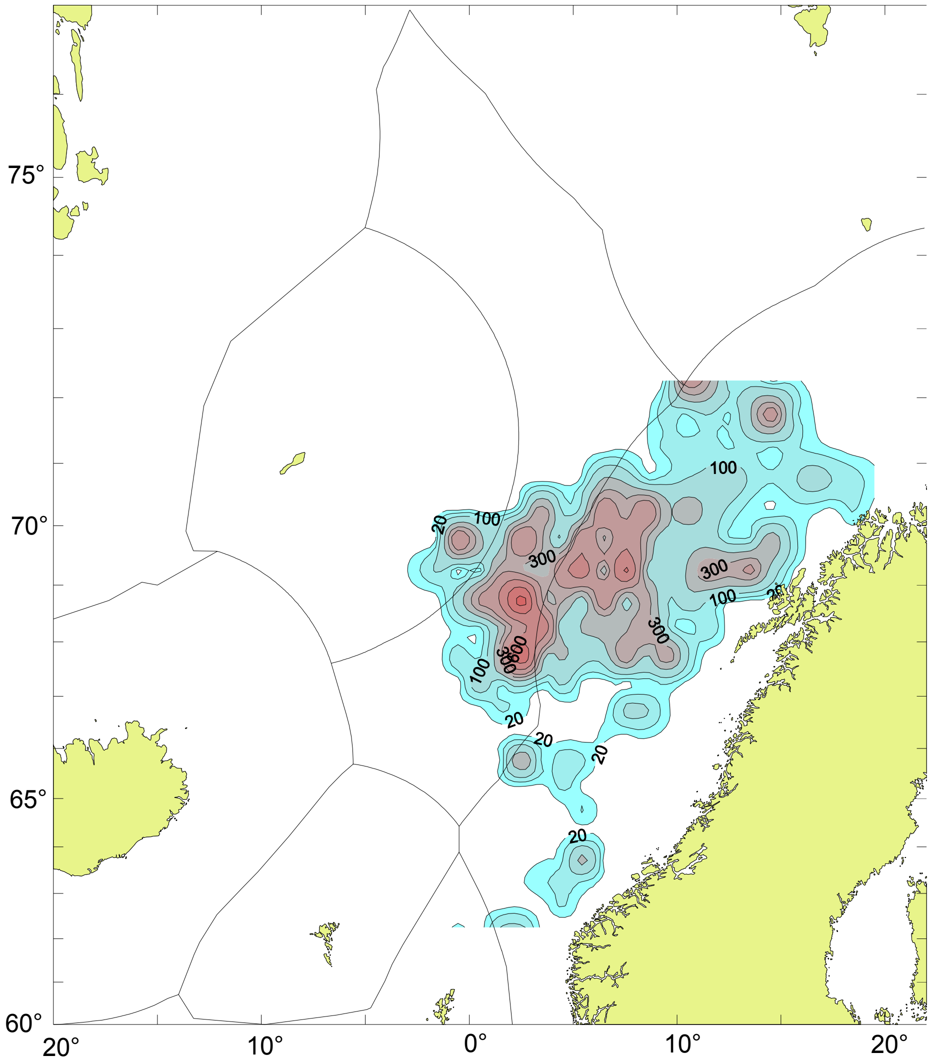


Figure 2.1.1.3. Distribution of Norwegian spring spawning herring as observed by R/V “Tridens”, R/V “Arni Fridriksson” and R/V “G.O.Sars” during the PGSPFN in May 2000.

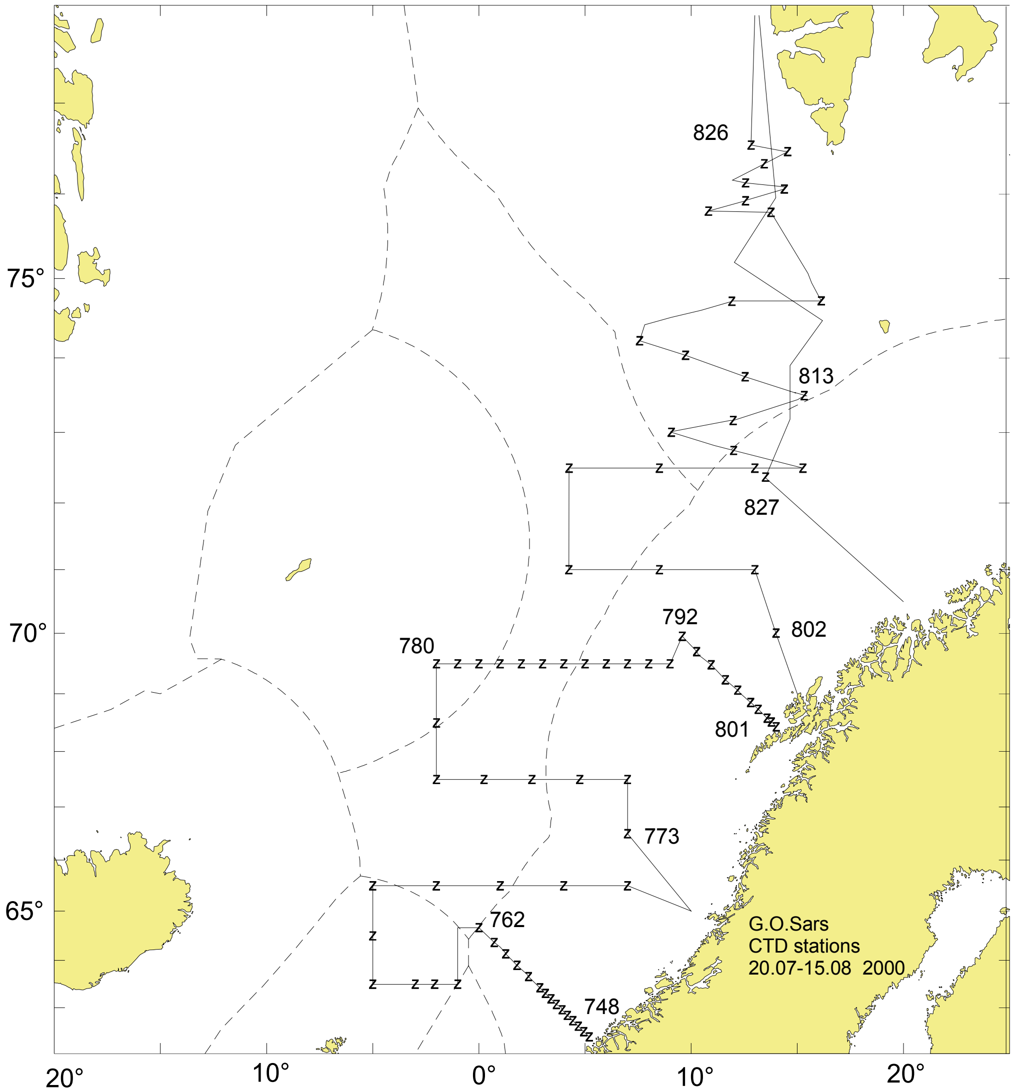


Figure 2.1.1.4. Transects of G.O.Sars during 20th July to 17th August 2000. CTD stations indicated.

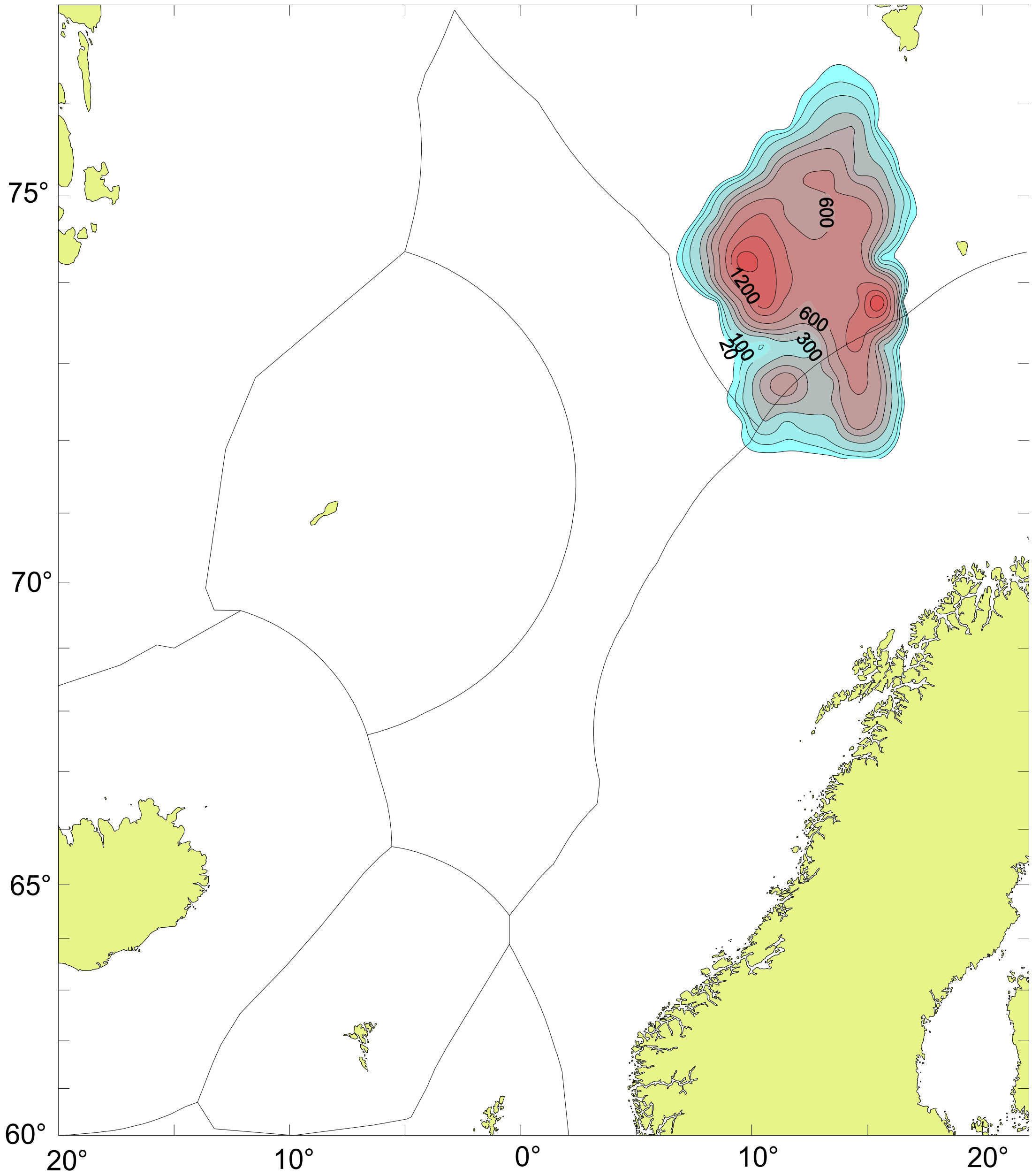


Figure 2.1.1.5. Distribution of Norwegian spring spawning herring as observed by G.O.Sars August 3rd-15th, 2000

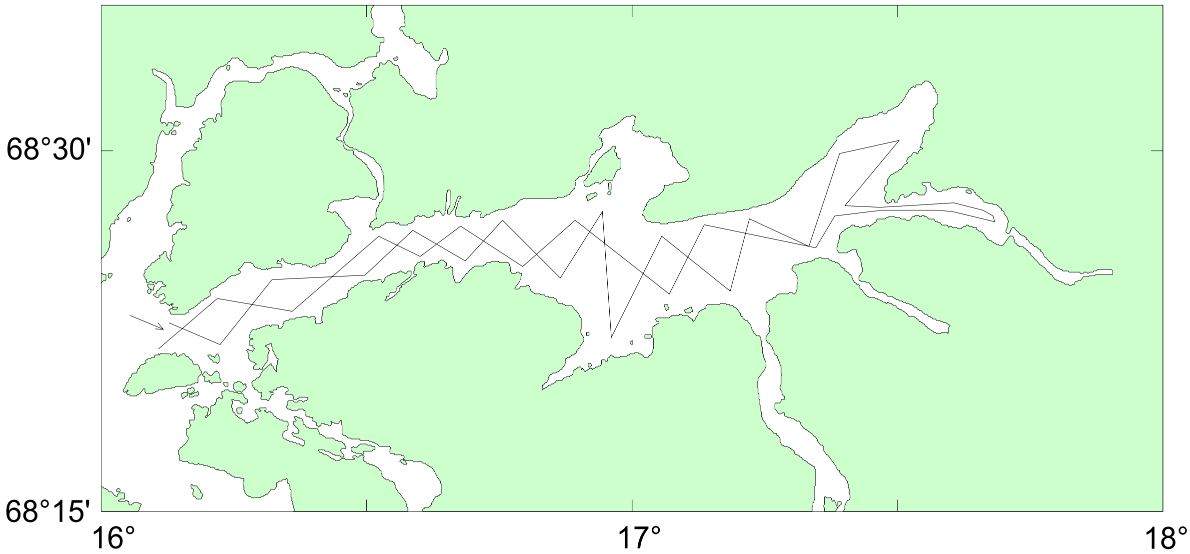


Figure 2.1.1.6 Survey tracks in the Ofotfjord 25.11 - 26.11, 2001

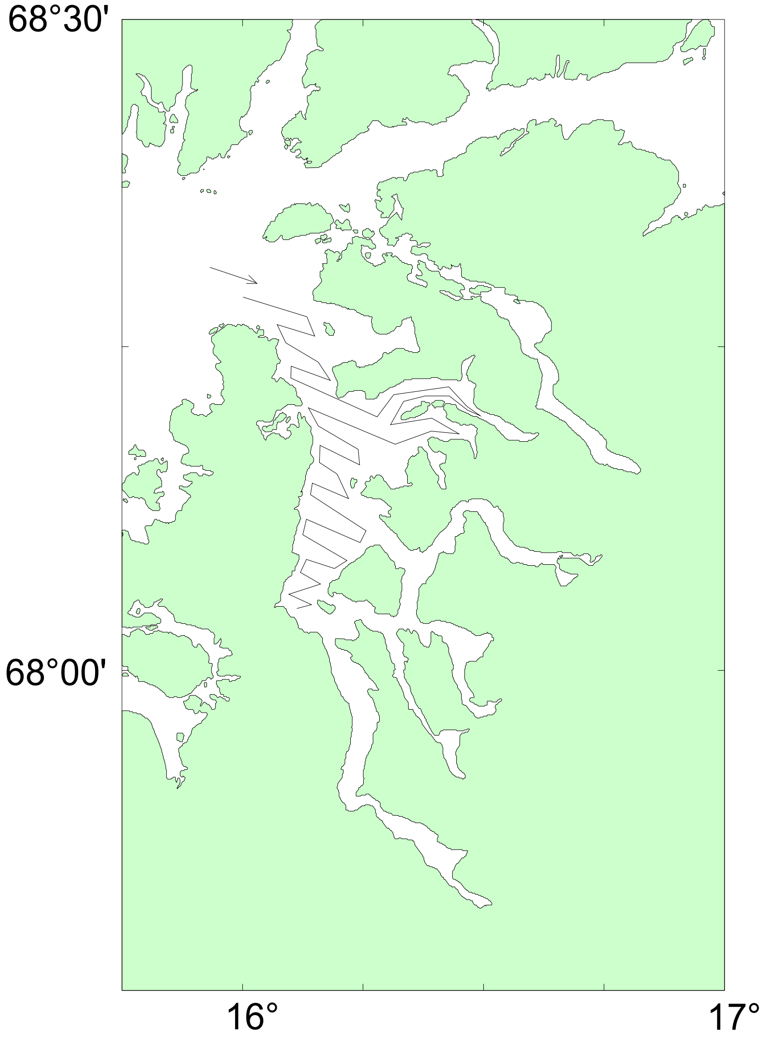


Figure 2.1.1.7 Surveytracks in the Tysfjord 26.11 2001.

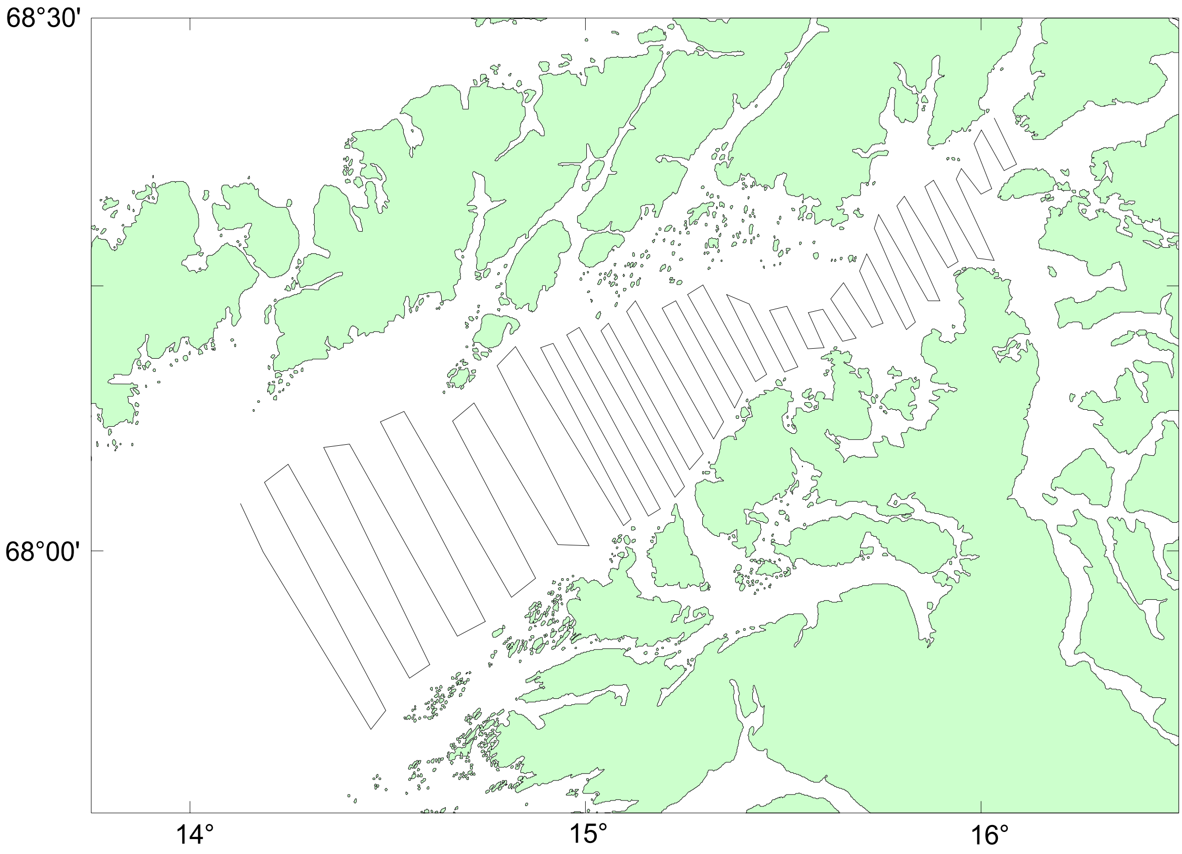


Figure 2.1.1.8 Survey tracks in the Vestfjord 27.11 – 29.11 2001.

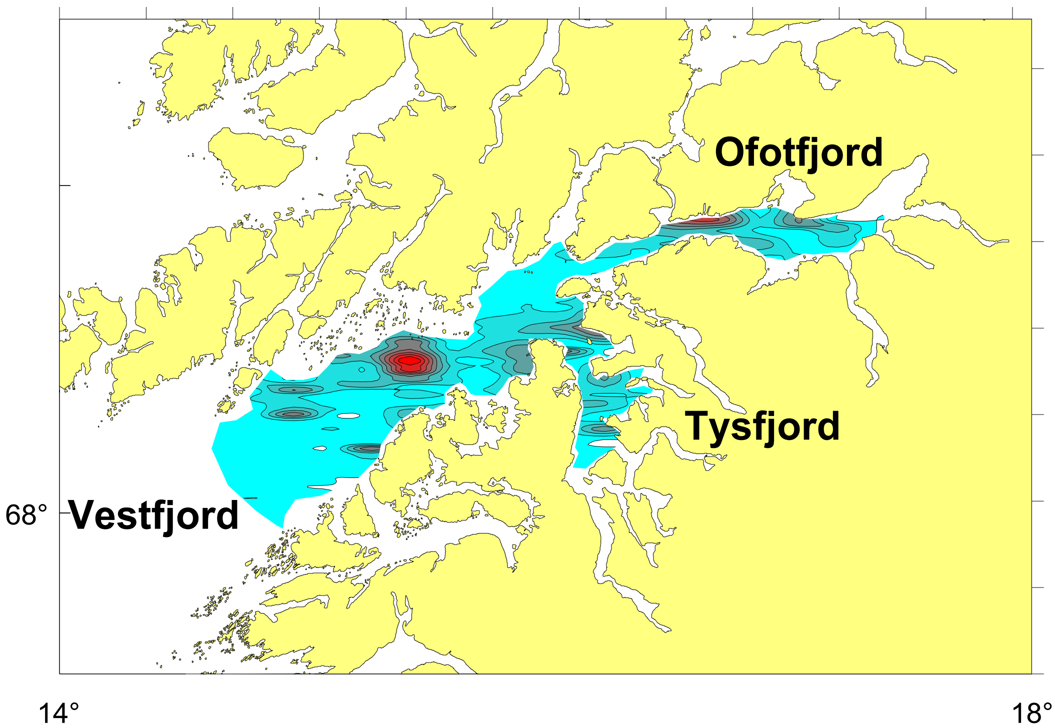


Figure 2.1.1.9. Herring distribution as measured by the RV Johan Hjort in the wintering area in November 1999

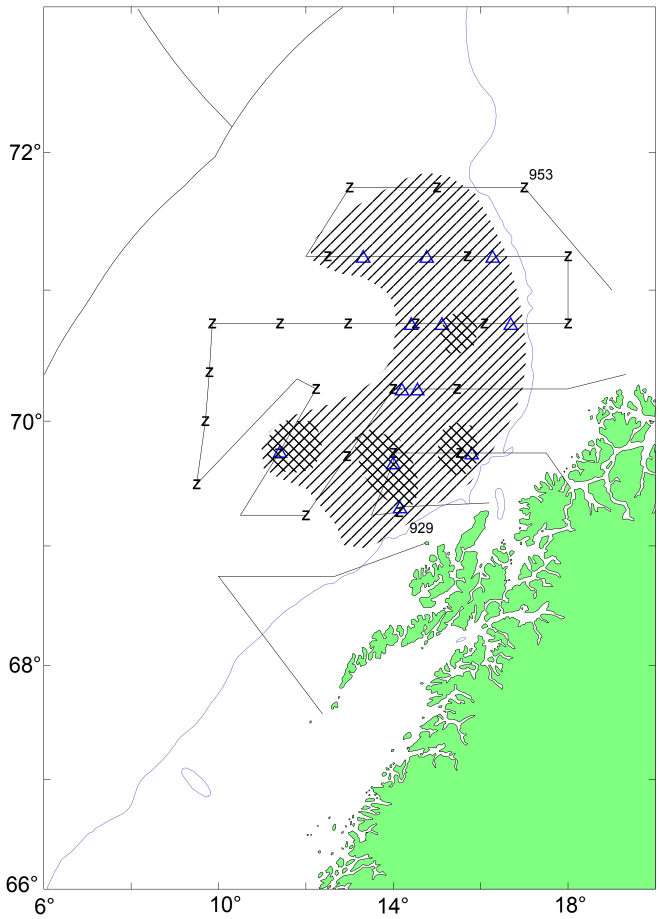


Figure 2.1.1.10 Distribution of Norwegian spring-spawning herring and survey tracks of RV Johan Hjort (north of 69N) and RV G.O.Sars (south of 69N) in December 2003. The total biomass was about 5.5 million tonnes in the offshore area while about 1.6 million tonnes was measured in the Tysfjord and Ofotfjord (ref. Figures 2.1.1.6-7).

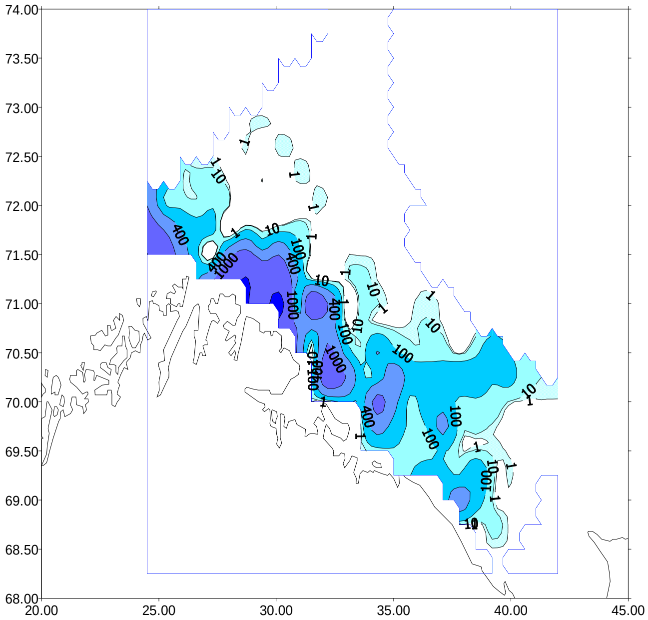


Figure 2.1.1.11. Distribution of herring in the Barents Sea in May-June, map of S A - values. R/V "Persei III", 23/5-12/6 2000.

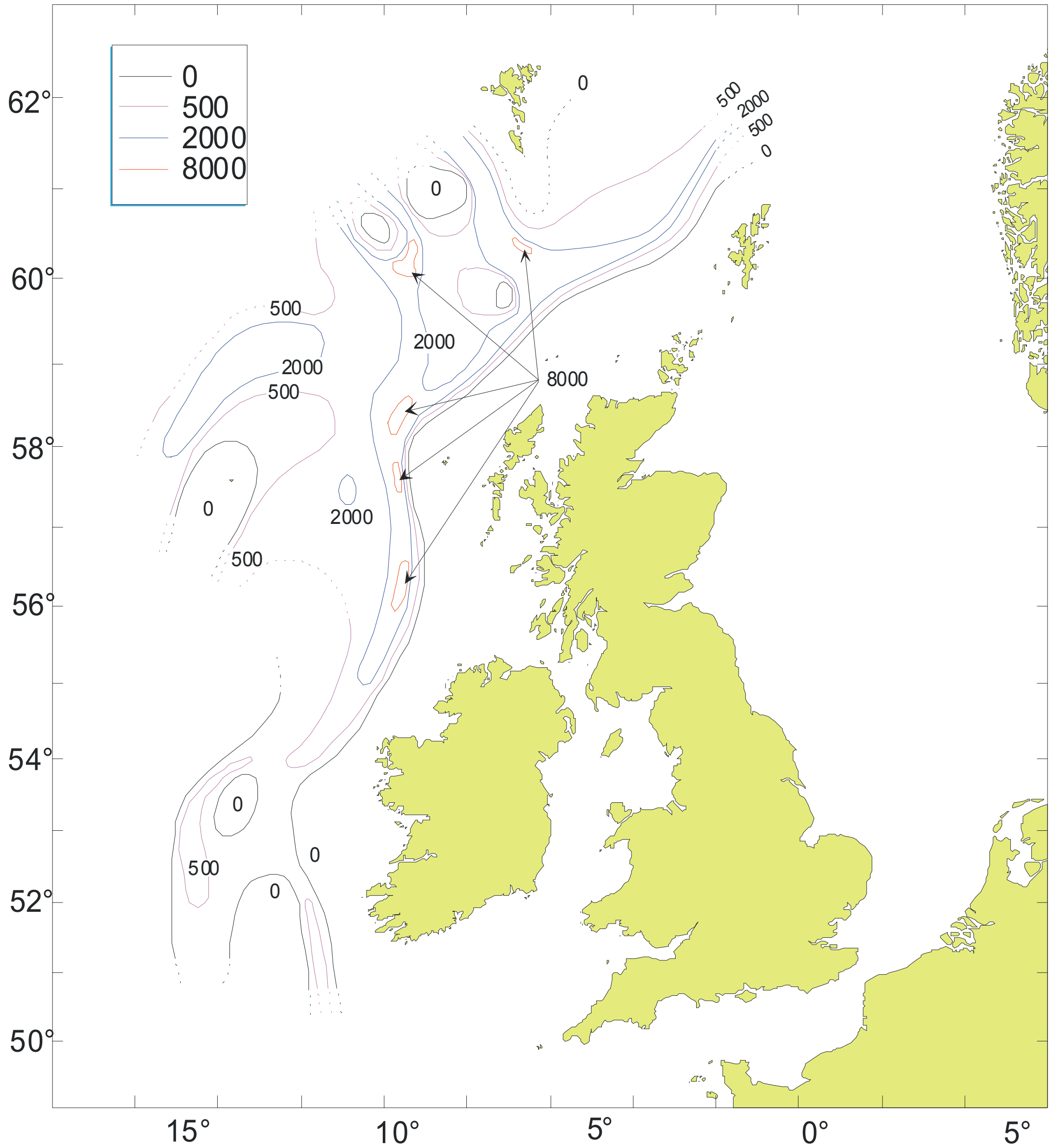


Figure 2.1.2.1. Measured relative densities of blue whiting in April 2003, Norwegian acoustic survey for blue whiting on the spawning grounds.

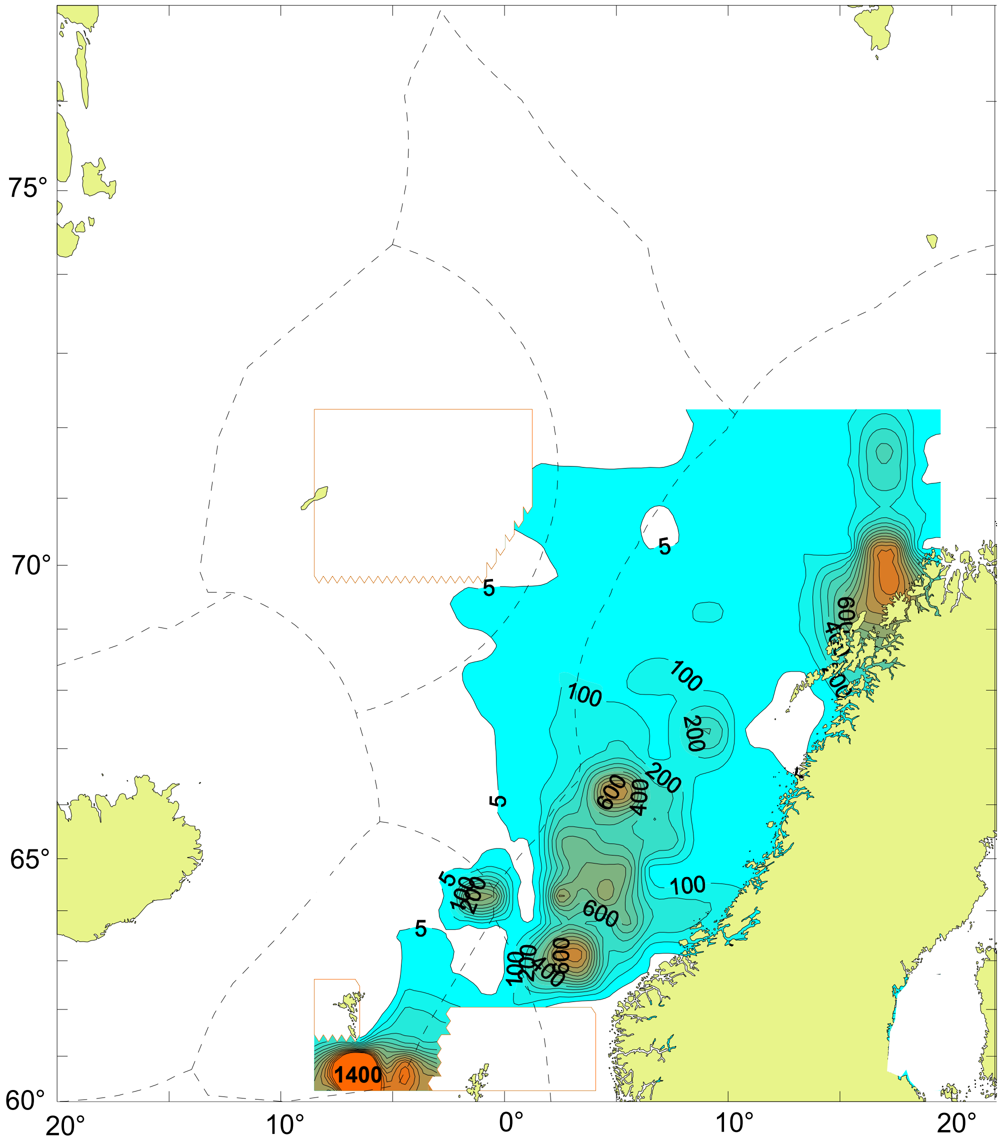


Figure 2.1.2.2. Distribution of blue whiting as observed by R/V "Tridens", R/V "Arni Fridriksson", R/V "G.O.Sars" and R/V "Magnus Heinason" during the international survey in May 2000.

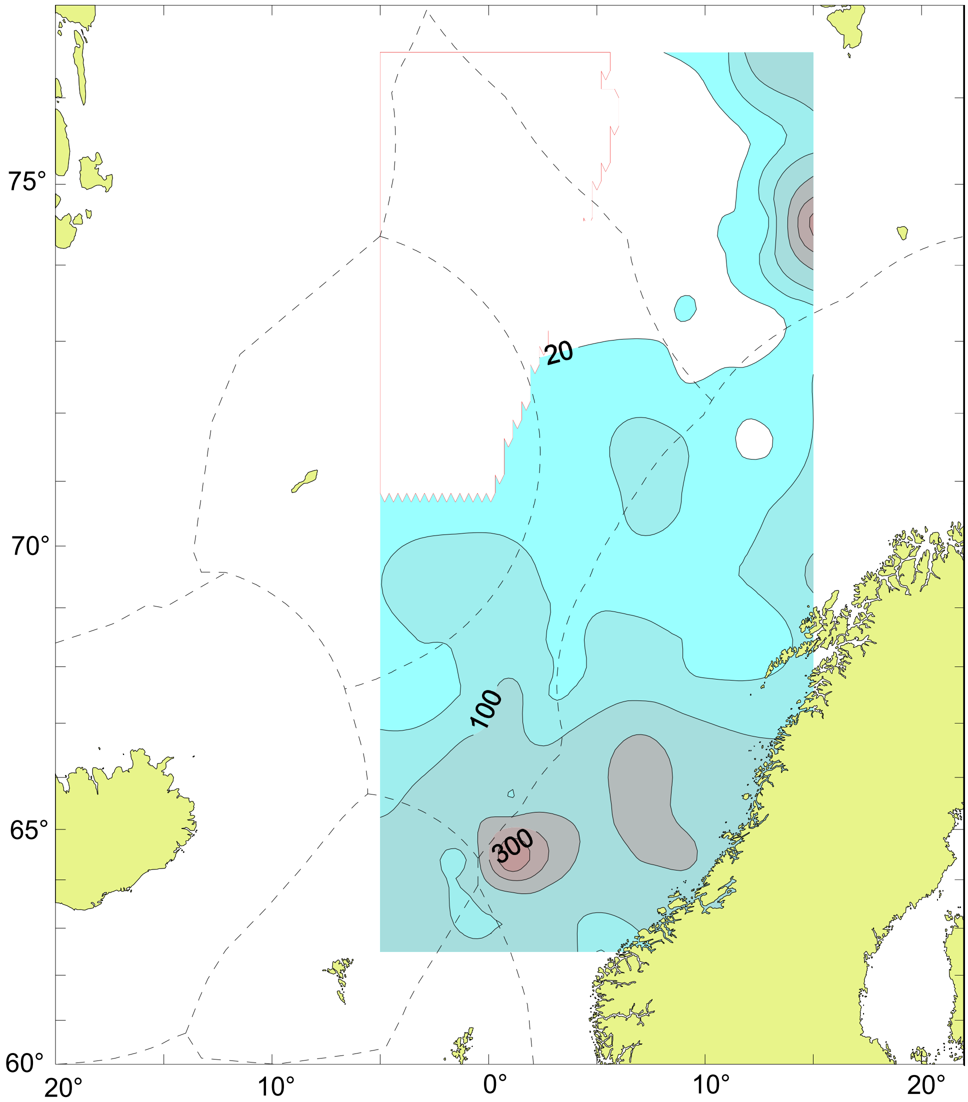


Figure 2.1.2.3. Distribution of blue whiting (Sa values) as observed by G.O.Sars 23rd July – 15th August , 2000.

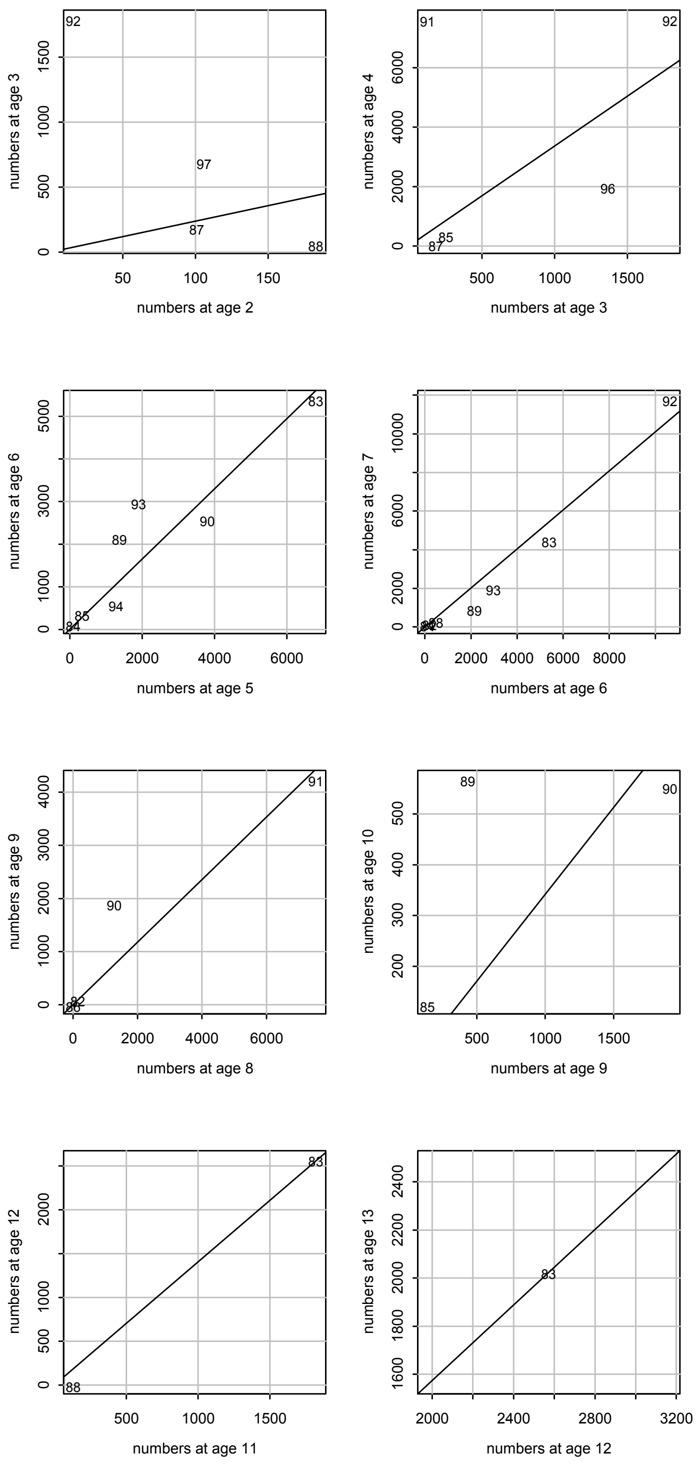


Figure 2.2.1.1 Norwegian Spring Spawn Herring. Correlation within the survey on the feeding grounds in February/March.

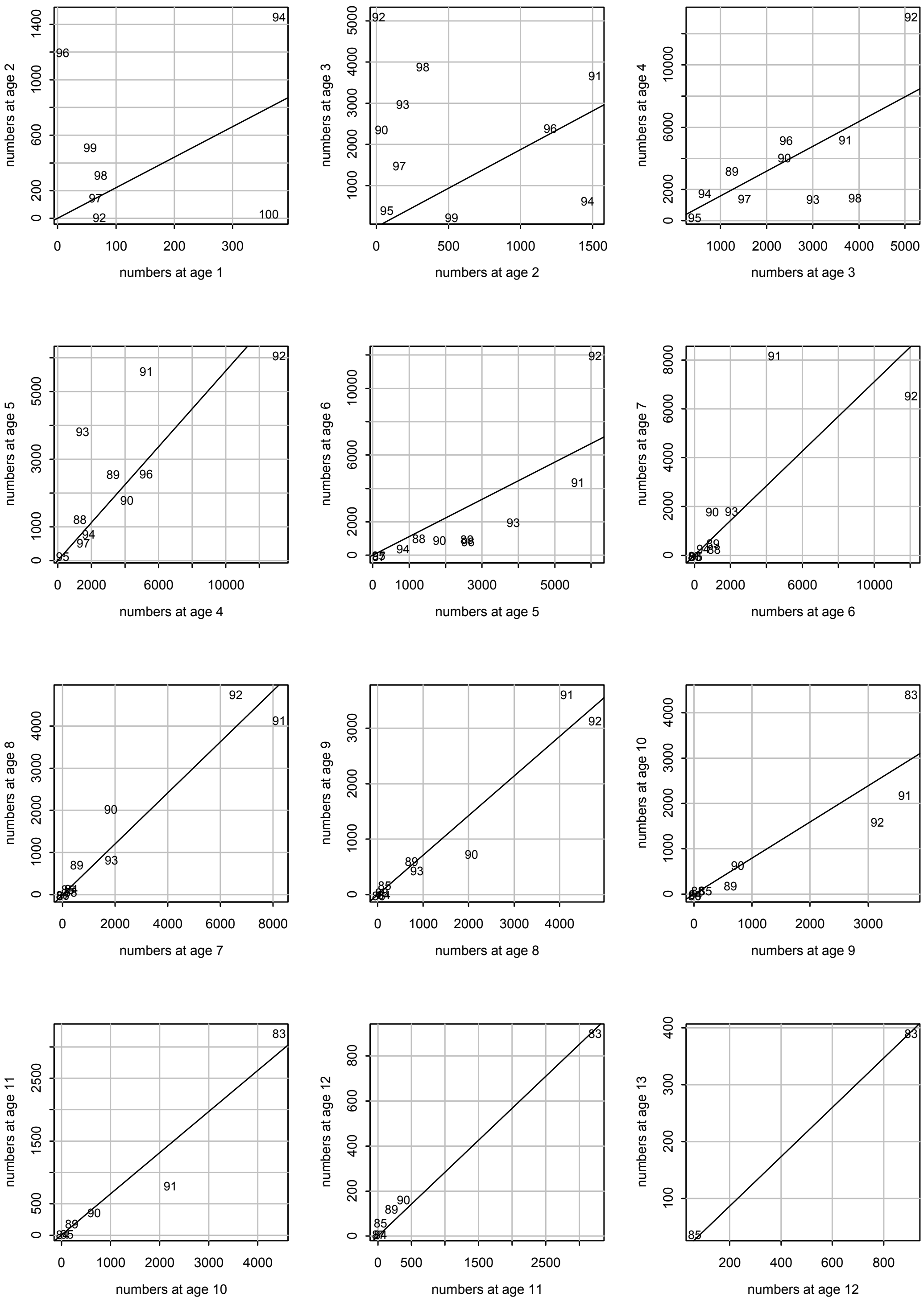


Figure 2.2.1.2 Norwegian Spring Spawn Herring . Correlation within the survey in the wintering area in November/December.

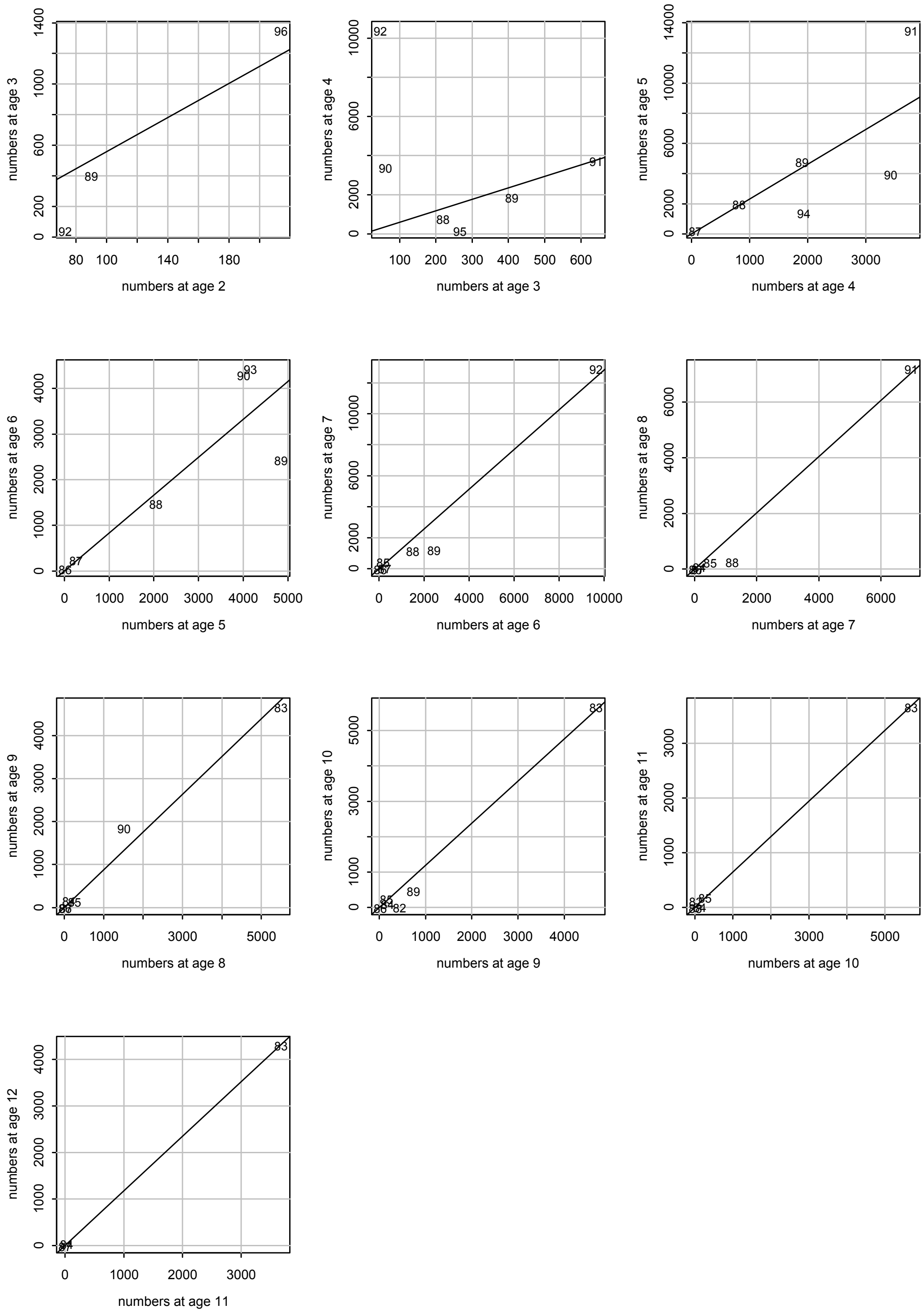


Figure 2.2.1.3 Norwegian Spring Spawn Herring. Correlation within the survey in the wintering area in January.

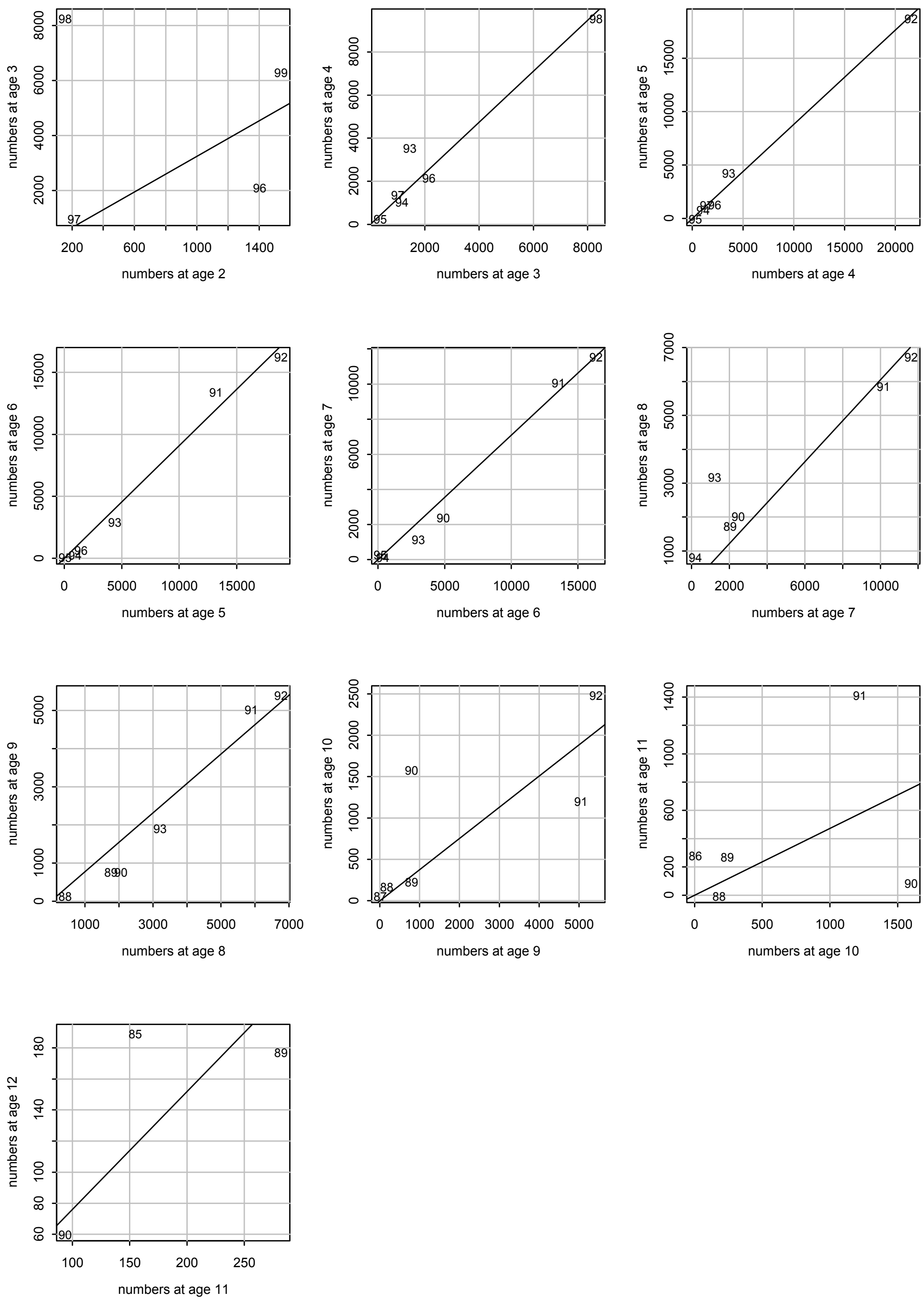


Figure 2.2.1.4 Norwegian Spring Spawn Herring. Correlation within the survey on the feeding grounds in May.

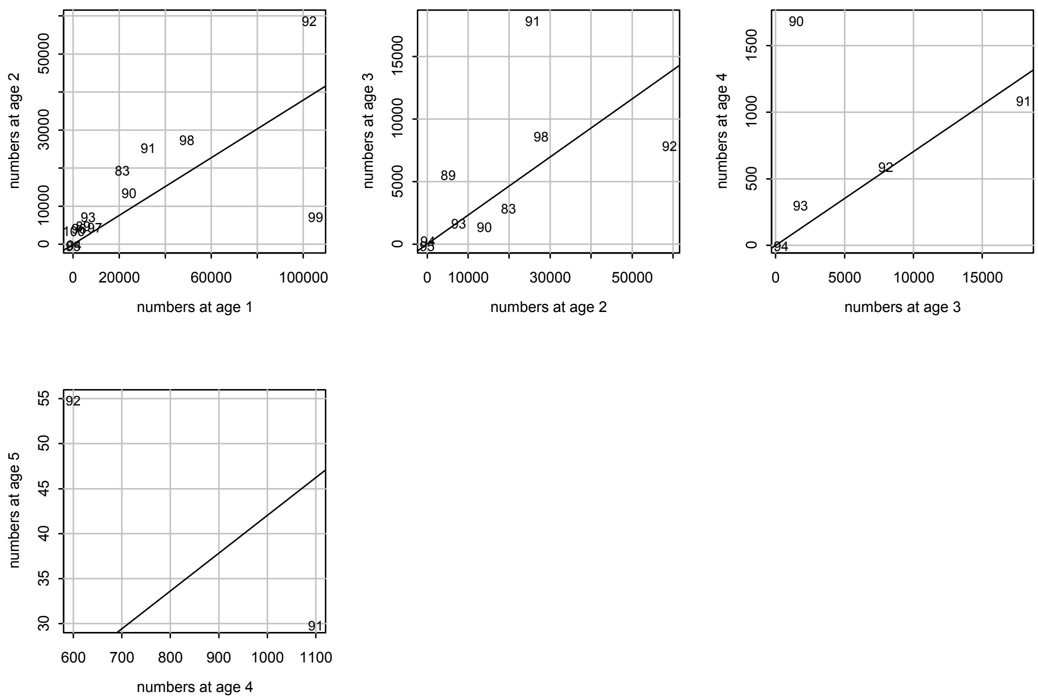


Figure 2.2.1.5 Norwegian Spring Spawn Herring. Correlation within the survey in Barents Sea in May.

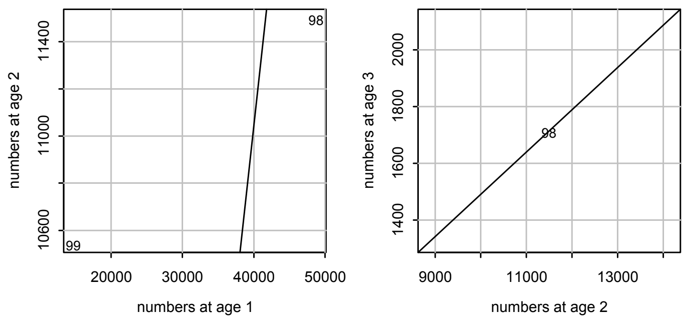


Figure 2.2.1.6 Norwegian Spring Spawn Herring. Correlation within the survey in Barents Sea in September.

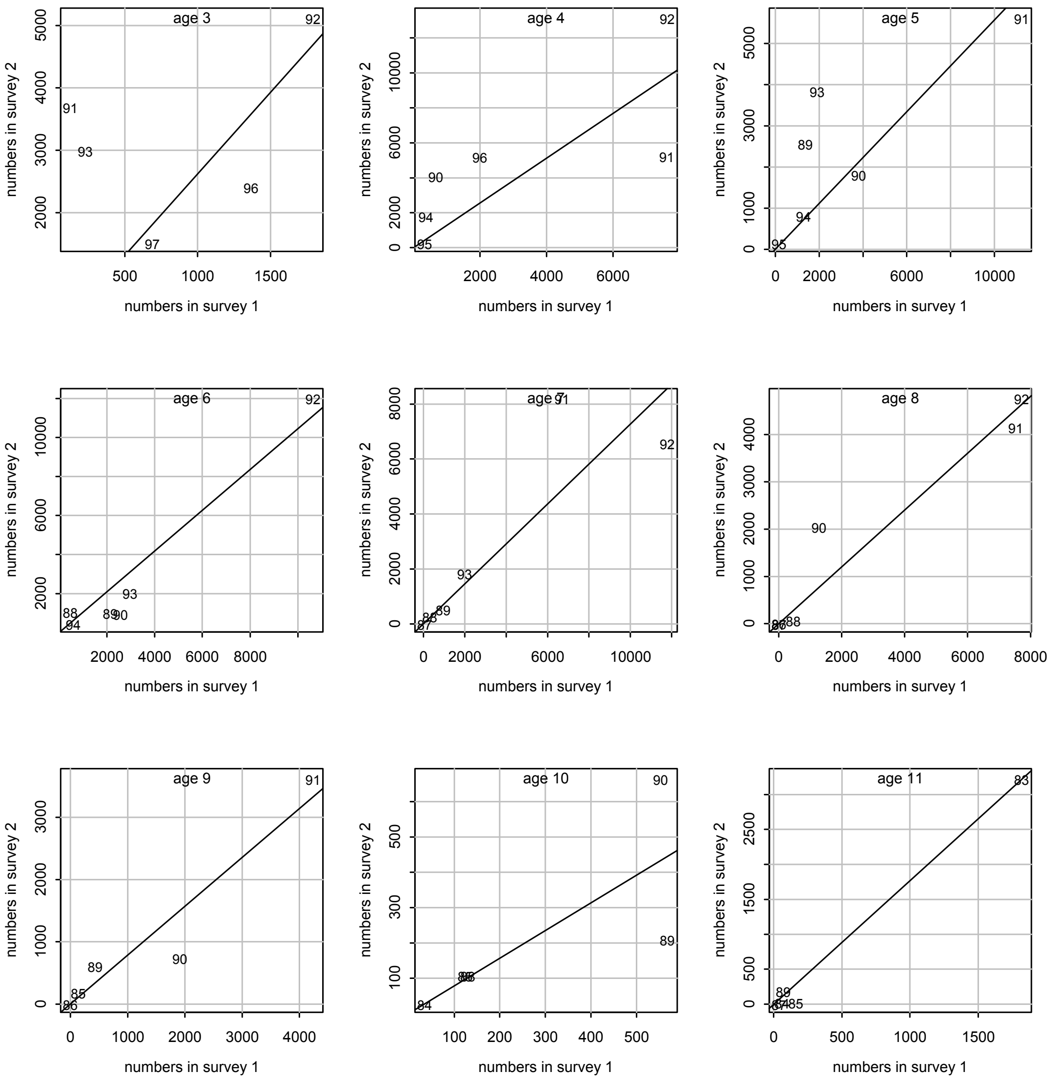


Figure 2.2.1.7 Norwegian Spring Spawn Herring. Correlation between the survey on the spawning grounds (survey1) and the survey in the wintering area in November (survey2).

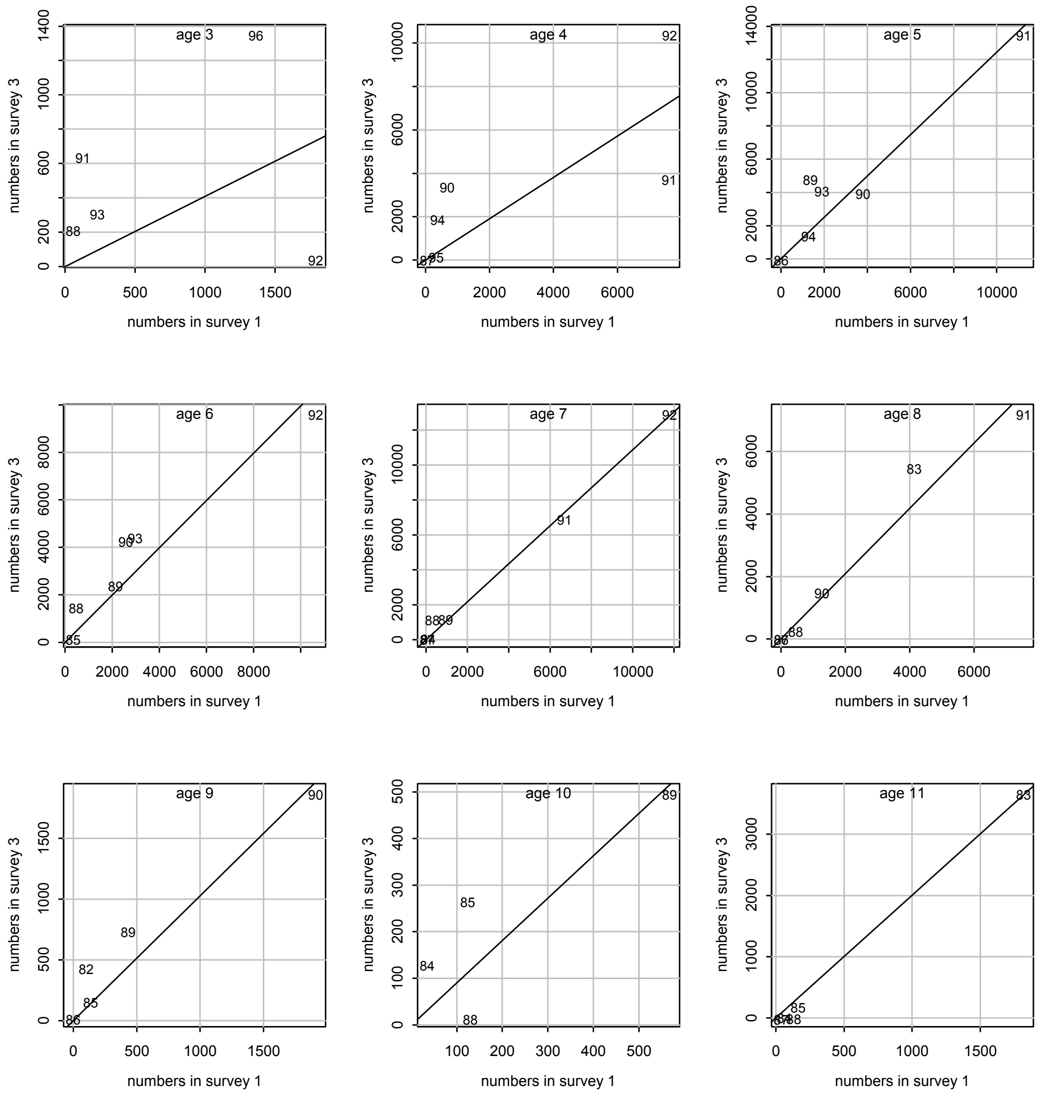


Figure 2.2.1.8 Norwegian Spring Spawn Herring. Correlation between the survey on the spawning grounds (survey1) and the survey in the wintering area in January (survey3).

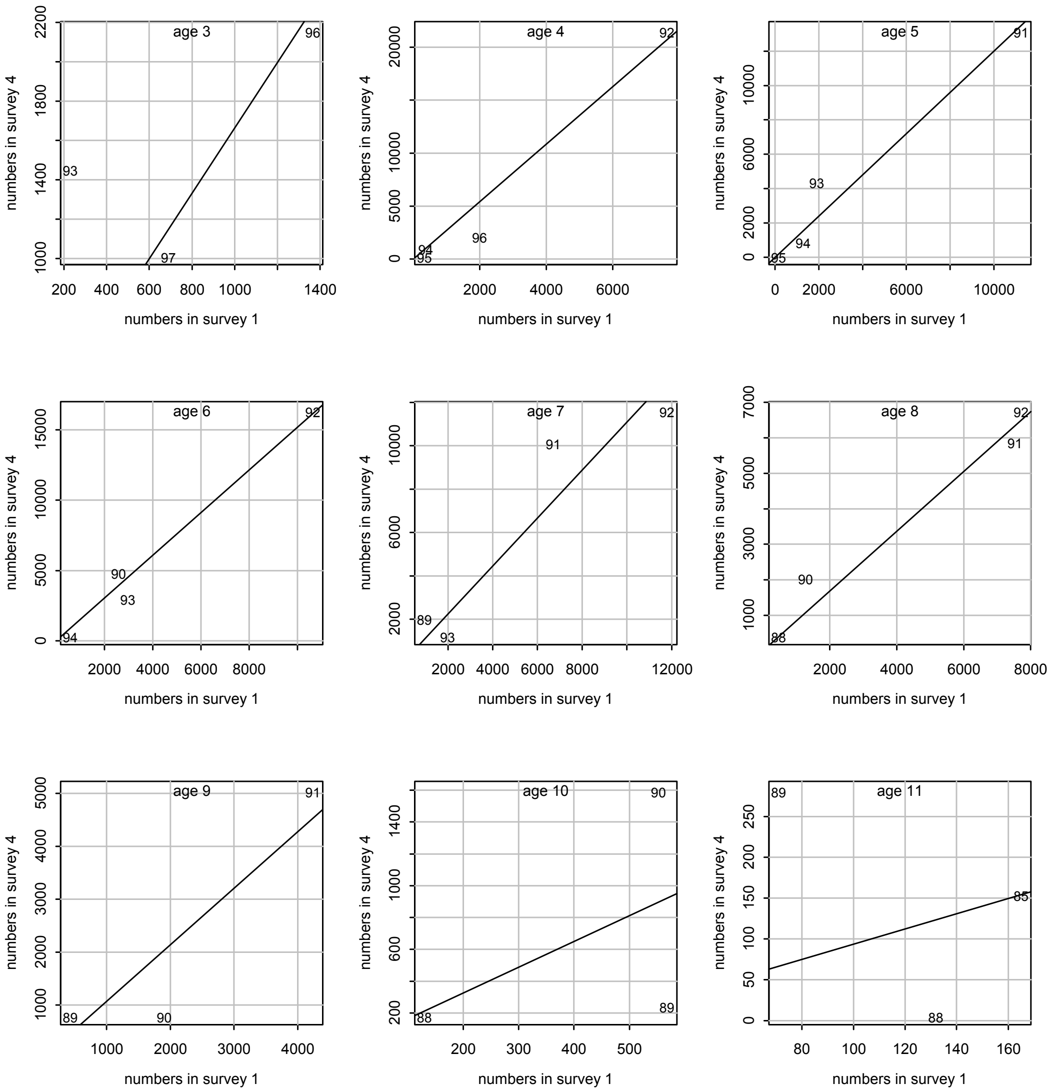


Figure 2.2.1.9 Norwegian Spring Spawn Herring. Correlation between the survey on the spawning grounds (survey1) and the survey on the feeding grounds (survey4).

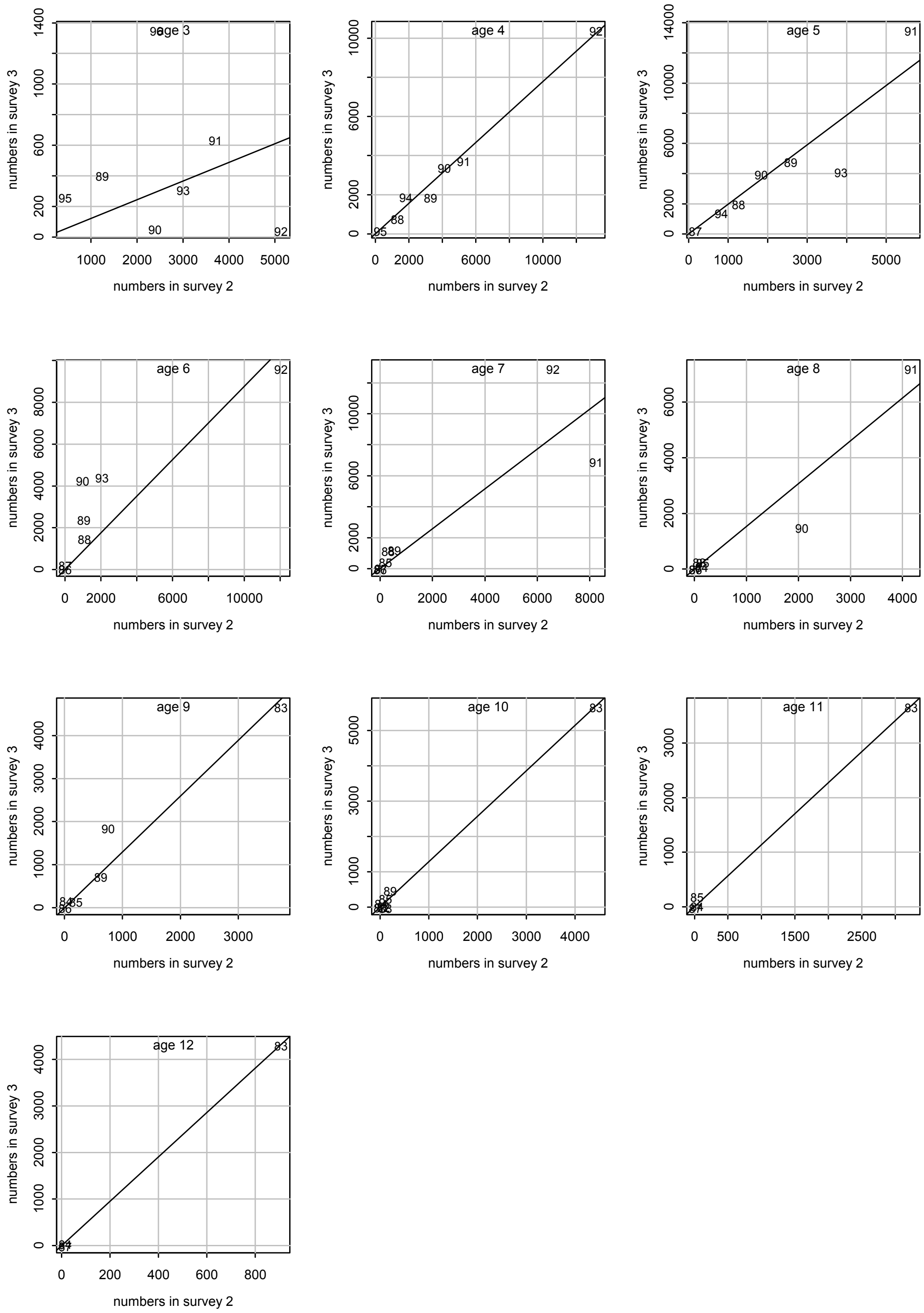


Figure 2.2.1.10 Norwegian Spring Spawn Herring. Correlation between the survey in the wintering are in November (survey2) and the survey in the wintering area in January (survey3).

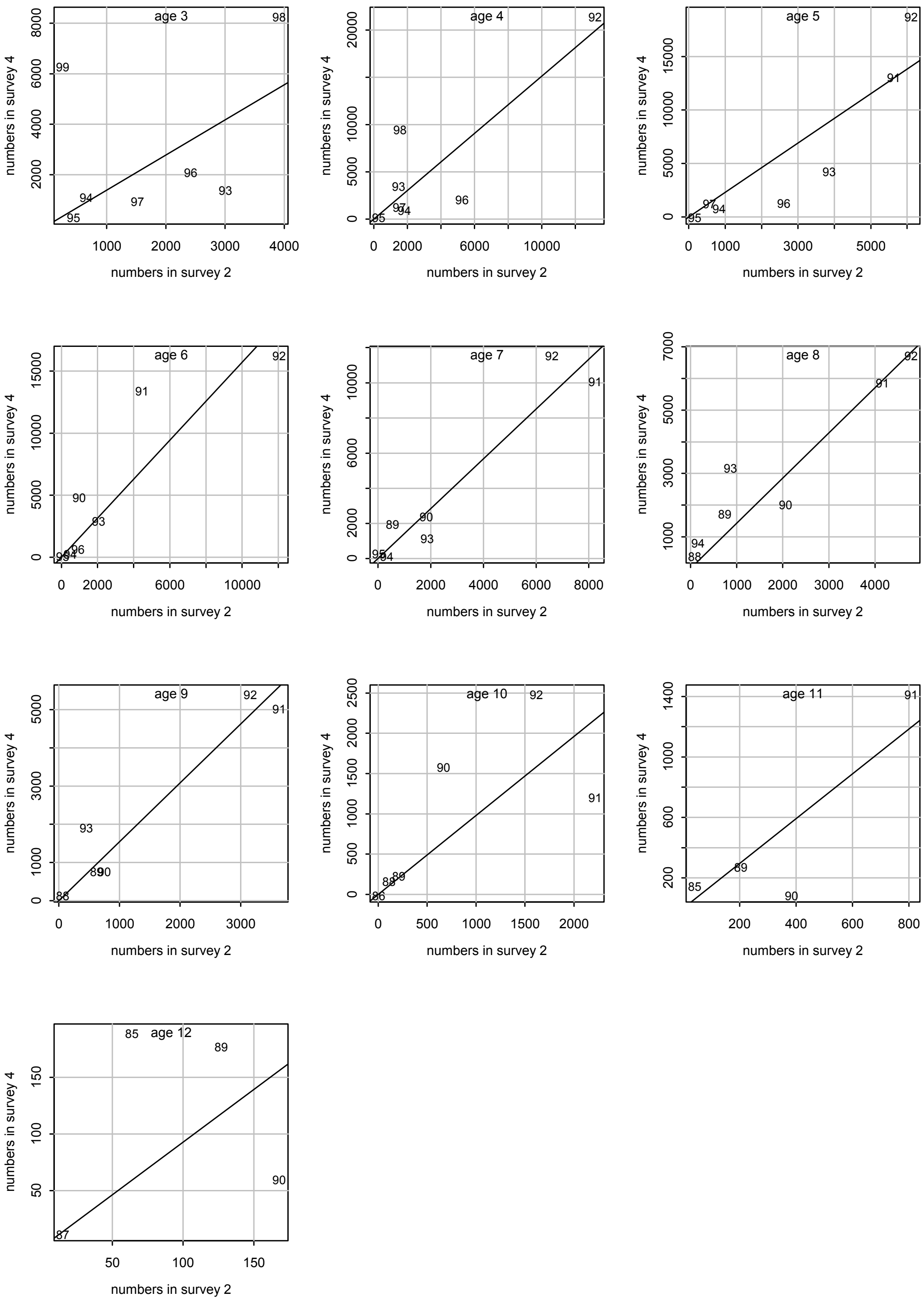


Figure 2.2.1.11 Norwegian Spring Spawn Herring. Correlation between the survey in the wintering area in November (survey2) and the survey on the feeding grounds (survey4).

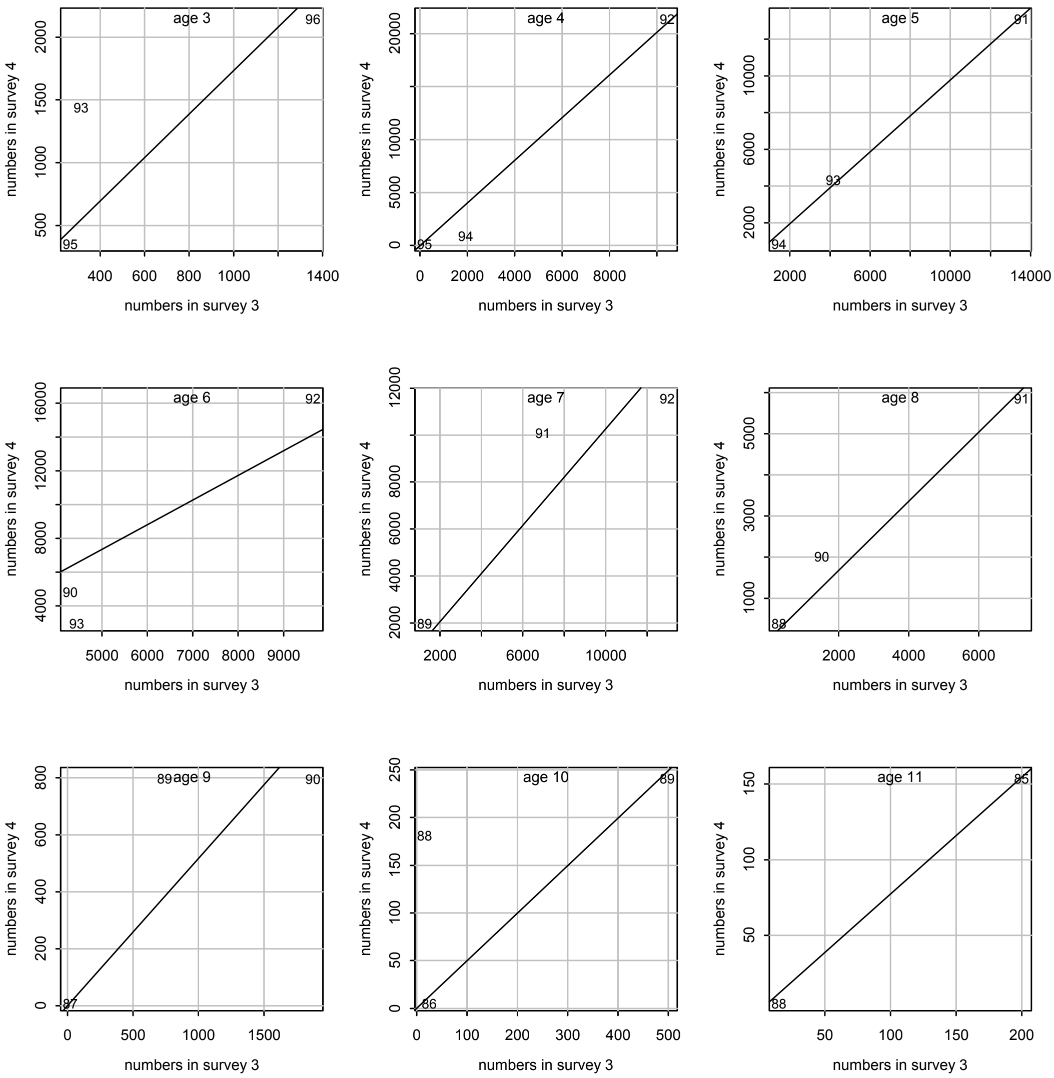


Figure 2.2.1.12 Norwegian Spring Spawn Herring. Correlation between the survey in the wintering are in January (survey3) and the survey on the feeding grounds (survey4).

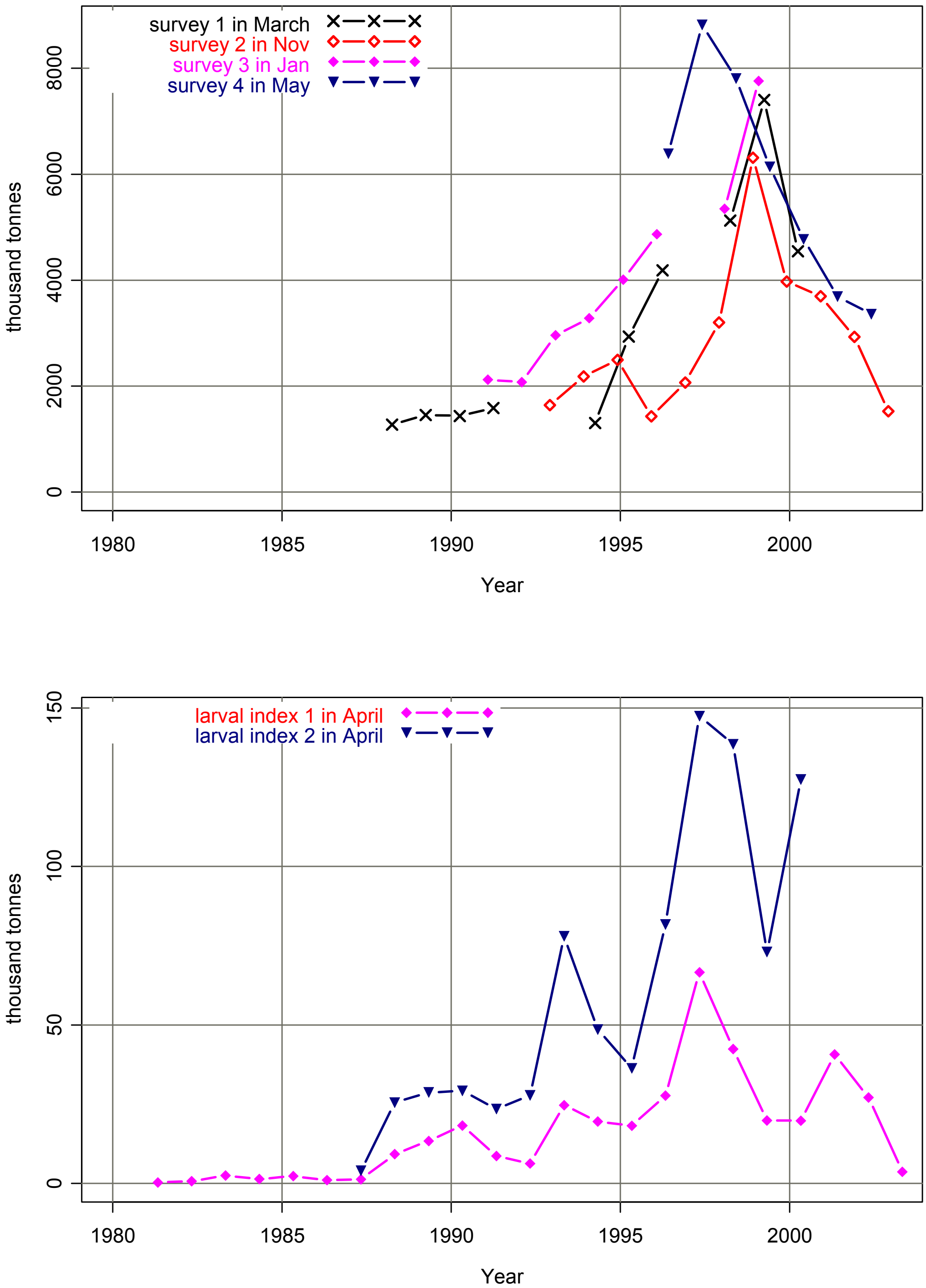


Figure 2.2.1.13 Norwegian Spring Spawn Herring. Biomass indices.

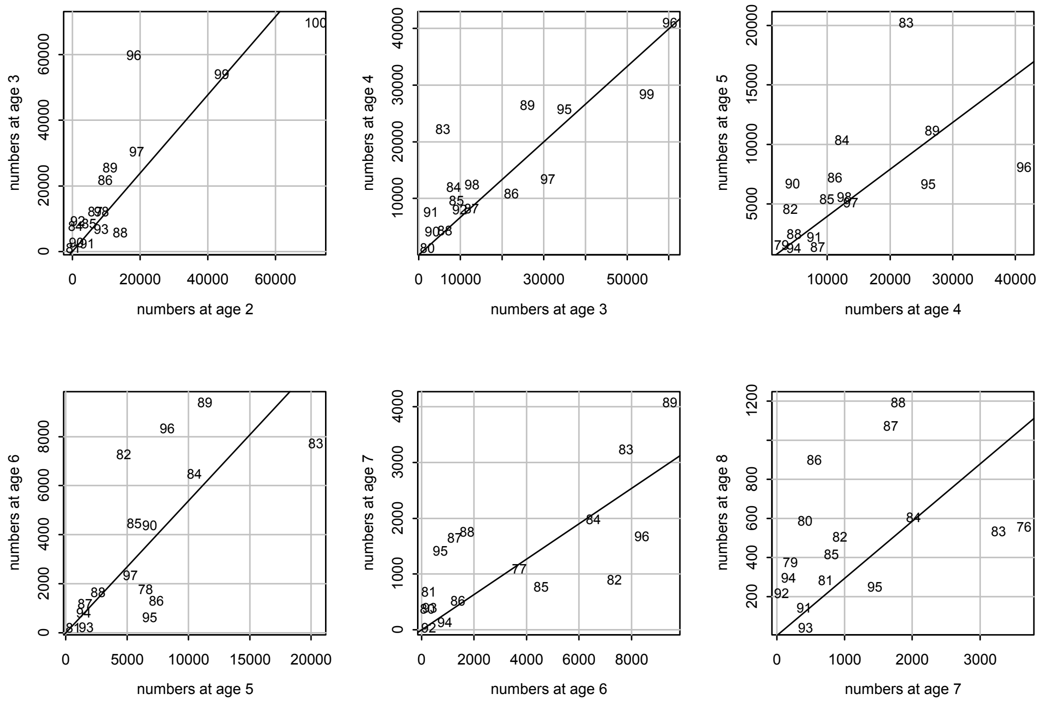


Figure 2.2.2.1 Blue whiting. Correlation within the Norwegian survey on the spawning grounds.

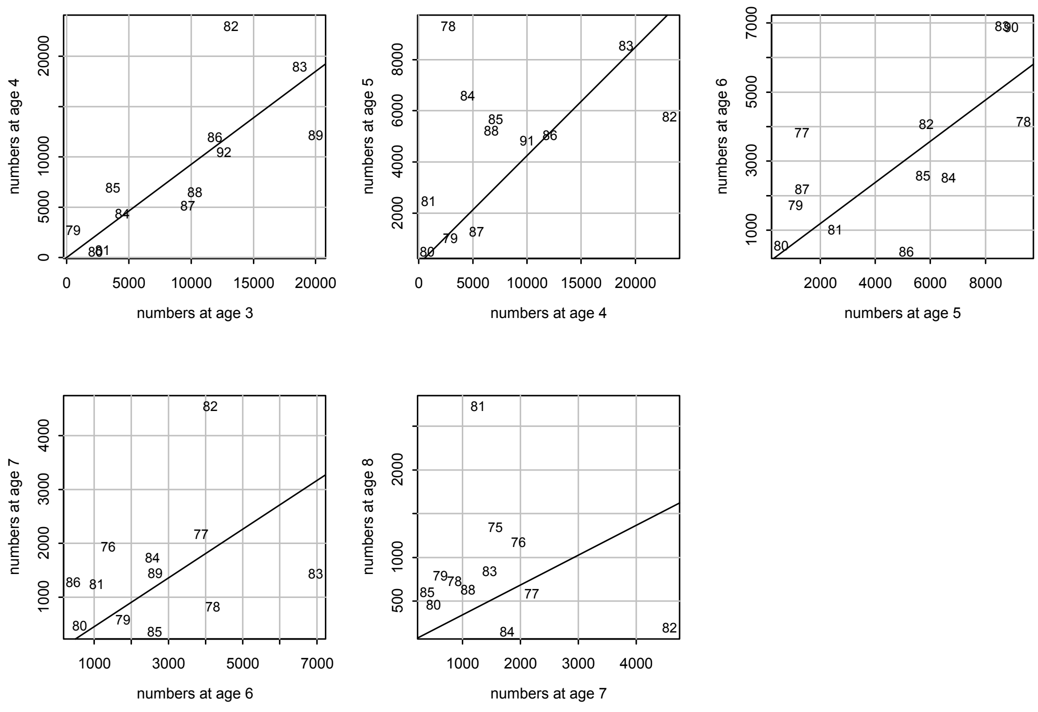


Figure 2.2.2.2 Blue whiting. Correlation within the Russian survey on the spawning grounds.

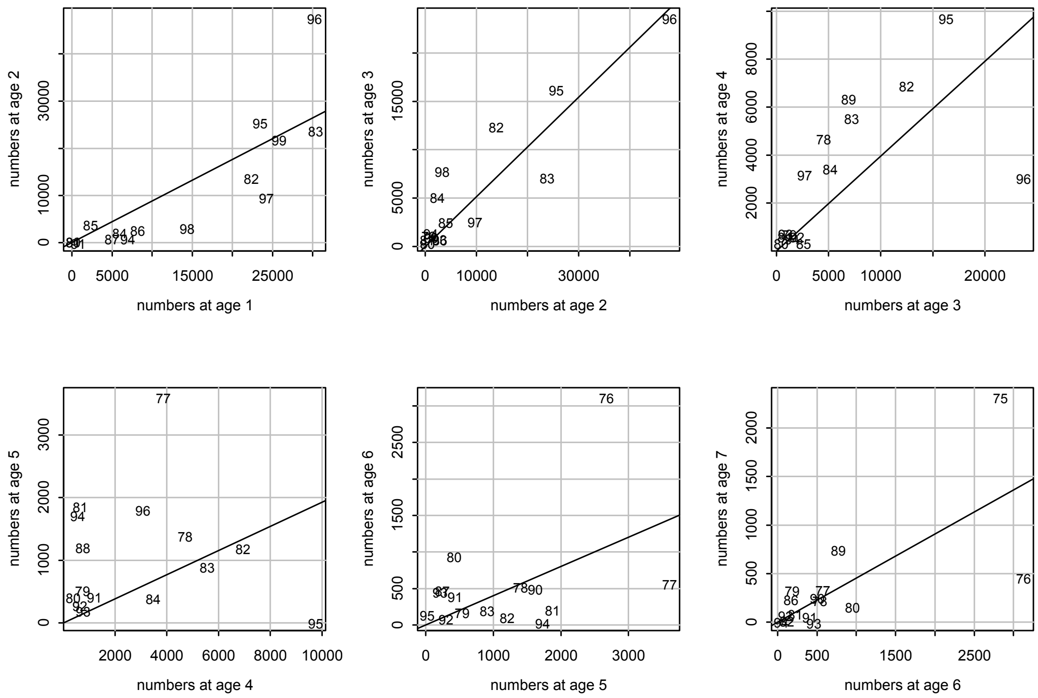


Figure 2.2.2.3 Blue whiting. Correlation within the Norwegian summer survey in the Norwegian Sea.

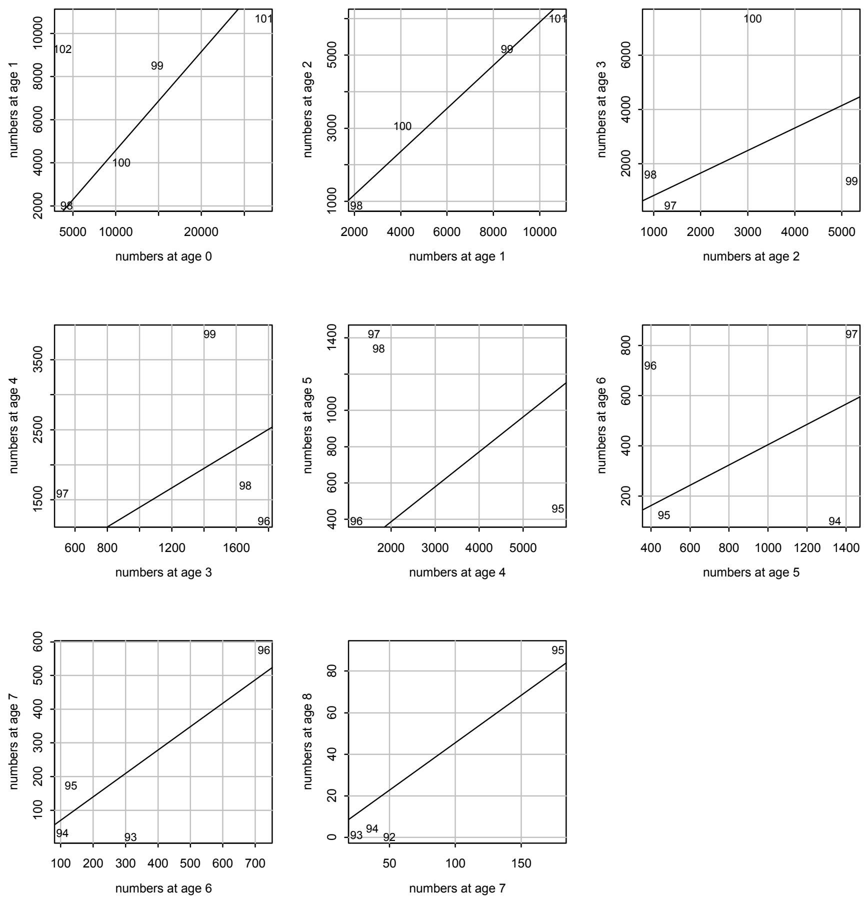


Figure 2.2.2.4 Blue whiting. Correlation within the Icelandic blue whiting survey in summer.

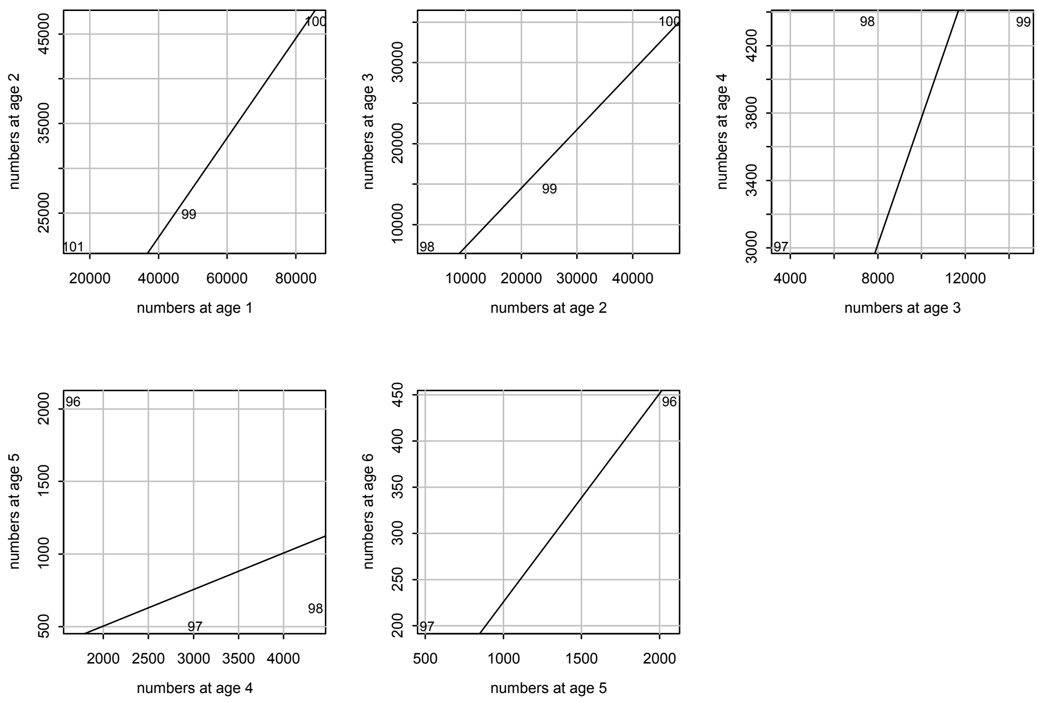


Figure 2.2.2.5 Blue whiting . Correlation within the International survey on the feeding grounds in the Norwegian Sea in May.

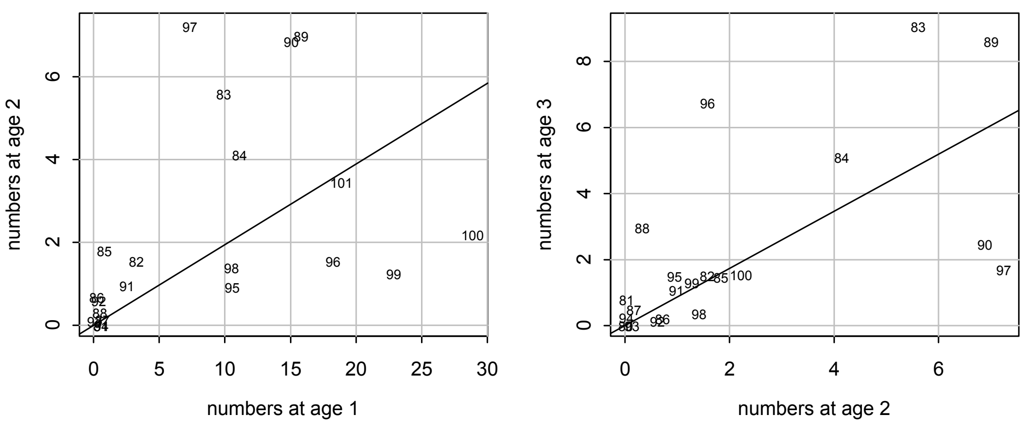


Figure 2.2.2.6. Blue whiting. Correlation within the Norwegian winter survey in the Barents Sea.

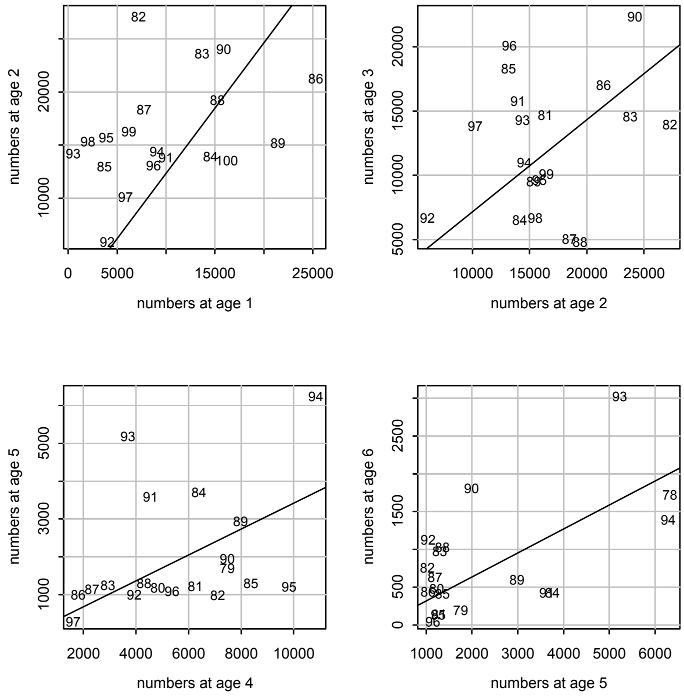


Figure 2.2.2.7 Blue whiting. Correlation within the Spanish pair trawl series (CPUE).

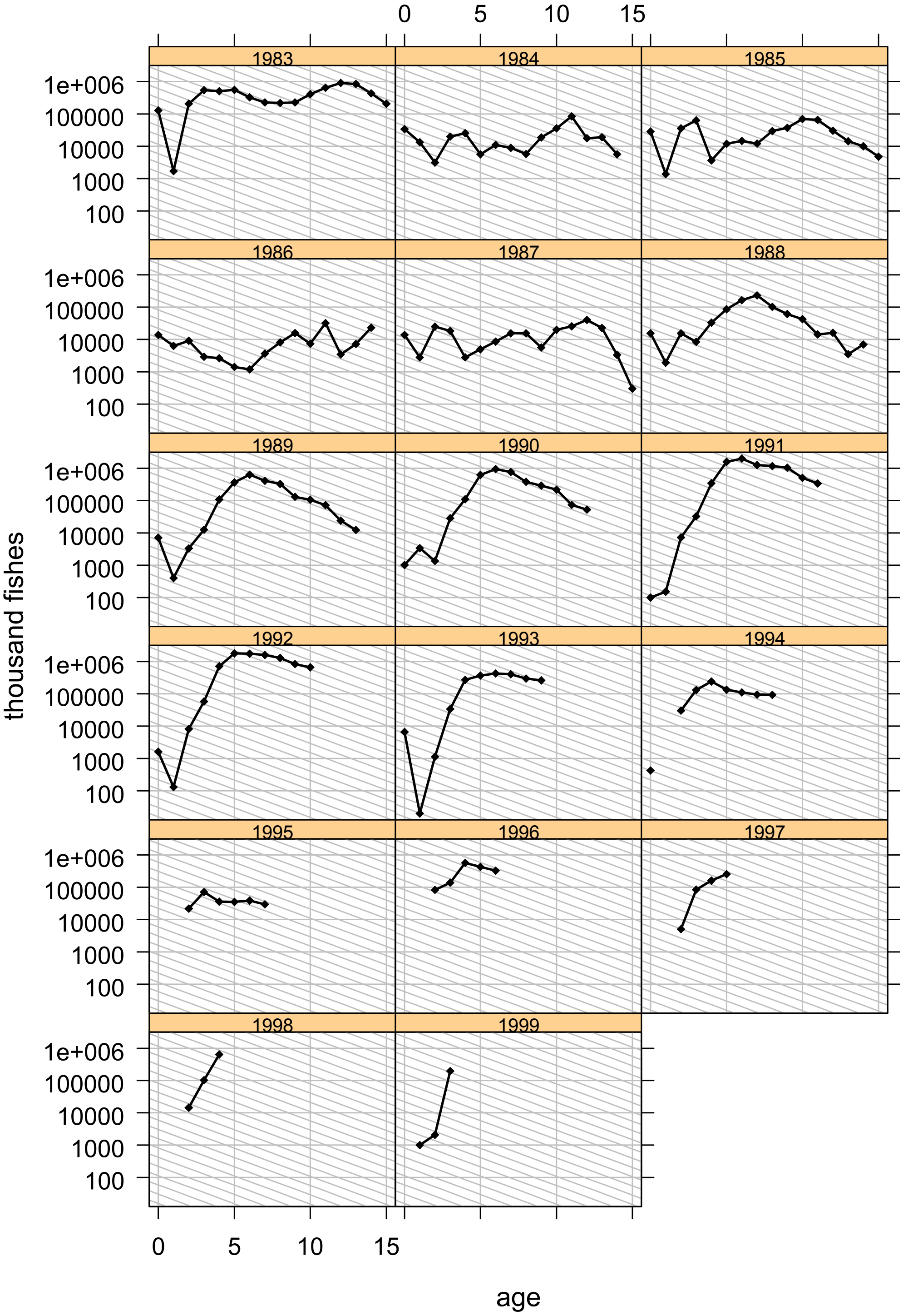


Figure 2.3.1.1 Norwegian spring spawn herring. Catch curves by yearclasses and age from the catch. Diagonal gray lines correspond to Z=0.4.

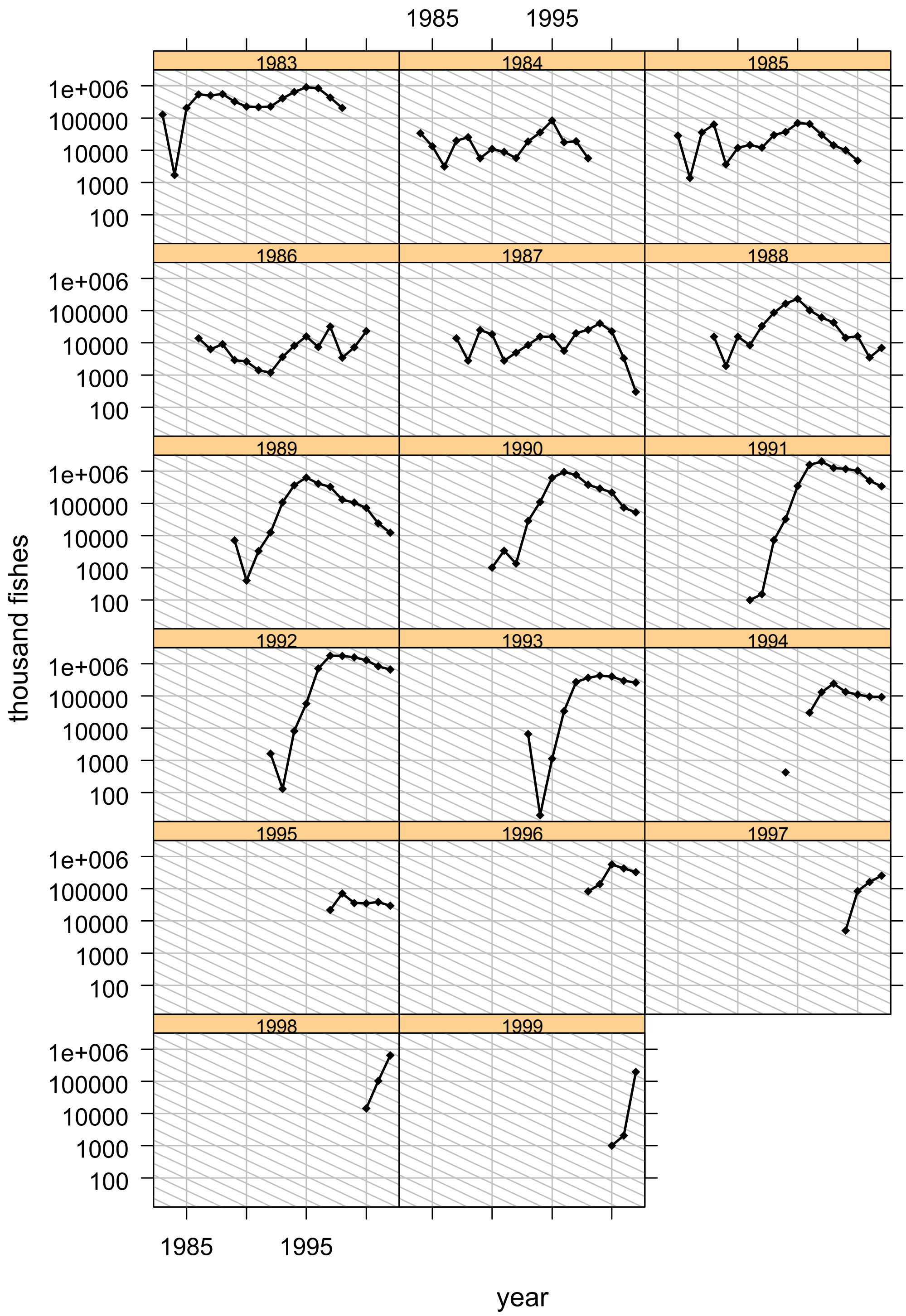


Figure 2.3.1.2 Norwegian spring spawn herring. Catch curves by year classes and years from the catch. Diagonal gray lines correspond to Z=0.4.

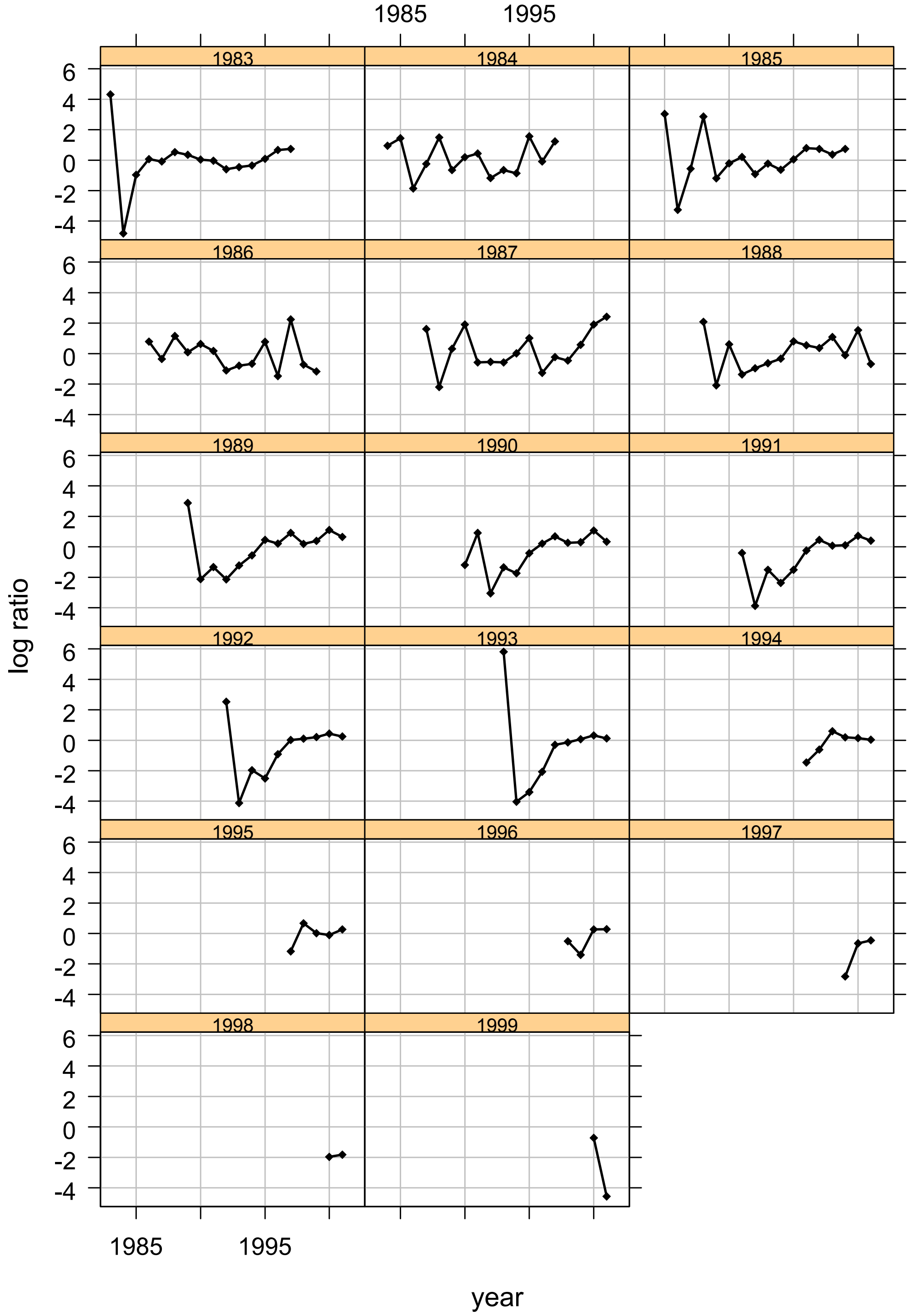


Figure 2.3.1.3 Log ratio (log(Cay/Ca+1,y+a) for Norwegian spring spawn herring based on catch data.

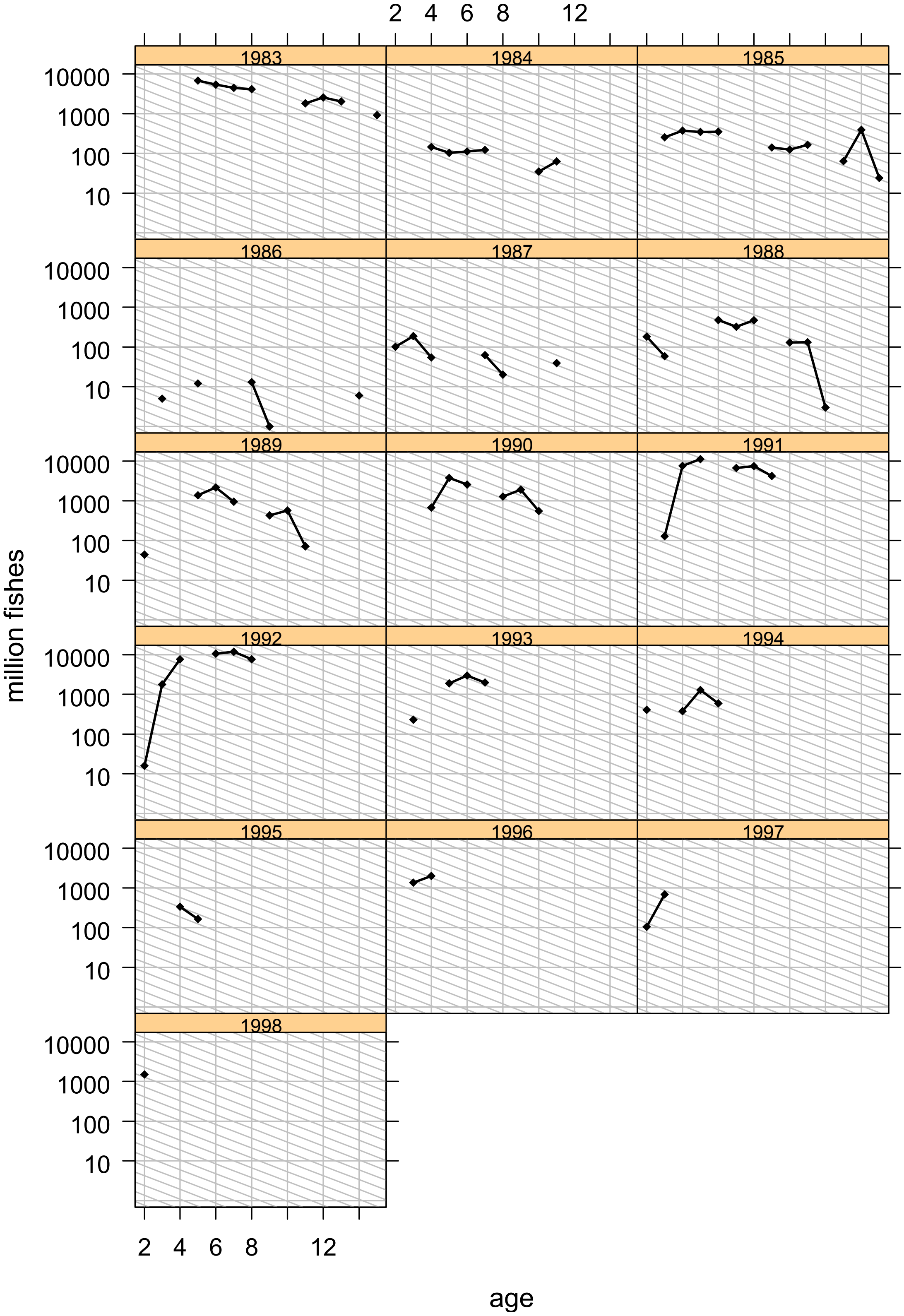


Figure 2.3.1.4. Catch curves for Norwegian spring spawn herring based on survey 1, acoustic survey on the spawning grounds in February/March. Diagonal gray lines correspond to Z=0.4.

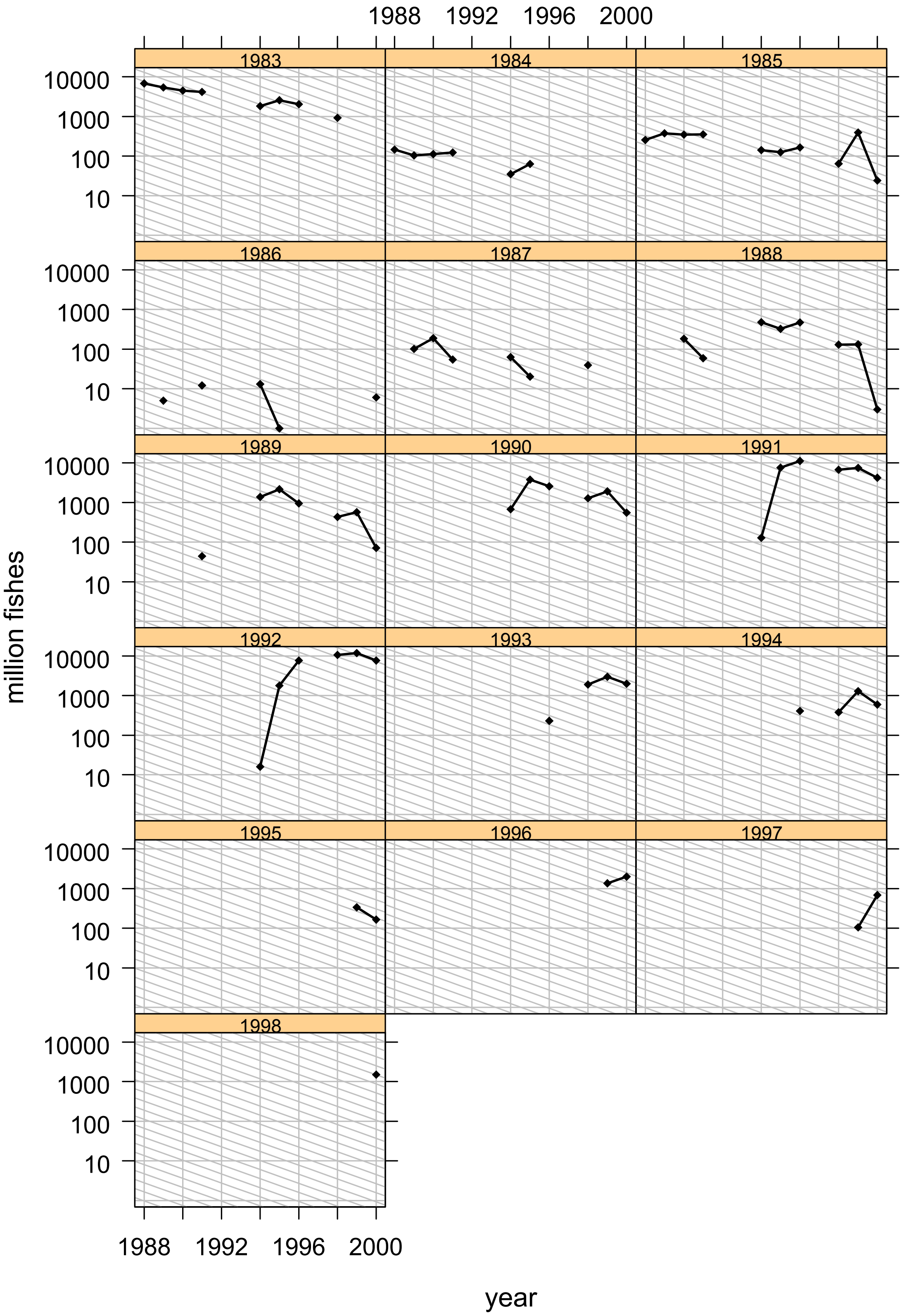


Figure 2.3.1.5. Catch curves for Norwegian spring spawn herring based on survey 1, acoustic survey on the spawning grounds in February/March. Diagonal gray lines correspond to Z=0.4.

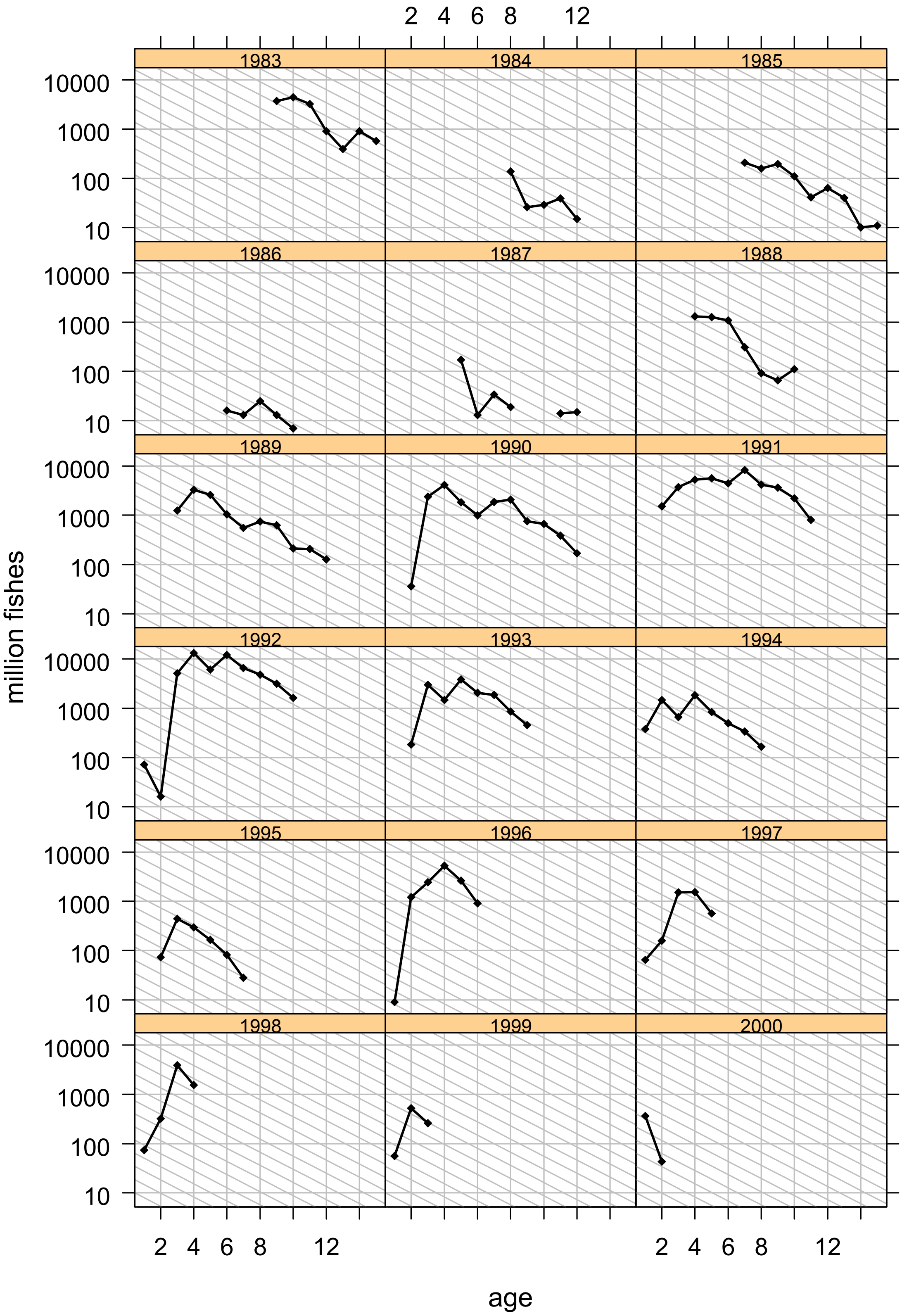


Figure 2.3.1.6. Catch curves for Norwegian spring spawn herring based on survey 2, acoustic survey in the wintering area in November/December. Diagonal gray lines correspond to Z=0.4.

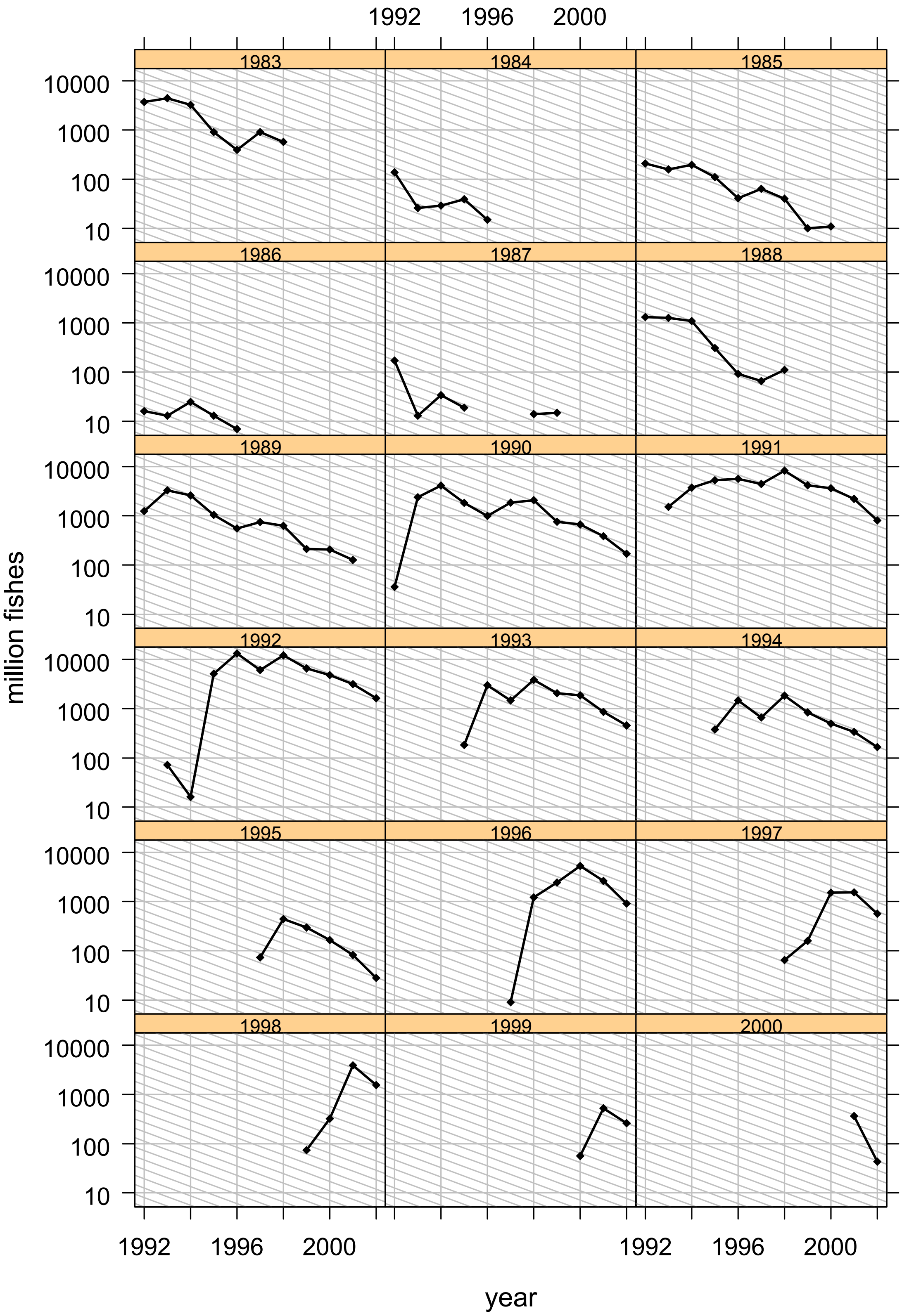


Figure 2.3.1.7 Catch curves for Norwegian spring spawn herring based on survey 2, acoustic survey in the wintering area in November/December. Diagonal gray lines correspond to Z=0.4.

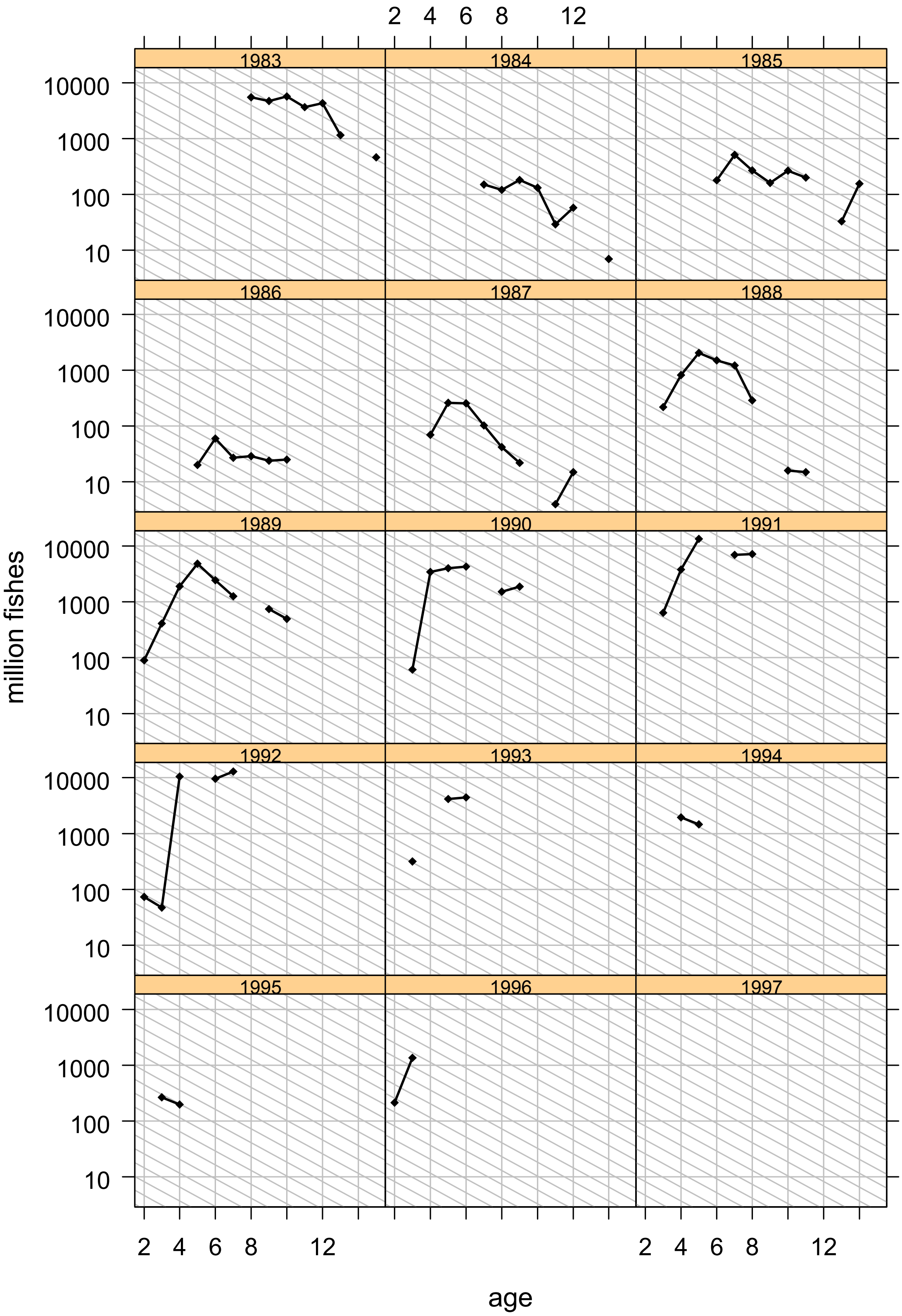


Figure 2.3.1.8 Catch curves for Norwegian spring spawn herring based on survey 3, acoustic survey in the wintering area in January. Diagonal gray lines correspond to Z=0.4.

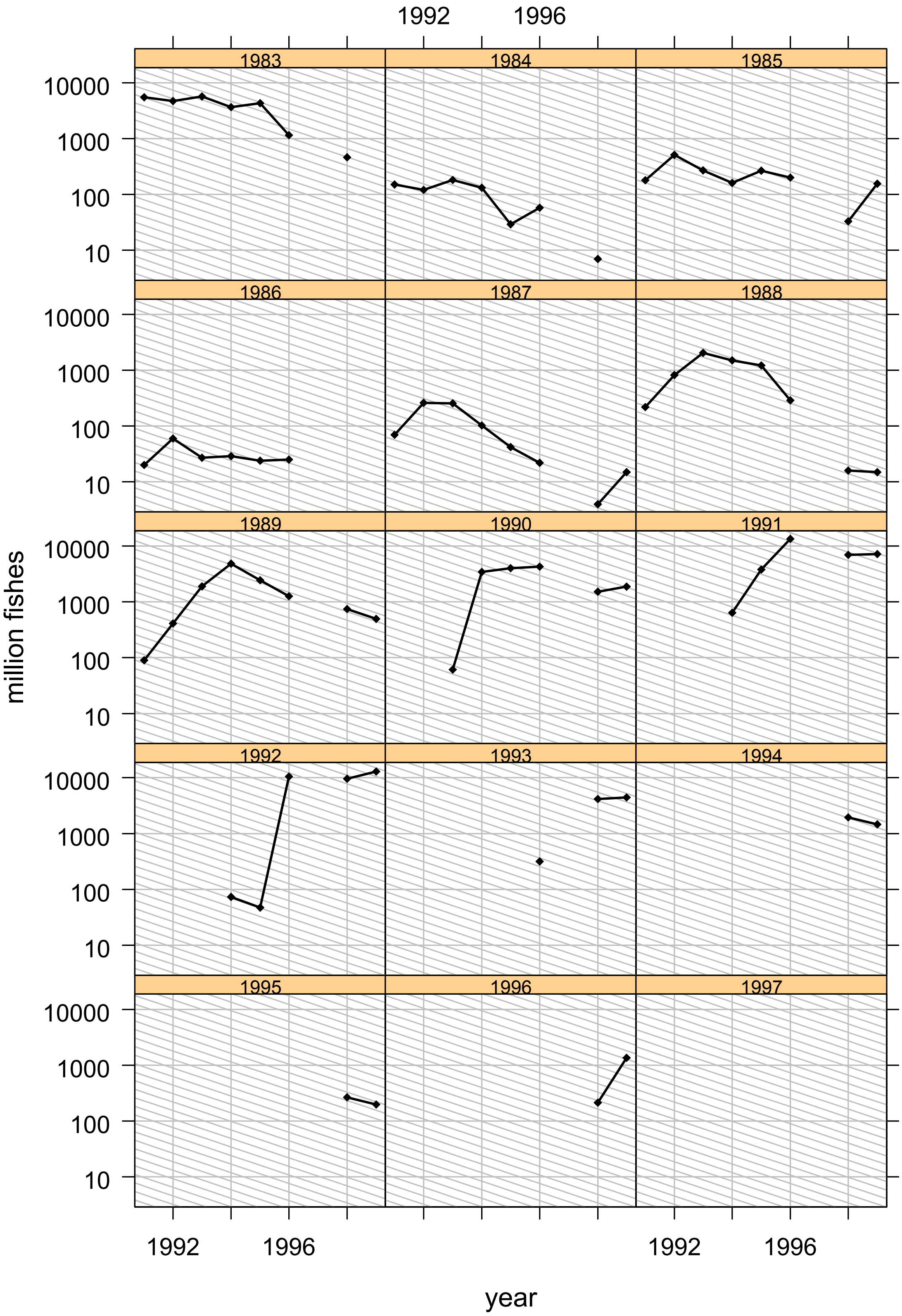


Figure 2.3.1.9 Catch curves for Norwegian spring spawn herring based on survey 3, acoustic survey in the wintering area in January. Diagonal gray lines correspond to Z=0.4.

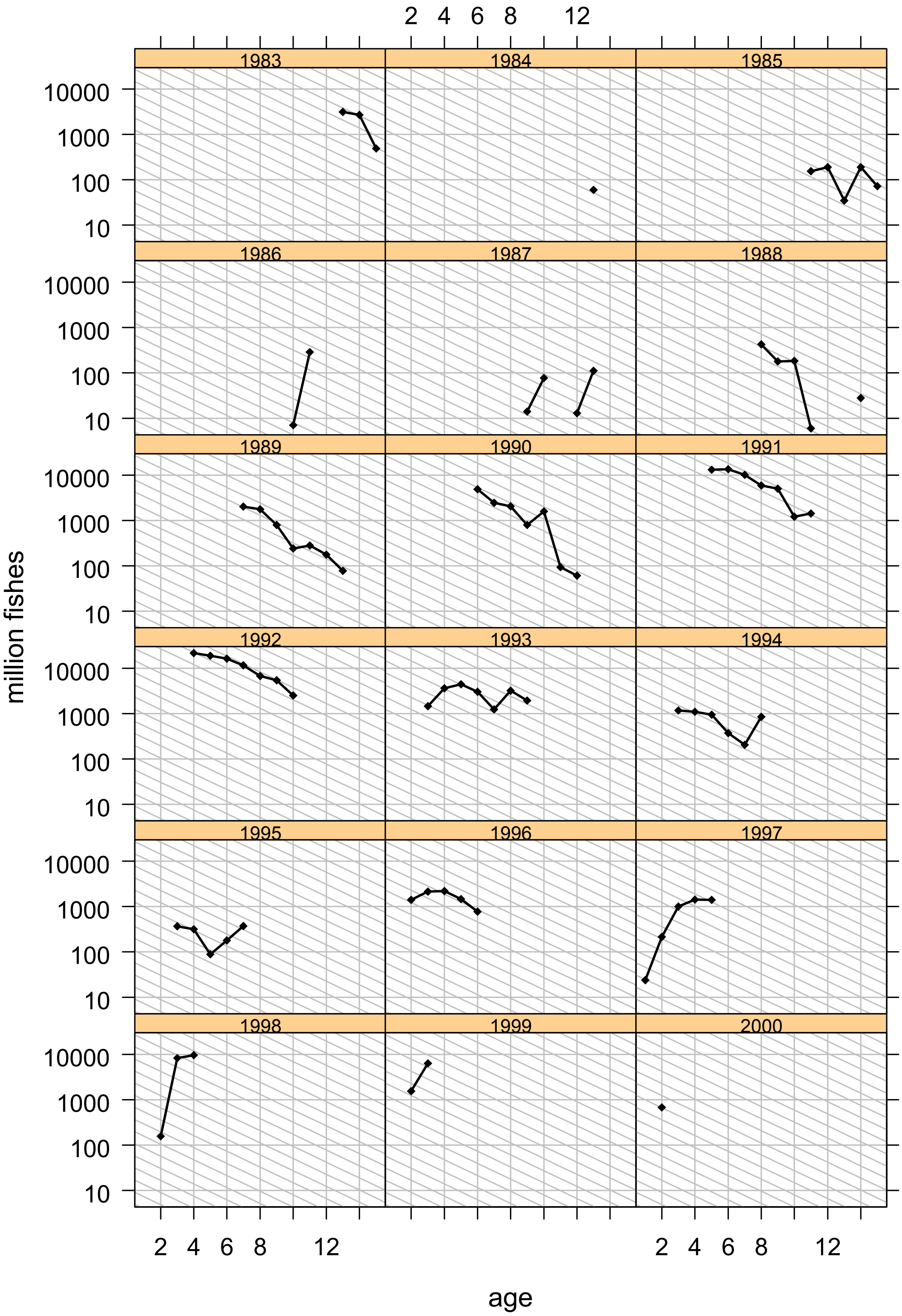


Figure 2.3.1.10 Catch curves for Norwegian spring spawn herring based on survey 4, acoustic survey on the feeding grounds in the Norwegian Sea in May. Diagonal gray lines correspond to Z=0.4.

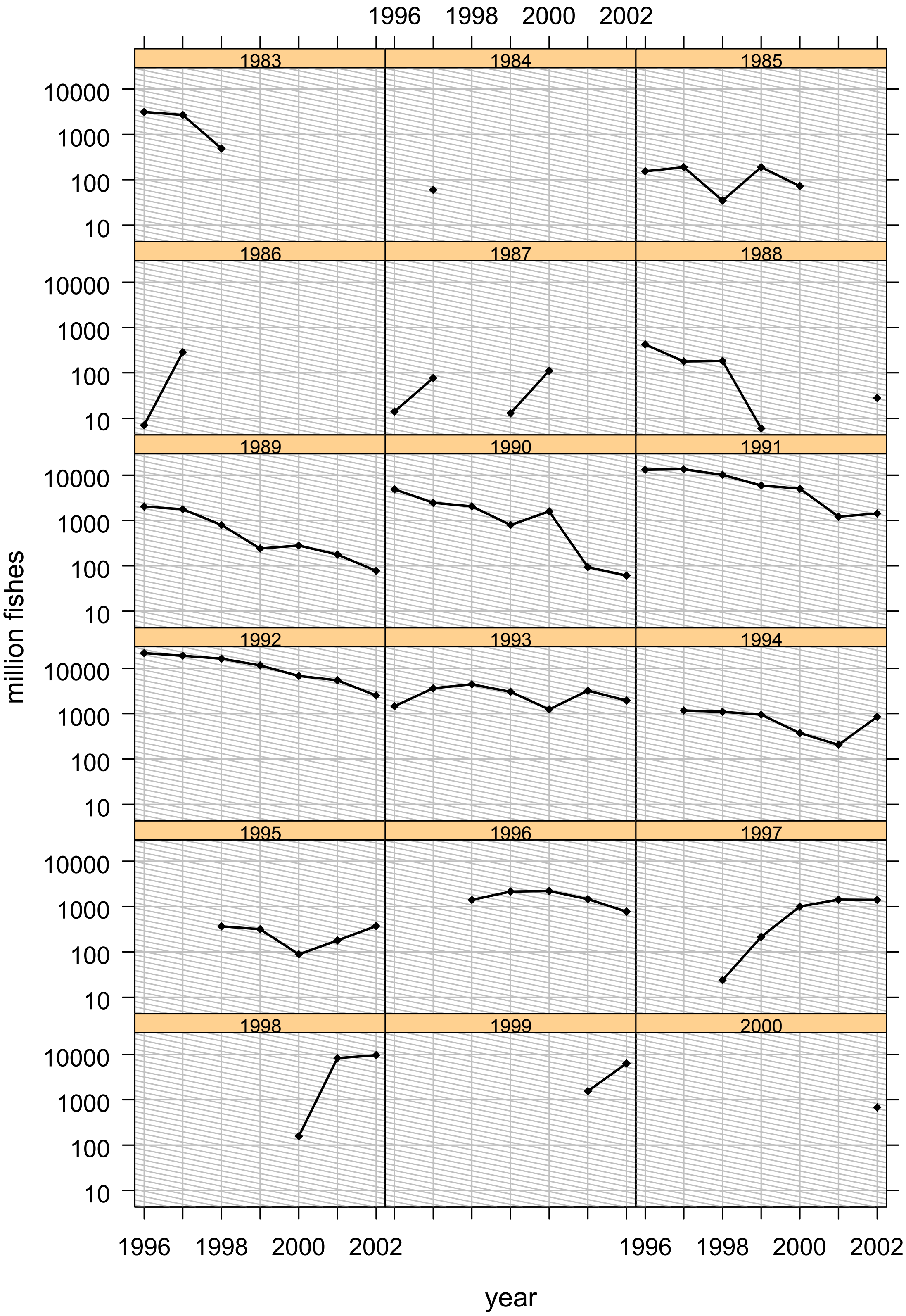


Figure 2.3.1.11 Catch curves for Norwegian spring spawn herring based on survey 4, acoustic survey on the feeding grounds in the Norwegian Sea in May. Diagonal gray lines correspond to Z=0.4.

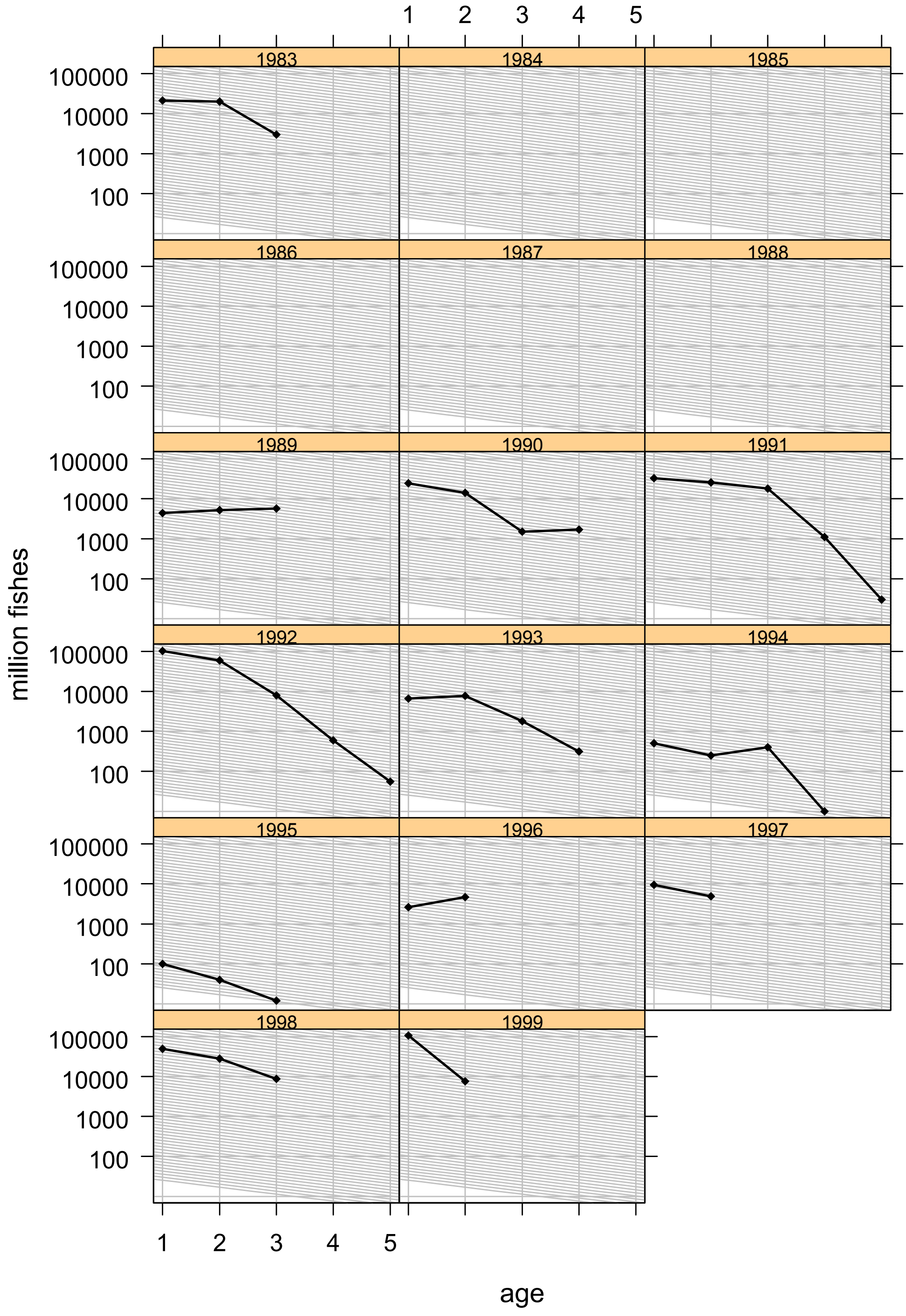


Figure 2.3.1.12 Catch curves for Norwegian spring spawn herring based on survey 5, acoustic survey in the Barents Sea in May. Gray lines correspond to Z=0.0.

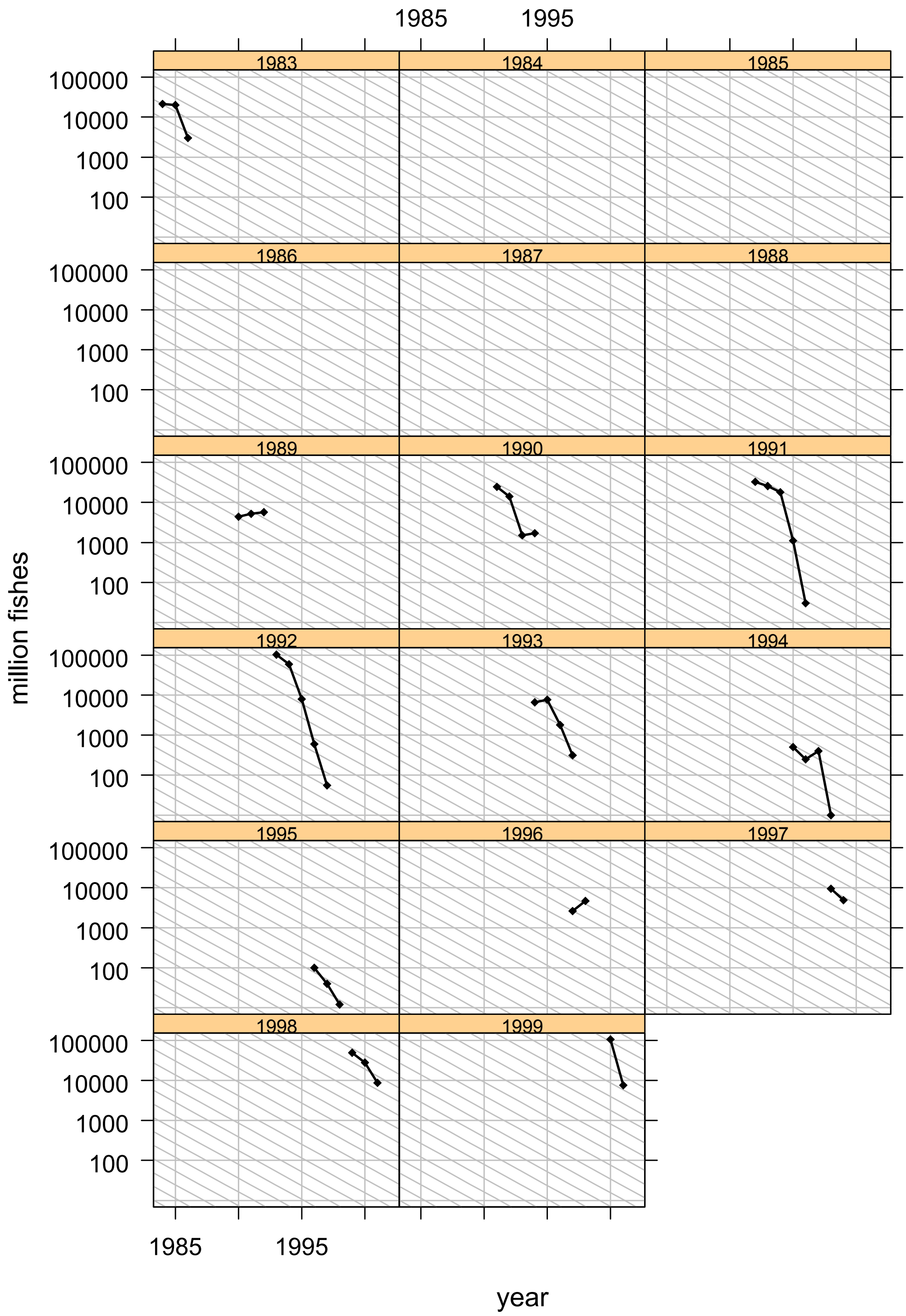


Figure 2.3.1.13 Catch curves for Norwegian spring spawn herring based on survey 5, acoustic survey in the Barents Sea in May. Grey lines correspond to Z=0.4.

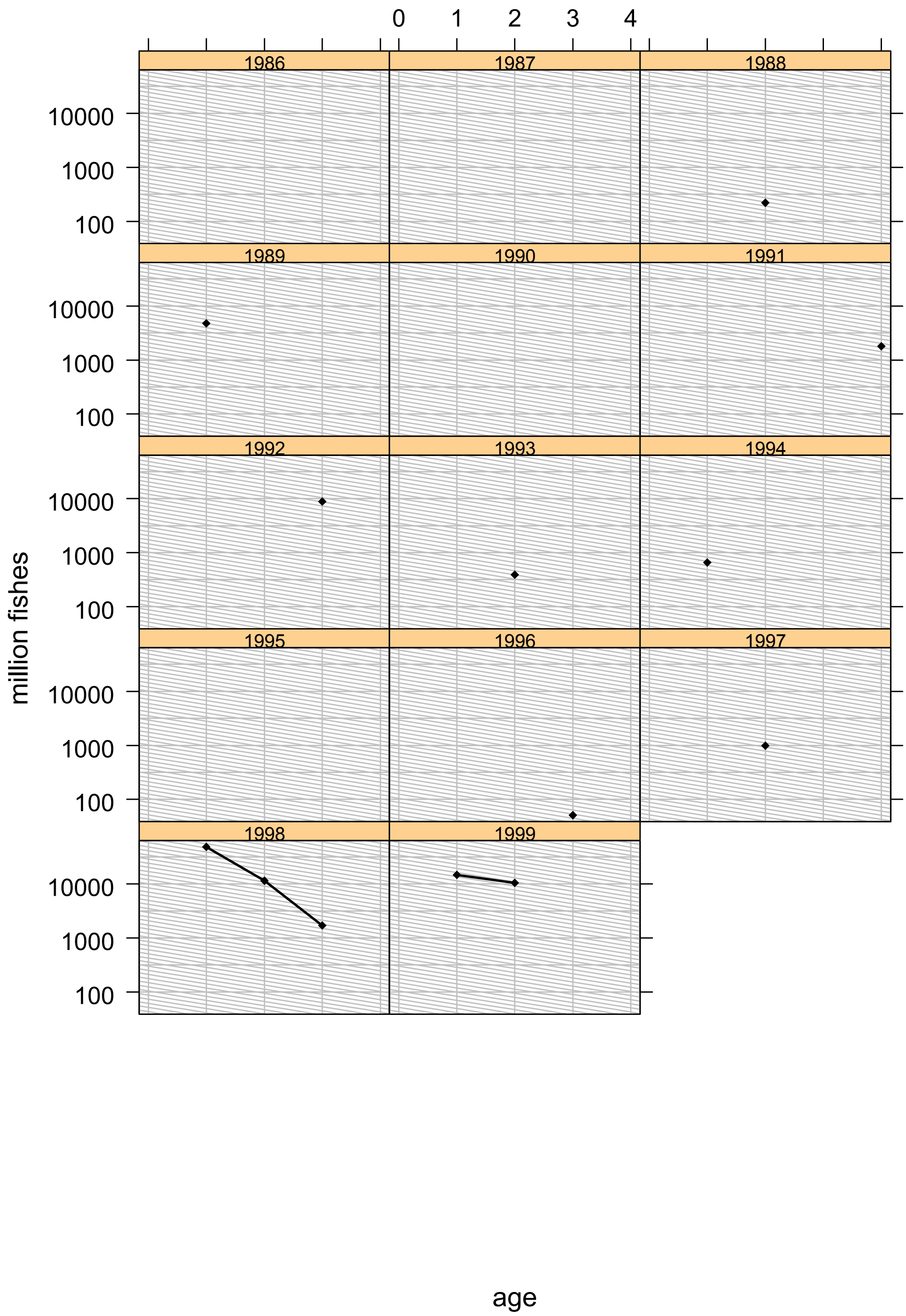


Figure 2.3.1.14 Catch curves for Norwegian spring spawn herring based on survey 6, acoustic survey in the Barents Sea in September. Gray lines correspond to Z=0.0.

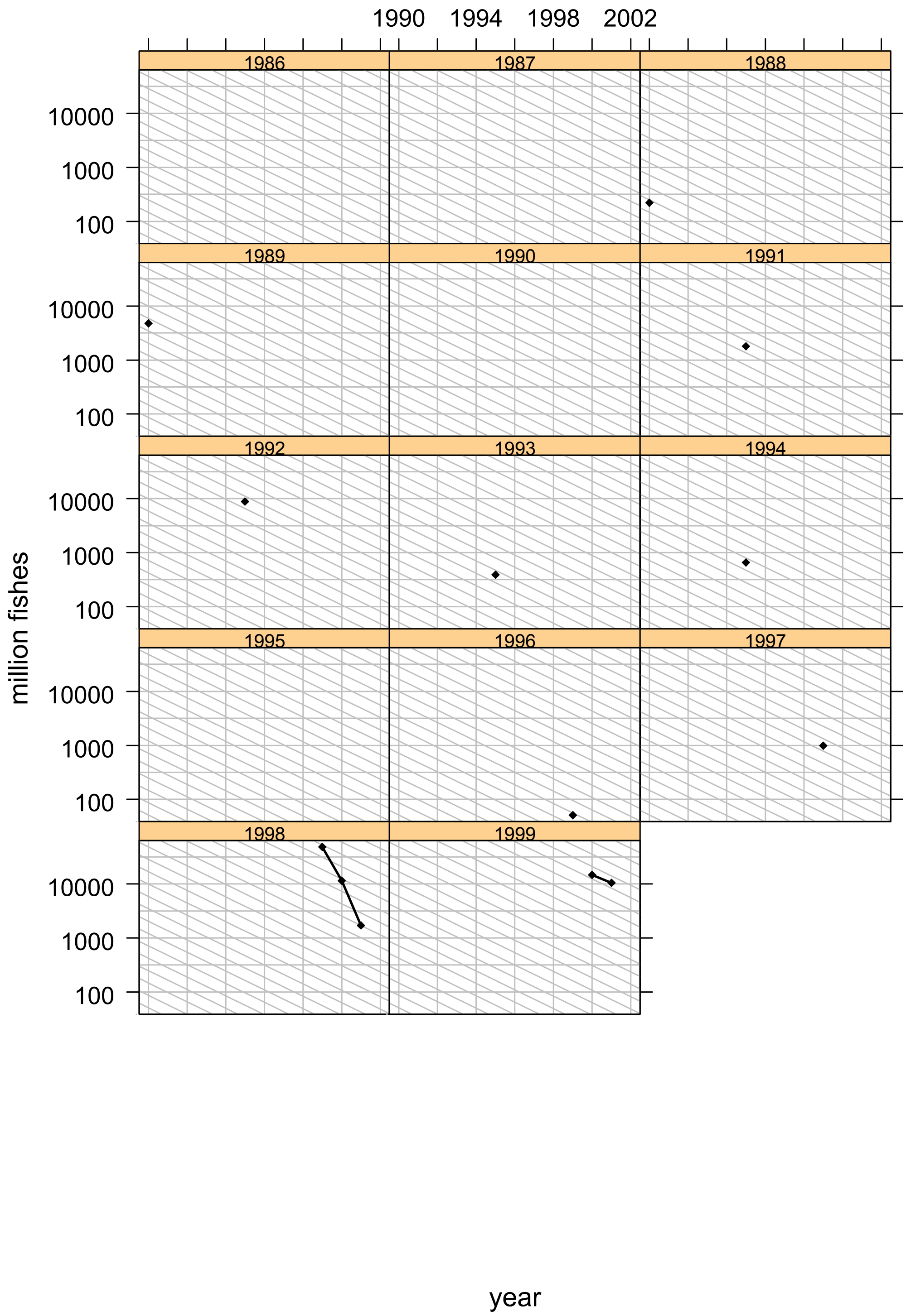


Figure 2.3.1.15 Catch curves for Norwegian spring spawn herring based on survey 6, acoustic survey in the Barents Sea in September. Diagonal gray lines correspond to Z=0.4.

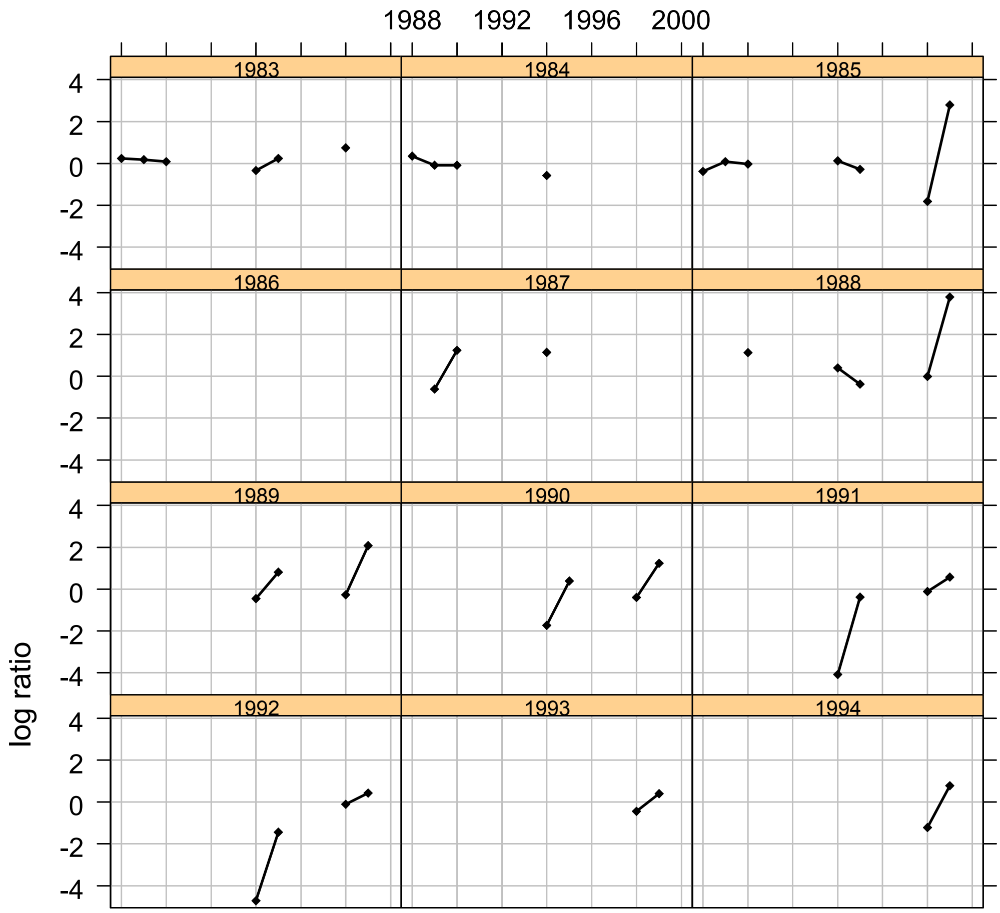


Figure 2.3.1.16 Log ratio for Norwegian spring spawn herring based on the survey on the spawning grounds.

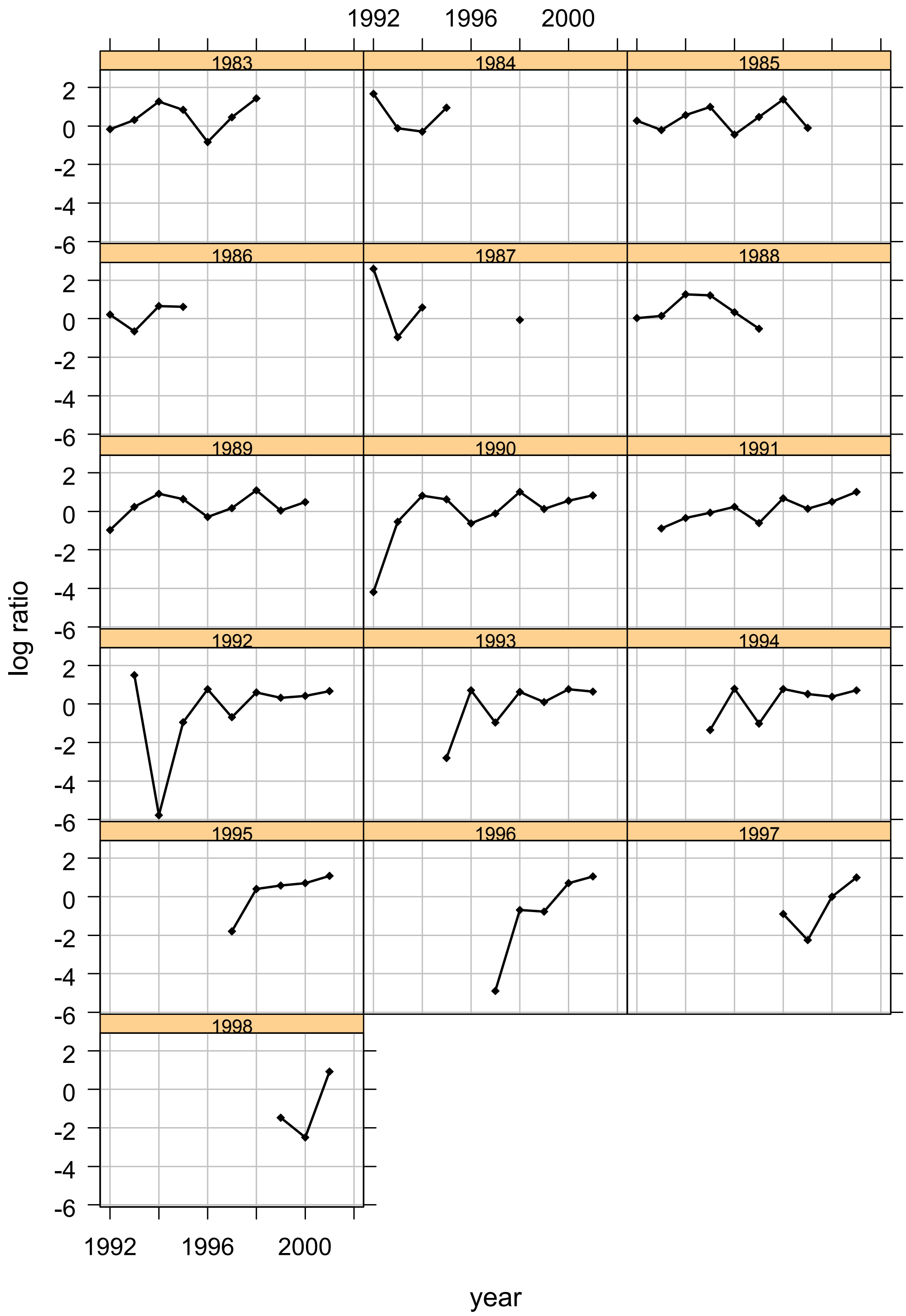


Figure 2.3.1.17 Log ratio for Norwegian spring spawn herring based on the survey on the wintering area in November.

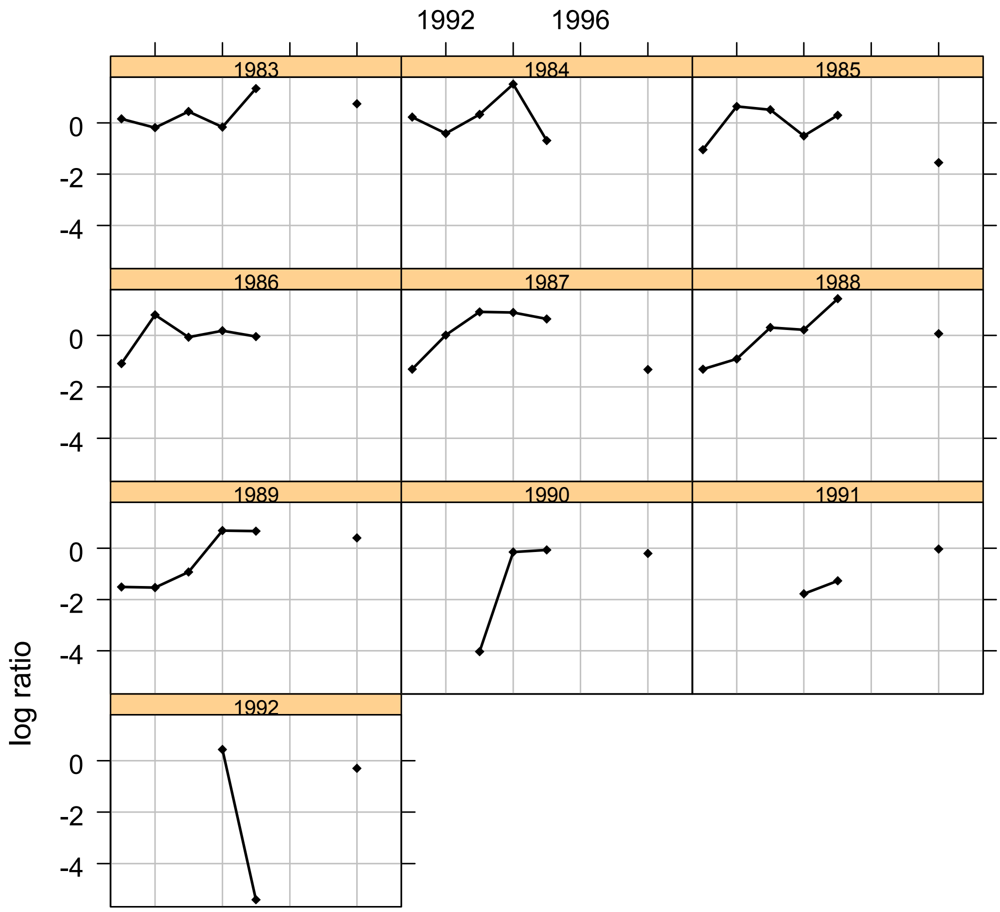


Figure 2.3.1.18 Log catch ratio for Norwegian spring spawn herring based on the survey on the wintering area in January.

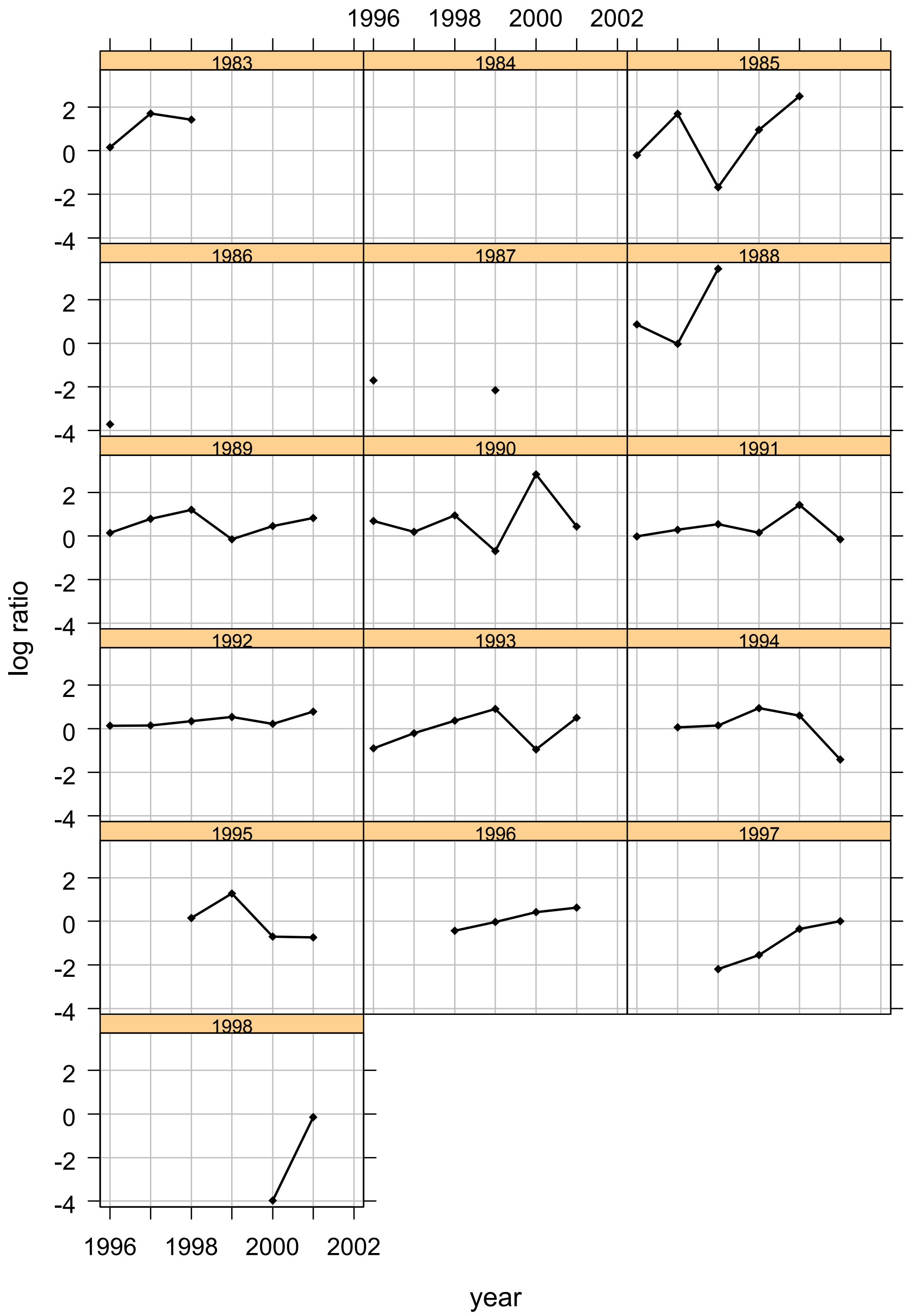


Figure 2.3.1.19 Log ratio for Norwegian spring spawn herring based on the survey on feeding area in May.

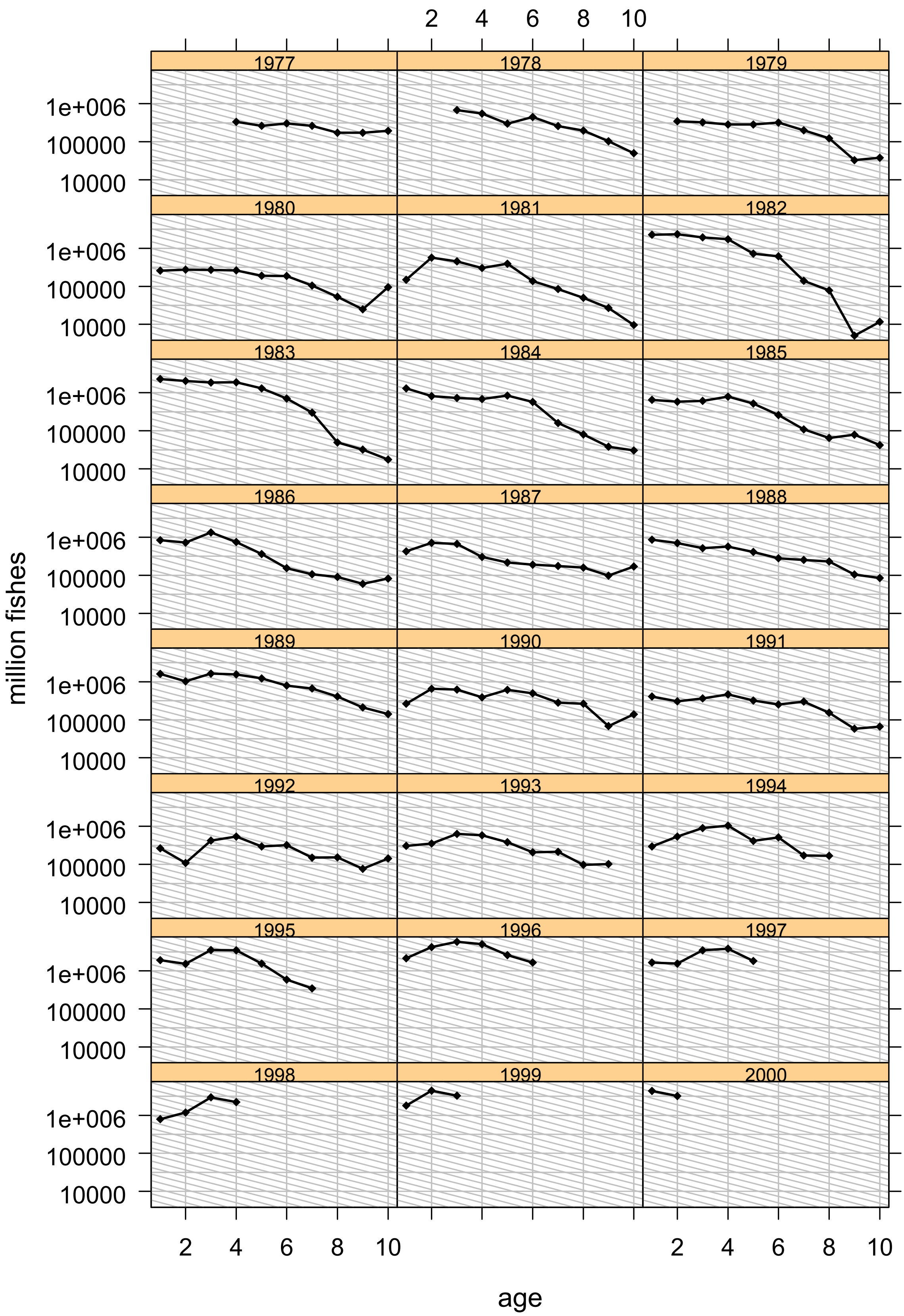


Figure 2.3.2.1. Blue whiting. Catch curves from the commercial catch.

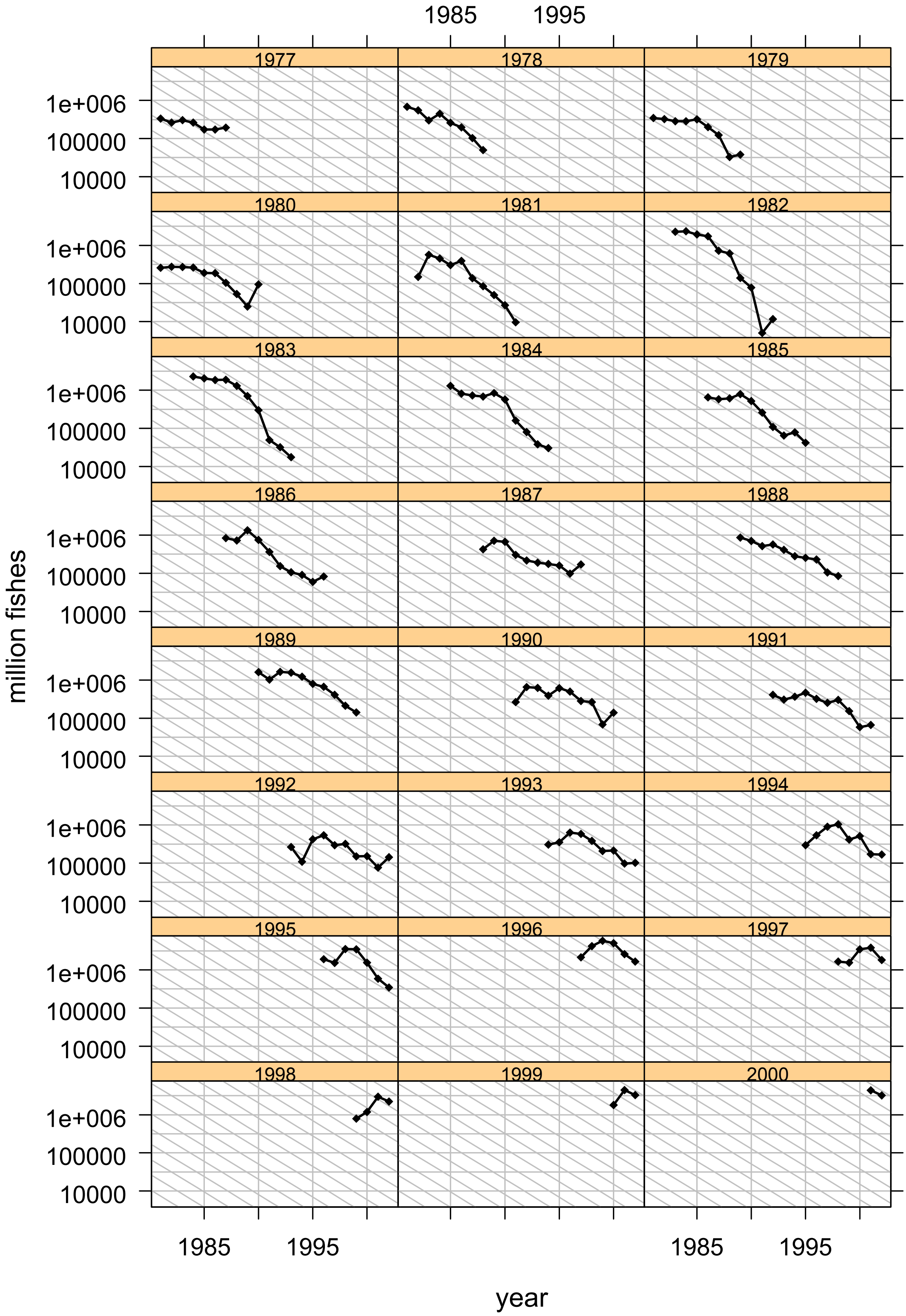


Figure 2.3.2.2. Blue whiting. Catch curves from the catch.

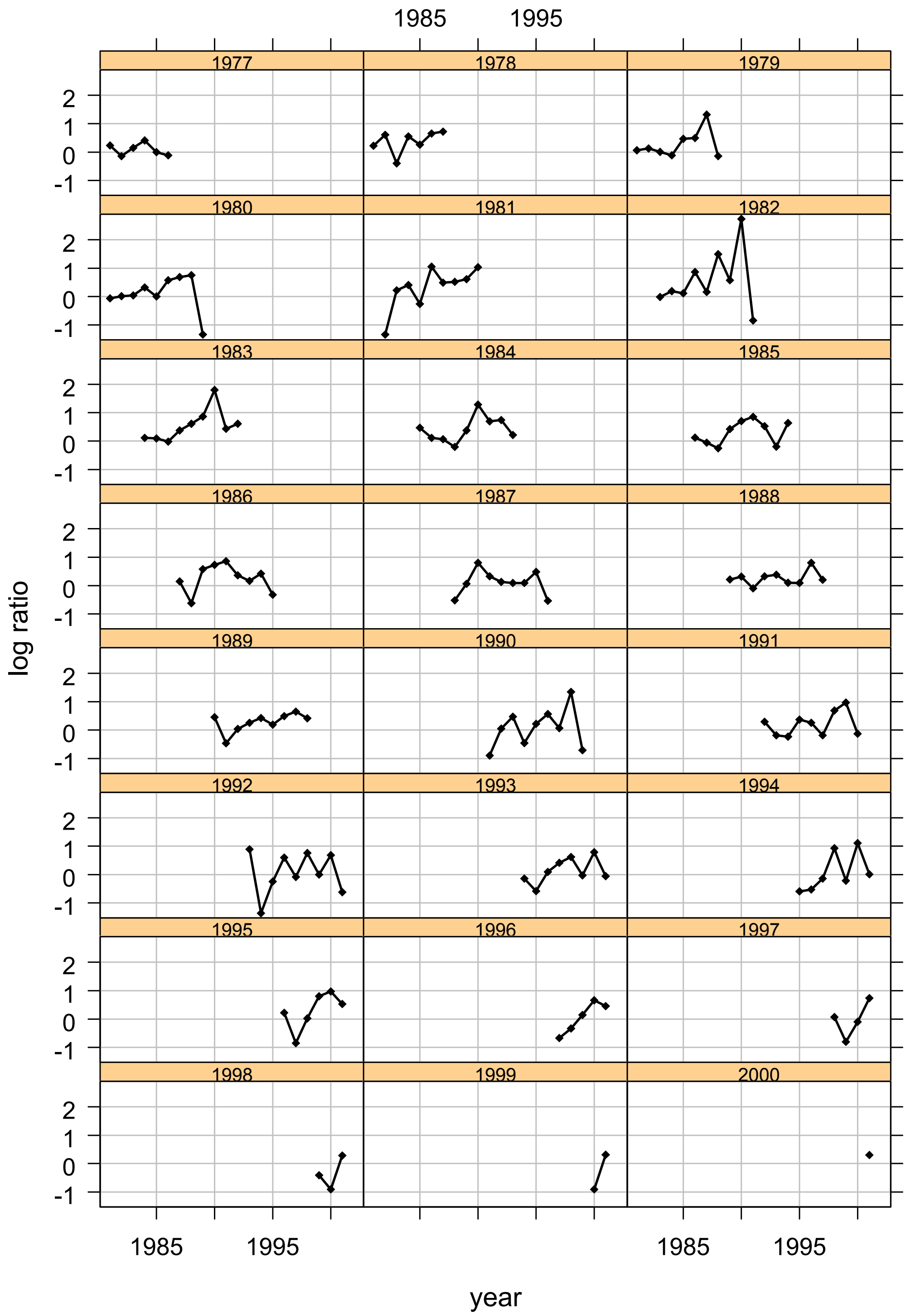


Figure 2.3.2.3. Blue whiting. Log catch ratios from the catch.

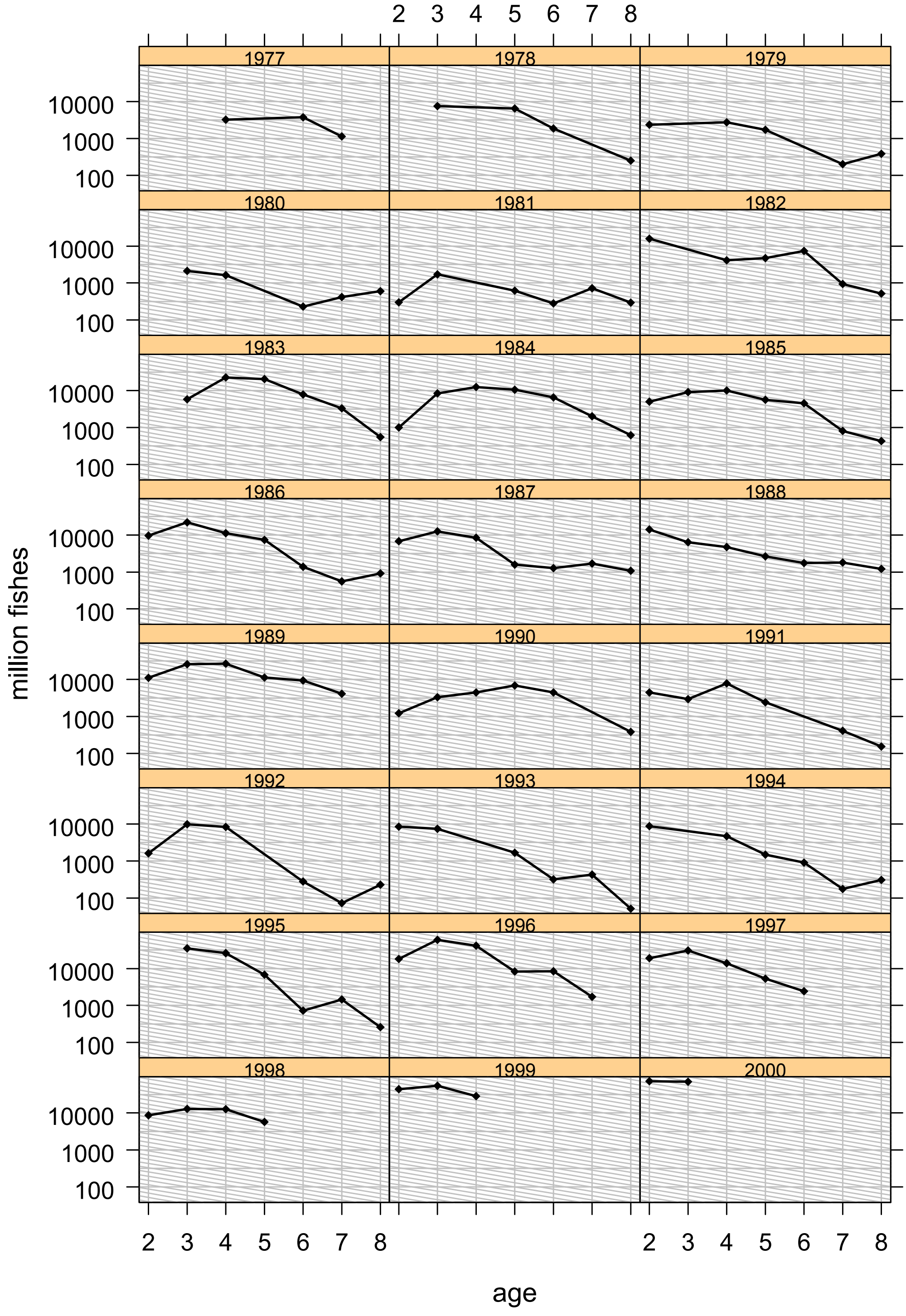


Figure 2.3.2.4. Blue whiting. Catch curves based on the Norwegian survey on the spawning grounds.

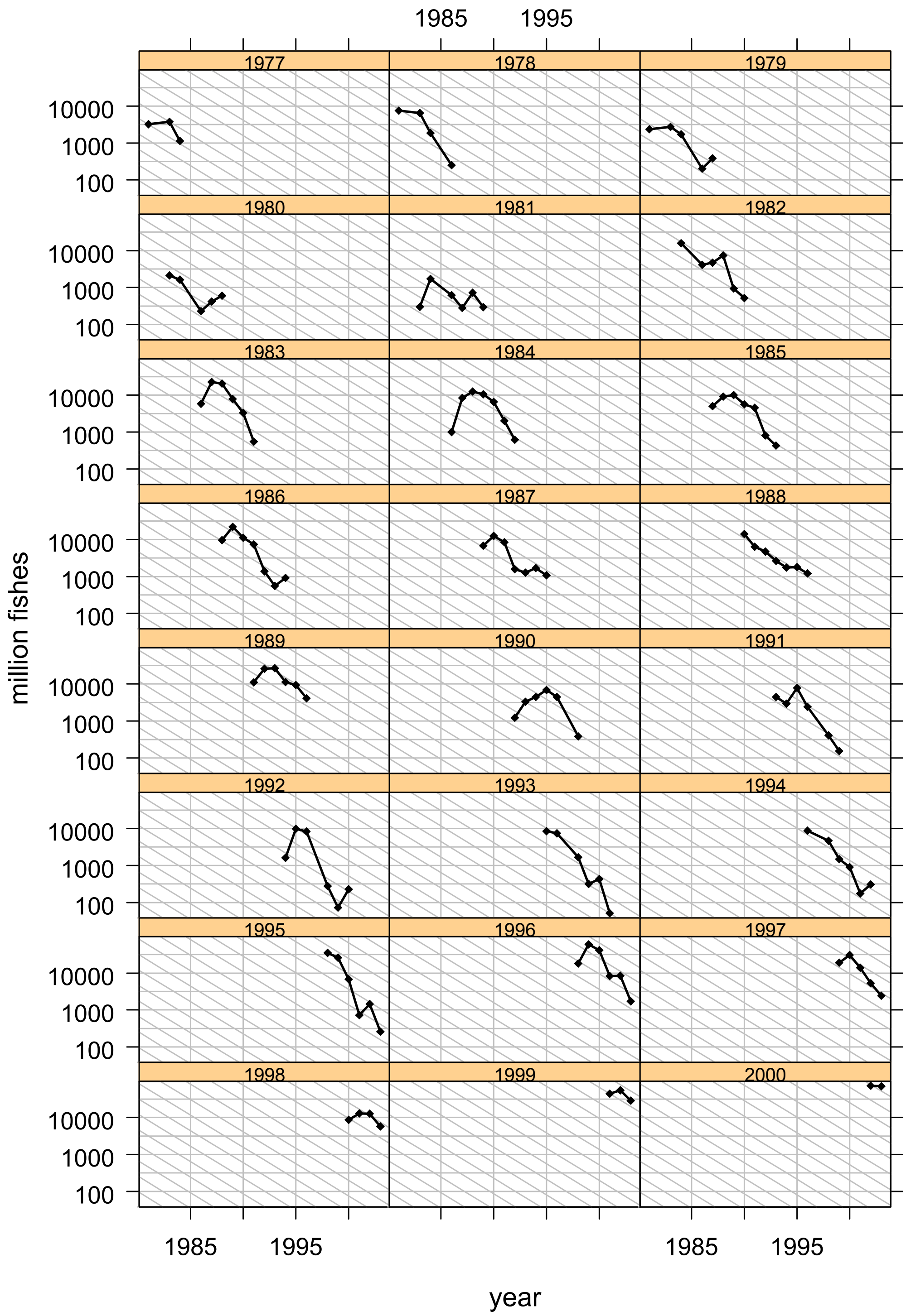


Figure 2.3.2.5. Blue whiting. Catch curves based on the Norwegian survey on the spawning grounds.

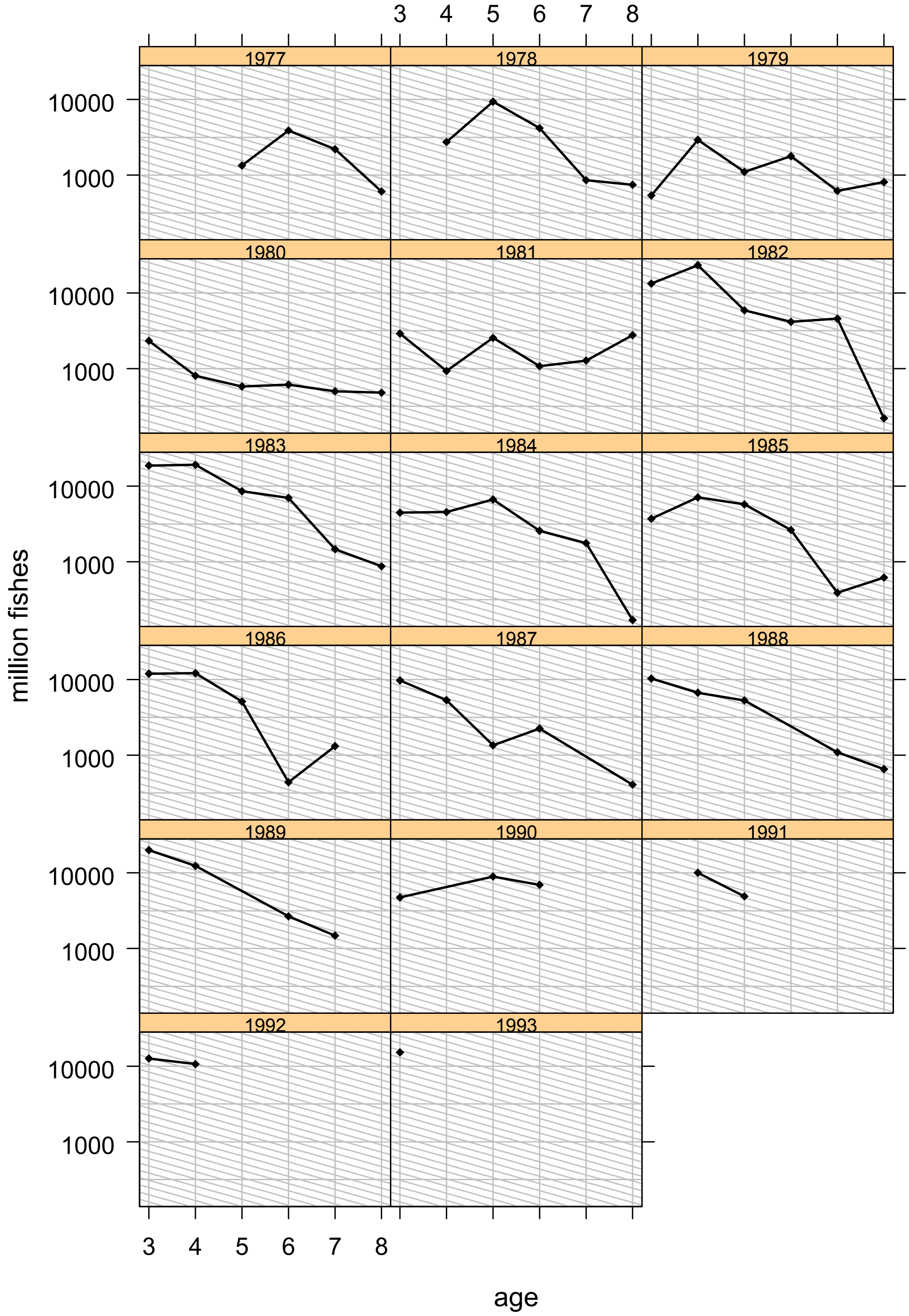


Figure 2.3.2.6. Blue whiting. Catch curves based on the Russian survey on the spawning grounds.

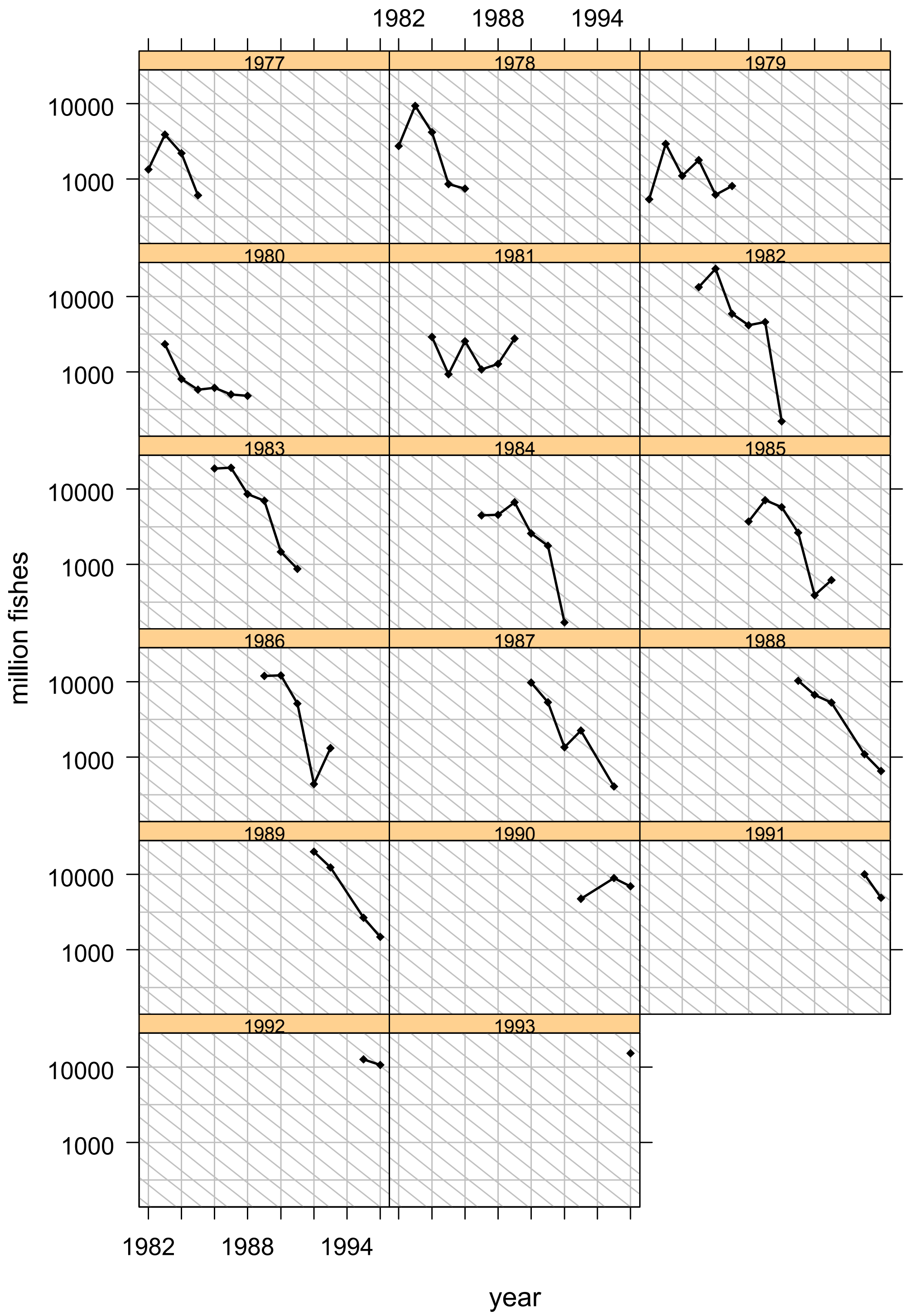


Figure 2.3.2.7. Blue whiting. Catch curves based on the Russian survey on the spawning grounds.

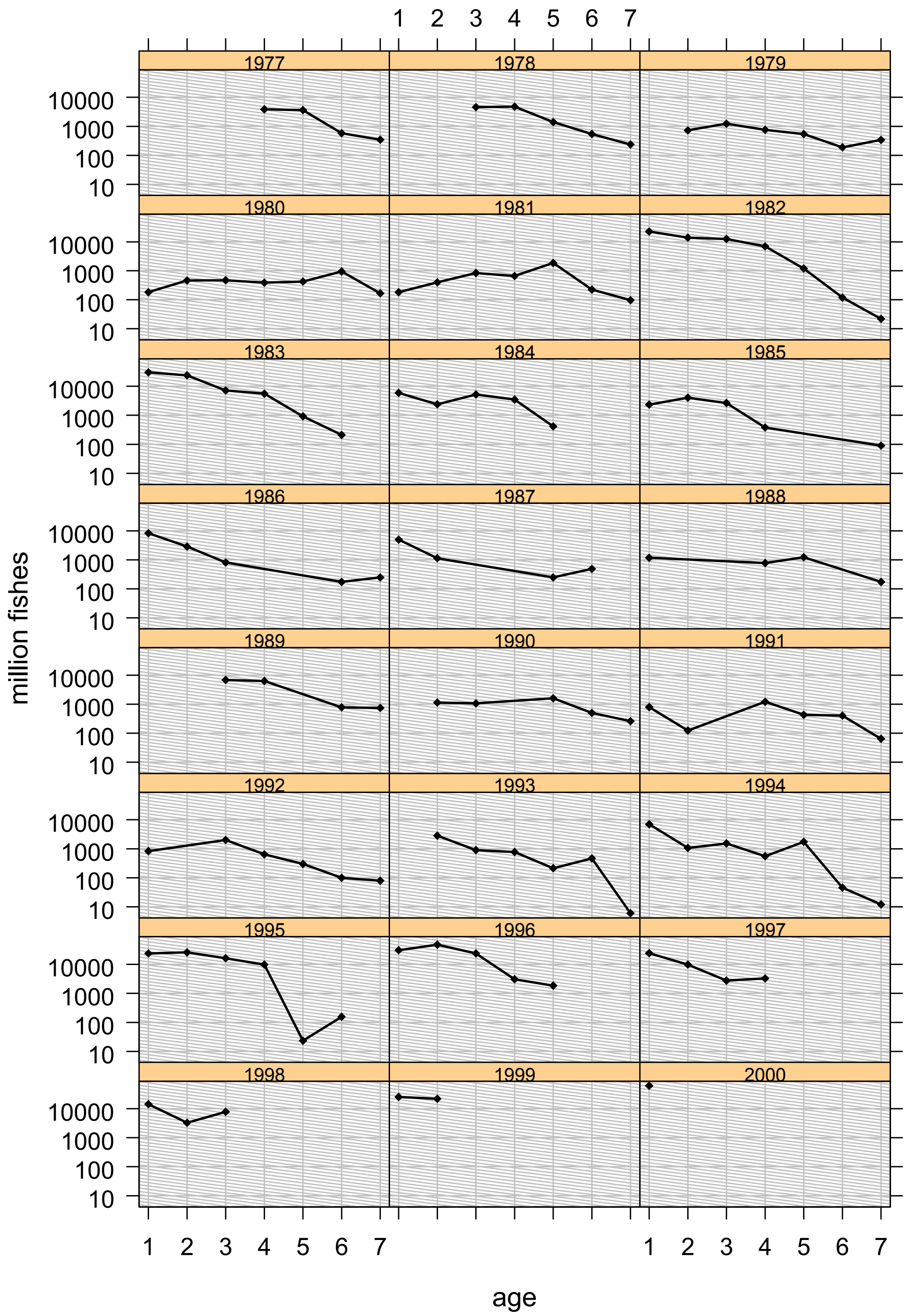


Figure 2.3.2.8. Blue whiting. Catch curves based on the Norwegian summer survey in the Norwegian Sea.

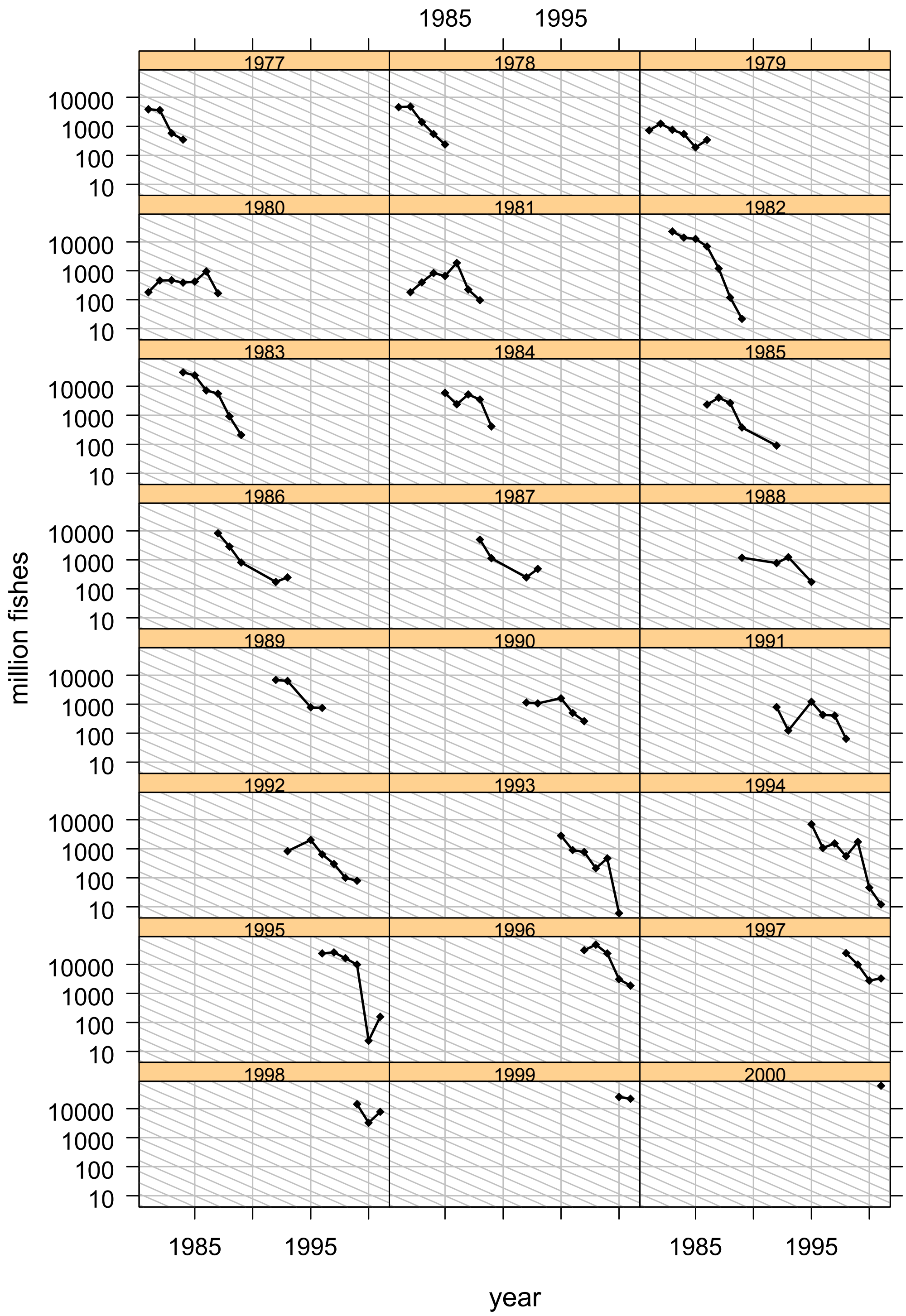


Figure 2.3.2.9. Blue whiting. Catch curves based on the Norwegian summer survey in the Norwegian Sea.

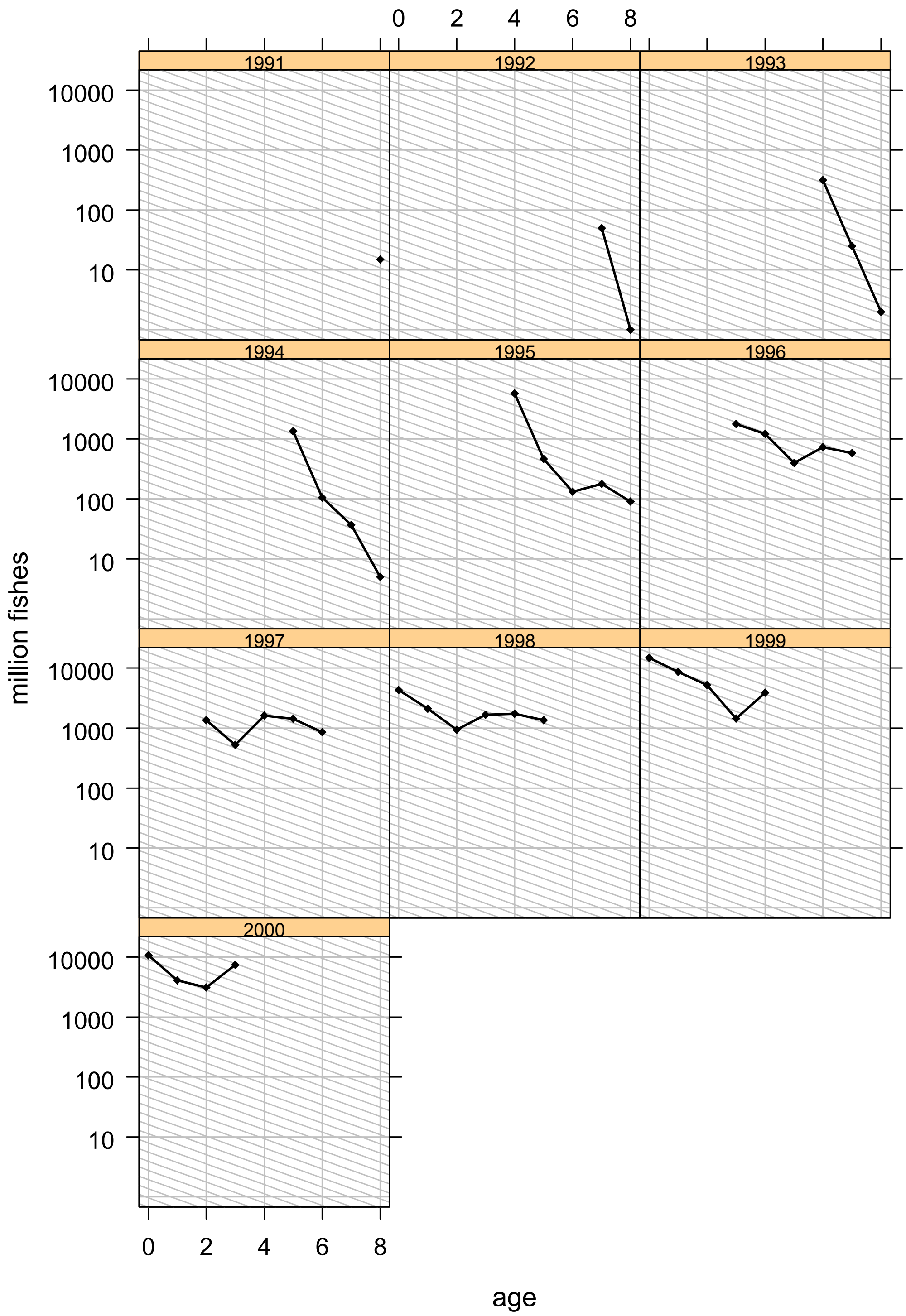


Figure2.3.2.10. Blue whiting. Catch curves based on the Icelandic blue whiting survey in summer.

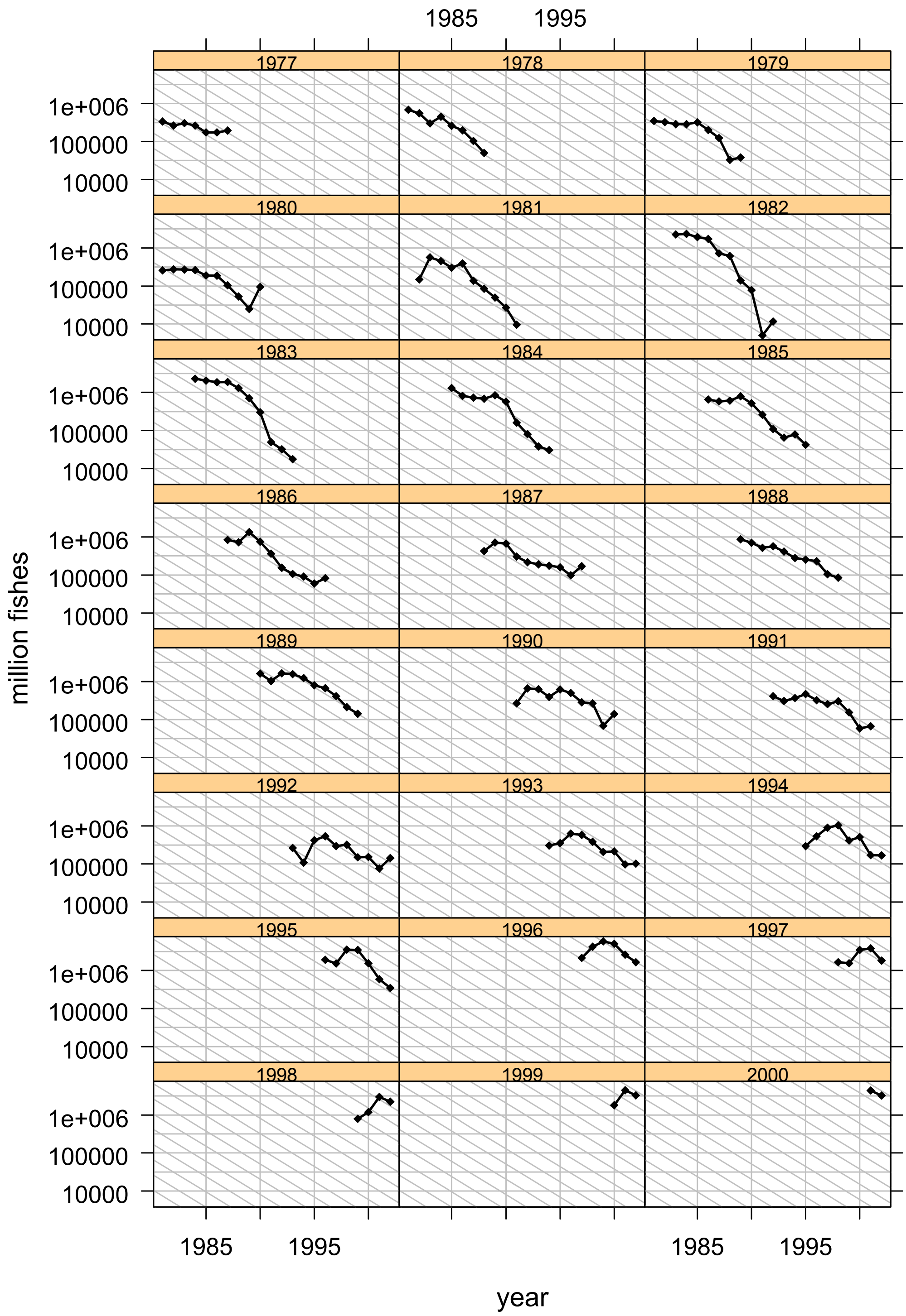


Figure2.3.2.11. Blue whiting. Catch curves based on the Icelandic blue whiting survey in summer.

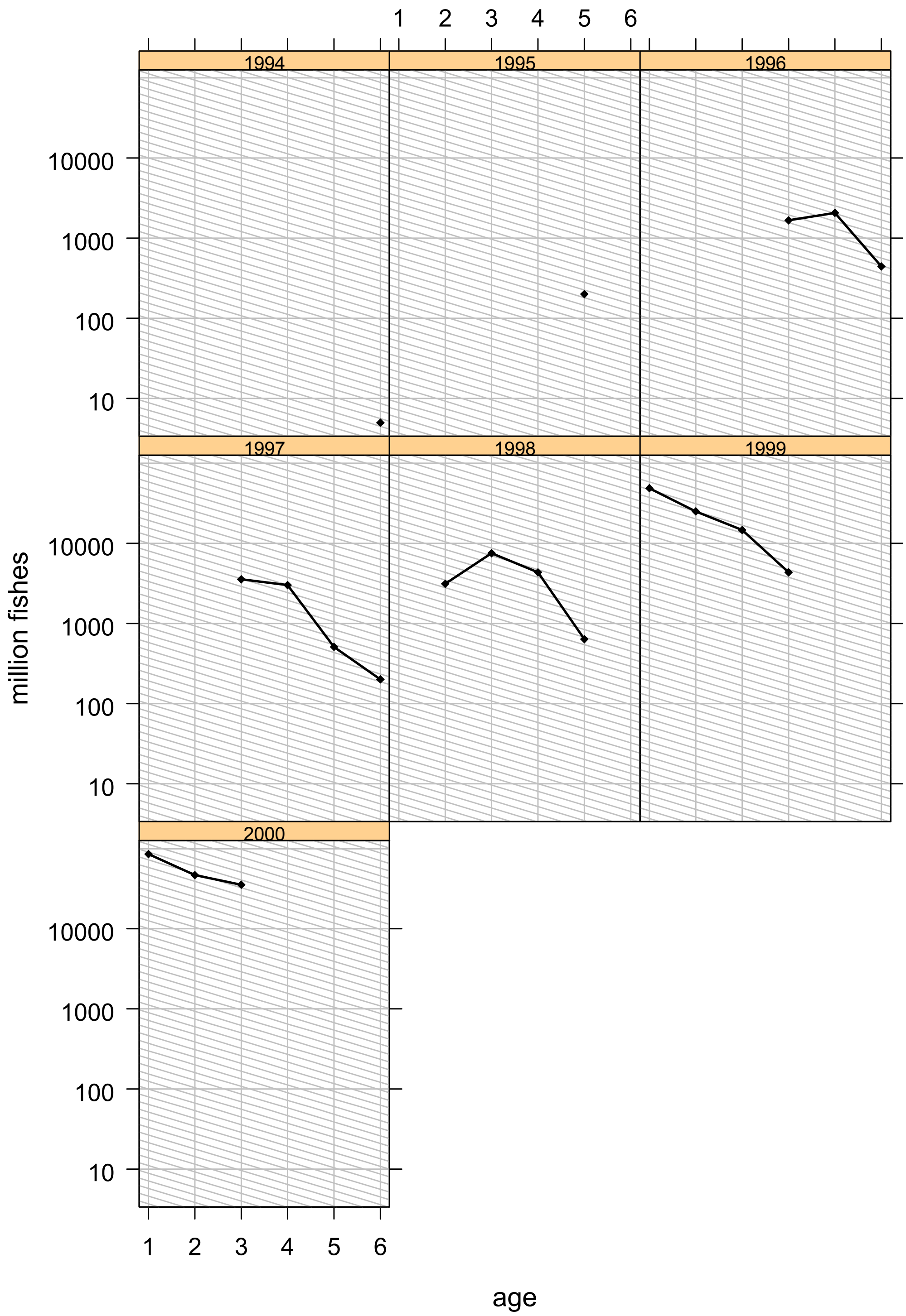


Figure 2.3.2.12. Blue whiting . Catch curves based on the International survey on the feeding grounds in the Norwegian Sea in May.

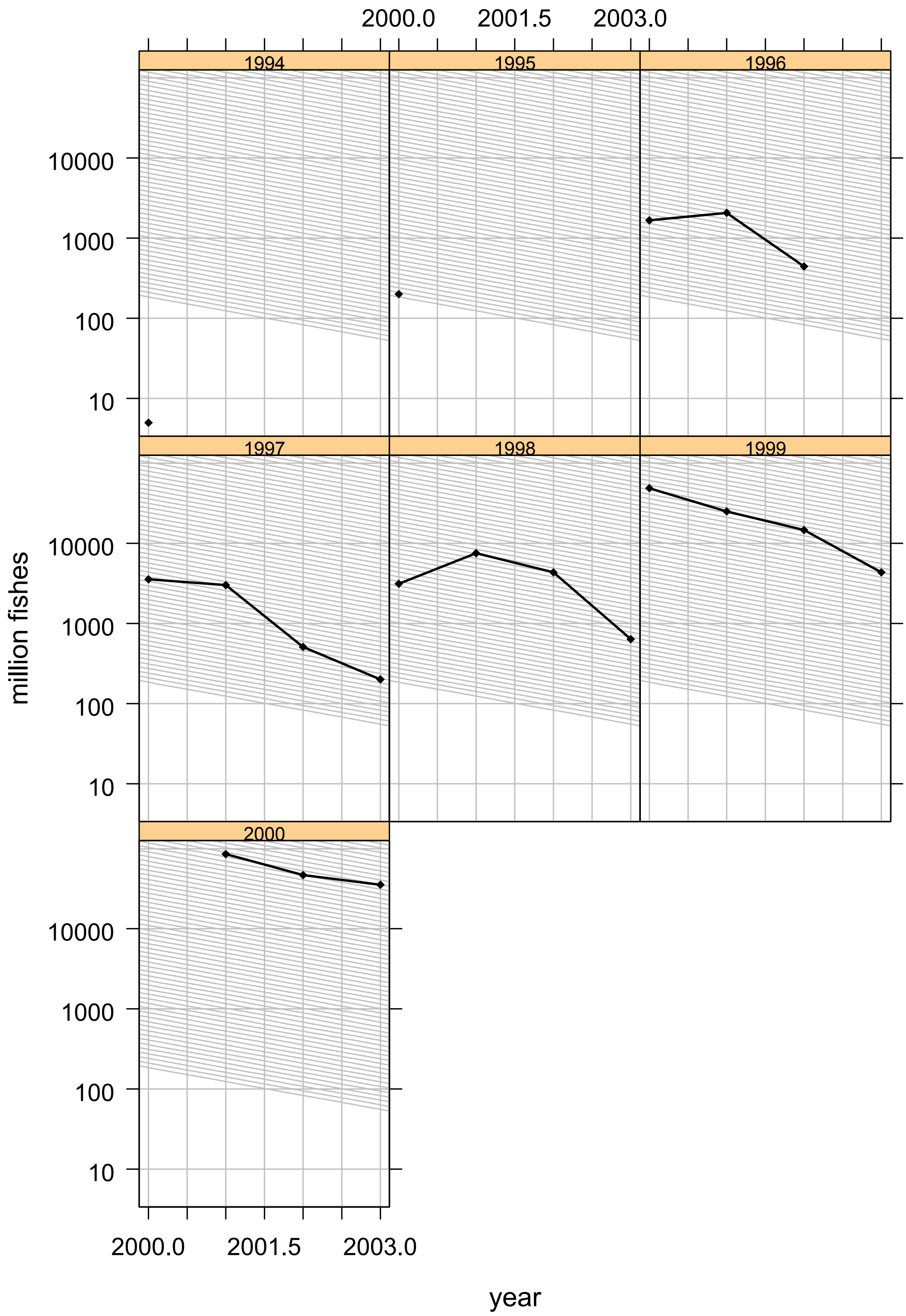


Figure 2.3.2.13. Blue whiting. Catch curves based on the International survey on the feeding grounds in the Norwegian Sea in the Norwegian Sea in May.

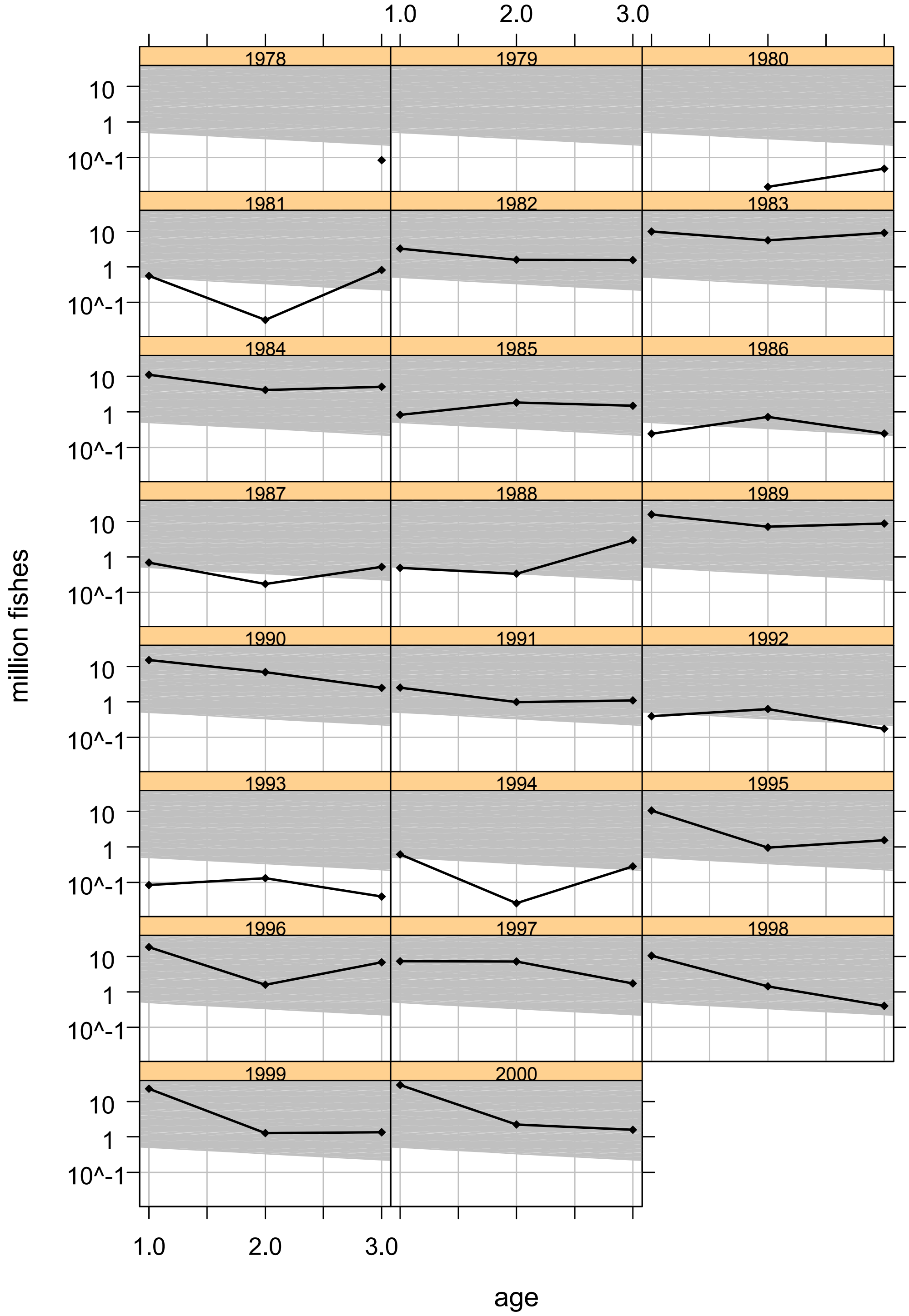


Figure 2.3.2.14. Blue whiting. Catch curves based on the Norwegian winter survey in the Barents Sea.

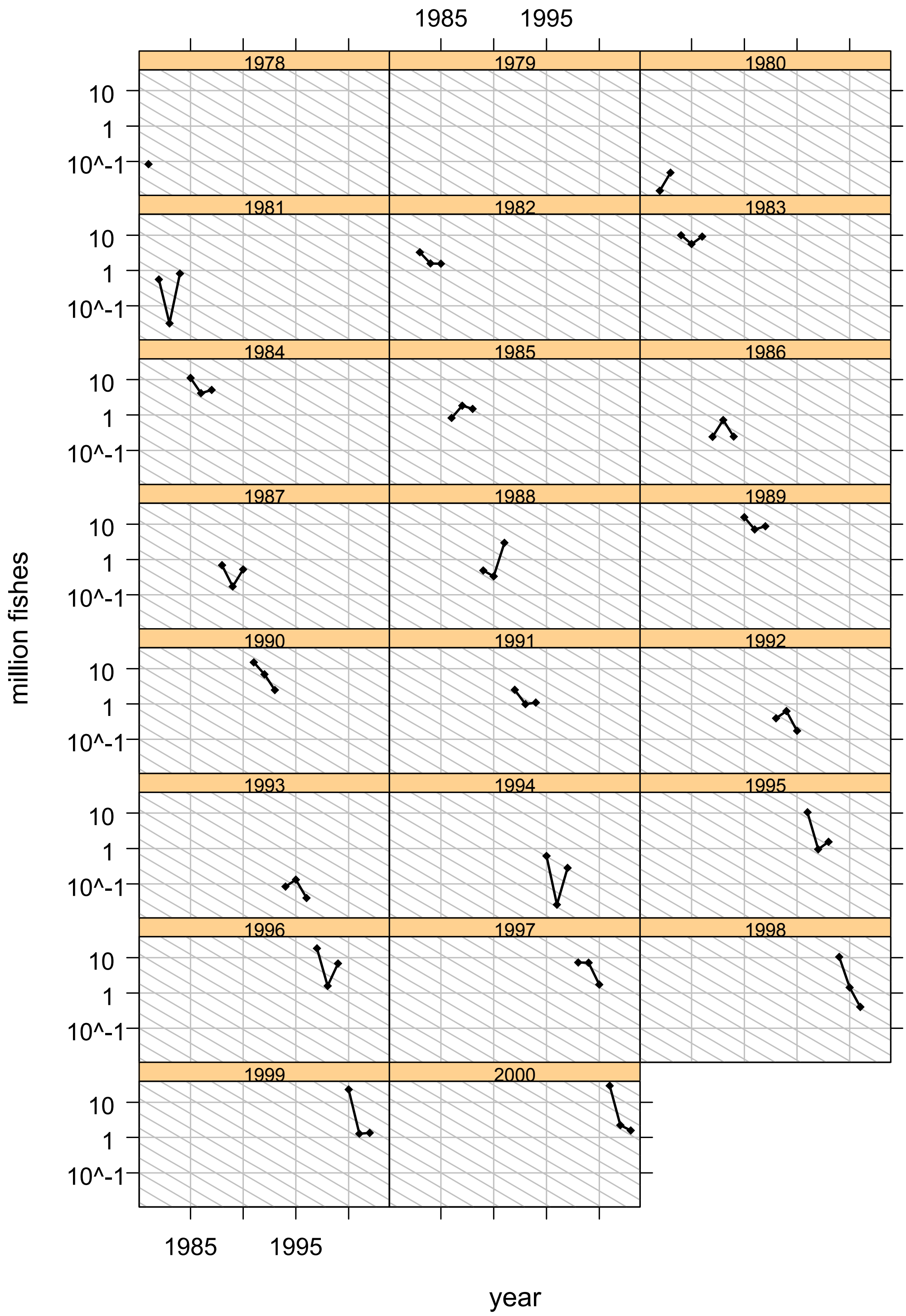


Figure 2.3.2.15. Blue whiting . Catch curves from the Norwegian winter survey in the Barents Sea.

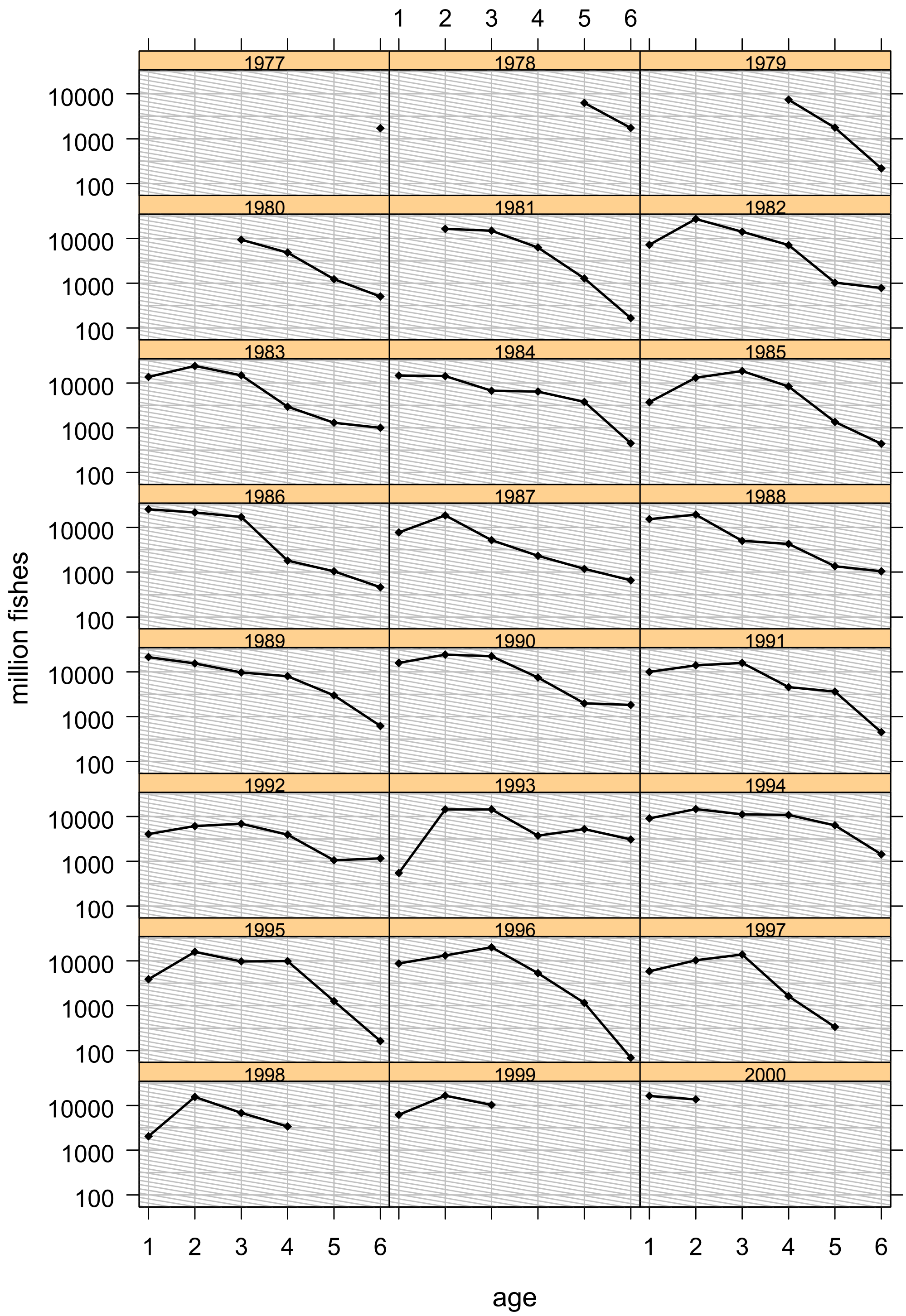


Figure 2.3.2.16 Blue whiting. Catch curves based on the Spanish pair trawl series (CPUE).

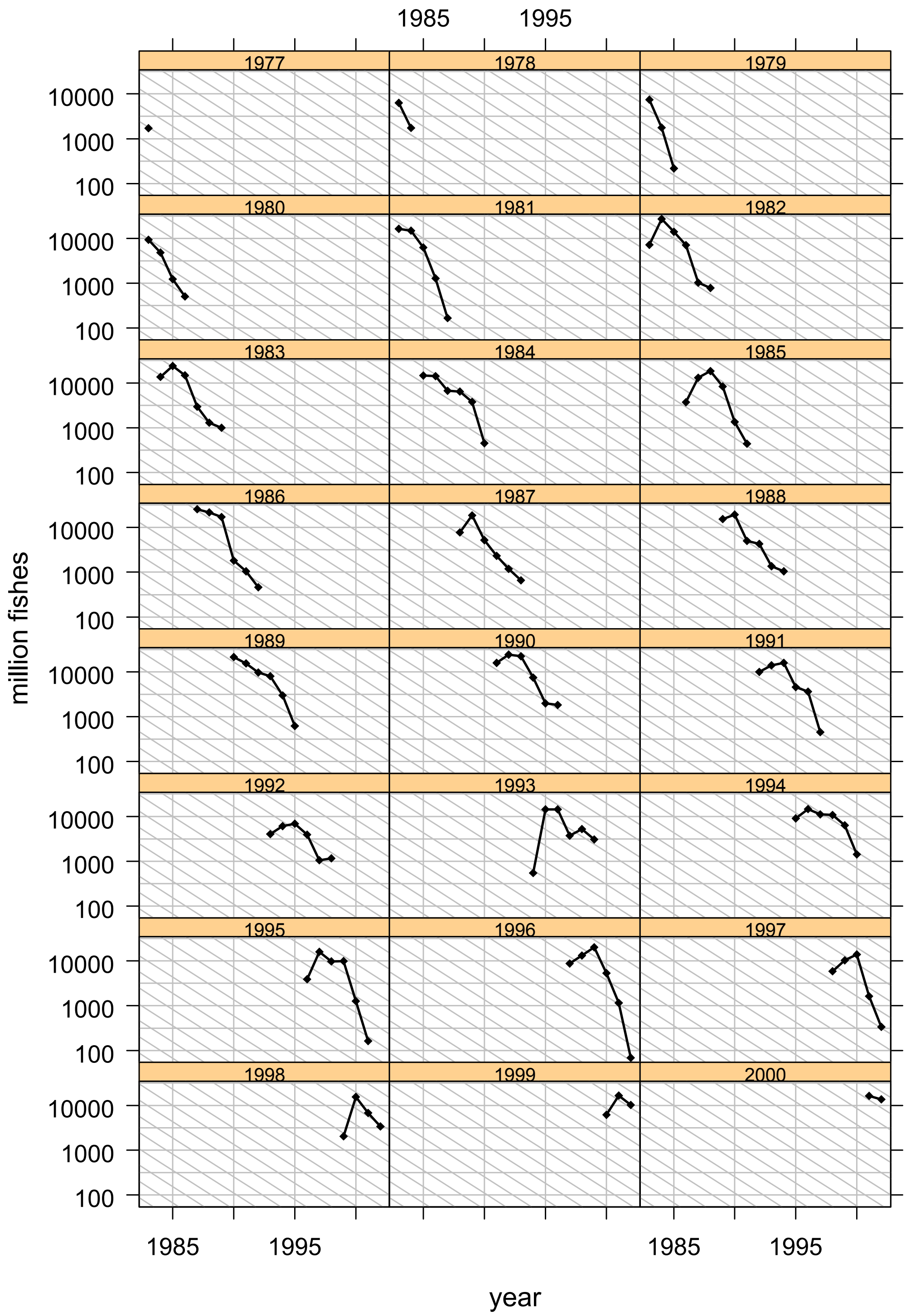


Figure 2.3.2.17 Blue whiting. Catch curves based on the Spanish pair trawl series (CPUE).

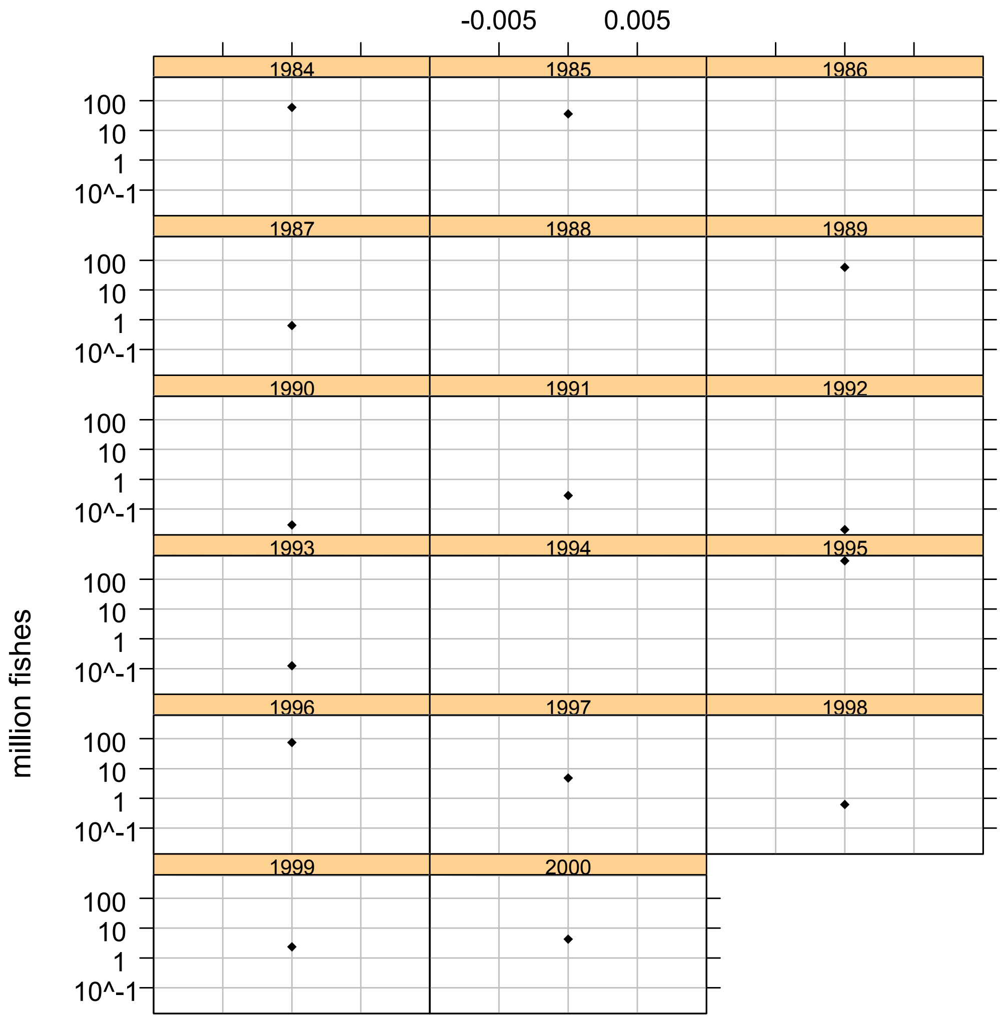


Figure. 2.3.2.18. Blue whiting. Catch curves based on the Norwegian shrimp survey in autumn.

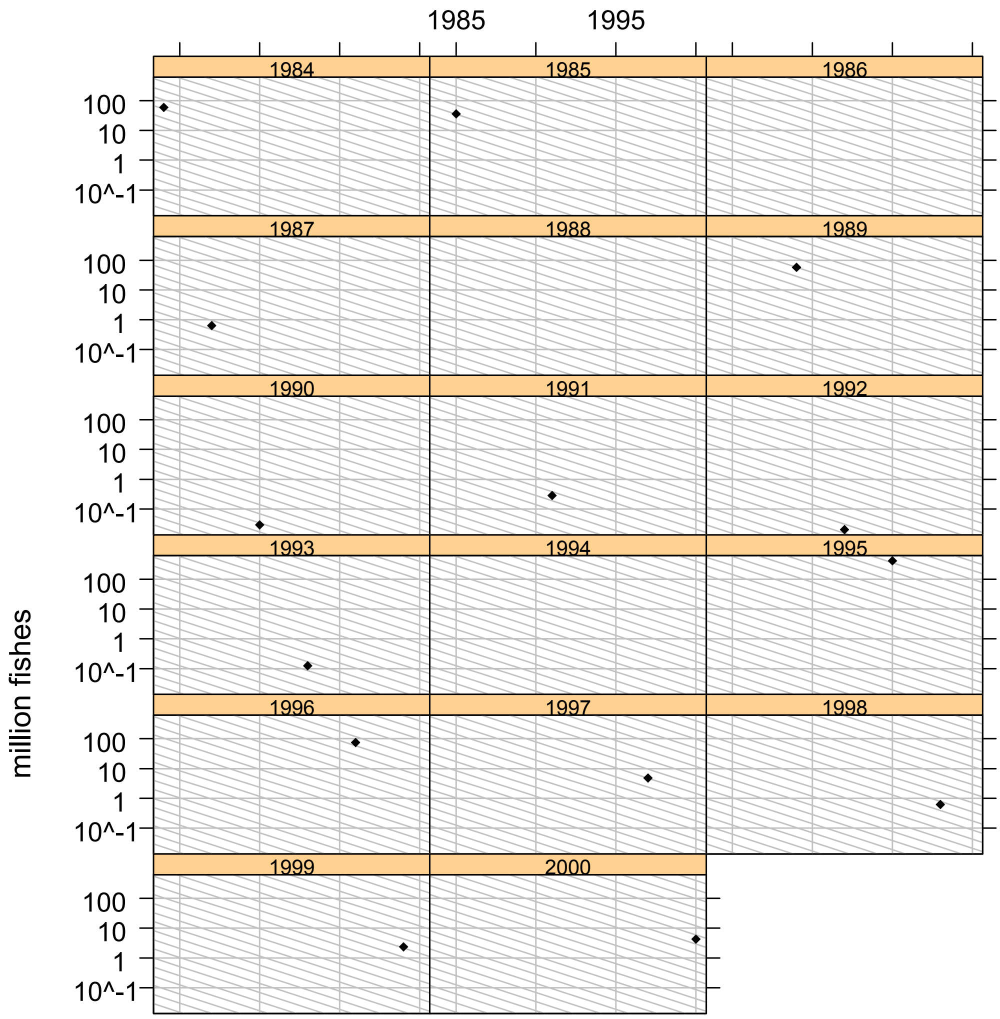


Figure. 2.3.2.19. Blue whiting. Catch curves based on the Norwegian shrimp survey in autumn.

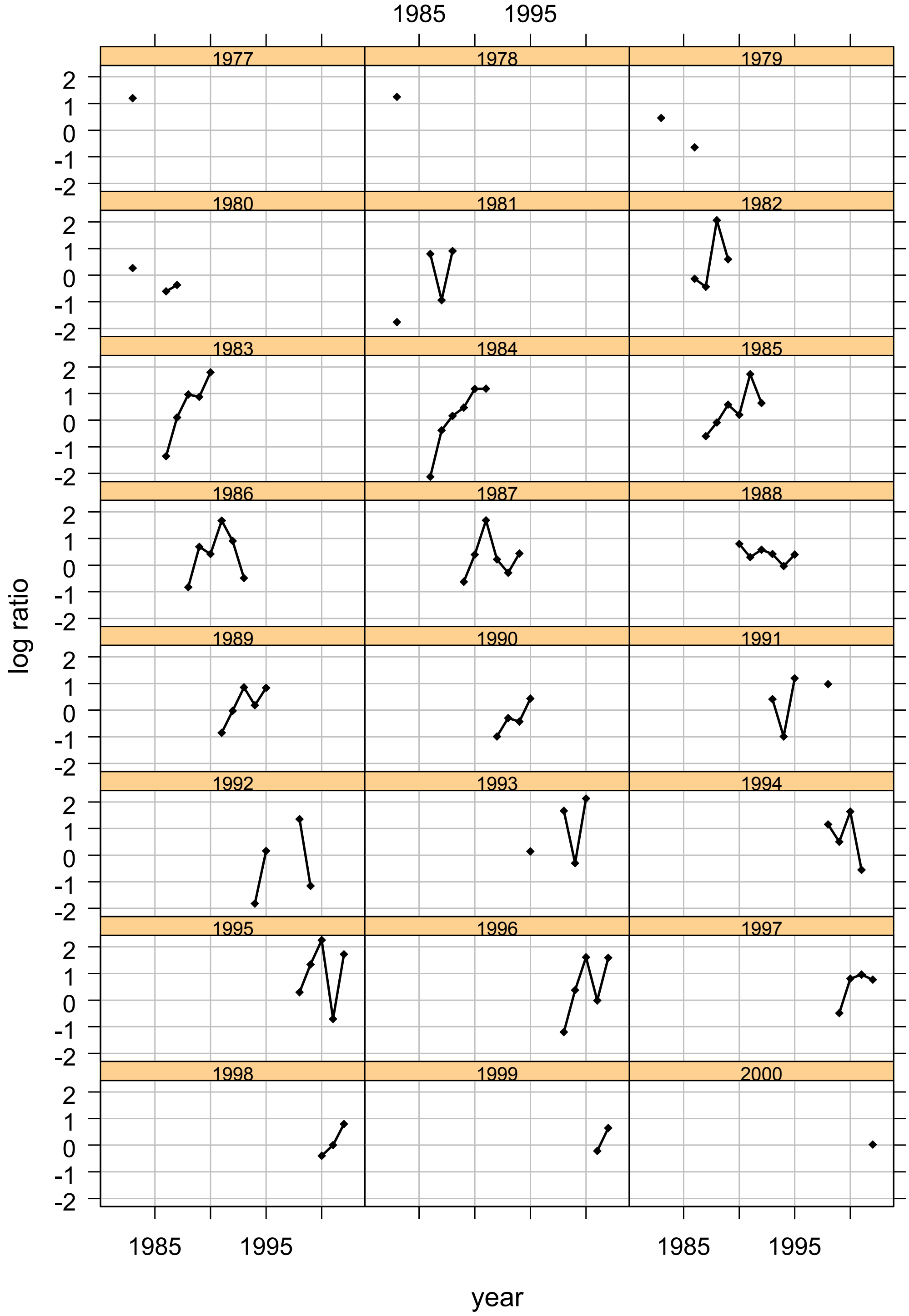


Figure 2.3.2.20 Blue whiting. Log catch ratios based on the Norwegian survey on the spawning grounds.

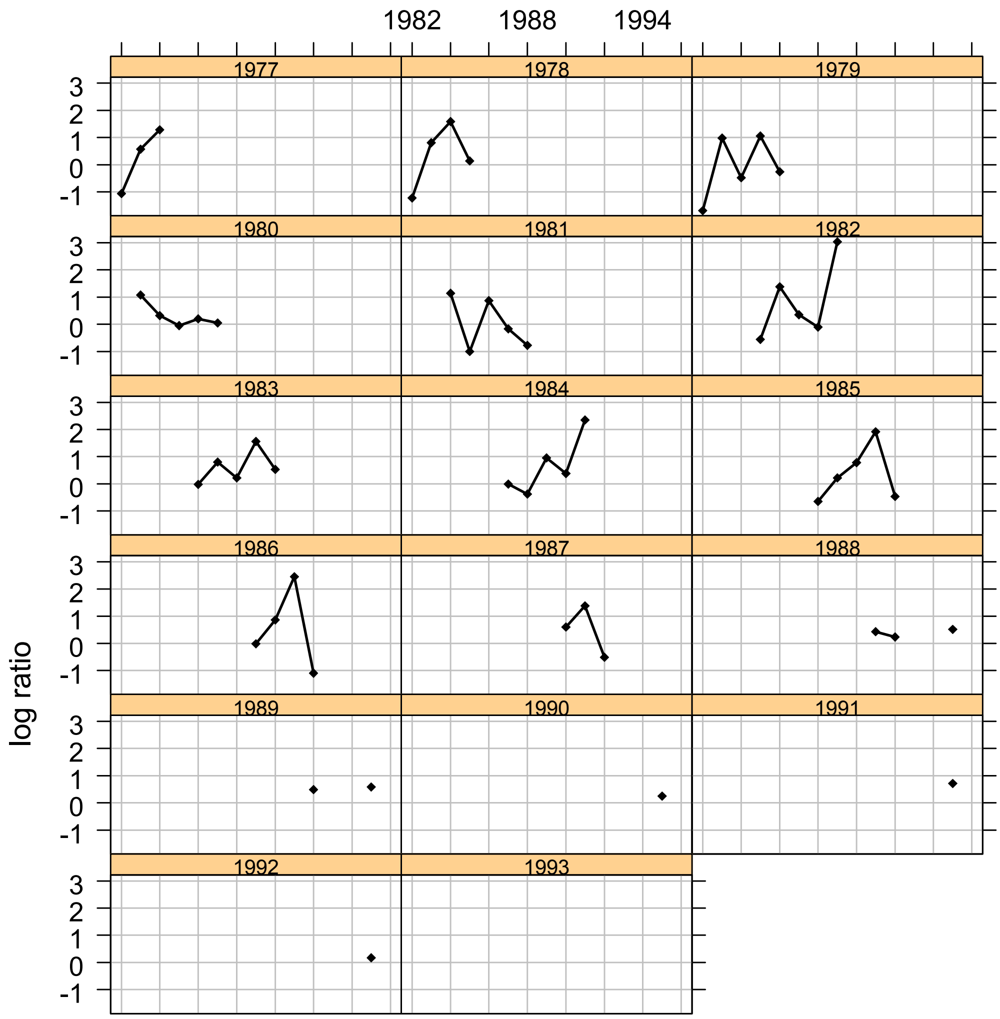


Figure 2.3.2.21 Blue whiting. Log catch ratios based on the Russian survey on the spawning grounds.

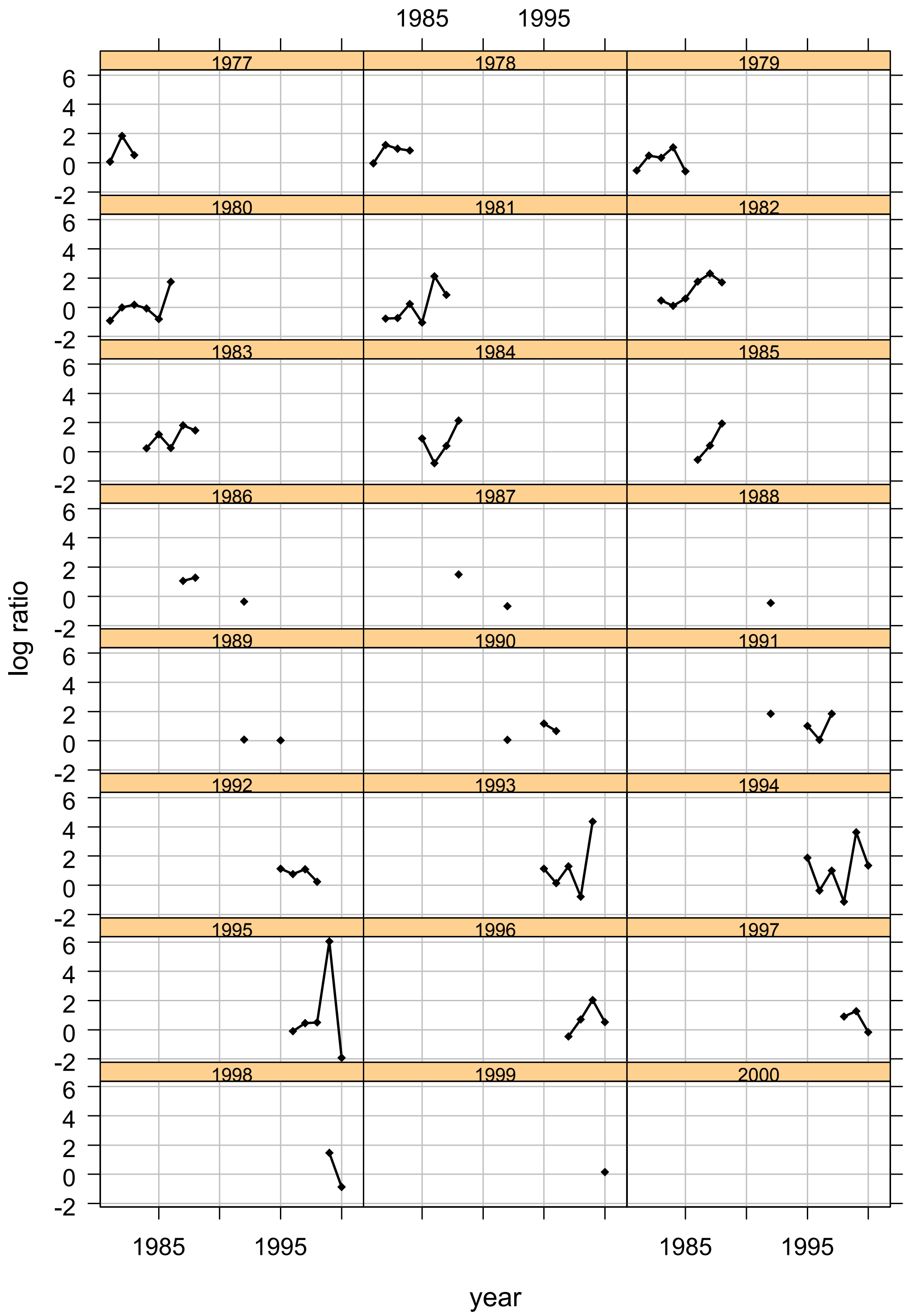


Figure 2.3.2.22 Blue whiting. Log catch ratios based on the Norwegian summer survey in the Norwegian Sea.

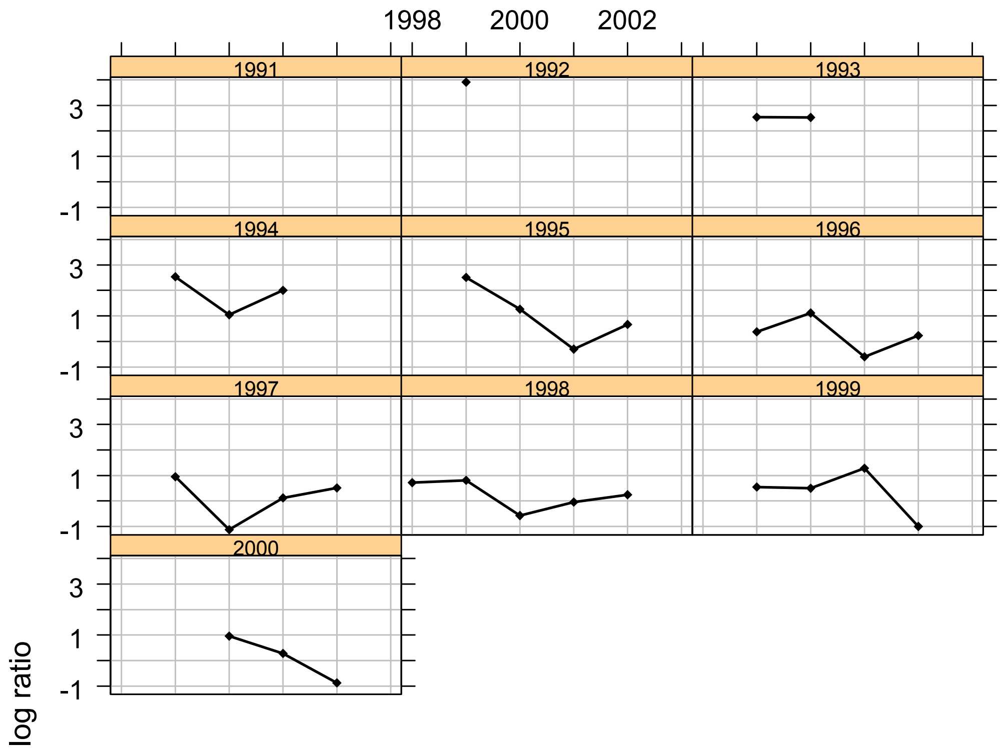


Figure 2.3.2.23 Blue whiting. Log catch ratios based on the Icelandic blue whiting survey in summer.

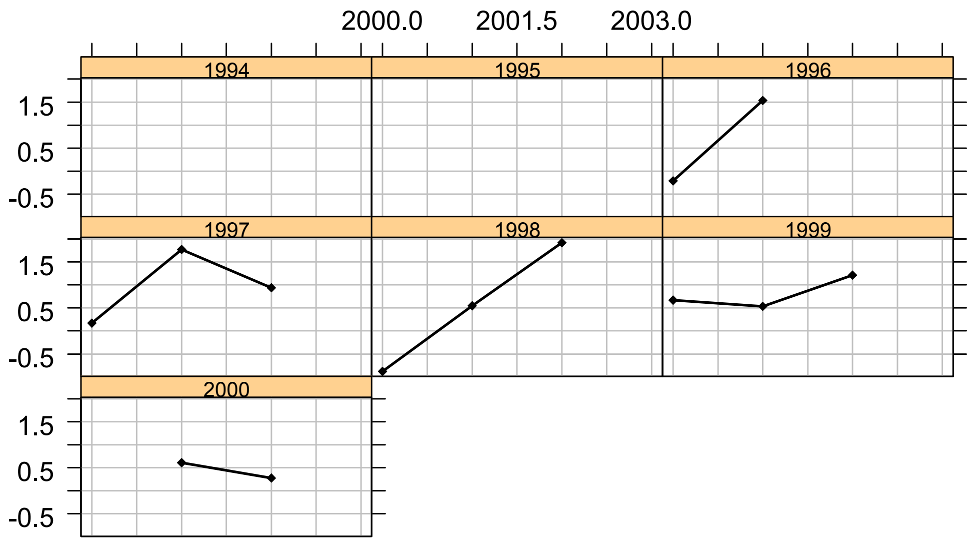


Figure 2.3.2.24 Blue whiting. Log catch ratios based on the International survey on the feeding grounds in the Norwegian Sea in May.

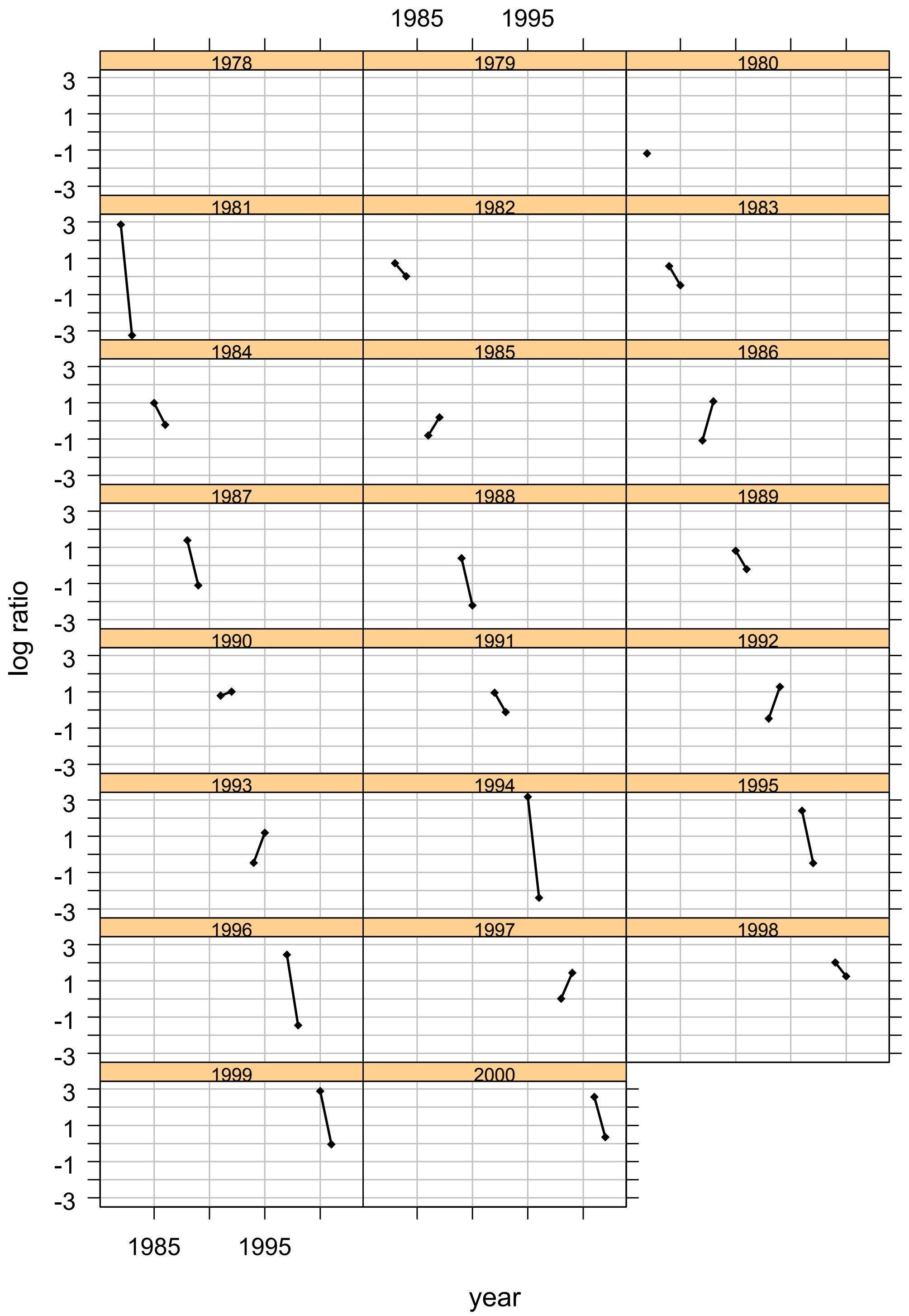


Figure 2.3.2.25. Blue whiting. Log catch ratios based on the Norwegian winter survey in the Barents Sea.

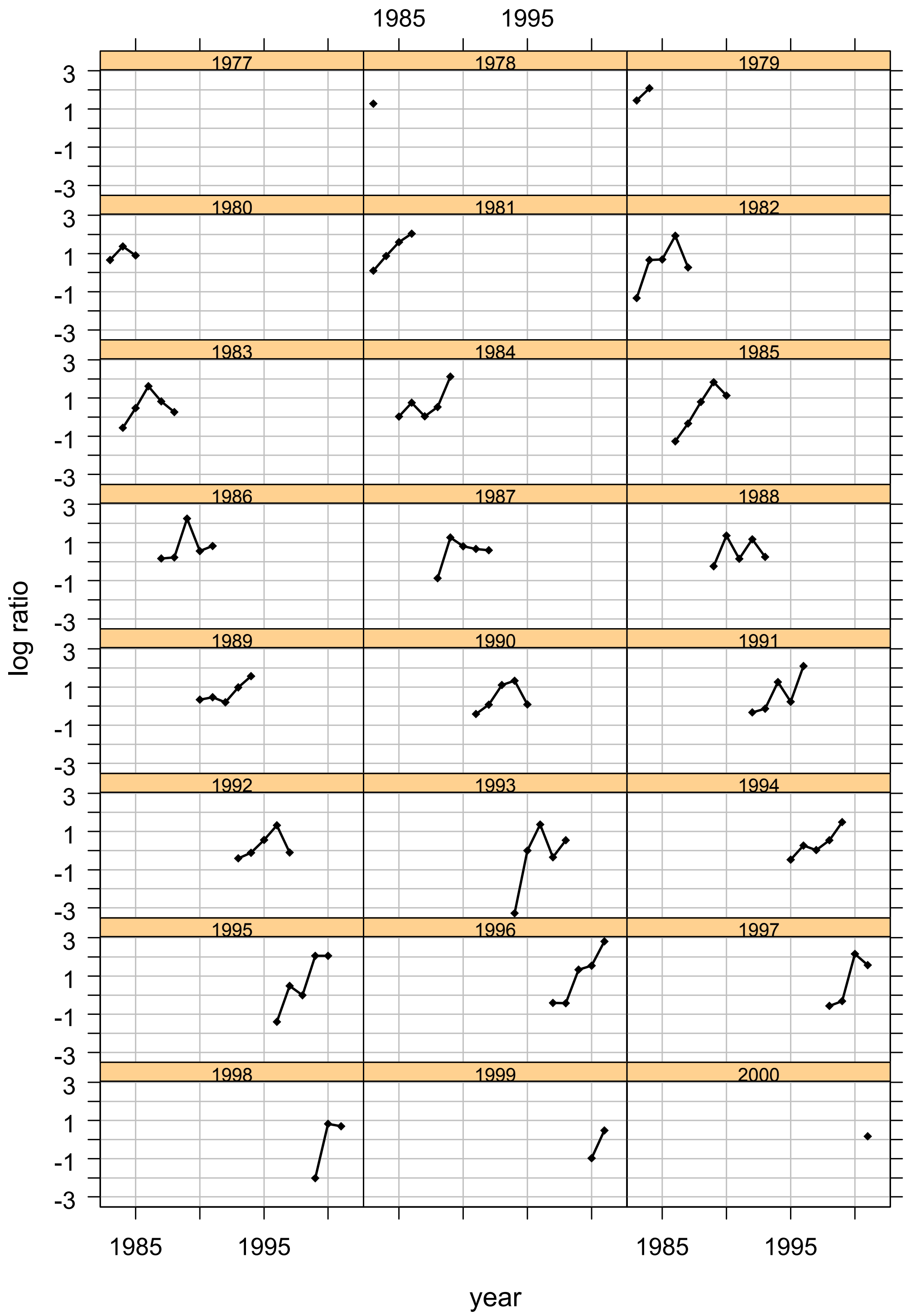


Figure 2.3.2.26 Blue whiting. Log catch ratios based on the Spanish pair trawl series (CPUE).

3 DESCRIPTION OF ASSESSMENT MODELS AND DIAGNOSTICS

This section of the report provides background on the three assessment modelling techniques that have been applied to Norwegian spring spawning herring and blue whiting. For Norwegian spring spawning herring, two models that have been applied by the working group are SeaStar and ISVPA (e.g., Figure 1.4.1), while for blue whiting, ISVPA and AMCI have been applied.

3.1 AMCI

The main building blocks in AMCI are:

1) **A population model** which projects the population forwards in time and distributes it on areas according to specified parameters - The usual exponential decay and Baranov catch equations are used, where fishing mortality is essentially separable. However, selectivity at age may be allowed to vary over time according to signals in the catches, and in the extreme this facility leads to a VPA-like algorithm.
2) **Observation models** that generate modelled counterparts to observed data (or data derived from observations) - Several options exist for modelling selection and catchability at age.
3) **Objective functions** that measure the fit of the modelled data to the observations, with optimisation routines to find the best fit - Components of the objective function could either have a maximum likelihood formulation or be based on more pragmatic measures of deviance. Flexibility exists for deciding which parameters to fix and which to estimate, allowing for judicious selection of estimable parameters based on the information content of the data. Evaluation of parameter uncertainty is based on either the delta method or bootstrapping techniques.

The design of AMCI places it closer to the category of 'statistical catch at age models', as proposed by Fournier and Archibald (1982), Methot (2000) and others, than to ‘tuned VPA’–type assessment tools such as XSA (Shepherd 1999) or ADAPT (Gavaris 1988). The methods used for modelling fishing mortalities and catchabilities have some similarity to those used for time series analysis (Gudmundsson 1994, Fryer 2002).

AMCI (source: ICES CM 2003/D:03)

Model	**AMCI**
Version	**2.3 (year: 2004)**
Model type	A separable model is applied to the whole assessment period. Selection can be allowed to change slowly according to the signal in the catches. The rate of change is determined by the user by specifying a gain factor for the influence of the current catch data. One extreme is then to keep the selection fixed. The population is projected forwards in time.
Selection	The selection at one age can be specified as the average over some other ages, but this specification cannot include any multiplier. The selection at oldest age is estimated unless it is linked by the user to some other age.
Estimated parameters	Recruitment, initial stock numbers, annual fishing mortalities, selection-at-age by year, catchability-at-age (and year), natural mortality, quarterly distribution of fishing, quarterly distribution of stock by area. The user decides upon which of these to estimate; the remainder are kept at fixed values.
Catchabilities	Catchabilities are in principle modelled as separable, but the age factor can be allowed to vary slowly using the same principle as for the selection-at-age in the catches. In practise, it will most often be kept fixed. Proportionality between index and stock abundance is always assumed. The proportionality can be fixed to the value one.
Plus group	The plus group is modelled as a dynamic pool. The fishing mortality assumed for the plus age can be estimated, or linked to some younger age. The fit of the modelled plus group is included in the objective function unless specified otherwise.
Objective function	There is a variety of objective functions available but most often, the weighted sum of squared log residuals is used. Weighting is decided by the user.
Variance estimates/ uncertainty	'Variances' of the parameter estimates can be derived from the Hessian, which is computed directly. There are also options for estimating uncertainty by parametric or non-parametric bootstrapping.
Other issues	AMCI allows the incorporation of tagging data and SSB indices as additional sources of data. It allows for multiple fishing fleets and multiple areas, defining local partial fishing mortalities. Distribution by area is specified as parameters but there is no migration model yet.
Program language	FORTRAN 77. No external libraries required.
References	Draft manual available but no formal publications yet.

3.2 ISVPA

ISVPA (Instantaneous Separable VPA) is essentially a separable VPA, but uses Pope's approximation of "instantaneous" catch (cohort analysis) instead of assuming a constant fishing mortality coefficient during the year as used in conventional VPAs. Models of the ISVPA group are similar in many respects to other separable cohort models and imply the existence of errors in catch-at-age data and in the separable representation of fishing mortality coefficients. However, their parameter estimation procedures are based on some principles of robust statistics that help to diminish the influence of error (noise) in the catch-at-age data on the results of the assessment. The solution can be guaranteed to be unbiased in a statistical sense. Special parameterisation of the model makes it unnecessary to use any preliminary assumptions about the age of unit selectivity and about the shape of selectivity pattern. This helps to obtain a unique solution in cases where catch-at-age data are noisy and auxiliary information provides conflicting signals or is not available. Otherwise, ISVPA may be used to estimate stock trends from catch-at-age data alone.

ISVPA (source: ICES CM 2003/D:03)

Model	**ISVPA**
Version	**Year:2004**
Model type	A separable model is applied to one or two periods, determined by the user. The separable model covers the whole assessment period
Selection	The selection at oldest age is equal to that of previous age; selections are normalized by their sum to 1. For the plus group the same mortality as for the oldest true age.
Estimated parameters	
Catchabilities	The catchabilities by ages and fleets can be estimated or assumed equal to 1. Catchabilities are derived analytically as exponents of the average logarithmic residuals between the catch-derived and the survey-derived estimates of abundance.
Plus group	The plus group is not modelled, but the abundance is derived from the catch assuming the same mortality as for the oldest true age.
Objective function	The objective function is a weighted sum of terms (weights may be given by user). For the catch-at-age part of the model, the respective term is: sum of squared residuals in logarithmic catches, or median of distribution of squared residuals in logarithmic catches MDN(M, fn), or absolute median deviation AMD(M, fn). For SSB surveys it is sum of squared residuals between logarithms of SSB from cohort part. For surveys; for age- structured indexes it is SS, or MDN, or AMD for logarithms of N(a,y) or for logarithms of proportions-at-age.
Variance estimates/ uncertainty	For estimation of uncertainty parametric conditional bootstrap with respect to catch-at-age, (assuming that errors in catch-at-age data are log-normally distributed, standard deviation is estimated in basic run), combined with adding noising to indexes (assuming that errors in indexes are log-normally distributed with specified values of standard deviation) is used.
Other issues	Three error models are available for the catch-at-age part of the model: errors attributed to the catch-at-age data. This is a strictly separable model ("effort-controlled version") errors attributed to the separable model of fishing mortality. This is effectively a VPA but uses the separable model to arrive at terminal fishing mortalities ("catch-controlled version") errors attributed to both ("mixed version"). For each age and year, F is calculated from the separable model and from the VPA type approach (using Pope's approximation). The final estimate is an average between the two where the weighting is decided by the user or by the squared residual in that point. Four options are available for constraining the residuals on the catches: Each row-sum and column-sum of the deviations between fishing mortalities derived from the separable model and derived from the VPA-type (effort controlled) model are forced to be zero. This is called "unbiased separabilization" As option 1, but applied to catch residuals. As option 1, but the deviations are weighted by the selection-at-age. No constraints on column-sums or row-sums of residuals.
Program language	Visual Basic
References	Kizner Z.I. and D.A.Vasilyev. 1997. Instantaneous Separable VPA (ISVPA). ICES Journal of Marine Science, 54 , N 3: 399-411 Vasilyev, D.A. (2001). Cohort models and analysis of commercial bioresources at information supply deficit. VNIRO Publishing: Moscow. Vasilyev D. 2003. Is it possible to diminish the impact of unaccounted time trends in age-structured surveys' catchability on the results of stock assessment by means of separable cohort models? ICES CM 2003/X:03. 13 pp.

3.3 SeaStar

The SeaStar (Stock Estimation with Adjustable Survey observation model and Tag Return data) assessment model is a conventional VPA using Pope's approximation where the cohorts may be tuned to survey data, tag return data, larval data (as proxy for the spawning stock) and 0-group data. SeaStar is fully statistically-based, i.e. there are no subjective weighting of the various tuning data. SeaStar is especially designed for Norwegian spring spawning herring, for which the recruitment dynamics is especially strong, resulting in larger dynamic range of year class strength than other fish stocks. Therefore, it is customary to use only the largest year classes in the tuning, linearly interpolating terminal F-values of the weaker year classes between those of the stronger. There is made no assumption about separability of the catches.

For Norwegian spring spawning herring it has traditionally been assumed that the survey selection is flat once a year class has recruited fully to an acoustic survey, although it is possible to define the survey selection in SeaStar rather freely. In connection with the present meeting a possibility for estimating a survey selection pattern along tuned cohorts was implemented.

Among the estimated parameters are catchabilities and distributional uncertainties of the tuning series', terminal F-values and initial tagging mortality. One may choose between normal, lognormal or gamma distributions for the tuning series'.

In order to assess how well the model can explain the data the total log-likelihood and number of terms in the likelihood, the fit between model and tuning series and ordered CDF values are used. In addition, the likelihood obtained by deleting one term at a time and the likelihood as function of each parameter keeping the other parameters constant has been diagnostics of usefulness in connection with the assessment of Norwegian spring spawning herring.

SeaStar (compiled by Sigurd Tjelmeland)

Model	**SeaStar**
Version	**Year: 2004**
Model type	Traditional VPA using Pope's approximation, but with an objective function based on maximum likelihood to include survey indices, larval indices, tagging data and an 0-group index.
Selection	The catches are not modelled
Estimated parameters	Terminal F values, catchabilities (survey and larval indices), tagging survival, initial tagging mortality, CV of assumed distribution (survey and larval indices).
Catchabilities	Catchabilities for survey and larval indices are estimable parameters within the model. There is considerable flexibility to, for example, specify catchabilities as functions of age or population abundance.
Plus group	The plus group is modelled as a dynamic pool. One may choose whether to include the plus group in the likelihood.
Objective function	The objective function is based on maximum likelihood, and comprises likelihood functions for survey and larval indices, and tagging data. Normal, lognormal or gamma distributions may be specified for survey and larval indices, and Poisson distributions for tagging data.
Variance estimates/ uncertainty	CVs of the assumed distributions for survey and larval indices are estimable parameters within the model. Analysis of assessment uncertainty is performed using parametric bootstraps.
Other issues	Can treat terminal Fs for only the strongest year classes as estimable parameters, with terminal Fs for the weaker year classes calculated by linear interpolation. Natural mortality can be constant or modelled as a function of predators. Sensitivity analysis can be performed by deleting one term at a time from the objective function.
Program language	Mathematica
References	www.assessment.imr.no

4 SIMULATION TESTING OF ASSESSMENT MODELS

The working group (WGNPBW) has previously evaluated the sensitivity of assessment outcomes to variations in input data and other assumptions about the models. However, in order to fully understand if the differences in assessment results for two models applied to the same stock are due to the models or nuances in the data, it is desirable to apply the models to exactly the same information, with known underlying characteristics (NRC 1998; ICES 2003b). By creating simulated data (with noise) using particular patterns of fishing mortality, selectivity characteristics of the fishery and surveys, and other attributes, including stock-recruitment relationships, natural mortality, and growth, the ability of a model to recover the "truth" can be evaluated. Previous testing with simulated data was undertaken by the Methods Working Group in 2003 (ICES 2003b), using ISVPA and AMCI, as well as some other standard assessment models applied elsewhere in ICES and North America. However, the methods Working Group only applied simulation tests to one set of data, and the SeaStar model was not subjected to that series of tests. In this section we describe the generation of three new sets of simulated data, and the application of AMCI, ISVPA and SeaStar to them. Additionally, in section 4.5 we describe the results of applying the SeaStar model to the simulated data used in earlier evaluations by the Methods Working Group (ICES 2003b).

Given the short time available to the working group (four days) it was difficult to generate consistent sets of simulated data, analyze those data using the three assessment techniques (e.g., three sets of data times three models = 9 outcomes) and have sufficient time to carefully analyze the outputs and diagnostics, relative to the problems at hand. Thus, this set of analyses must be considered somewhat provisional. If additional simulation testing is desired, then the machinery now exists with which to develop data series, to apply them, and to apply standard sets of analysis and diagnostic tests to evaluate output.

4.1 Simulation Data Sets

Three different data sets we simulated to provide challenges for the assessment models to the three main assessment model currently used for Norwegian spring spawning herring and blue whiting assessments. The data sets shared a common time period from 1974 to 2003 and age range from 0 to 16 years. The data were provided to the analysts in modified Lowestoft format data files for recorded catch at age, weight at age, fraction mature at age, two surveys covering different age ranges 1, 2, and 5—16. In addition to the output files containing data with errors true numbers, surveys and fishing mortality values are provided for reference and comparisons. The mean weights at age and fraction mature were fixed and are given in Table 4.1.1.

Table 4.1.1 Mean weights at age and fraction mature used for all the simulated data sets.

Age	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Maturity	0	0	0	0	0.3	0.9	1	1	1	1	1	1	1	1	1	1	1
Weight	0.001	0.018	0.025	0.075	0.15	0.223	0.24	0.264	0.283	0.315	0.345	0.386	0.386	0.386	0.382	0.382	0.407

4.1.1 Method used to generate the simulated data sets

Three different data sets were generated, one without any noise (data set 1), one with noise both in survey data and commercial catches (data set 2) and one in which the selection pattern of surveys and catches varied over time over the last 10 years, along with noise in survey and catch data (data set 3). The characteristics of the functions used to develop the data sets are provided in Figures 4.1.1-4.1.5. Thus, data sets 1-3 provided increasing complexity of challenge to the models, and data set 3 was a particularly diabolical case, but one probably very similar to a number of real world applications, including those faced by the working group. The underlying stock model was made close to the dynamics of the Norwegian spring spawning herring, and had the following characteristics:

Recruitment

Rather than being drawn from an underlying parametric stock-recruitment relationship, simulated recruitments were drawn from a multinomial distribution, with different mean-variance characteristics for "good" vs. "poor-average" year classes.

The probability for "good" recruitment was 1 in 8, and in the case of good recruitment the recruitment was drawn at random with a uniform probability between 50 and 100 units. In cases of poor to average recruitment, the recruitment was drawn at random with a uniform probability between 1 and 10 units.

Natural mortality
The instantaneous rate of natural mortality (M) is set to 0.15 for all age groups, which is input as a given (not estimated) by the three assessment programs.

Fishing mortality
In all cases, fishing mortality was assumed stable (with or without noise) for the first 15 years of the time series, followed by an increasing trend. The same selection as for the survey data (see below) was used with the addition that the selection at age 4 is set to 0.8, at age 3 to 0.6, at age 2 to 0.02 and at age 1 to 0.01. The selection pattern is multiplied with a scaling factor (0.15 before the change starts) and which in the last 10 years is interpolated linearly with time to 0.25 in the final year. The scaling factor is applied in all three data sets.

During the 2003 meeting of the NPBWWG one of the problems connected to the assessment of Norwegian spring spawning herring was identified as whether this stock is separable. Recruitment of large year classes from the Barents Sea to the fishery in the Norwegian Sea may generate non-separability due to the fishing fleet operating on different age components of the stock. The Norwegian fleet operates along the Norwegian coast, thus largely missing newly recruited year classes, while vessels from other countries operate in the Norwegian Sea. This effect was incorporated in data set 3.

Suppose there is a constant fishery of C_N individuals in the Norwegian Sea and C_C individuals along the Norwegian coast. Suppose that in a certain year a new year class of N_{young} individuals is recruited to the Norwegian Sea fishery but not to the fishery along the Norwegian coast. The number of older fish is denoted N_{old}. Disregarding natural mortality, for a fishery that is small compared to the stock the fishing mortality can be expressed as $F = \frac{C}{N}$. Suppose the young and older fish in the Norwegian Sea is perfectly mixed, then the ratio of F_{young} to F_{old} is :

$$\frac{F_{young}}{F_{old}} = \frac{\frac{C_N}{N_{young}} \frac{N_{young}}{N_{young} + N_{old}}}{\frac{C_C + C_N \frac{N_{old}}{N_{young} + N_{old}}}{N_{old}}} = \frac{N_{old}}{\frac{C_C}{C_N} N_{young} + N_{old} + N_{old}}$$

In the data generator C_C / C_N is set to 1.5, F_{old} is set to the F-value of 6 year old fish and F_{young} is calculated for ages 3-5. Figure 4.1.1 shows the ratio F_{young} / F_{old} as a function of the ratio of N_{young} to N_{old} using the construction above. Also shown as red dots are the realised F_{young} / F_{old} ratios for ages 3-5 in data set 3. The dynamic range was 0.15 – 0.40. In practice, the period during which newly recruited year classes are more available to the non-Norwegian fishing fleet than to the Norwegian fishing fleet may be variable. Here, it is assumed that this period lasts for 3 years.

In data sets 2 and 3 the catch is generated with assumed sampling error The error in catches was generated from the true catch using a gamma distribution with a CV of 0.3 with a probability of 0.9 and a CV of 0.9 with a probability of 0.1.

Generation of survey data
One survey for the adult stock (age 5-16) and one survey for the juvenile stock (age 1-2) were simulated. For the adult stock survey, the following model for the selection-at-age parameters for the survey were used:

+ 0.3 (1.0 – exp (-(age – 5))), age <= 5
2.3 – exp (0.072 (age – 10)), age > 5

R is interpolated linearly from 1.0 when the change starts to 0.0 at the end of the period. The effect is to gradually remove the downward part of the selection for older fish. The survey data were generated using an expected value of the true stock values multiplied by the above selection. Figure 4.1.2 shows this selection pattern, the dashed line is the selection pattern at the end of the period (data set 3). For survey 2 a flat selection pattern was used.

For data sets 2 and 3 a gamma distribution with a CV of 0.2 was used to generate noise in the survey data. Also, a year effect was added by drawing a random number uniform in (0.7,1.3) that was applied to all age groups in each survey. The year effects in each survey were independent. Additionally, the age distribution was multiplied by a normalised age distribution obtained by drawing from the multinomial distribution with N = 1000 and expected values equal to the original normalized age distribution. The rationale for this is that the relative uncertainty tends to be larger for weak year classes where the determination of age-disaggregated abundance relies on only a few scales or otoliths. This created not only additional variance, but also a bias that was nonlinearly dependent on abundance, (see Figures 4.1.3 and 4.1.4 for the relationship between abundance and bias).

Simulation
The underlying stock was started with a uniform age structure and run over 30 years to yield an age structure complying with the chosen recruitment model. During the subsequent 30-year data generation period, first the two survey numbers at age were generated.

The stock was simulated forward by using the true catch selection, (i.e. the catches were assumed to taken gradually during the year). The plus group was updated by adding the oldest true age group and the stock was reduced by natural mortality and fishing mortality, no senescent mortality was assumed.

With a probability of 0.1 the standard deviation of perceived catches and surveys were multiplied with 3 (outliers).

4.1.2 Data sets provided

Four data were supplied in 4 sub-directories (Set1, Set2, Set3 Set4). Only data sets 1,2 and 3 were used at the WG. The data files are in Lowestoft format the file names are as follows:-

CC	Catch in numbers at age
MAT	Fraction maturity in numbers
SC1	Survey 1 catch at age
SC1NoNoise	Survey 1 without survey noise added
SC2	Survey 2 catch at age
SC2NoNoise	Survey 2 without survey noise added
TrueCatch	True Catch without recording errors
TrueFvalues	True F exploitation on the stock
TrueStock	True stock values

The data used for the three simulations are shown in the Figures 4.1.2.1-4.2.1.22. These data sets are provided as a zip file along with the WG report. The figures that illustrate the data are given in the text table below.

	Set1	Set2	Set3
True Numbers at age	Figure 4.1.2.1	Figure 4.1.2.7	Figure 4.1.2.16
True Catch at age	Figure 4.1.2.2	Figure 4.1.2.8	Figure 4.1.2.17
Exploitation / Selection	Figure 4.1.2.3	Figure 4.1.2.9	Figure 4.1.2.18
Recorded Catch at age	Figure 4.1.2.4	Figure 4.1.2.10	Figure 4.1.2.19
Catch errors		Figure 4.1.2.11	Figure 4.1.2.20
Survey 1 with errors	Figure 4.1.2.5	Figure 4.1.2.12	Figure 4.1.2.21
Survey 1 without errors		Figure 4.1.2.13	
Survey 2 with errors	Figure 4.1.2.6	Figure 4.1.2.14	Figure 4.1.2.22
Survey 3 without errors		Figure 4.1.2.15	

Data set 3 is included for reference as the data used and reported in the WG report.

4.2 Assessment Model Application to Simulated data and Results

During the meeting of the Study Group, three teams of analysts developed stock assessments using data provided in sets 1, 2, and 3. Standard methods for comparison of assessment results with "true" output were developed, and we attempted to provide consistent diagnostics. However, the various models have different objective functions and fitting methods, so standard fitting diagnostics could not be developed. We have attempted to standardize the output as much as possible for the purposes of comparison. For each assessment model and data set, at a minimum, we provide:

1) comparisons of the "true" vs. calculated average fishing mortality rates (age 5-10 unweighted average), spawning stock biomass (SSB), and age 1 recruitment
2) log residuals of the calculated catch vs. the "true" catch by year and age
3) log population size (numbers, N) residuals, vs. "true" population sizes, by year and age,
4) log survey catch at age residuals (vs. "true" survey catches) for surveys 1 and 2.
5) Additionally, where appropriate the fitting surfaces, calculated exploitation patterns and a few other output comparison results of interest are provided.

The primary basis for comparing the three models and three data sets is the assessed vs. "true" F, SSB and recruitment, and overall evaluations of the performance of models with respect to these and other characteristics are presented in section

4.2.1 Application of AMCI to data sets

MODEL FORMULATION

Somewhat different choices of configurations were made for different sets of data, partly because different people did this in parallel since three implementations of AMCI were developed for application to the three simulated data sets.

A common problem was very slow convergence, and the choices made were to some extent attempts to improve the convergence of the model fit. The cause of this slow convergence is not clear. Some people decided to remove age 0, as there was no information about that in the data. Apart from this, the overall approach was to avoid more constraints on the model than absolutely necessary.

Normally, one would recommend to start with analysing the data as described in Section 2, and to formulate model and constraints according to the results from that analysis. Following that, further refinement of the model would have been made according to the diagnostics. Due to time constraints, this procedure was not followed, and the choices made here are not necessarily those that would have been made in a real assessment situation. The text table below shows the main settings for each data set.

Data	Ages	Constraints	Gain surveys	Gain catches
Set 1 (perfect)	1-16+	Survey 2 catchability at age 2 = age 1	No (constant catchability)	No (constant selection)
Set 2 (noisy)	1-16+	Survey 1 catcability at age 16 = age 15 Survey 2 catchability at age 2 = age 1	No (constant catchability)	No (constant selection)
Set 3 (dirty)	0-11+	F at +group is average of previous 2 ages	No (constant catchability)	0.3

MAIN RESULTS:

Re sults for the AMCI model are summarized in figures 4.2.1.1-4.2.1.14.

Set 1.

The age 0 was left out of the analysis to enable a more rapid convergence, since there is no information at this stage. Likewise, catchability at age 1 was set equal to catchability at age 2 for Survey 2. The assumptions made (constant catchability and constant selection, correct natural mortality) are all in accordance with the specification of the data. Under these conditions, a near perfect fit was obtained, the objective function (SSQ) being $7*10^{-6}$. The stock numbers were also reproduced almost exactly, with an over-estimate of stock numbers at old age and in late years, of up to 0.8%. Thus, when conditioned correctly, AMCI could reproduce noise-free data. No plotting of comparisons with "true" values or other diagnostics are provided because the data were fit so well.

Set 2.

This data set was analysed with the same model assumptions as for set 1, but the catchability at age 16 was constrained to be equal to that of age 15 in survey 1. The results were rather close to the true values (Figure 4.2.1.1). However, some of the noise in the data carried over to the fishing mortalities, and some year classes were either systematically over – or underestimated.

The catch residuals did not show any particular pattern (Figure 4.2.1.2), and the outliers in the data showed as large residuals. The residuals for Survey 1 had some year class effects in terms of positive residuals (Figure 4.2.1.4), while most of the residuals outside these year classes were negative without any particular pattern. Catchabilities for Survey 1 show a declining trend (Figure 4.2.1.6).

The year class patterns in the survey residuals should cause some suspicion. The catch residuals were not alarming. The probability response curve when screening over a range of terminal fishing mortalities show a shallow minimum from F at 0.2 to somewhat above 0.25.

Set 3.

Due to problems with getting the software running, the time to explore the dataset and the number of possible runs was restricted. The plus group in the data was reduced from age 16 to 11. Since there were no catches at age 0, age 1 was selected as the first age in the assessment. The fishing mortality at the plus group was set as the mean of the two last true age groups in order to reduce the number of parameters to be estimated by the model. The selection pattern in the first year was estimated by the model and a gain factor of 0.3 was allowed on the selection pattern in the following years. The catchability assumed in the surveys was constant. Both surveys were given weight 1 in the objective function (a survey has weight equal to one year of catch data).

In the first run using set 3, the survey data were excluded. These run gave unrealistic results.

FISHING MORTALITY AND EXPLOITATION PATTERN: Fishing mortality F(5-10) is fluctuating between years around 0.15 with no trend until the mid nineties and increased in the last 10 years to 0.25 in 2003 (Figure 4.2.1.8). The estimated exploitation pattern (Figure 4.2.1.13) is rather constant over the whole period showing a sharp increase between ages 5 and 8. The dip in the exploitation pattern at ages is not real and does not correspond to the true exploitation pattern and might be due to overparameterisation.

SPAWNING STOCK BIOMASS: A comparison of the estimated SSB with the true SSB is given in Figure 4.2.1.8. Spawning biomass peaks in 1979 due to a very strong year class 1973. The trends in stock biomasses estimated by AMCI are similar to those in the true stock. There is some tendency for AMCI to overestimate stock biomass.

STOCK NUMBERS AND RECRUITMENT: Residuals from estimated and true stock numbers indicate no particular problem areas in the model fit despite some year class effects (Figure 4.2.1.10). There are only a few very large residuals in the estimates of the recruiting year classes in the last 2 years. This can be expected since there is little information available to estimates these year classes; furthermore this information is conflicting (low survey index in the juvenile survey and large catch at age 1).

CATCH AND SURVEYS: The final run for set 3 includes both surveys. Residuals between estimated and observed catches were noisy but showed no trends in ages, years and year classes (Figure 4.2.1.9). No age or year effects are detected in the residuals of survey 1 (Figure 4.2.1.11). However there seem to appear trends in a number of year-classes in earlier years (Figure 4.2.1.11). The origins of these effects are not clear but may be related to outliers. Survey 2 (juveniles) shows large residuals which indicate very noisy data (Figure 4.2.1.12) and seems to give no consistent information.

SCAN OVER F. The minimum sums of squares between observed and estimated catch and survey data was found at a fishing mortality in the last year of F_{2003} = 0.30.

CATCHABILITY: Estimated survey catchability of survey 2 is given in Figure 4.2.1.14 and shows a strong decreasing trend from age 8 onwards.

COMPARISON WITH TRUE DATASET: This assessment provided a reasonably good estimate of SSB, level of Fishing mortality and Recruitment with the exception of the recruitment in the two most recent years. In a real assessment, recruitment in the most recent years would have received more scrutiny. The exploitation pattern at the older ages shows a decrease, which is absent in the true data.

Experience gained: In this case, reducing the plus group and fixing the fishing mortality on the plus group stabilized the result.

Experience gained

The problem that AMCI allows the user to over-parameterise the model is well known. There are no diagnostics that reveal this clearly. Slow convergence should be a warning. If there is a very close correlation between parameters the Hessian will become singular. The problem is wider than that, however, because it includes situations where an optimum of the objective function can be found, but almost entirely determined by the way residuals due to noise are balanced.

The sensitivity to deviations from the separable hypothesis was hardly explored in this study. The problem was partly avoided by excluding the oldest ages in data set 3. This was primarily done to reduce the number of parameter and improve the convergence in the optimisation, but would also be a sensible thing to do if there are doubts about e.g. sampling variance and age determination at old age, and the cohorts still are long enough to be adequately converged.

The plus-group poses a special problem in AMCI, since it is modelled as a dynamic pool with mortality and catchability that can be estimated in the optimisation process. The mortality signal in the plus group is embedded in the time trend of that group, and if this signal is weak compared to the noise, there is hardly any information about the mortality or the year classes entering the plus group. Any level of the plus group can be made compatible with the catches and survey data by adjusting the selection and catchability at that age. Therefore, normally, one would link both selection and catchability to that at younger age. This may not always be correct, however, and may also be misleading if the natural mortality in reality is different in the plus group compared to other ages. The problem is related to that experienced with two stage models (Collie and Sissenwine, 1983) as discussed in relation to the CSA model (Mesnil, 2004) at the Methods WG in 2004. In such models the ratio between the catchabilities at the young and old stage apparently cannot be determined within the model, and the choice of this parameter has a strong effect on the final stock estimates.

A note on quality aspects of assessments.
On the practical level this study confirmed that AMCI is very demanding with respect to formats of input files, which often makes setting up AMCI for a new data set a time consuming procedure where mistakes are easily made.

Different assessment programmes often need input data in a different format. Bringing the data in the required format frequently leads to mistakes, sometimes discovered at a late stage or may be not discovered at all. It is desirable to design a uniform data format to be accepted by all programmes, while maintaining sufficient flexibility for model specific data- or parameter requirements. Also it is desirable that all programmes should print all input- and estimated data used in the final assessment in order to be able to check that these were correct. Designing a menu-structured shell which controls the input and output of the programmes is expected to enhance the quality of the assessment process greatly.

Finally, it is clear from these studies that using AMCI properly as an exploratory tool is a substantial process, that cannot be accomplished in a short meeting.

4.2.2 Application of ISVPA to data sets

Results from the application of ISVPA are provided in Figures 4.2.2.1-4.2.2.19.

In the initial runs, ISVPA was applied to the three simulated sets in its most simple version. Further changes in the ISVPA settings resulted from attempts to get better signals from catch-at-age and both surveys (in cases when there were no distinct minima in partial loss functions of the model) and are described below.

Data set 1
The first ISVPA run was made with the following settings:

1) Catch-controlled version of the model attributing residuals in logarithmic catch-at-age to violations of separability assumption,
2) Condition of unbiased separabilizasion is applied.
3) Single selection pattern for all years (1974-2003) and all age groups in the model (1-16+). Zero age group was not used in the analysis since it does not include any catches.
4) Minimization of sum of squared residuals between logarithms of observed and theoretical catches.
5) Minimization of sum of squared residuals between logarithms of observed and model-derived abundances-at-age for survey 1.
6) Catchability-at-age q(a)=1 for all ages for survey 1
7) Minimization of sum of squared residuals between logarithms of observed and model-derived abundances-at-age for survey 2.
8) Catchability-at-age q(a)=1 for all ages for survey 2

Profiles of partial loss functions for these settings revealed minima in somewhat different positions for catch-at-age data and surveys (Figure 4.2.2.1). This was the reason for an additional run which included estimation of catchability-at-age and minimization of squared residuals between logarithms of observed (surveys) and model-derived age-proportions for surveys 1 and 2. In this case all sources of information gave similar signals about the stock (Figure 4.2.2.1).

The solution was found giving equal weights for partial loss functions (after bringing values to the same scale) in the overall loss function. ISVPA-derived estimates of F, SSB and R(1) are compares to the “truth” in Figure 4.2.2.2. Figures 4.2.2.3-4.2.2.6 present residuals of model approximation of catch-at-age, numbers-at-age and survey data.

Data set 2
The first ISVPA run for data set 2 was made with the following settings:

1) Catch-controlled version of the model attributing residuals in logarithmic catch-at-age to violations of separability assumption,
2) Condition of unbiased separabilizasion is applied.
3) Single selection pattern for all years (1974-2003) and all age groups in the model (1-16+). Zero age group was not used in the analysis since it does not include any catches.
4) Minimization of sum of squared residuals between logarithms of observed and theoretical catches.
5) Minimization of sum of squared residuals between logarithms of observed and model-derived abundances-at-age for survey 1.
6) Catchability-at-age q(a)=1 for survey 1 was estimated within the model
7) Minimization of sum of squared residuals between logarithms of observed and model-derived abundances-at-age for survey 2.

8) Catchability-at-age q(a)=1 for all ages for survey 2

Profiles of partial loss functions for the model settings listed above revealed a rather good minimum only for survey 1, while for catch-at-age data and survey 2 minima were deteriorated (Figure 4.2.2.7).

In order to make them better by diminishing the influence of noise in the data on the signal, for catch-at-age data and survey 2 minimization of SSE was substituted by minimization of the median of distribution of squared logarithmic residuals (MDN). This gave a minimum for catch-at-age data coherent to signal from survey 1, but for survey 2 the signal was still deteriorated (Figure 4.2.2.7).

The signal from survey 2 appeared, and it was coherent to signals from catch-at-age and survey 1, when for survey 2 minimization of absolute median deviation (AMD) was applied. (AMD is the median of distribution of deviations of residuals in logarithmic abundance-at-age from their median value). AMD in some cases (asymmetric distributions) is referred to be more robust with respect to MDN, but can be less sensitive. Profiles of partial loss functions for such a settings of the model are shown in Figure 4.2.2.7.

The solution was found giving equal weights to partial loss functions (after bringing values to the same scale) in the overall loss function. ISVPA-derived estimates of F, SSB and R(1) for data set 2 are compares to the "truth" in Figures 4.2.2.8. Figures 4.2.2.9-4.2.2.12 present residuals of model approximation of catch-at-age, numbers-at-age and survey data.

Data set 3

The following ISVPA settings were used for the first run:

1) Catch-controlled version of the model attributing residuals in logarithmic catch-at-age to violations of separability assumption,
2) Condition of unbiased separabilizasion is applied.
3) Single selection pattern for all years (1974-2003) and all age groups in the model (1-16+). Zero age group was not used in the analysis since it does not include any catches.
4) Minimization of sum of squared residuals between logarithms of observed and theoretical catches.
5) Minimization of sum of squared residuals between logarithms of observed and model-derived abundances-at-age for survey 1.
6) Catchability-at-age q(a)=1 for survey 1 was estimated within the model
7) Minimization of sum of squared residuals between logarithms of observed and model-derived abundances-at-age for survey 2.
8) Catchability-at-age q(a)=1 for all ages for survey 2

As it was observed in the case of data set 2, profiles of partial loss functions for the model settings listed above revealed a minimum only for survey 1,while for catch-at-age data and survey 2 minima were deteriorated (Figure 4.2.2.13). Application of median minimization (MDN) for catch-at-age data and survey 2 resulted in appearance of minima only for catch-at-age data. Main (deepest) minimum was accompanied by a number of local more flat minima (Figure 4.2.2.13). For median measures it is a rather common situation in dealing with noisy data. Appearance of signal from survey 2 resulted from application of AMD as a criterion of goodness of fit (Figure 4.2.2.13). Signals from catch-at-age now were coherent, while signal from survey 1 corresponded to somewhat lower values of fishing mortality in the terminal year.

As in previous cases, the solution was found giving equal weights to partial loss functions (after bringing values to the same scale) in the overall loss function. ISVPA-derived estimates of F, SSB and R(1) for data set 3 are compares to the truth in Figure 4.2.2.14. Figures 4.2.2.15-4.2.2.18 present residuals of model approximation of catch-at-age, numbers-at-age and survey data.

Residuals in survey 1 for all data sets contain cohort effects caused by more complex structure of simulated data in comparison to the model, but this should not be overestimated because in the model survey data are acting "integrally" determining the choice of the value of terminal fishing mortality coefficient.

ISVPA-derived estimates of selection pattern for the three data sets are presented in Figure 4.2.2.19 (the values are normalized to unit by their sum).

4.2.3 Application of SeaStar to Simulated Data

Results for SeaStar fits to the three simulated data sets are provided in Figures 4.2.3.1-4.2.3.12.

SeaStar has only been used for Norwegian spring spawning herring. For this species, where fish recruit to the surveys by migration and not by growth, it has been assumed that the survey catchability is independent of age once each year class has recruited fully to the survey. Consequently, age dependent catchability has not been implemented as such. However, there is flexibility in constructing a user-defined catchability, which was used to define a parabolic dependence of catchability on age during the meeting. This did not work very well for data set 2, for which the estimated parabola was realised as a monotonically decreasing function, which is not realistic. After the meeting adaptation to a possible age-dependent selection was implemented by fitting a third-order polynomial along the tuned cohorts. This approach gave a much better fit between model and survey data for the tuned cohorts and also a greatly increased likelihood. However, this approach can only be used when there are sufficiently many data points along all the tuned cohorts, i.e. only for data set 2.

In order to avoid that large relative uncertainties connected to weak year classes of Norwegian spring spawning herring propagate through the assessment it is customary to represent only the strongest year classes with free terminal F-values. The terminal F-values for non-tuned year classes are interpolated between those of the strongest year classes. Terminal F-values for year classes younger than the youngest tuned year class are interpolated to 0 at age -1. The selection of year classes to be tuned

Because SeaStar has a survey model only along tuned cohorts, residual plots for surveys are not presented, rather plots of modelled vs measured survey data are used as indicators of goodness of fit.

The CVs in the assumed gamma distribution have been estimated separately for the two surveys.

Data set 1

Tuned year classes: 1973, 1974, 1977, 1978, 1979, 1981, 1982, 1984, 1986, 1988, 1992, 1994, 1997

Figure 4.2.3.1 shows the comparison between the true and perceived (modeled) Fs, SSB and recruitments. There is a slight overestimate of mean F. There is a slight underestimate of SSB in the middle of the time series. The model reproduces the recruitment exactly, except for a small underestimate the first year and an appreciable underestimate the last two years. A deviance the latest years is trivial and would be expected since the youngest year classes are not among the tuned year classes. Figures 4.2.3.2 and 4.2.3.3 show the perceived historic stock and perceived data along the tuned cohorts for surveys 1 and 2, respectively. The correspondence is good. Figure 4.2.3.4 shows the difference (not log-scale) between true and estimated historic stock. The residuals are generally small, but there is a considerable deviance along the 1973 cohort, the reason for this is unknown.

Data set 2

Tuned year classes: 1972, 1973, 1977, 1980, 1985, 1988, 1990, 1992, 1994, 2000.

This data set contains several large year classes and seems to be closer to the dynamics of Norwegian spring spawning herring than the other two.

Figure 4.2.3.5 shows the comparison between the true and perceived Fs, SSBs, and recruitments. During the period of constant F the perceived F-value exhibits an increasing trend while fluctuating around the true value during the period of increasing true F. There is an underestimate for the lowest SSB values, but the correspondence is generally good. There are deviances for the largest year classes and the expected deviance for the youngest non-tuned year classes. Figures 4.2.3.6 and 4.2.3.7 show the perceived historic stock and perceived data along the tuned cohorts for surveys 1 and 2, respectively. The model seems to capture the survey trends well. Figure 4.2.3.8 shows the difference (not log-scale) between true and estimated historic stock.

Data set 3

Tuned year classes: 1967, 1972, 1973, 1991, 1998, 2000, 2001, 2002

Figure 4.2.3.9 shows the comparison between the true and perceived Fs, SSBs, and recruitments. Fishing mortality rates are overestimated by the model, but the trend tracks the “truth” fairly well. The model tracks the true SSB trend reasonably well, with a slight underestimate the latest years. There is good correspondence in recruitment, except for the untuned 2003 year class. Figures 4.2.3.10 and 4.2.3.11 show the perceived historic stock and perceived data along the tuned cohorts for surveys 1 and 2, respectively. The correspondence is good, except for the 1967 year class, for which there is an underestimate for the youngest ages. Figure 4.2.3.12 shows the difference (not log-scale) between true and estimated historic stock.

4.3 Influence of perturbations in input variables on assessment results

A series of perturbation tests were undertaken to examine the effects of variations in input data on the results of the analyses. Because of the press of time in the meeting, these perturbation tests were not extensive (across the various data), nor were they all together consistent. In order to simulate the effects of increasing catches, a 10% increase in the terminal fishing mortality rate was evaluated to see how the residuals for the various data sets responded. Results from these perturbations are provided in Figures 4.3.1-4.3.12.

AMCI

These tests were run for AMCI results for series 2 and 3. Increasing the Fishing mortality in the terminal year by 10% induced little change in the residuals in the last year and on the older ages in the most recent years (Figures 4.3.1-4.3.6). Data set 3 appeared to show more change than set 2, and this was particularly visible in the pattern of survey residuals.

ISVPA
Figures 4.3.7-4.3.9 provide the perturbation test results for ISVPA with data set 3. In this case there was a rather strong response in catch residuals with a significant effect across the 1988 year class. Similar to AMCI results, there were obvious impacts on survey residuals for surveys 1 and 2.

SEASTAR
SeaStar results for perturbation tests are provided for all three data sets (Figures 4.3.10-4.3.12). In this case, the perturbation results are only provided for the response in estimated numbers at age. Clearly, the addition of catches in the terminal year has strong and trended implications for the assessment results in all three data sets, with the results restricted primarily to the upper right hand corner of the numbers at age matrix for set 3.

These analyses, although only preliminary in nature and not entirely consistent, demonstrate that even minor (e.g. 10%) differences in catches can have significant implications for the veracity of assessment results. A more formal analysis of perturbations requires that multiple sources of error be investigated. Importantly, such investigations should not be undertaken by perturbing one variable (e.g. catch, M, weights at age) at a time, since synergy between small aberrations likely leads to unintended and non-linear effects on assessments. Isolating these effects is an interesting academic exercise, but understanding how all of these potential changes interact is more important for real-world problems.

If additional perturbation analyses should be undertaken by the Working or study groups, it is recommended that a formal factorial design be used to guide such studies, and that they be well coordinated ahead of time.

4.4 Inferences from Simulation Studies using Assessment Models

The Study Group undertook a considerable amount of work in a relatively short time span to attempt to put the three models through their paces and to see if the conflicting assessment results were somehow a result of deficiencies in one or more of the modelling approaches. The simulations attempted were of increasing complexity, incorporating a number of diabolical and confounded properties, including noise in survey and catch data (reflective of sampling variation), bias in the survey due to year effects, and time trends in fishing mortality. Recruitment was quite variable in the simulations, and a number of strong and corresponding weak year classes were simulated.

Overall, the three model approaches (AMCI, ISVPA, and SeaStar) were able to recover the "true" trends in fishing mortality, spawning stock biomass and recruitment rather well. There were some obvious differences between model performances. Below we provide some detailed comparisons between simulation results:

Set 1:
Set 1 comparisons are based on Figures 4.2.2.2 and 4.2.3.1 (ISVPA and SeaStar). No corresponding figure is presented from AMCI, as the estimated values tracked the true values very closely. Both ISVPA and SeaStar tended to slightly overestimate F for most of the time series, although ISVPA underestimated terminal F very slightly. Additionally, both ISVPA and SeaStar slightly underestimated SSB in the middle part of the time series as the stock declined from the large year classes in the early part of the series. Both models estimated the large year class at the beginning of the series, and generally correctly identified the recruitments. ISVPA slightly overestimated recruitment in the terminal year, whereas SeaStar underestimated R in the terminal year.

Set 2:
Set 2 comparisons are based on Figures 4.2.1.1, 4.2.2.8, and 4.2.3.5. All of the assessment models overestimated fishing mortality slightly, but were able to recover the trend of increasing fishing mortality. In the case of AMCI, in 17 of 29 cases, estimated Fs exceeded the truth, while for ISVPA estimated Fs exceeded true Fs in 24 cases, and for SeaStar in 21 cases. All three models underestimated terminal F. AMCI provided relatively precise estimates of SSB throughout the series, with ISVPA and SeaStar showing the proper overall trend but some underestimation of SSB following the very string year classes in the earlier part of the series.

Recruitments were generally well estimated by all three methods. For terminal recruitment, AMCI slightly overestimated R, while ISVPA estimated the true terminal R well, and SeaStar underestimated R.

Set 3:
All three models were able to detect the change in mean fishing mortality beginning in 1994 (Figures 4.2.1.8; 4.2.2.14; 4.2.3.9). For AMCI, the estimated Fs exceeded the true Fs in 10 of 29 cases, but the estimated values fluctuated above and below the true values without apparent trend. The terminal estimate of F from AMCI was below the true F. For ISVPA the estimated F was below the true F in all but 3 of the 29 years, and the terminal F was about 1/3 below the true value. For SeaStar, most of the estimated F values exceeded the true F, and the terminal F value exceeded the true terminal F by a modest amount.

SSB patterns were consistent with the F patterns identified above. For AMCI there was a slight tendency to overestimate SSB in the last several years, but differences were slight. ISVPA overestimated SSB throughout the series, but not greatly, and the trend was preserved. SeaStar provided close approximations of SSB for the entire series, with a very slight tendency to underestimate in the last four years.

All three models provided relatively good estimates of recruitment. In the terminal year AMCI and SeaStar underestimated recruitment while ISVPA overestimated recruitment.

Overall all three models correctly recovered the trends in F and SSB and were able to reconstruct the year class structure in set 3. There were some differences in the absolute levels of F and SSB, and performance in terminal year recruitment.

Population abundance at age in the terminal year for the three model outcomes and the three data series are provided in Figure 4.4.1. The only important differences in these estimates are for recruitment and age 2 abundance, which (as noted above) tends to be underestimated in SeaStar but variably estimated by the other models.

All three models were able to recover the essential details of the three simulation experiments, with some interesting differences among them. For the most part, no one model seemed to under- or over-estimate F or SSB on a consistent basis, and thus the differences seemed to be associated with how the various models handled the details. For example, while ISVPA overestimated SSB and underestimated F in set 3, the opposite was true for set 2. There is perhaps an indication that SeaStar may underestimate recruitment in the terminal year, but this is by no means a definitive result.

These simulation results, and those undertaken by the Methods Working Group (see section 4.5 below), all indicate that the source of disparity in model results that have plagued the working group in recent years are not likely the result of models that are inherently disposed to providing under- or overestimates of management parameters. Rather, it seems clear that the differences are likely due to how the models (and their analysts) interact with imperfect and conflicting data sets. Thus, attention to the details of data coherence (section 2 of this report) seems more important than to resolving model differences. That being said, there are a number of important issues related to the three models that were revealed in the study group's travails.

The work of the Study Group was not sufficient to reveal the exact source of the disparities between assessments that were the source of disagreement between model results in the working group (e.g., Figures 1.4.1 and 1.4.2). The simulation experiments were probably too simple to reveal the underlying source of the problems encountered. If it is desired by ACFM and others to isolate these issues in more detail, then more realistic simulations, incorporating multiple tuning indices at age, variable degrees of correlation between the tuning indices, and noise in catch-at-age data need to be tested. In the end, we suspect that these are the sources of the disparities. One potential way to reveal such conflicts would be to conduct the assessments using each index one-at-a-time. This would provide information on the amount of leverage exerted by each index taken into the assessments. The problem with such an approach using the actual data from the assessments, is that the analysts would, *a priori*, know the overall result and the results from the comparison between the methods. Since the assessment results involve complex interactions between software and assumptions by analysts, the results would not necessarily isolate differences due to the algorithms used. If such experiments need to be pursued, it is suggested that considerable thought and attention be devoted to the development of appropriate test data sets prior to any additional meetings on the subject.

None of the three models are considered "main-stream:" assessment approaches and so do not have wide use and dissemination in the stock assessment community. Resultantly, the models required considerable interaction with their creators in order to undertake these analyses. We suspect that others wishing to undertake analyses with these models would encounter similar problems. Part of the confusion regarding assessment results stems from the fact that analysts and peer reviewers are generally unfamiliar with these models, and so cannot appreciate how they are working, and, most importantly, how subjective decisions by the analyst change assessment outcomes.

With respect to the Northern Pelagic and Blue Whiting Working Group, there is a need to adopt assessment approaches that are flexible, well documented and supported, and transparent. The desire to use non-standard assessment inputs (e.g. tagging data) means that standard VPA type analyses are inadequate for one or more assessment uses. This need for flexibility must be preserved in any approach that may succeed the methods currently in use. While the Study Group cannot provide definitive evidence that one of the three tools considered provides better performance than the others, it considered all approaches somewhat deficient with respect to one or more of the above criteria. Accordingly, the study group recommends that work commence on the development of a flexible, integrated assessment approach that allows more transparency, training, documentation and ease up the learning curve. An exploratory analysis using AD Model Builder is provided in the Appendix. This is not the only such flexible modelling approach, but is one that is well supported and has many implementations world-wide.

4.5 Extension of the 2003 Methods Working Group Simulations to the SeaStar Model

SeaStar was run on the same simulated (noisy) data set as was used for ISVPA and AMCI at the 2003 Methods WG meeting. The survey selection was modeled as a third order polynomial, different for each of the two surveys. The natural mortality was set to 0.2 for all ages. Only the largest year classes as subjectively judged from the catch and survey data were tuned and the terminal F for younger year classes was linearly interpolated between those of the largest year classes. The survey errors were assumed to follow the gamma distribution with a constant CV, which was estimated independently for the two surveys.

Comparisons of SeaStar model results for fishing mortality and SSB with "true" data provided in the Methods Working Group Report (ICES 2003b) are provided in Figures 4.5.1 and 4.5.2. [Note that the "true" data were drawn from the Methods Working Group Report by eye].

Overall, SeaStar slightly underestimated fishing mortality and overestimated SSB, relative to the "true" data. This pattern was evident for the three simulations, with the non-trended data set showing the closest correspondence for fishing mortality and the tuning with both fleets combined providing the closest representation of the SSB data. The trends in F and SSB were well represented by SeaStar in all analyses with simulated data.

Based on comparison with results provided by the Methods WG (ICES 2003b) for ISVPA and AMCI, all three models had very similar patterns of slightly underestimating F and slightly overestimating SSB (see Figures 4.5.1 and 4.5.2 in this report and Figures 8.3.3 and 8.3.4 in ICES 2003b). All models captured the proper time trends in "true" SSB and F.

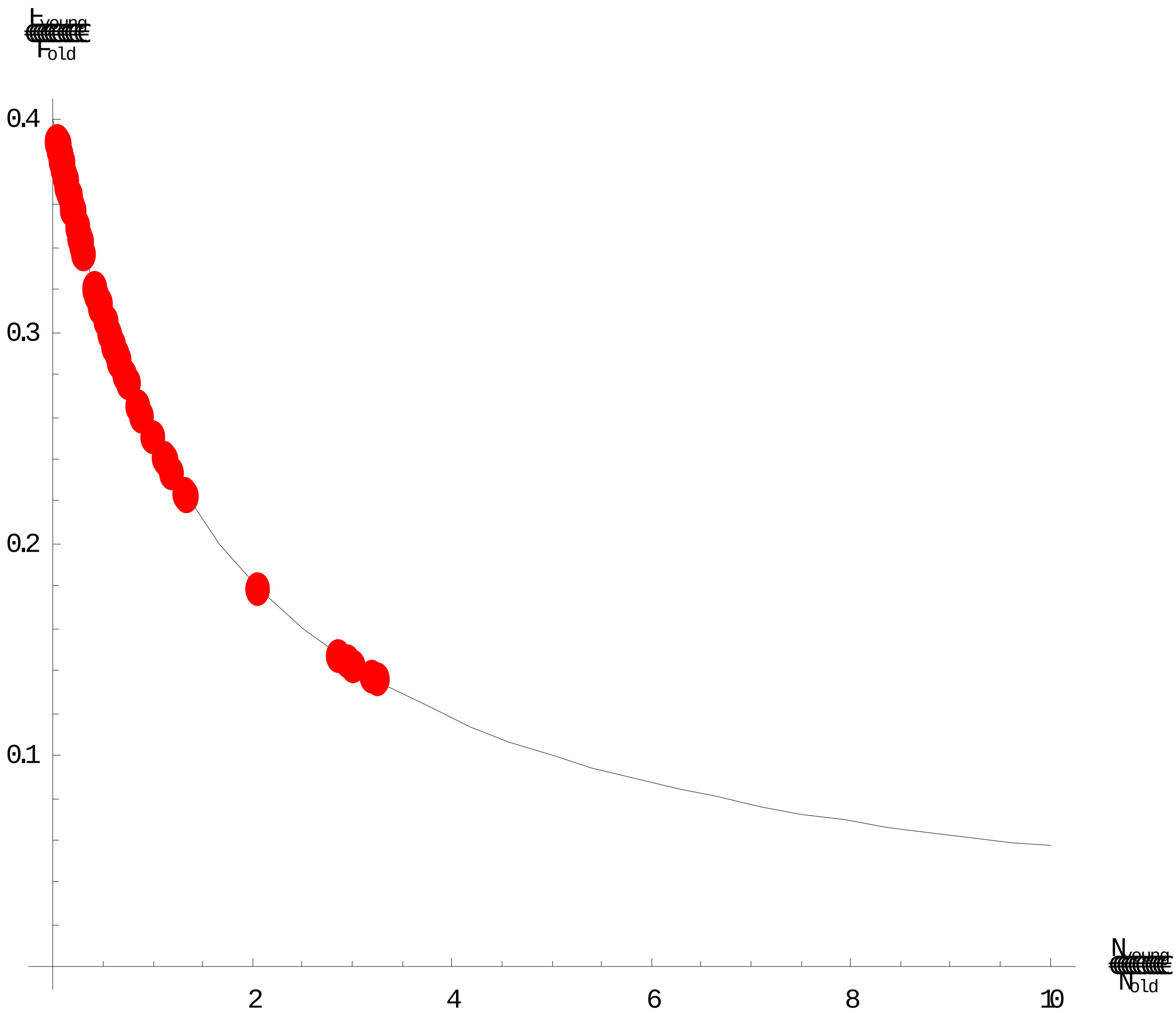


Figure 4.1.1. Assumed reduction of F-values for fish of ages 3-5 in relation to fish of age 6 as function of abundance in relation to total abundance of fish of age 6 and older. Red dots show the values that were realised in data set 3.

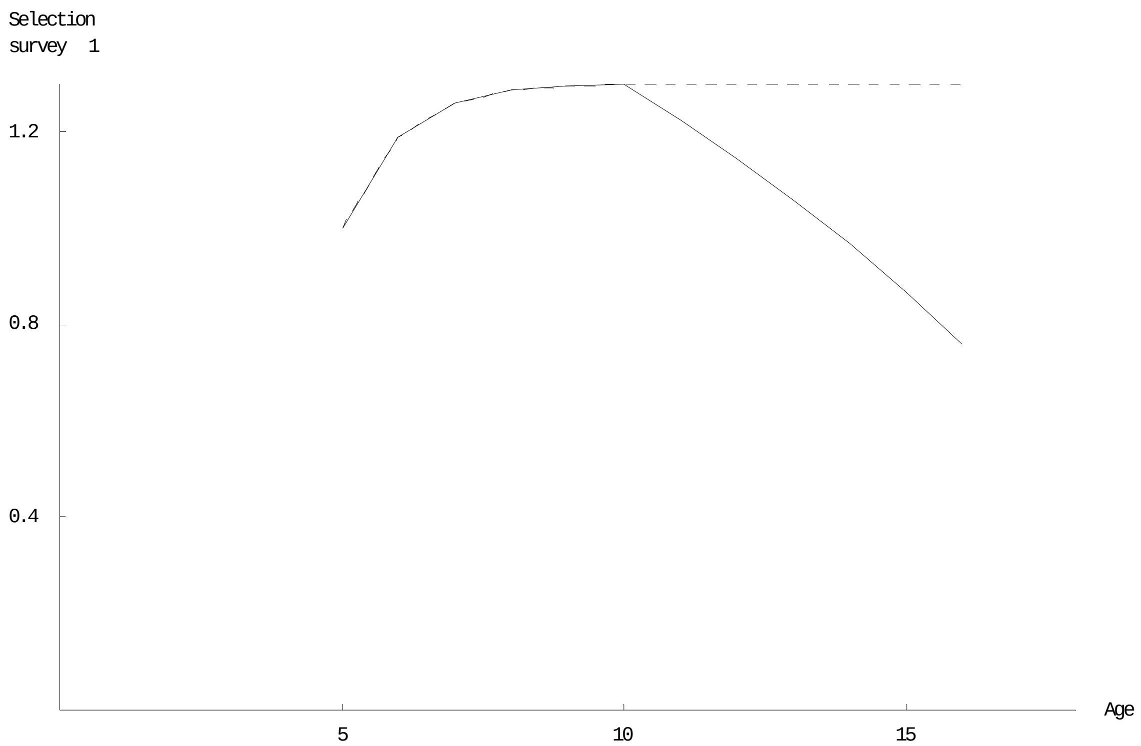


Figure 4.1.2 Selection pattern for survey 1. The dashed line shows the selection pattern at the end of the historic period (data set 3).

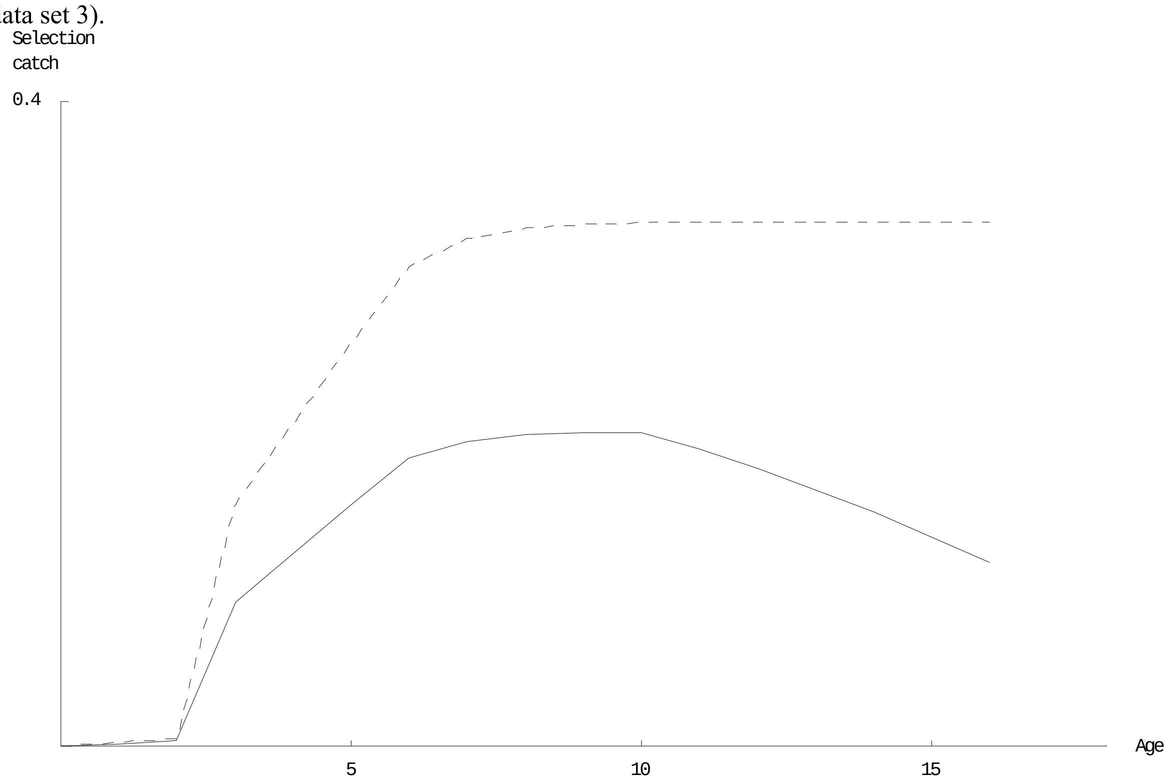


Figure 4.1.3. Selectivity of catches. The dashed line shows the selectivity at the end of the period (data set 3).

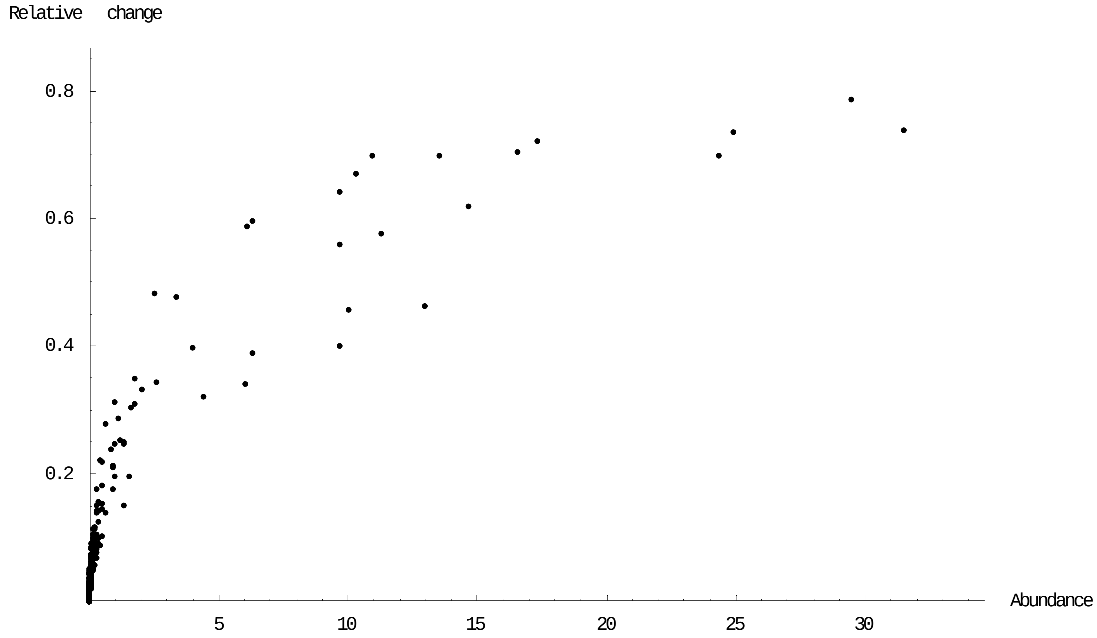


Figure 4.1.4. Ratio of survey abundances before and after the final step in uncertainty was applied (draw to the multinomial distribution), survey 1.

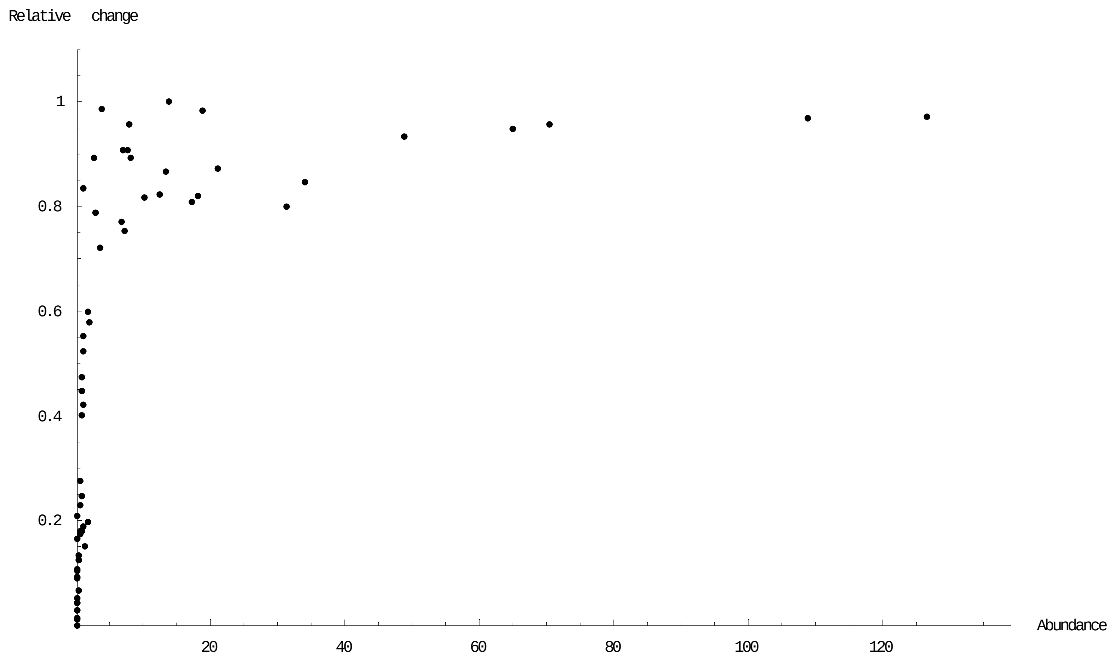
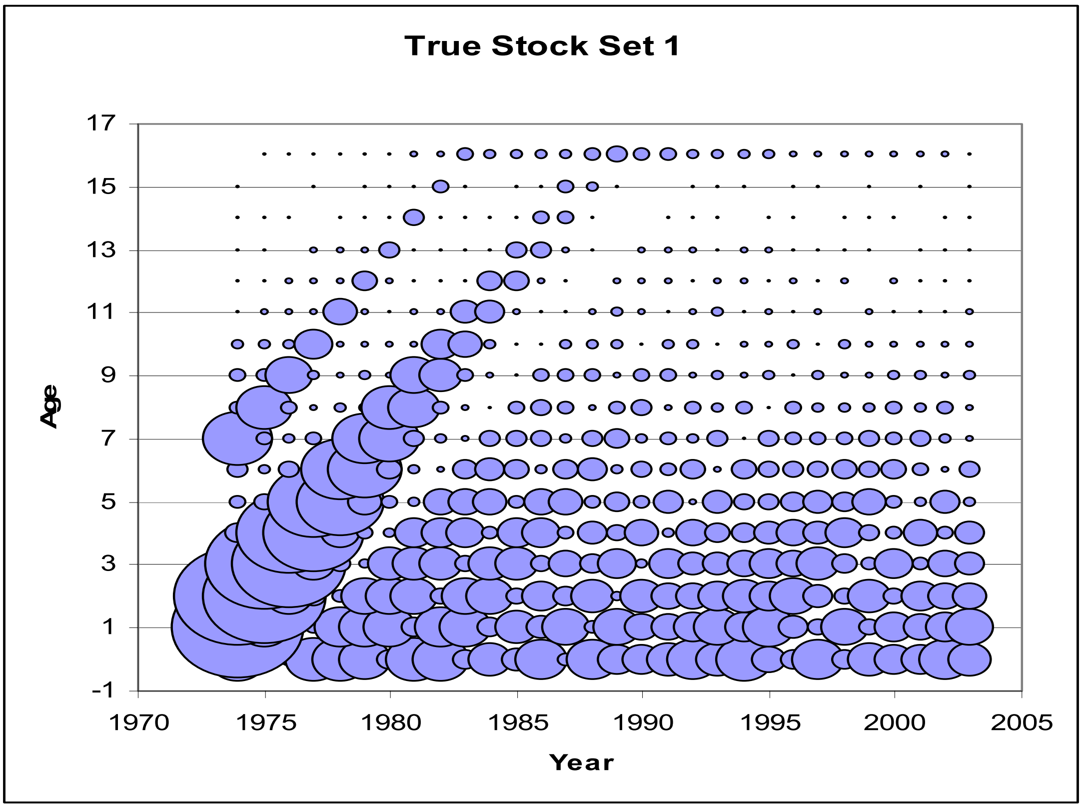


Figure 4.1.5. Ratio of survey abundances before and after the final step in uncertainty was applied (draw to the multinomial distribution), survey 2.



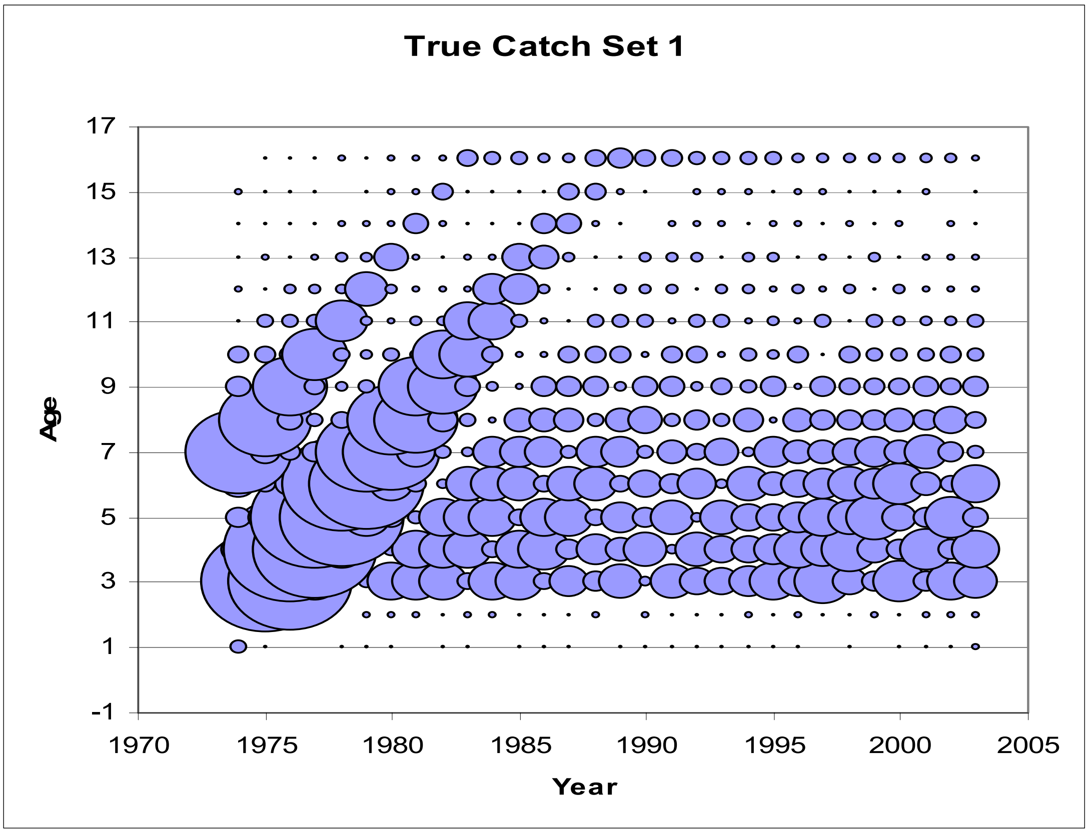


Figure 4.1.2.2 True catches taken in simulated data set 1.

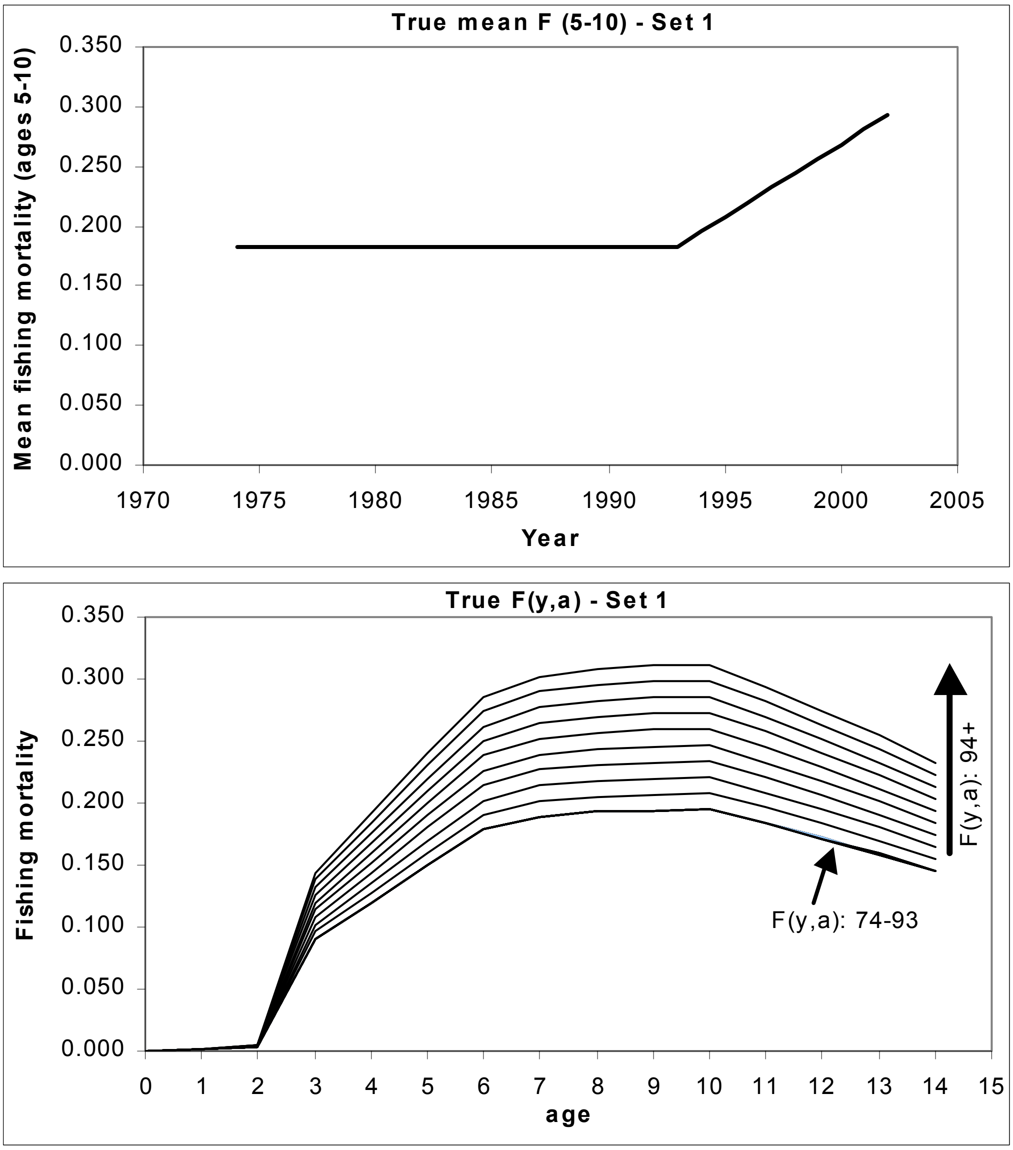


Figure 4.1.2.3 True Fishing Mortality and selection pattern for simulated data set 1.

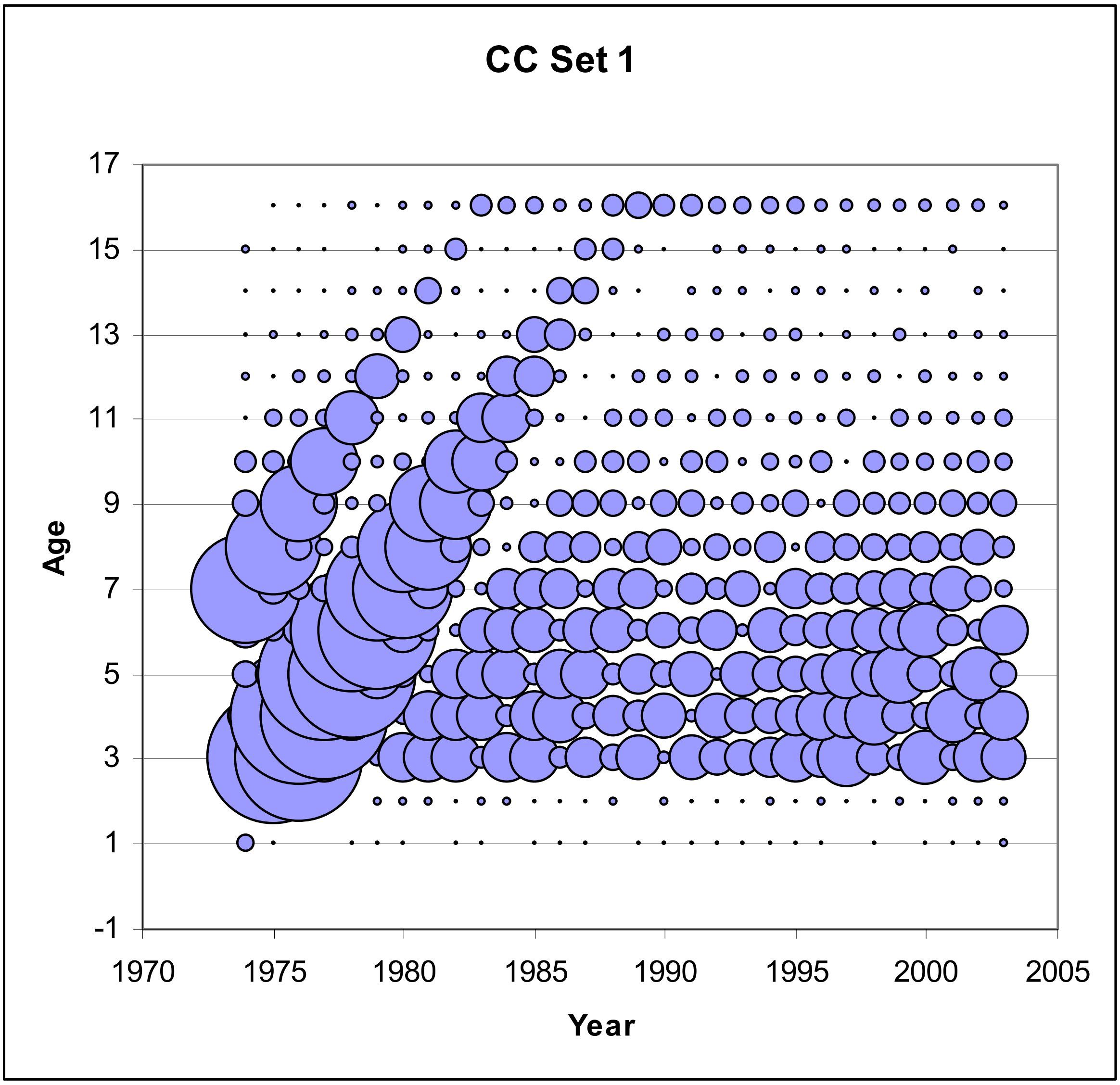


Figure 4.1.2.4 Recorded Catch for test set 1 (without error).

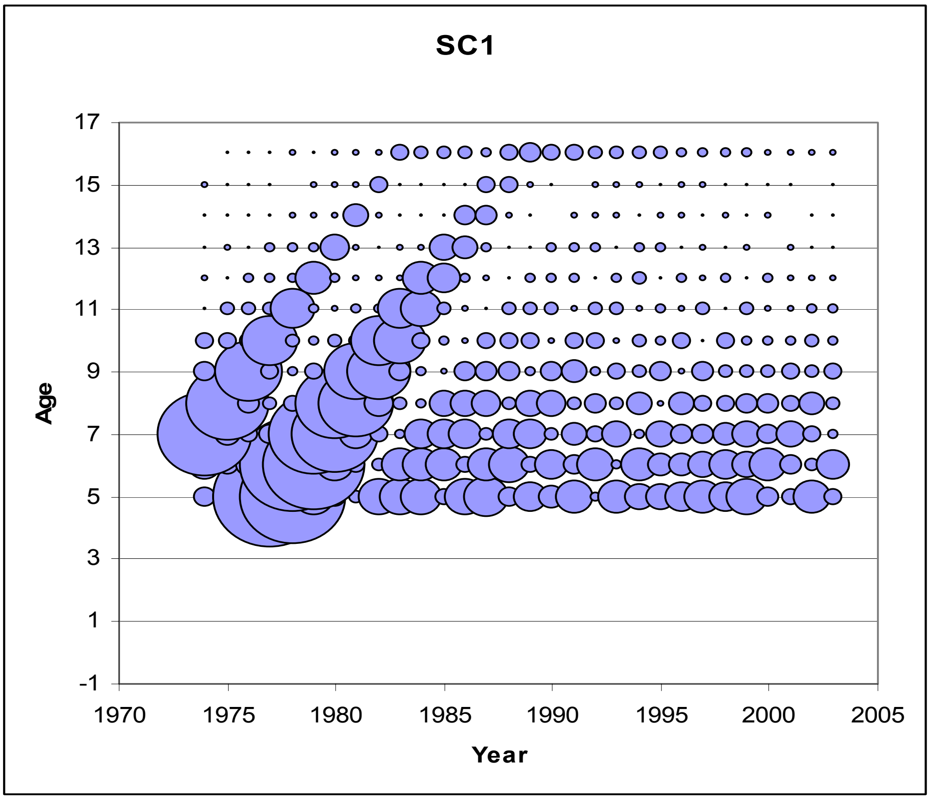


Figure 4.1.2.5 Adult Survey 1, with error data set 1.

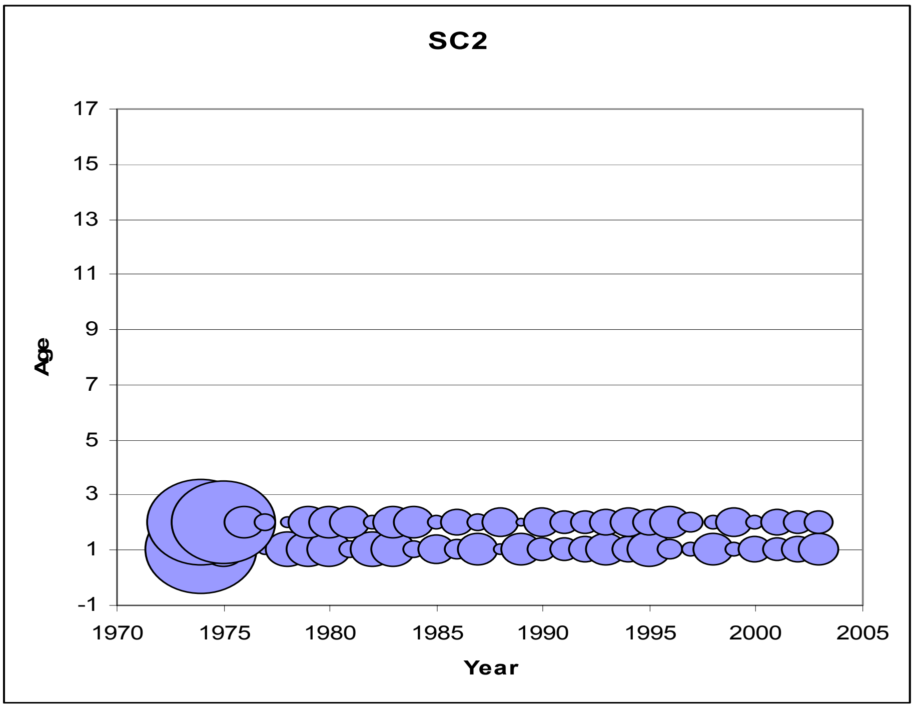


Figure 4.1.2.6 Survey 2 Juvenile survey with error data set 1.

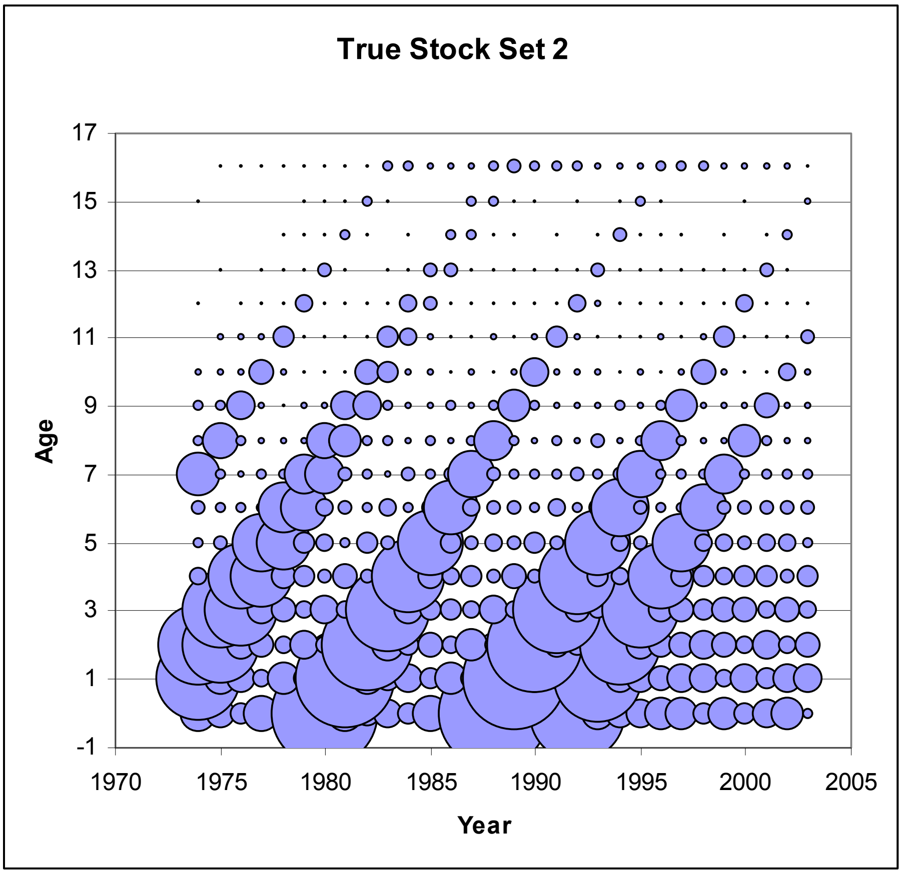


Figure 4.1.2.7 True stock in numbers for simulated data set 2.

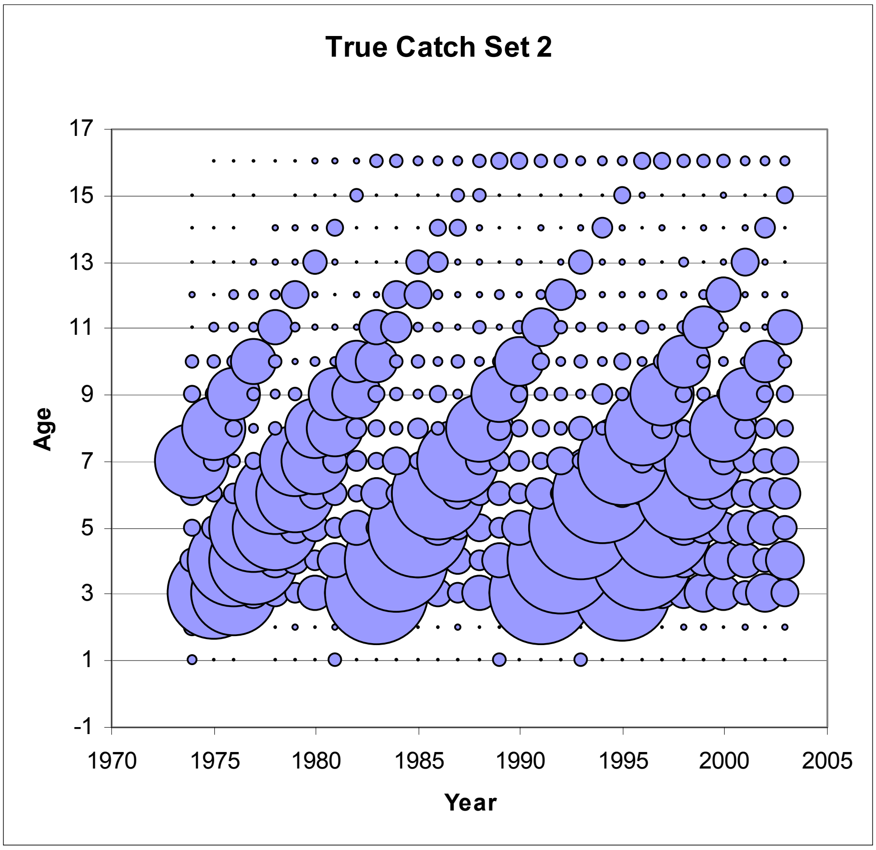


Figure 4.1.2.8 True catches taken in simulated data set 2.

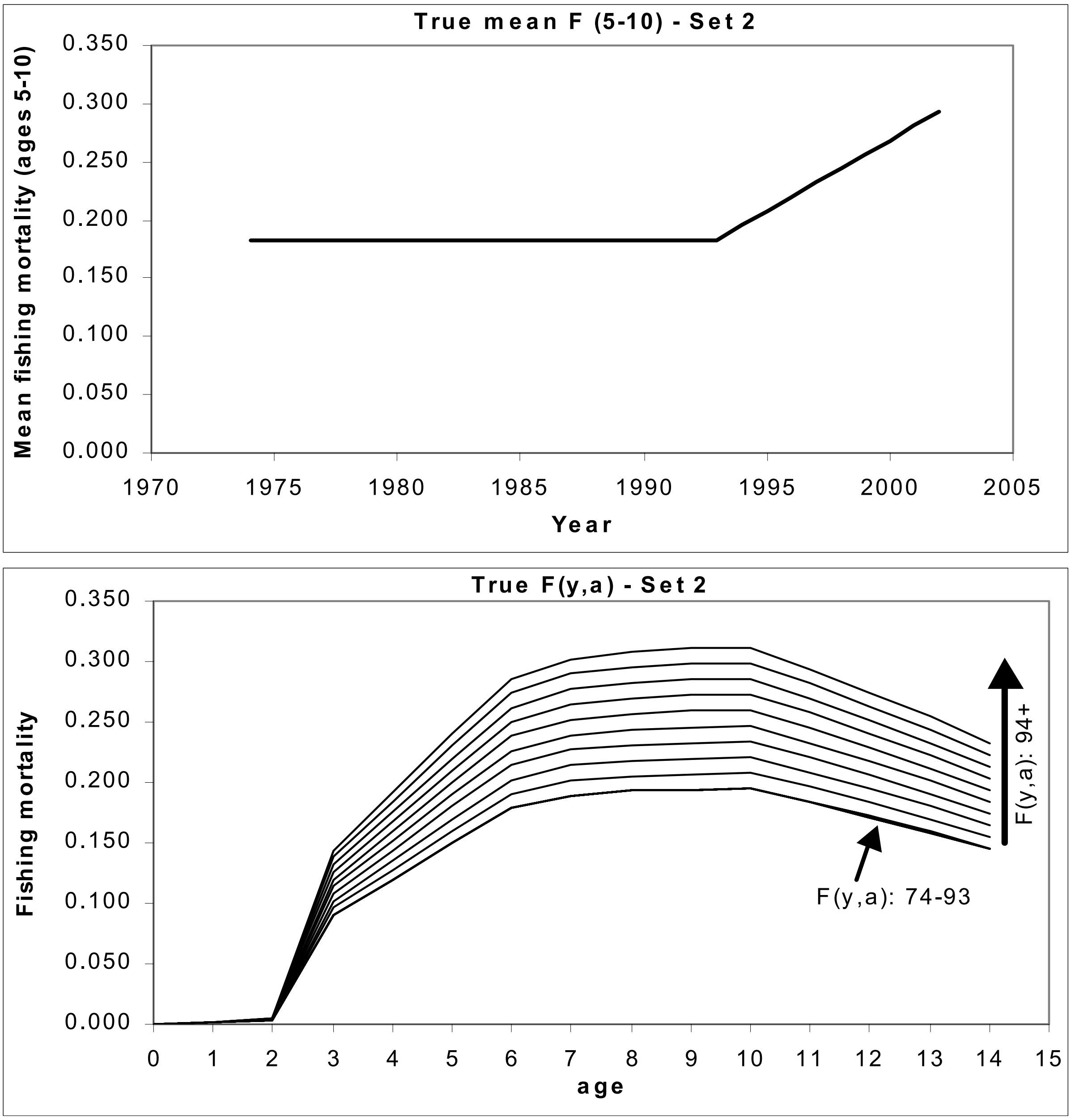


Figure 4.1.2.9 True fishing mortality and selection pattern for data set 2.

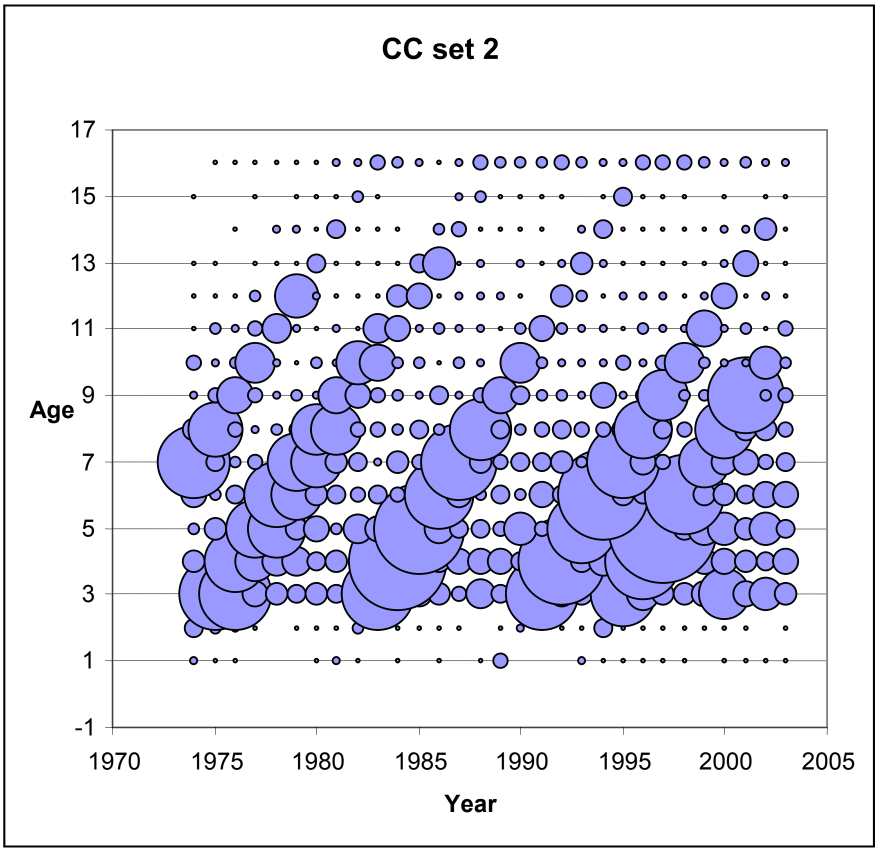


Figure 4.1.2.10 Recorded catches for data set 2 (with error).

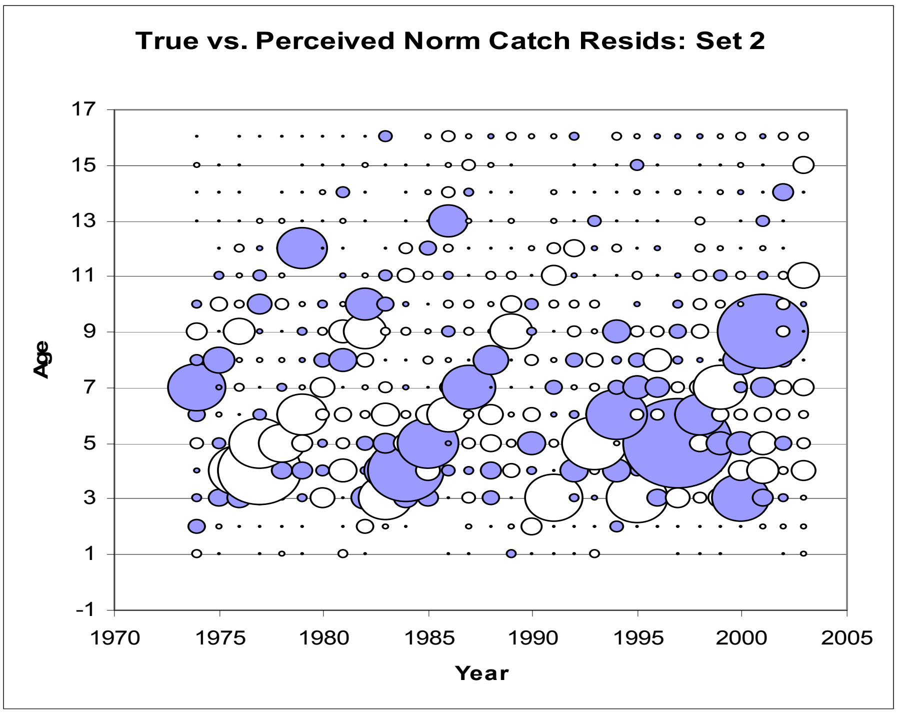


Figure 4.1.2.11 Error in catch recording expressed as arithmetic deviation from true catch (filled circles atm +ve, empty circles atm –ve) for set 2.

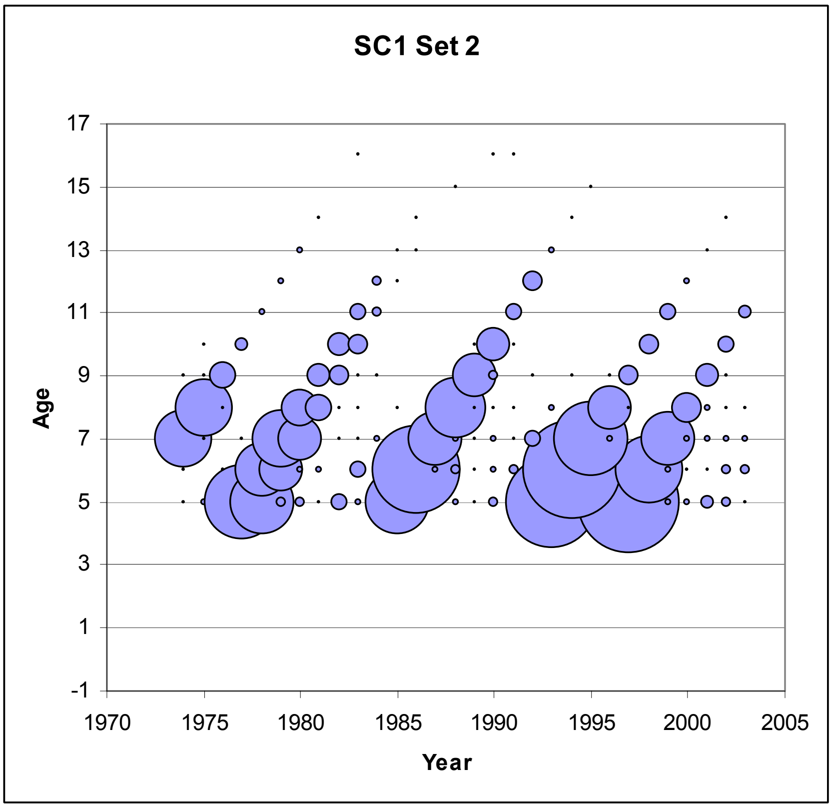


Figure 4.1.2.12 Survey 1 adult survey including error for data set 2.

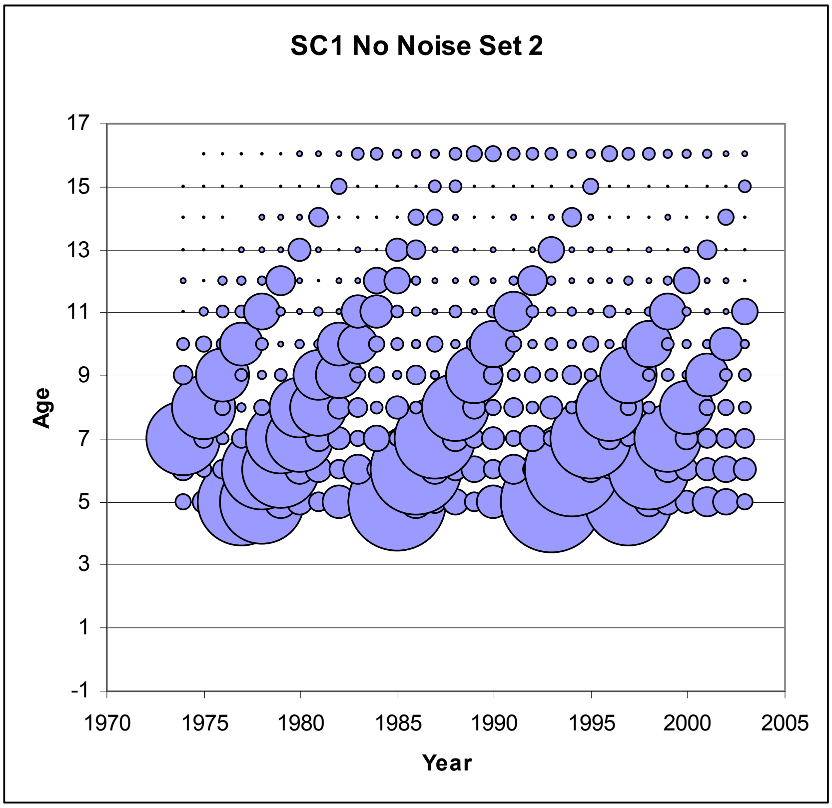


Figure 4.1.2.13 Survey 1 adult survey without error for data set 2.

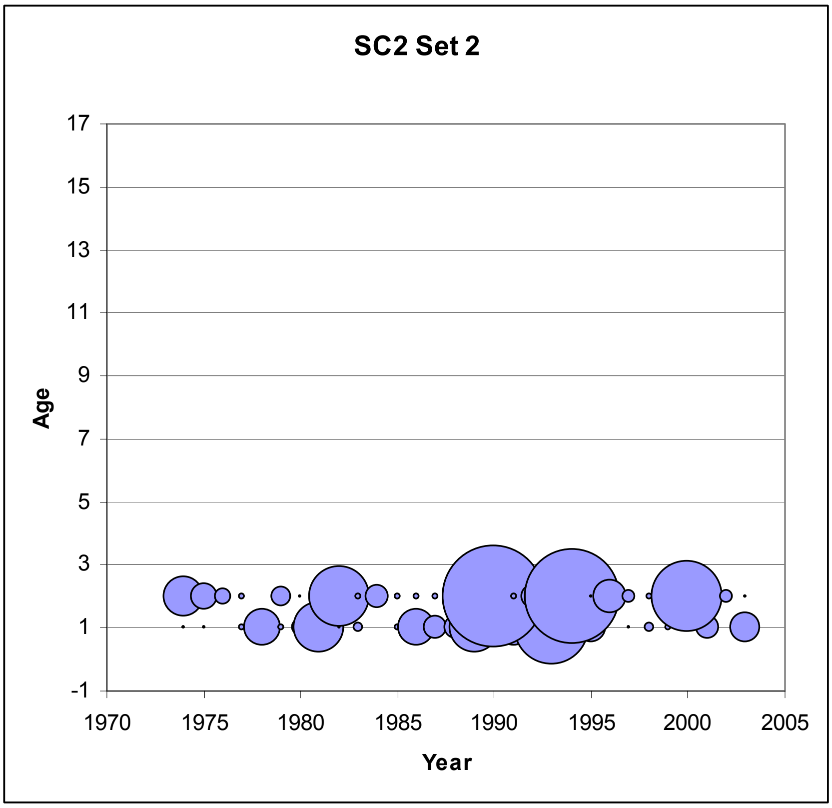


Figure 4.1.2.14 Survey 2 juvenile survey including error for data set 2.

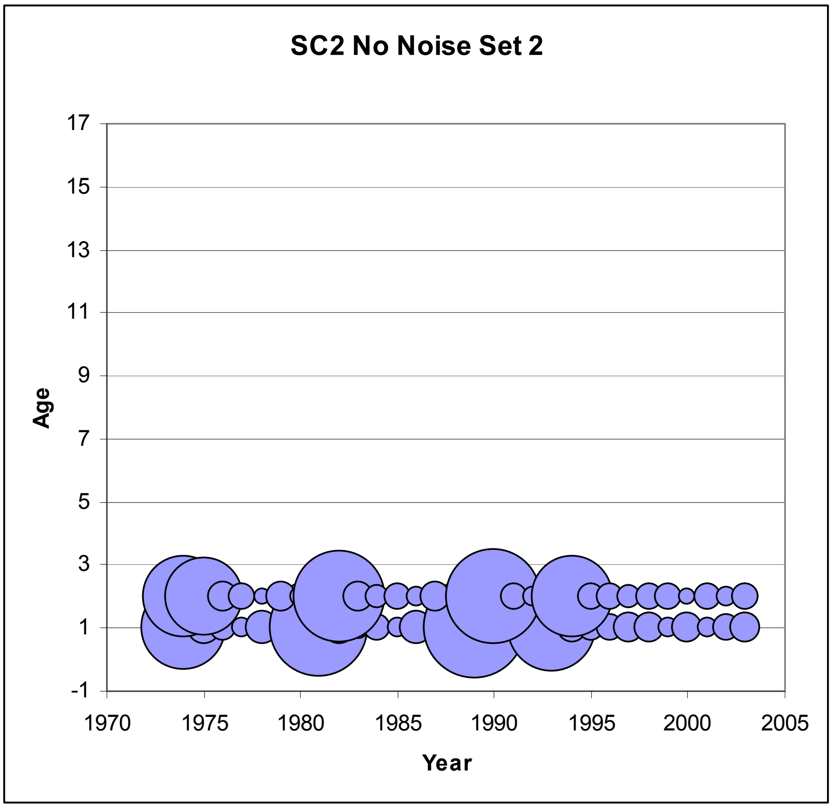


Figure 4.1.2.15 Survey 2 juvenile survey without error for data set 2.

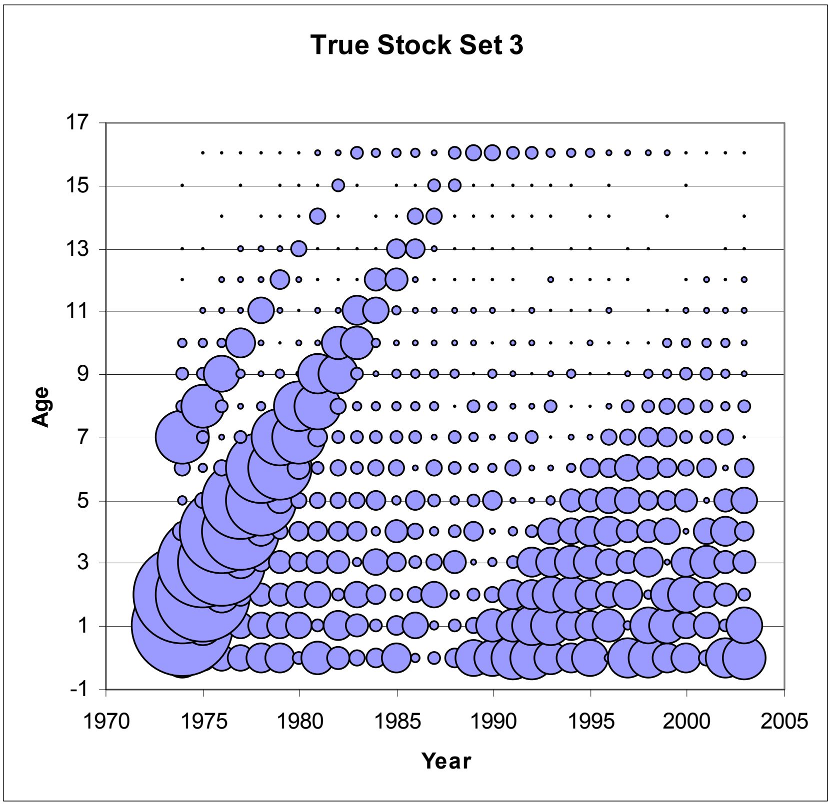


Figure 4.1.2.16 True stock in numbers for simulated data set 3.

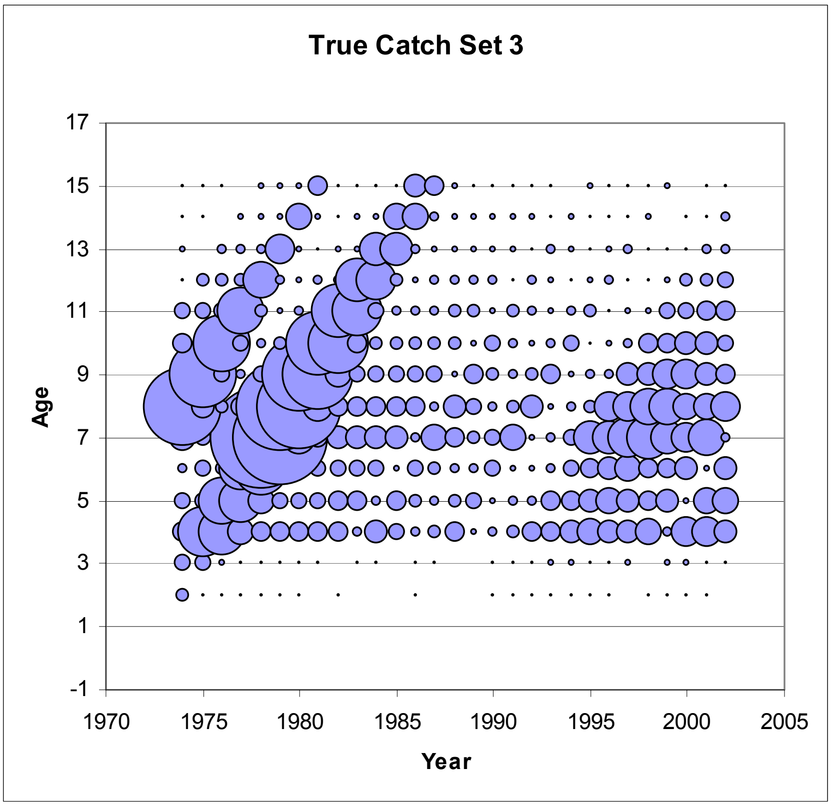
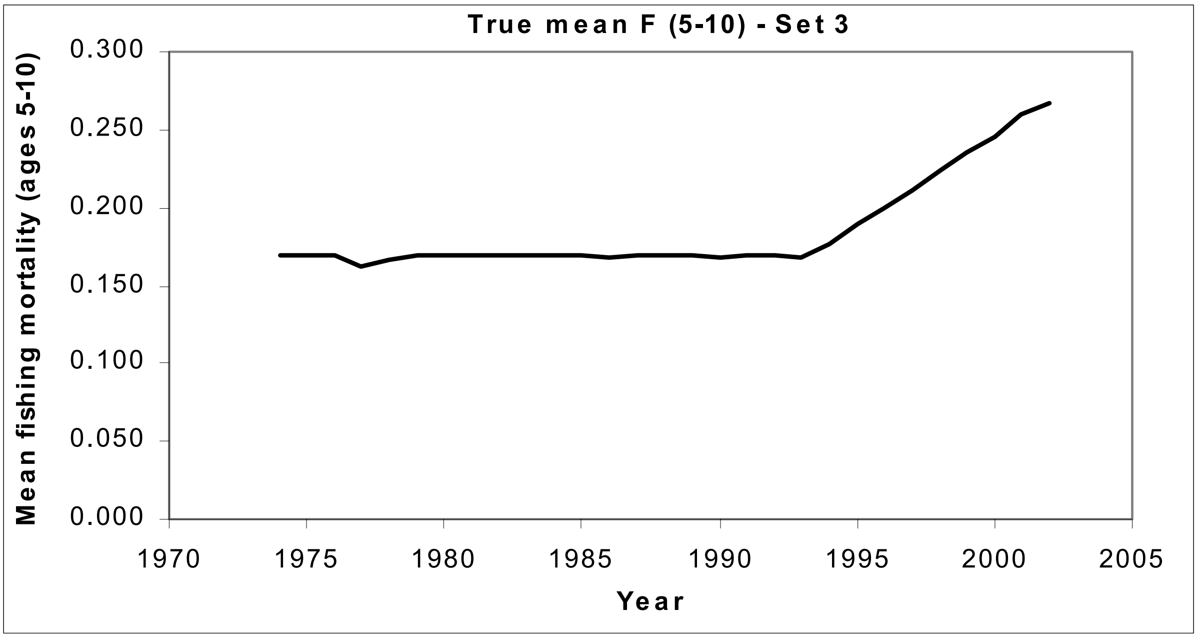


Figure 4.1.2.17 True catch in numbers for simulated data set 3.



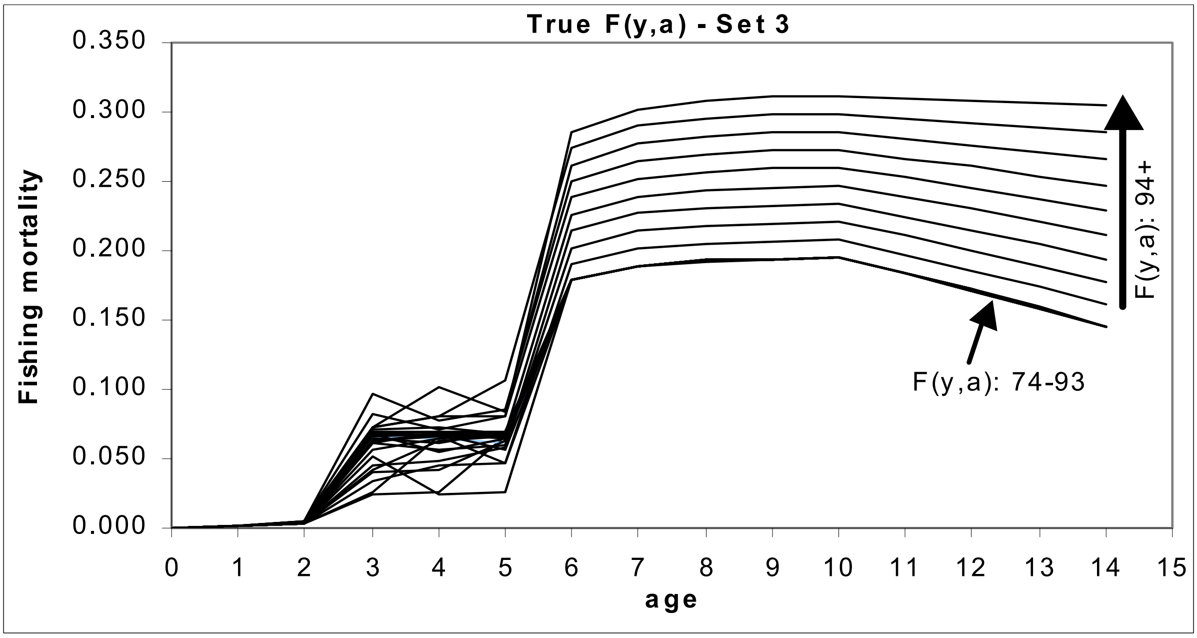


Figure 4.1.2.18 True fishing mortality and exploitation pattern for simulated data set 3.

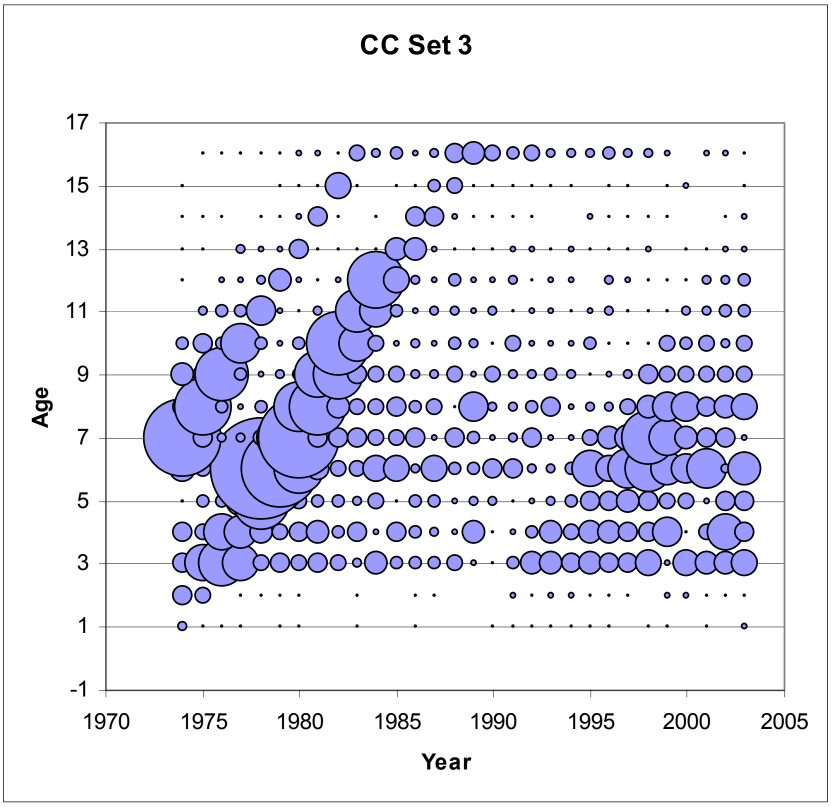


Figure 4.1.2.19 Reported Catch for data set 3.

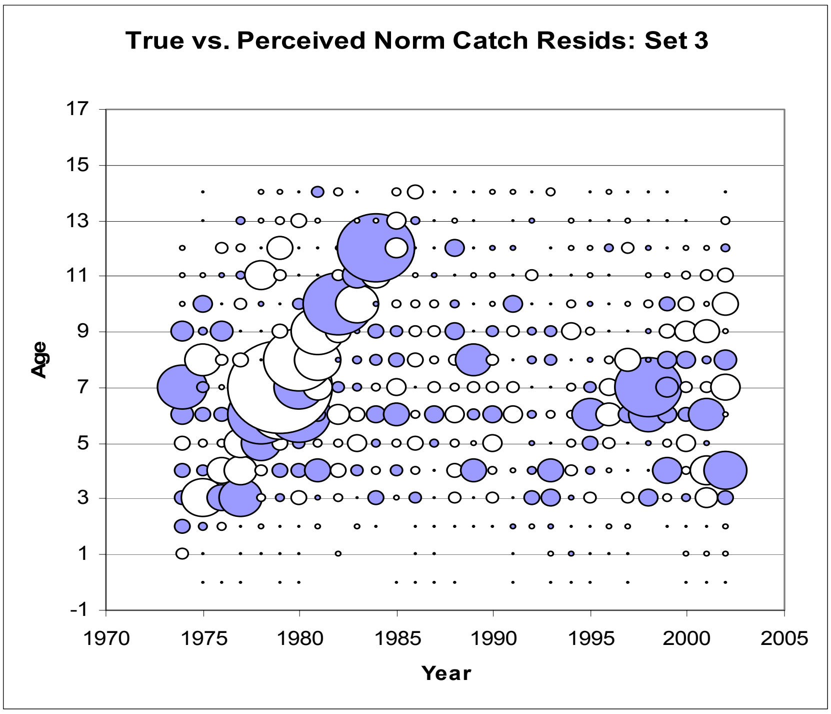


Figure 4.1.2.20 Error in catch recording for data set 3 expressed as arithmetic deviation from true catch (filled circles atr +ve, empty circles atr –ve).

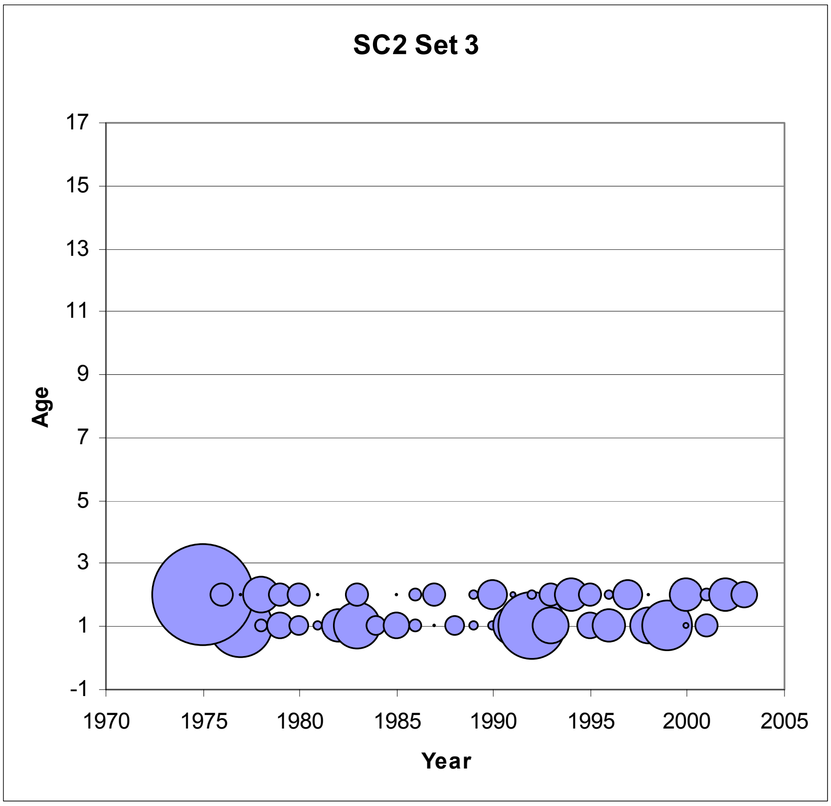


Figure 4.1.2.21 Survey 1, juveniles including error data set 3.

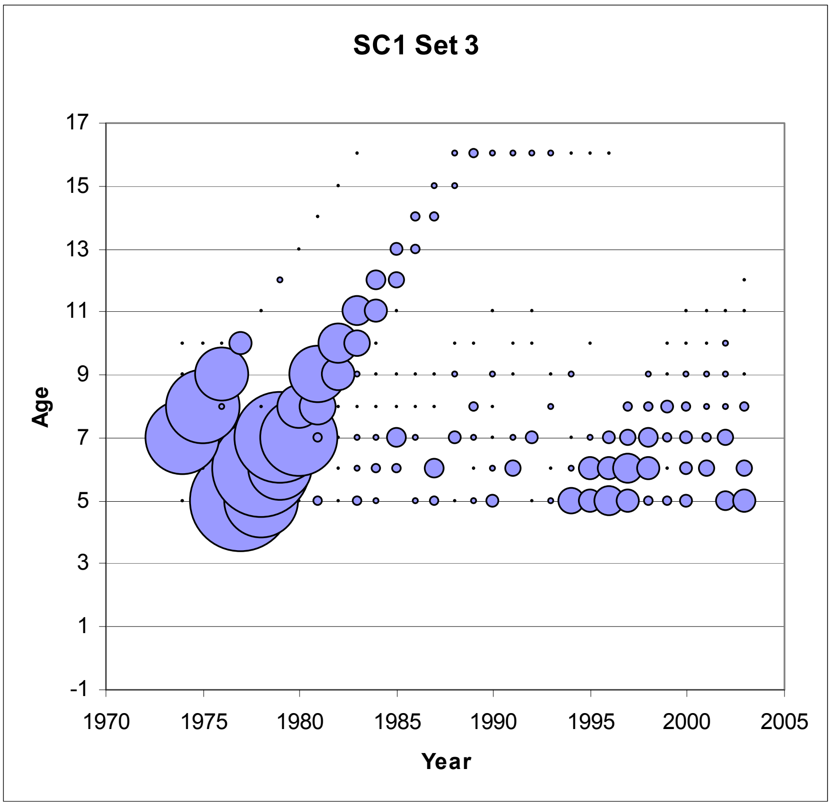
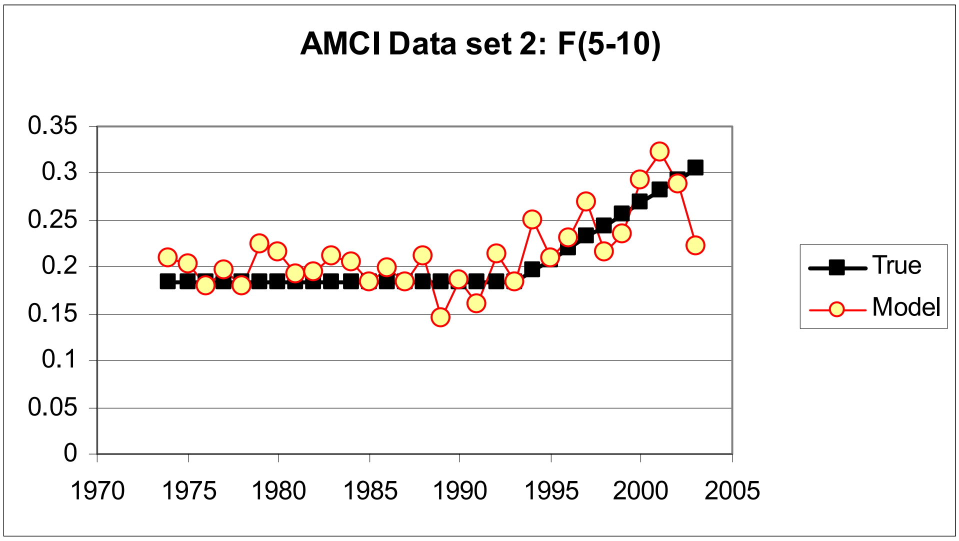
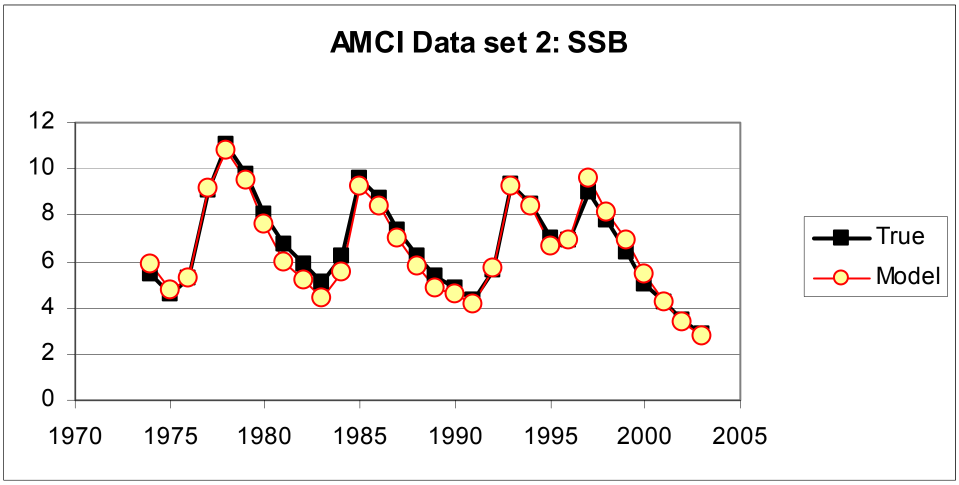


Figure 4.1.2.22 Survey 2, adult including error data set 3.





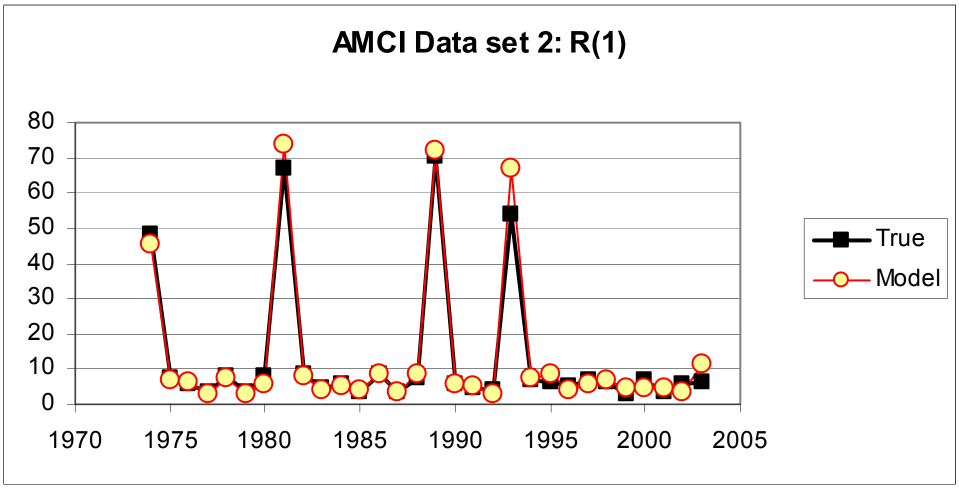


Figure 4.2.1.1. Estimated vs. "true" F (mean, age 5-10), SSB and age 1 recruitment using AMCI with simulated data set 2.

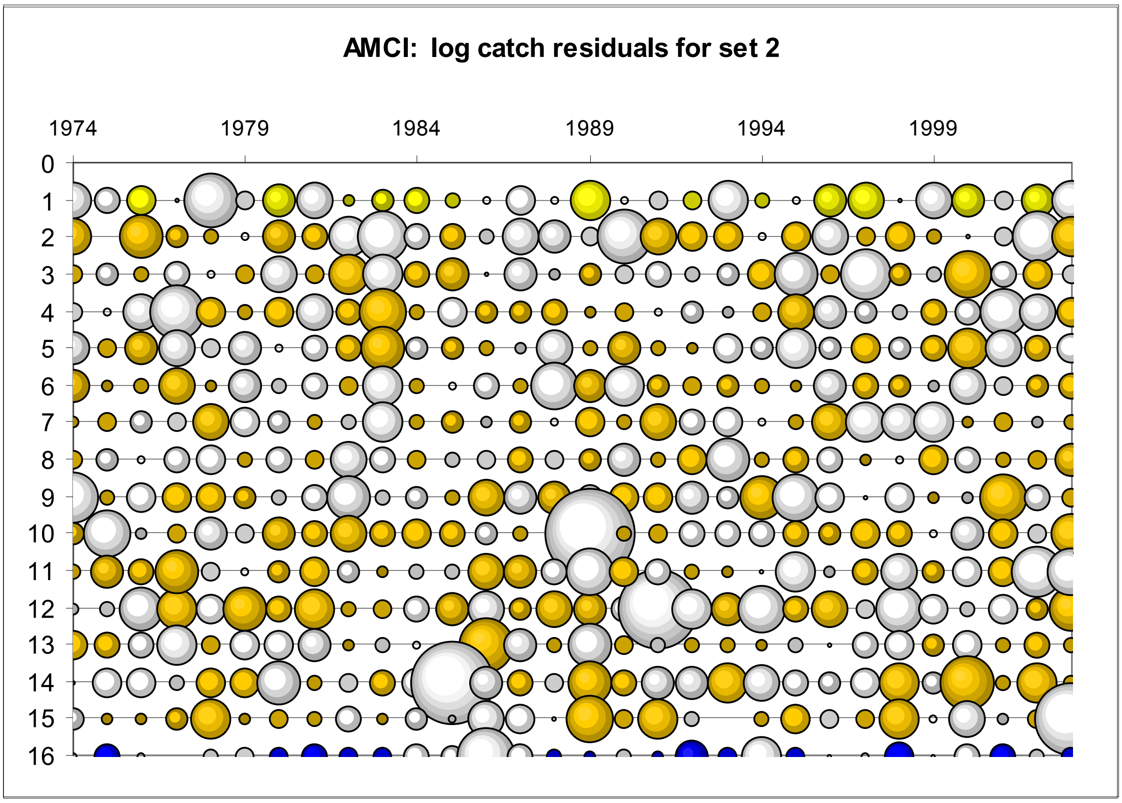


Figure 4.2.1.2. Log catch residuals for AMCI fit using simulated data set 2.

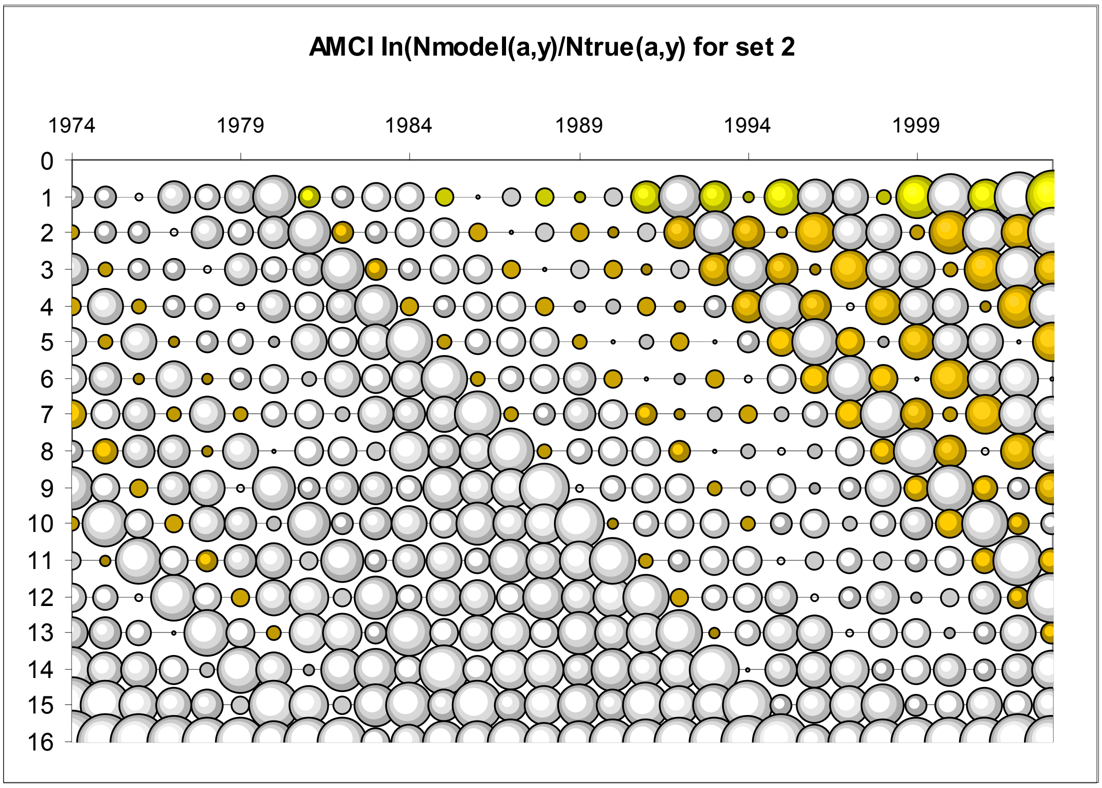


Figure 4.2.1.3. Log population size in numbers (N) residuals by year and age for AMCI fit to simulated data set 2.

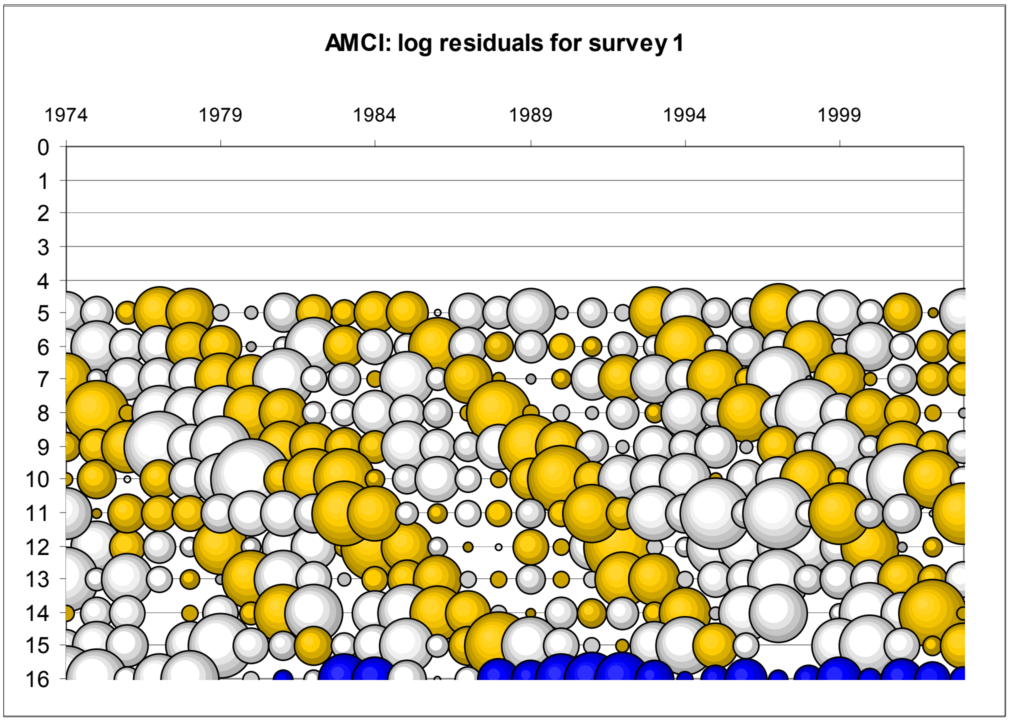


Figure 4.2.1.4. Log survey residuals (survey numbers vs. predicted) for AMCI fit with data set 2 for survey 1. Residuals are given by age and year.

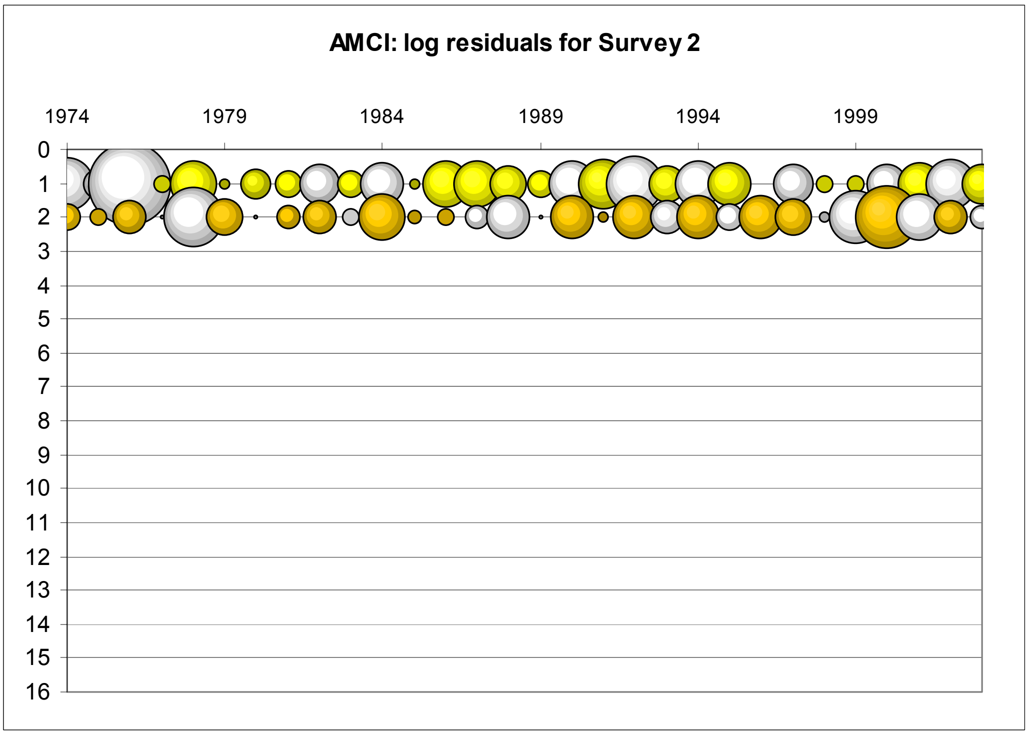


Figure 4.2.1.5. Log survey residuals (survey numbers vs. predicted) for AMCI fit with data set 2 for survey 2. Residuals are given by age and year.

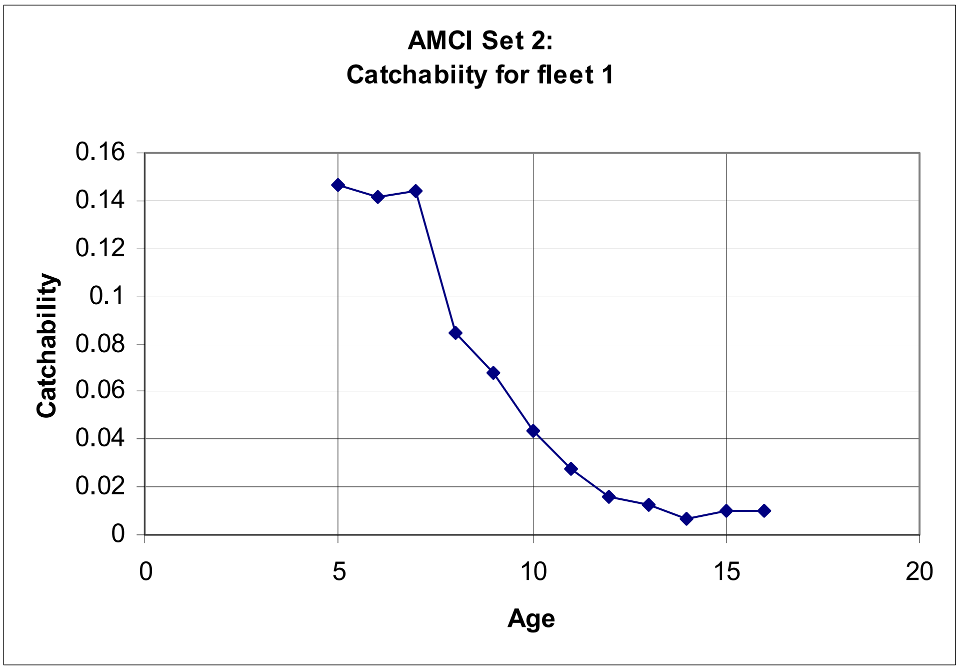


Figure 4.2.1.6. Calculated catchability at age for AMCI fit to simulated data set 2, for Survey 1.

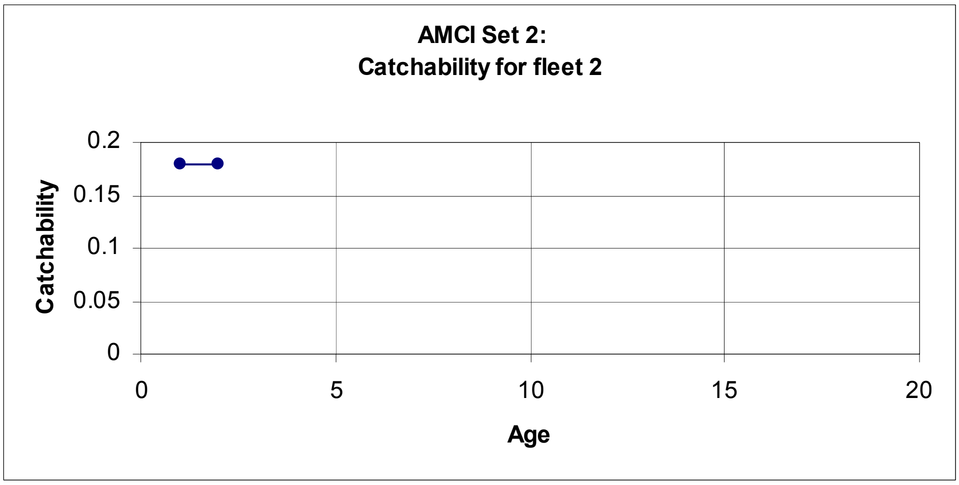


Figure 4.2.1.7. Calculated catchability at age for AMCI fit to simulated data set 2, for Survey 2.

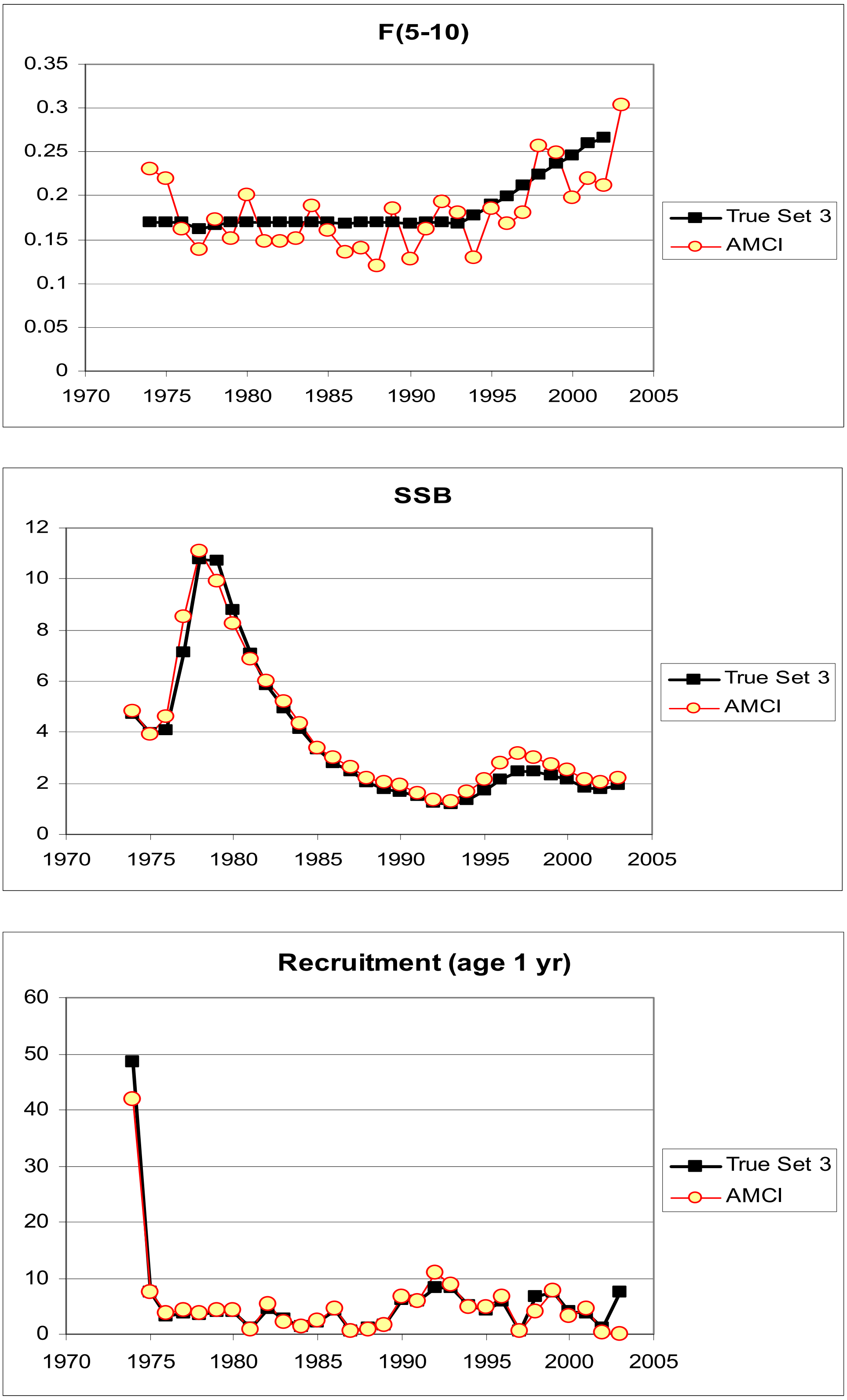


Figure 4.2.1.8. Estimated vs. "true" F (mean, age 5-10), SSB and age 1 recruitment using AMCI with simulated data set 3.

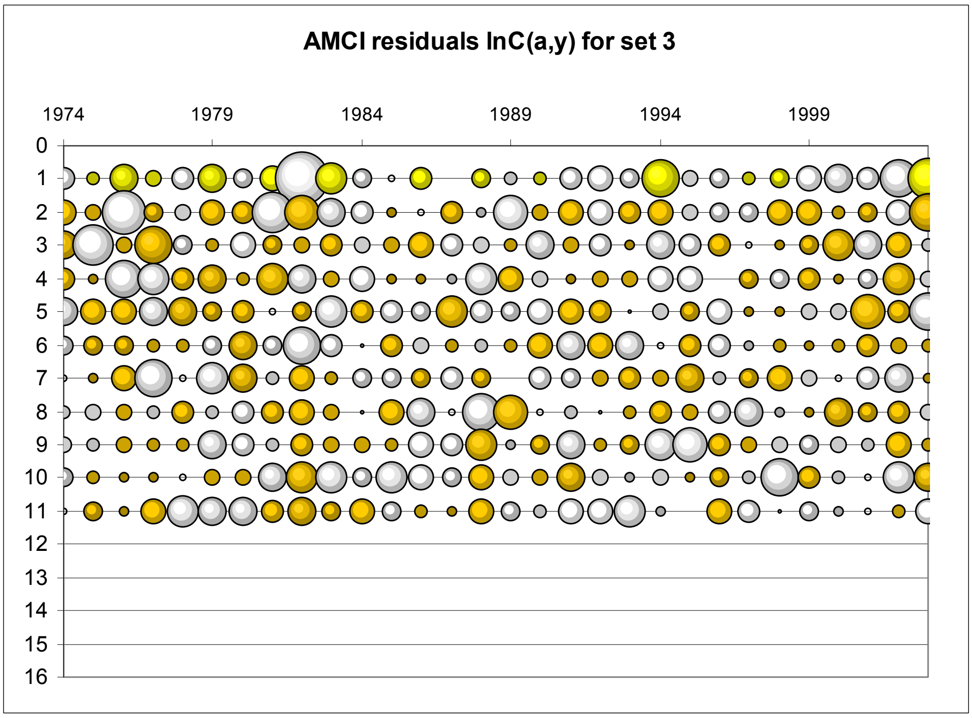


Figure 4.2.1.9. Log catch residuals for AMCI fit using simulated data set 3.

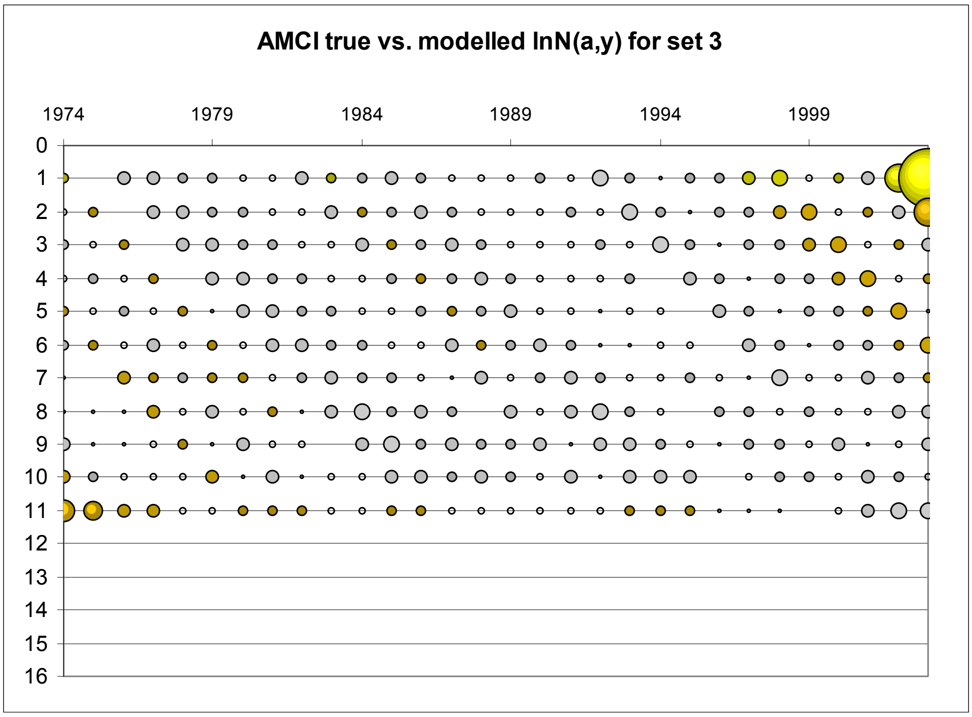


Figure 4.2.1.10. Log population size in numbers (N) residuals by year and age for AMCI fit to simulated data set 3.

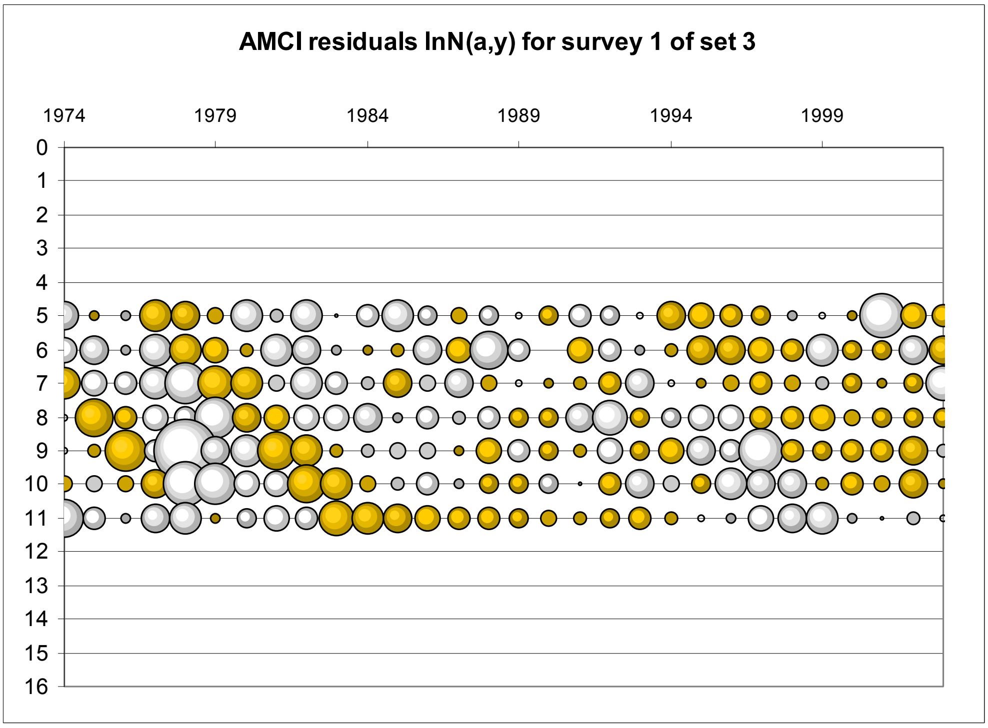


Figure 4.2.1.11. Log survey residuals (survey numbers vs. predicted) for AMCI fit with data set 3 for survey 1. Residuals are given by age and year.

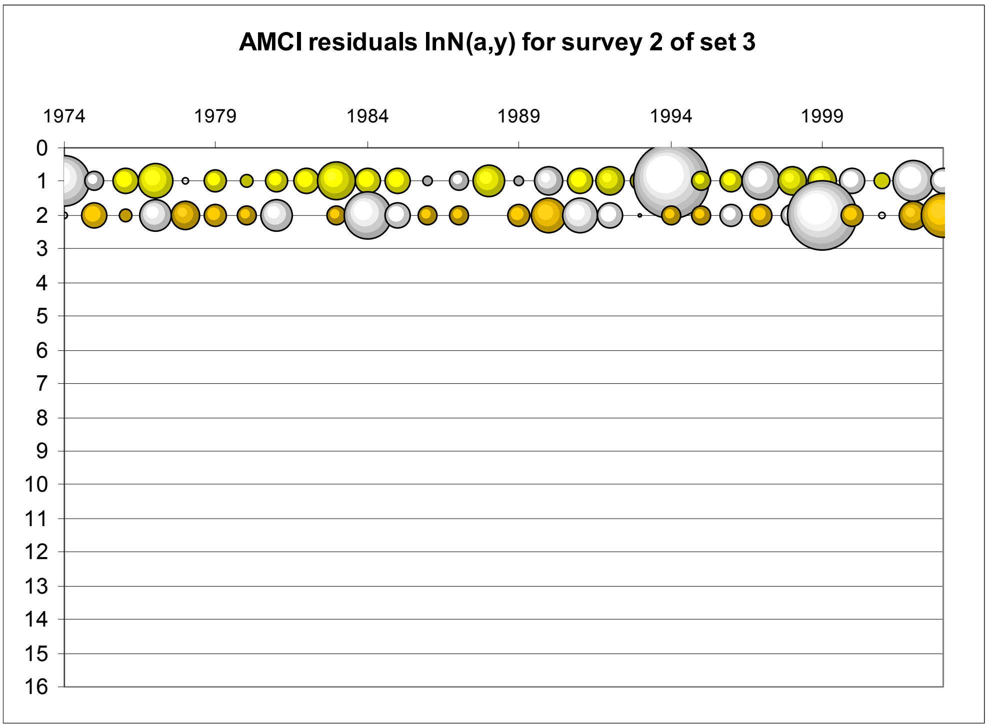


Figure 4.2.1.12. Log survey residuals (survey numbers vs. predicted) for AMCI fit with data set 3 for survey 2. Residuals are given by age and year.

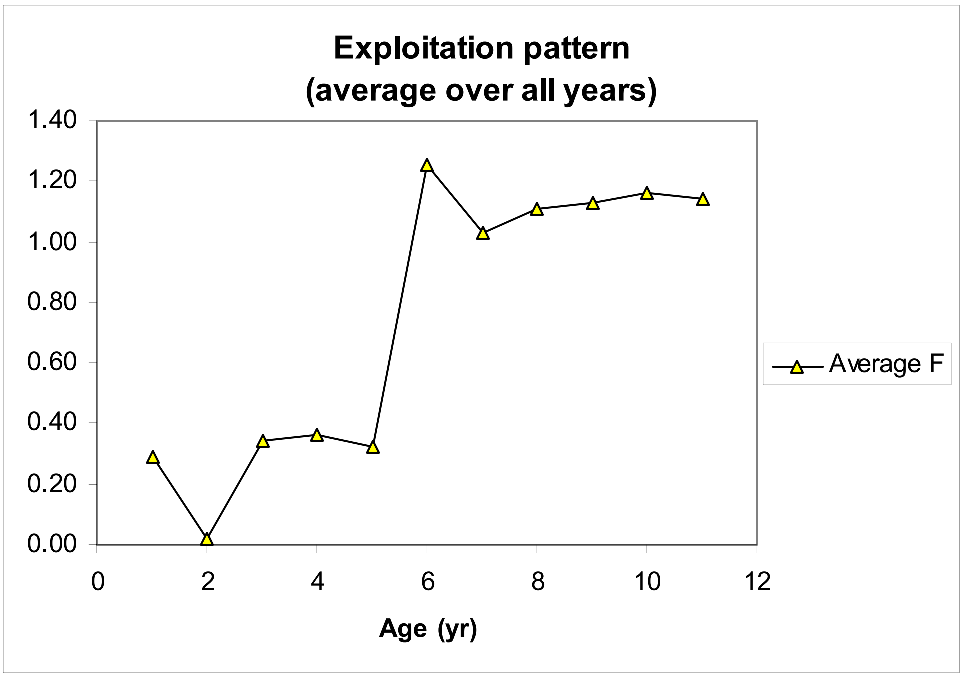


Figure 4.2.1.13. Calculated exploitation pattern at age for AMCI fit to simulated data set 3, for Commercial catches.

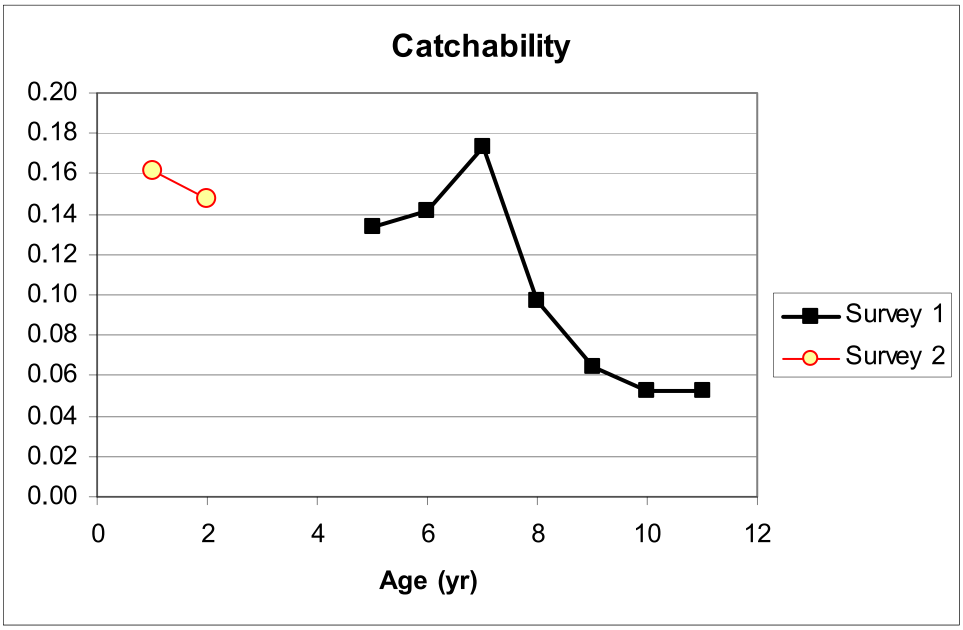


Figure 4.2.1.14. Calculated catchability at age for AMCI fit to simulated data set 3, for Surveys 1 and 2.

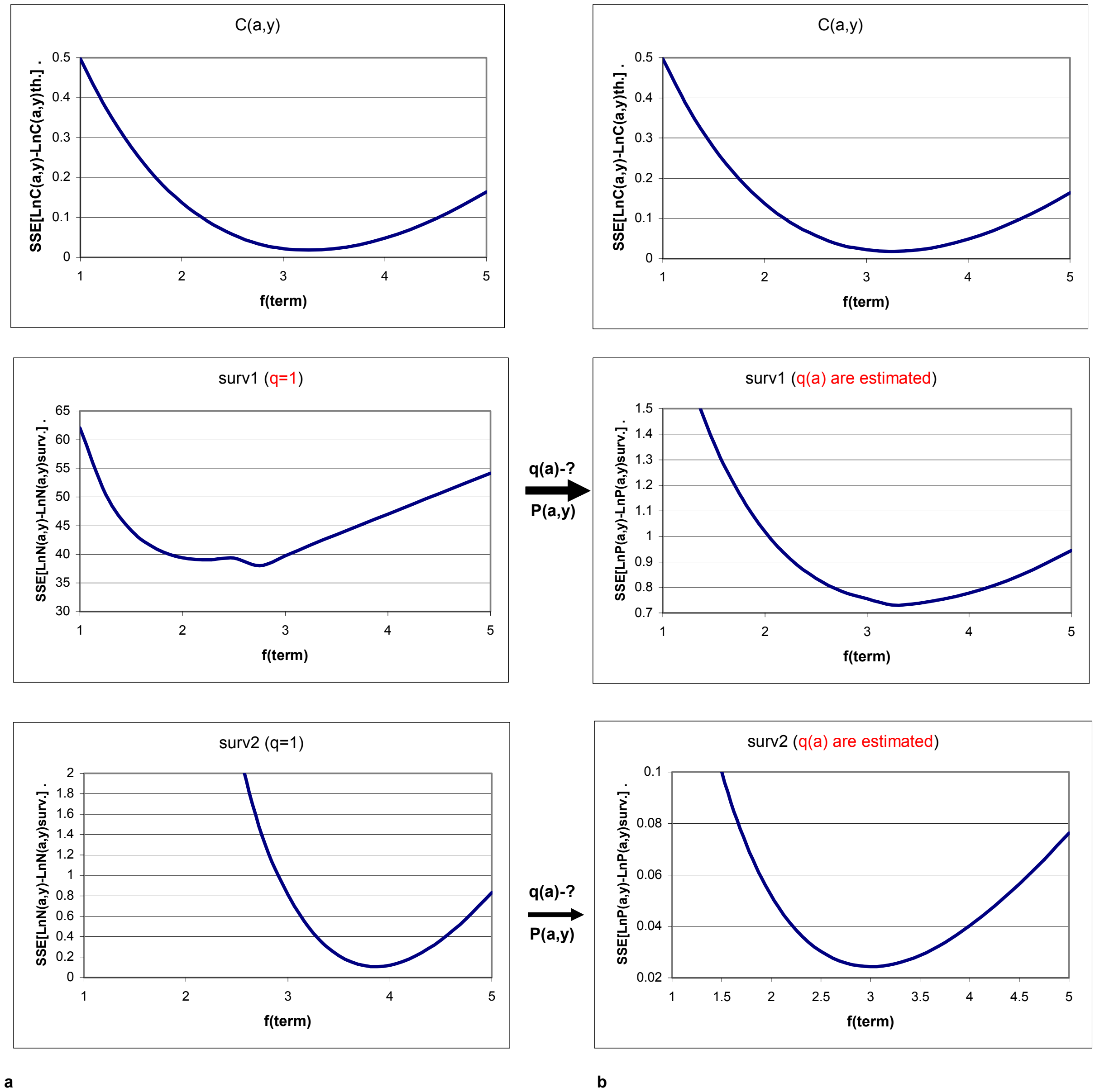


Figure 4.2.2.1. ISVPA SSE fitting surfaces for catch and survey indices for simulated data set 1.

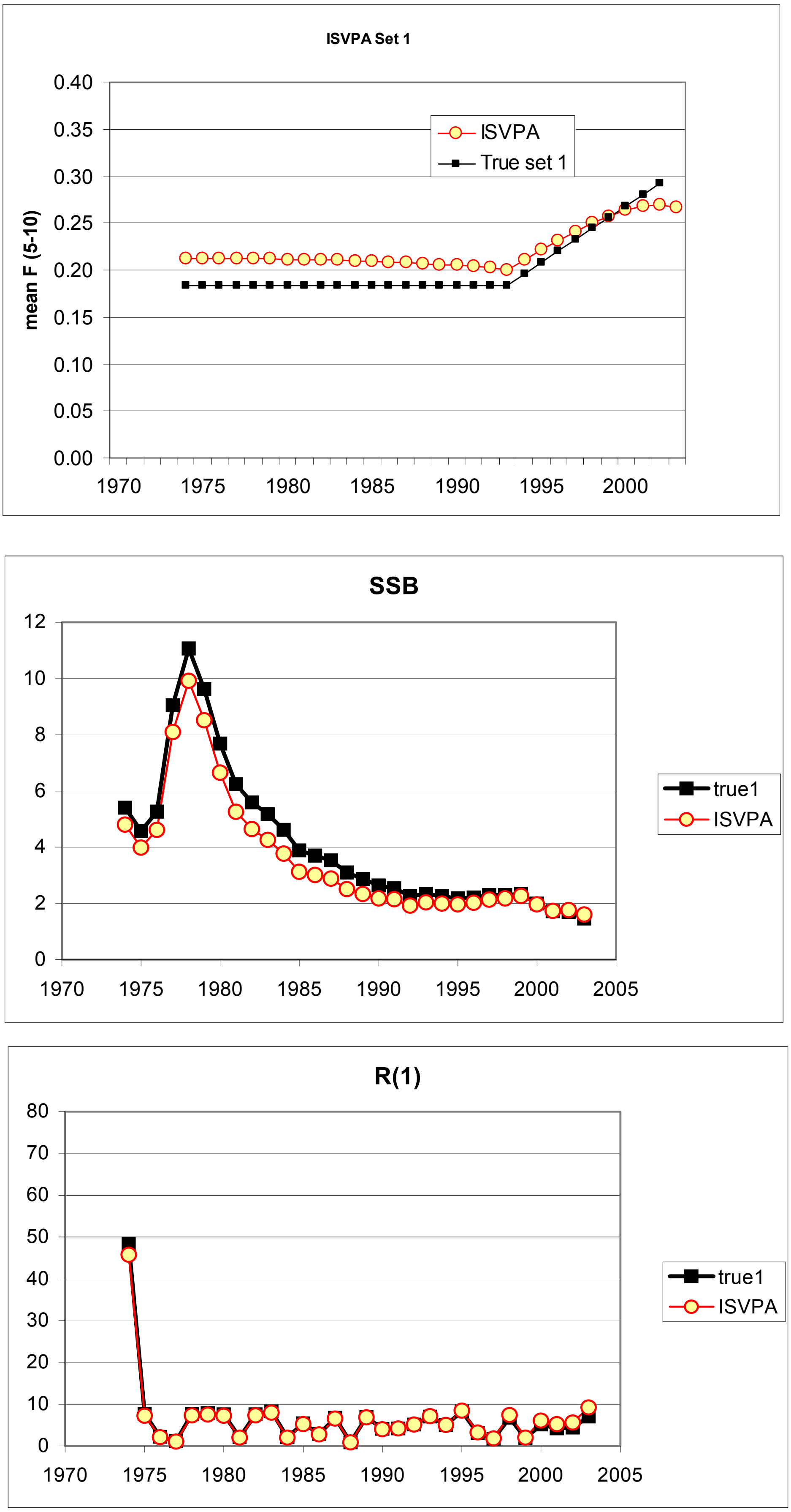


Figure 4.2.2.2. Estimated vs. "true" F (mean, age 5-10), SSB and age 1 recruitment using ISVPA with simulated data set 1.

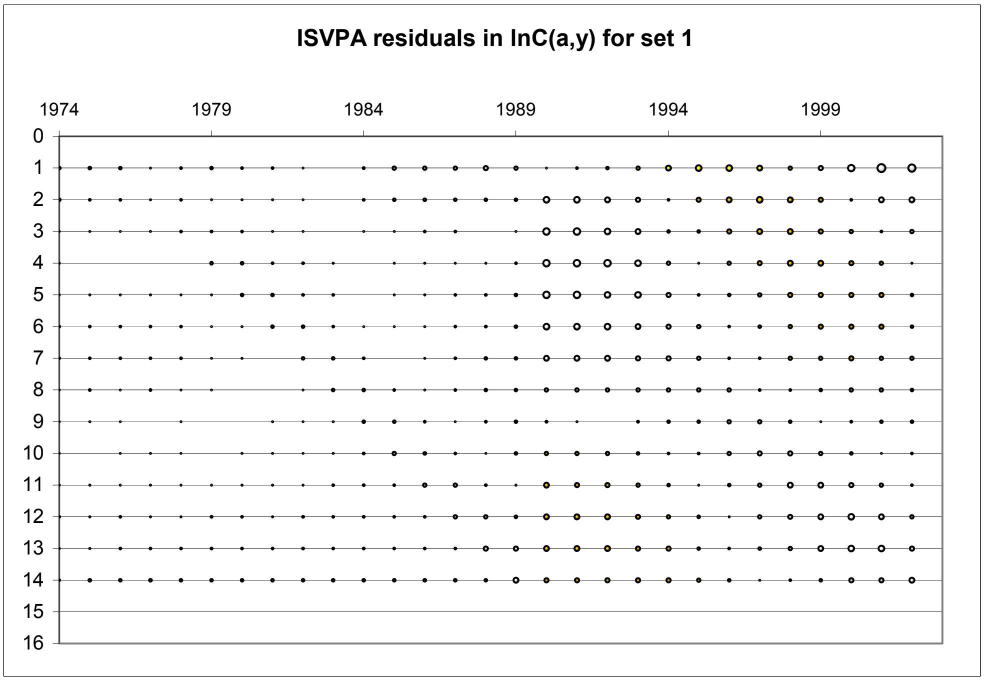


Figure 4.2.2.3. Log catch residuals for ISVPA fit using simulated data set 1

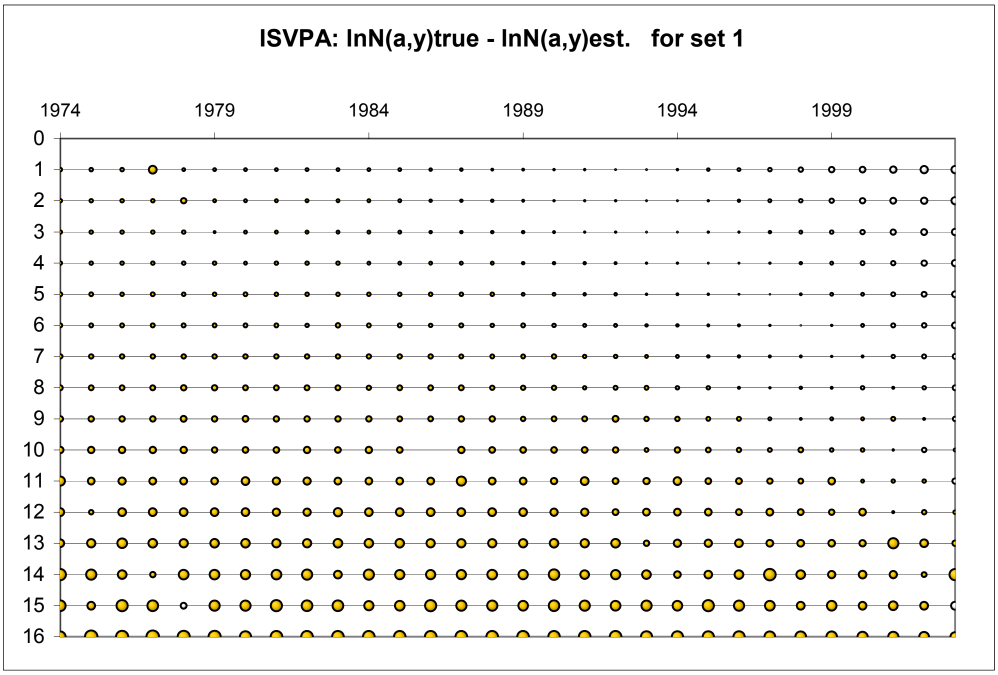


Figure 4.2.2.4. Log population size in numbers (N) residuals by year and age for ISVPA fit to simulated data set 1

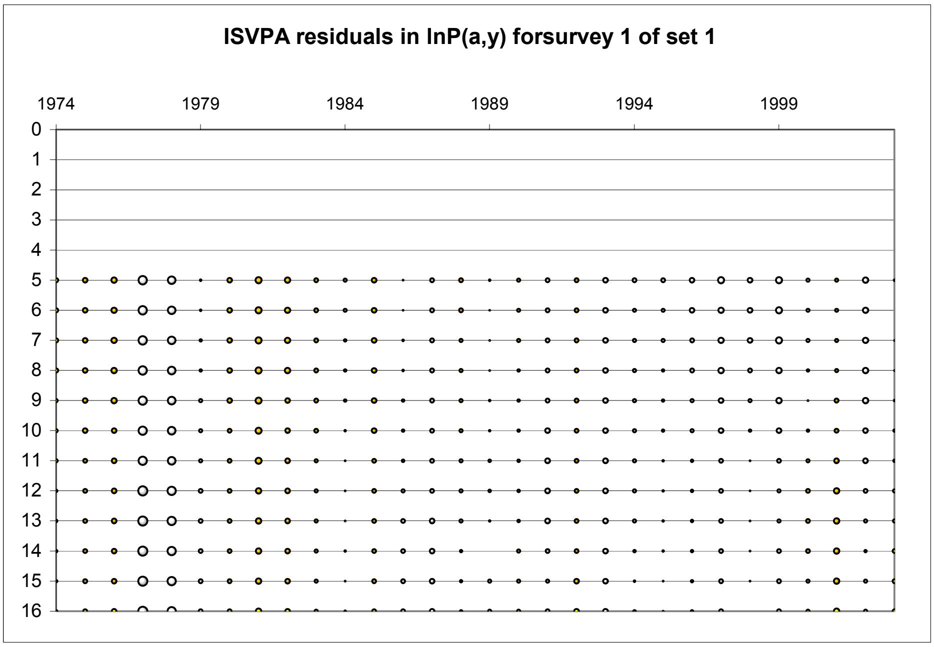


Figure 4.2.2.5. Log survey residuals (survey numbers vs. predicted) for ISVPA fit with data set 1 for survey 1. Residuals are given by age and year

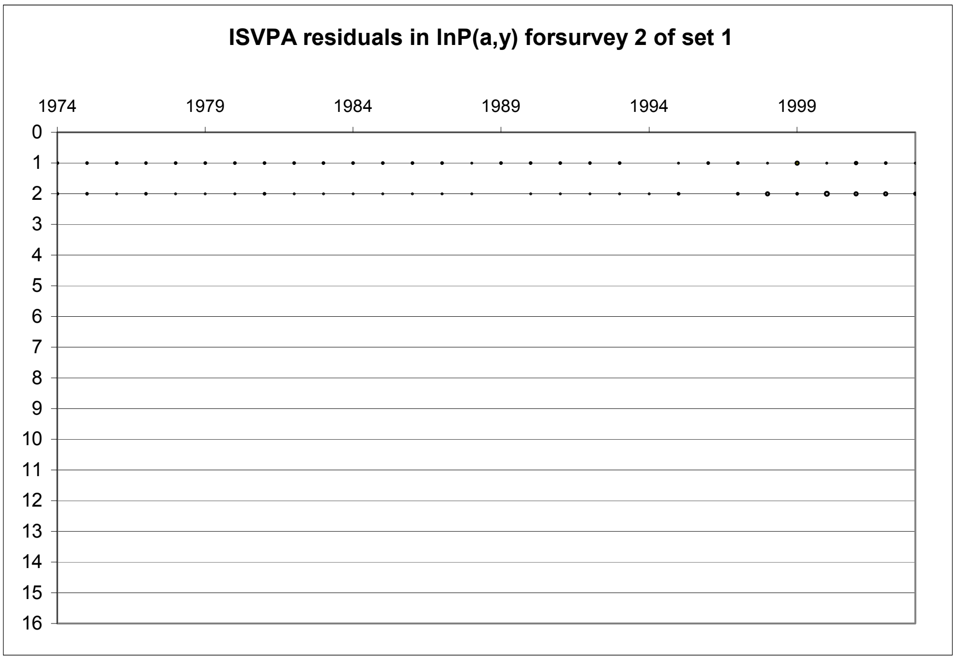


Figure 4.2.2.6 Log survey residuals (survey numbers vs. predicted) for ISVPA fit with data set 1 for survey 2. Residuals are given by age and year

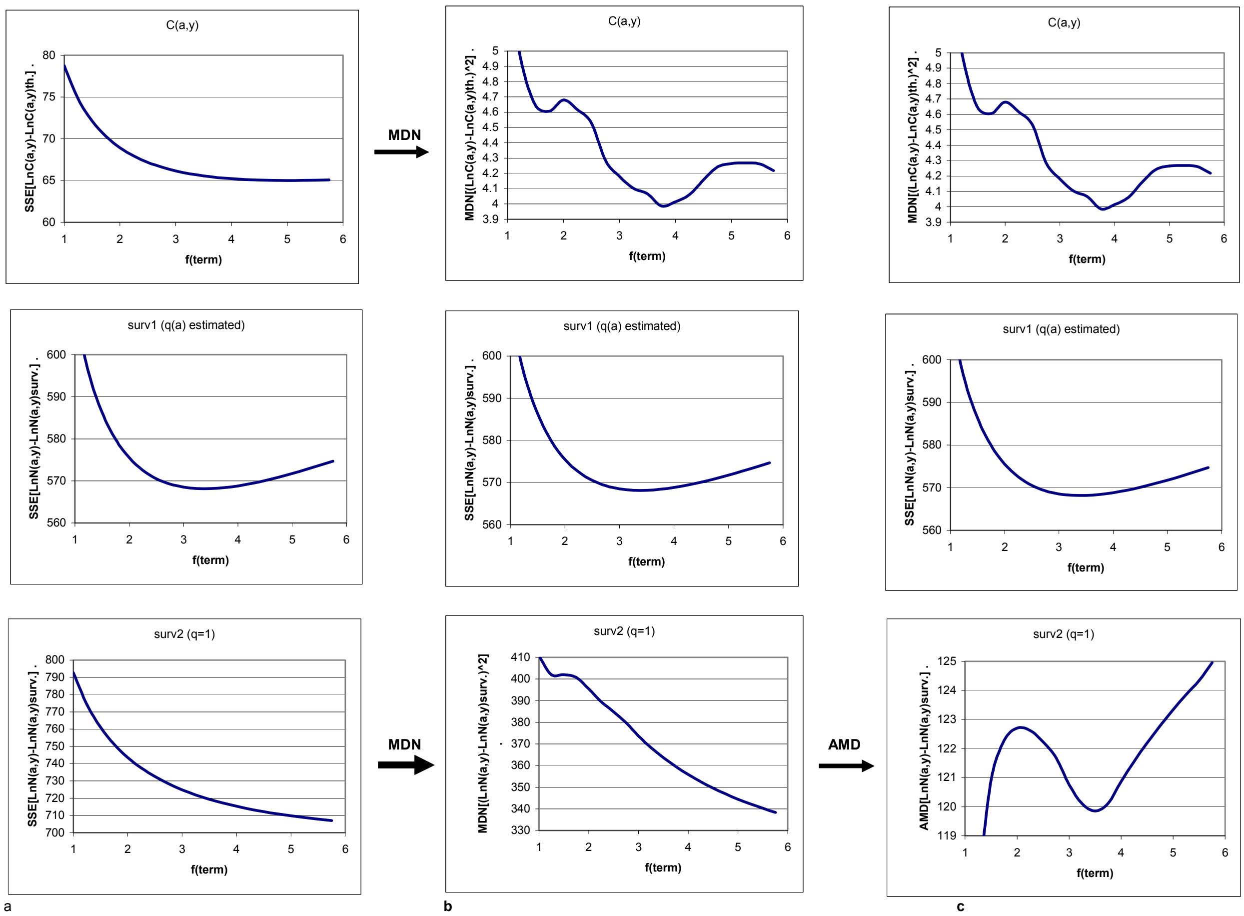


Figure 4.2.2.7. ISVPA SSE fitting surfaces for catch and survey indices for simulated data set 2.

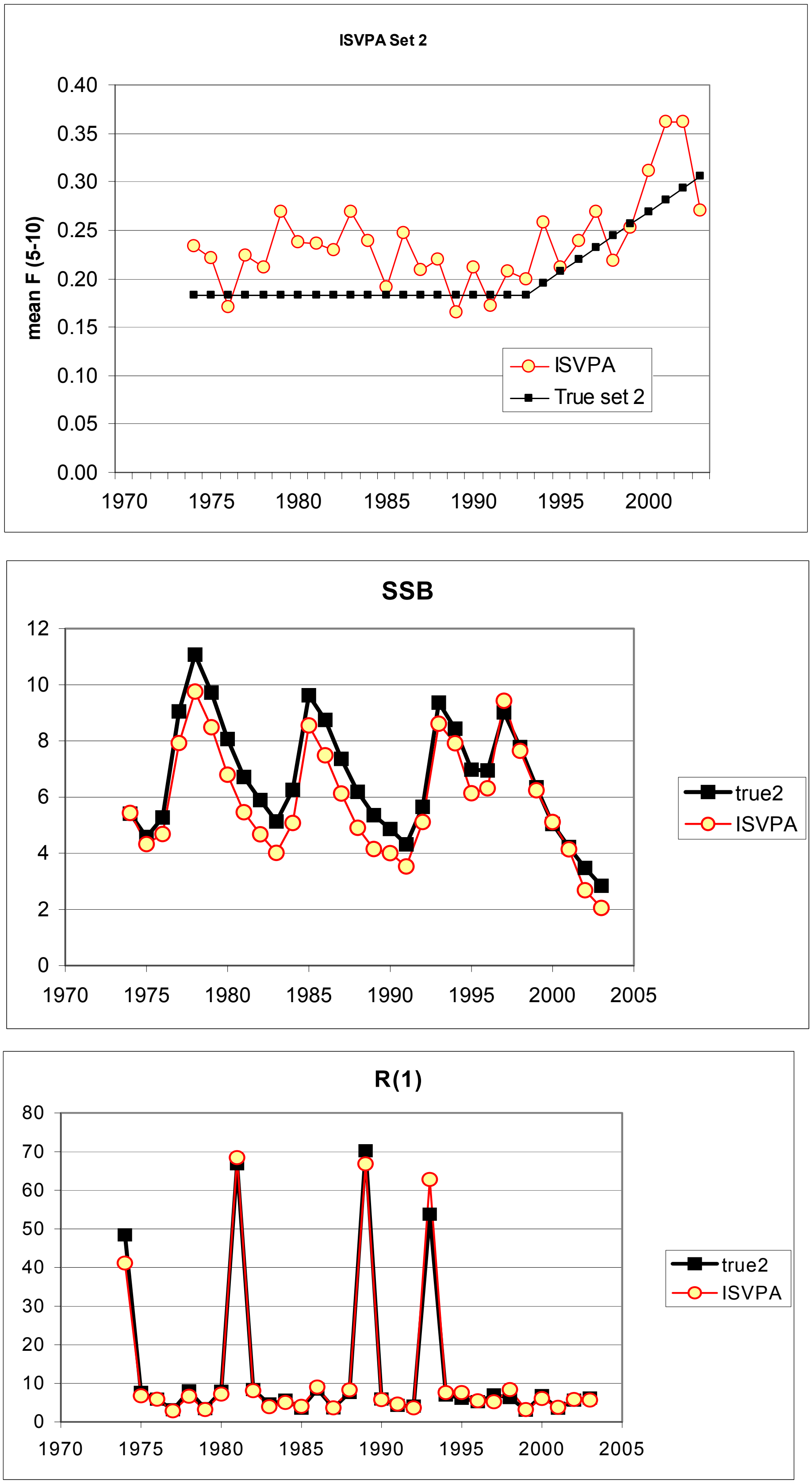


Figure 4.2.2.8. Estimated vs. "true" F (mean, age 5-10), SSB and age 1 recruitment using ISVPA with simulated data set 2.

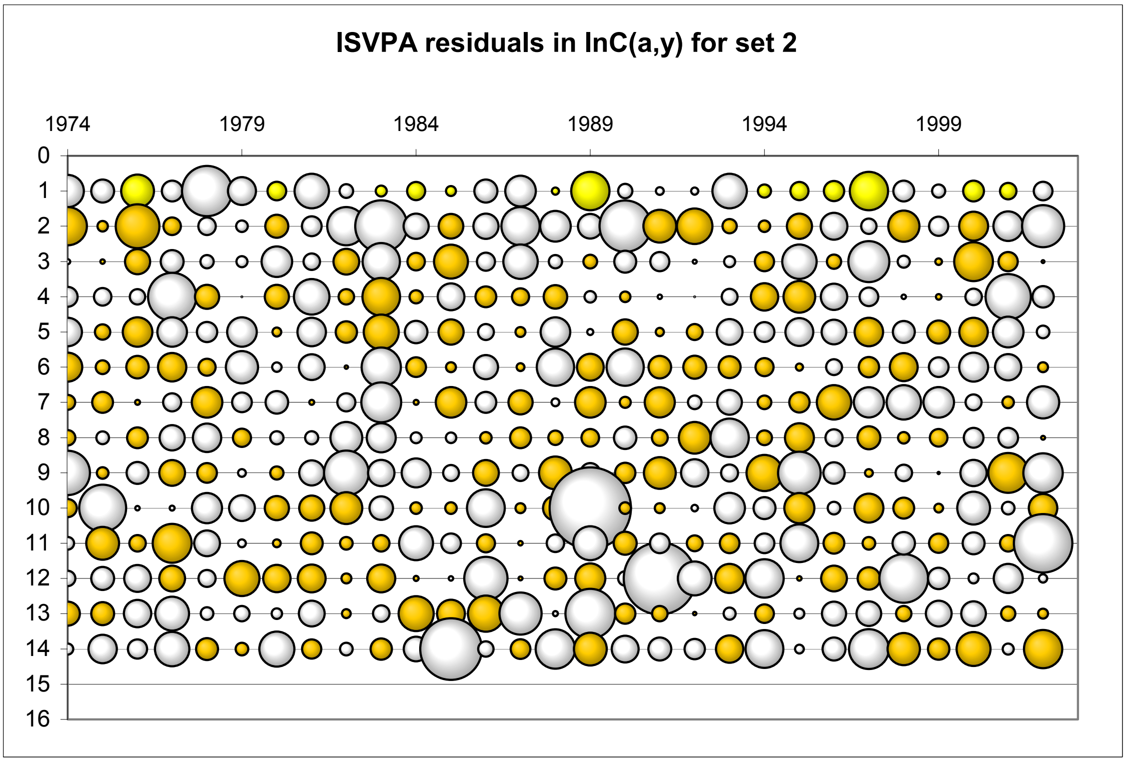


Figure 4.2.2.9. Log catch residuals for ISVPA fit using simulated data set 2.

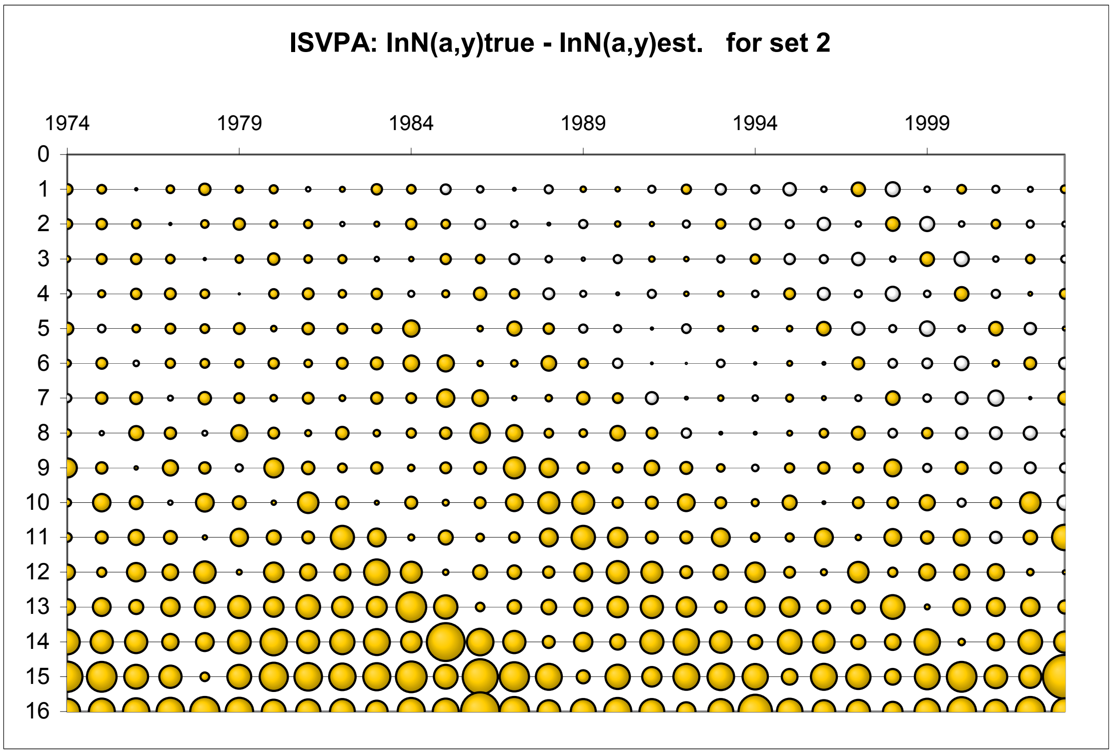


Figure 4.2.2.10. Log population size in numbers (N) residuals by year and age for ISVPA fit to simulated data set 2.

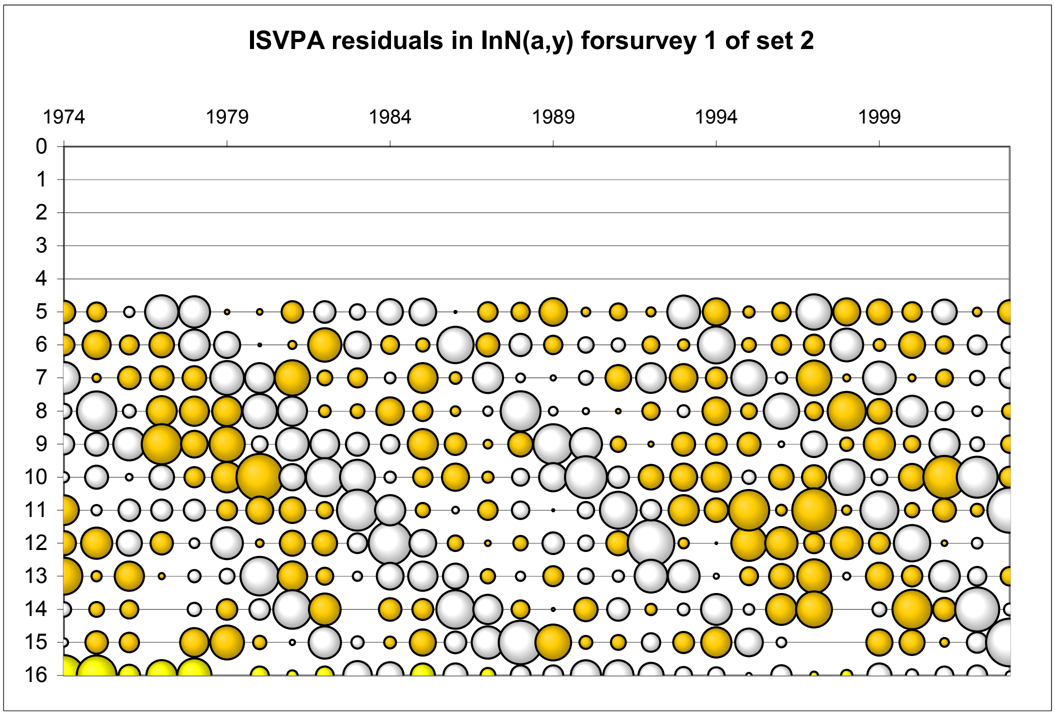


Figure 4.2.2.11. Log survey residuals (survey numbers vs. predicted) for ISVPA fit with data set 2 for survey 1. Residuals are given by age and year.

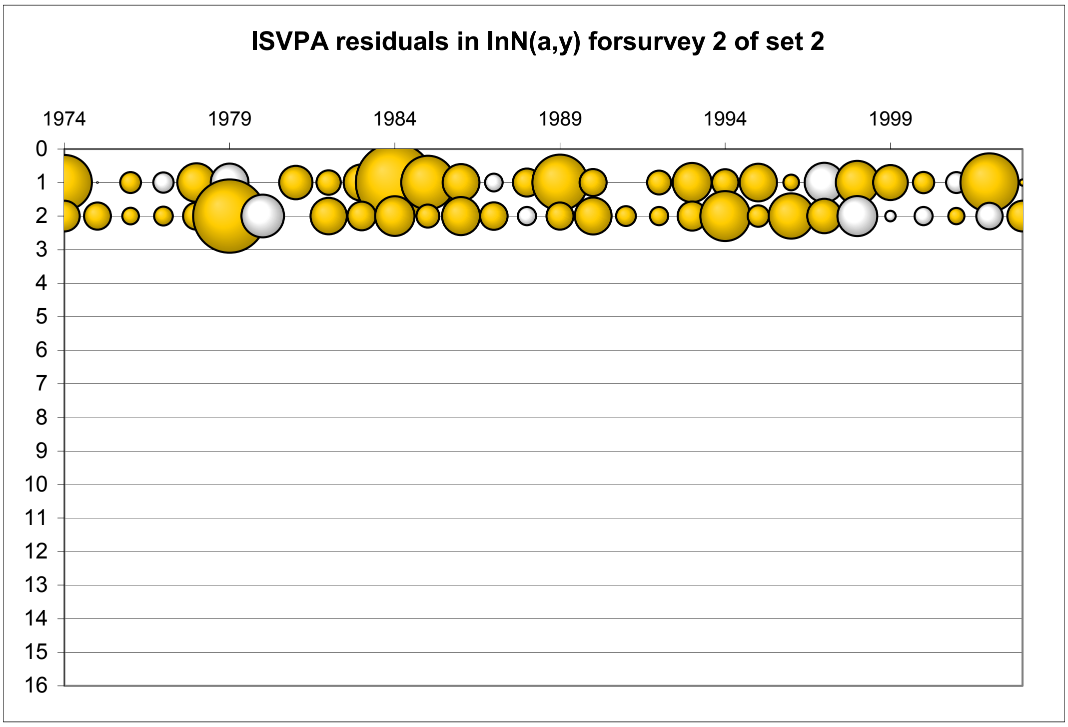


Figure 4.2.2.12. Log survey residuals (survey numbers vs. predicted) for ISVPA fit with data set 2 for survey 2. Residuals are given by age and year.

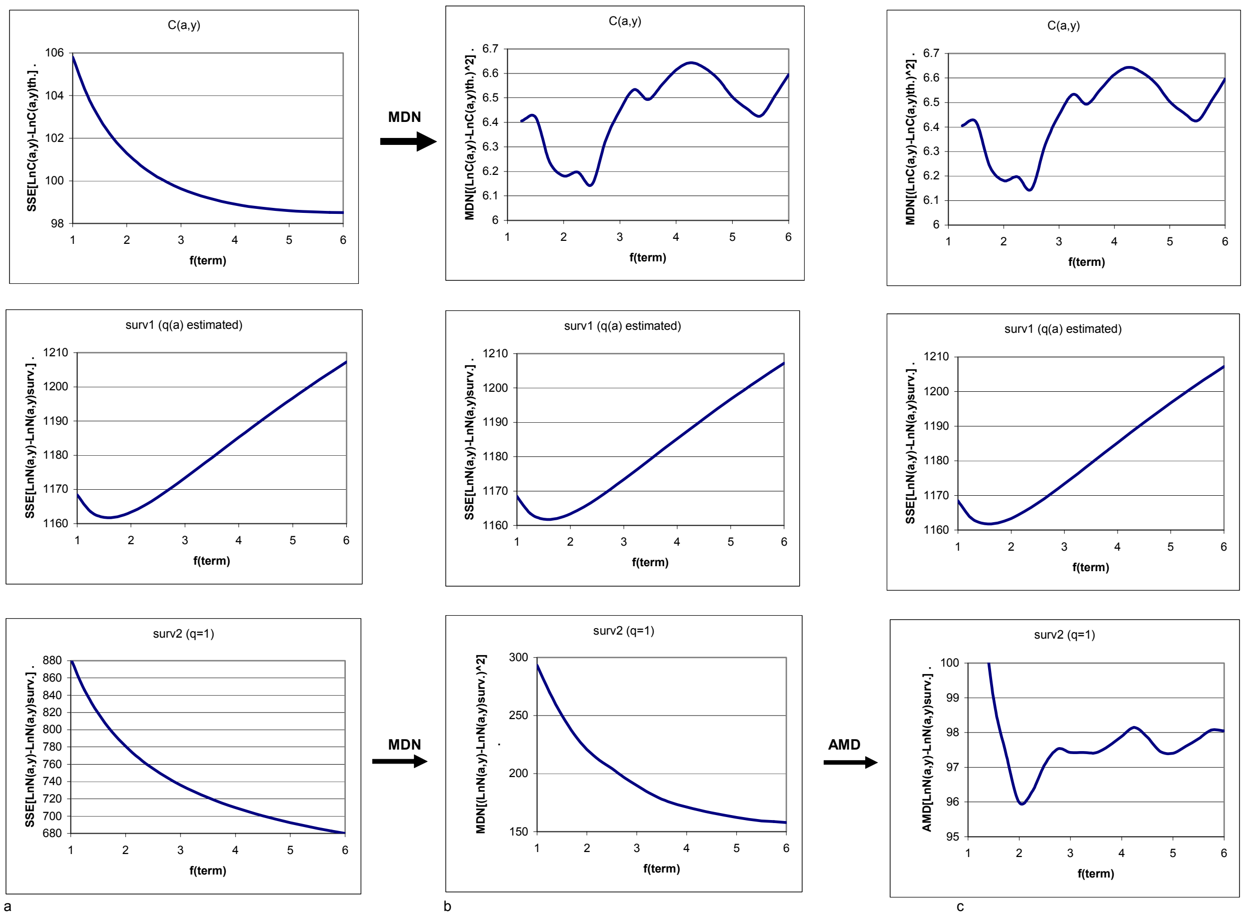


Figure 4.2.2.13. ISVPA SSE fitting surfaces for catch and survey indices for simulated data set 3.

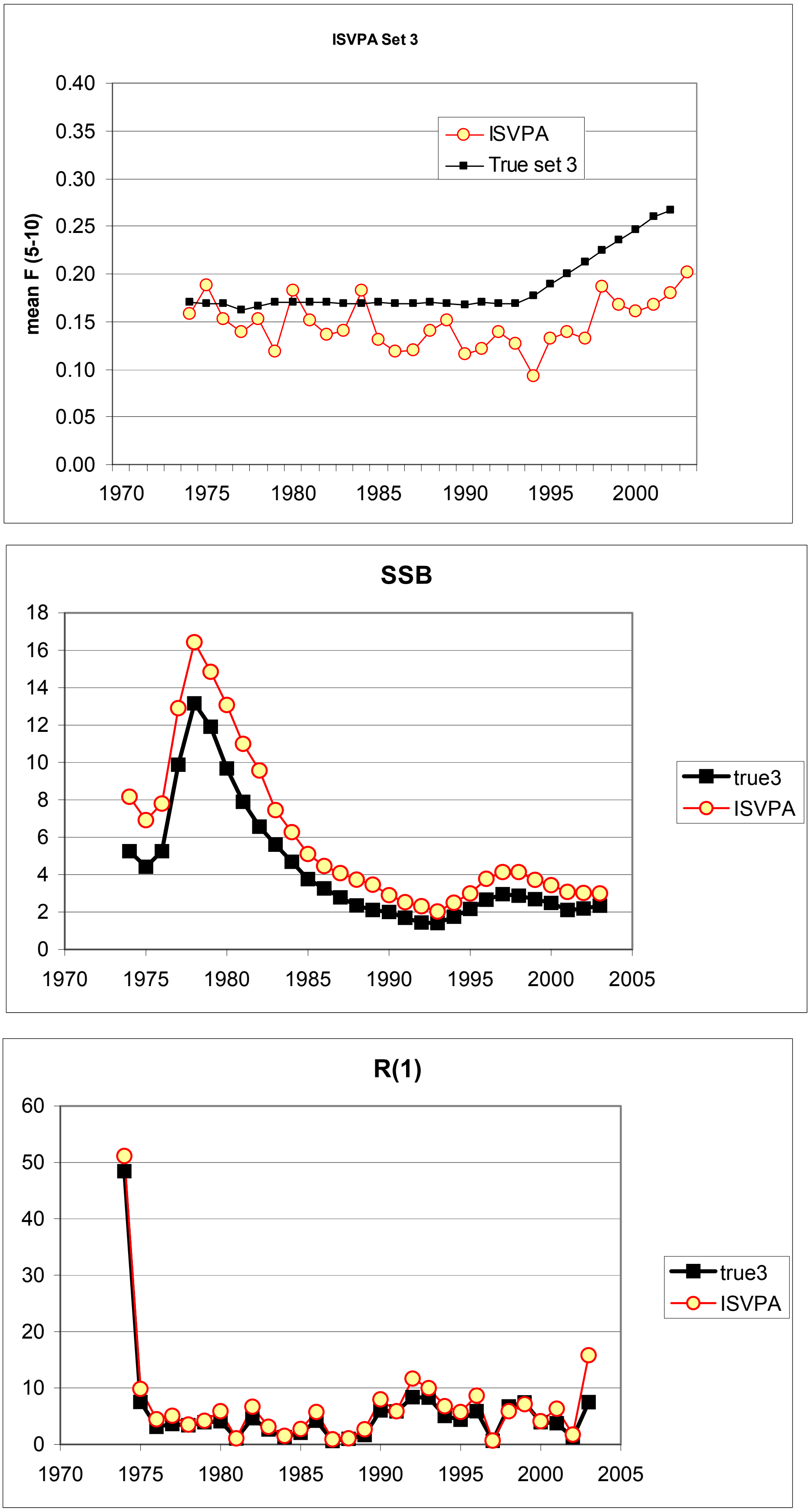


Figure 4.2.2.14. Estimated vs. “true” F (mean, age 5-10), SSB and age 1 recruitment using ISVPA with simulated data set 3.

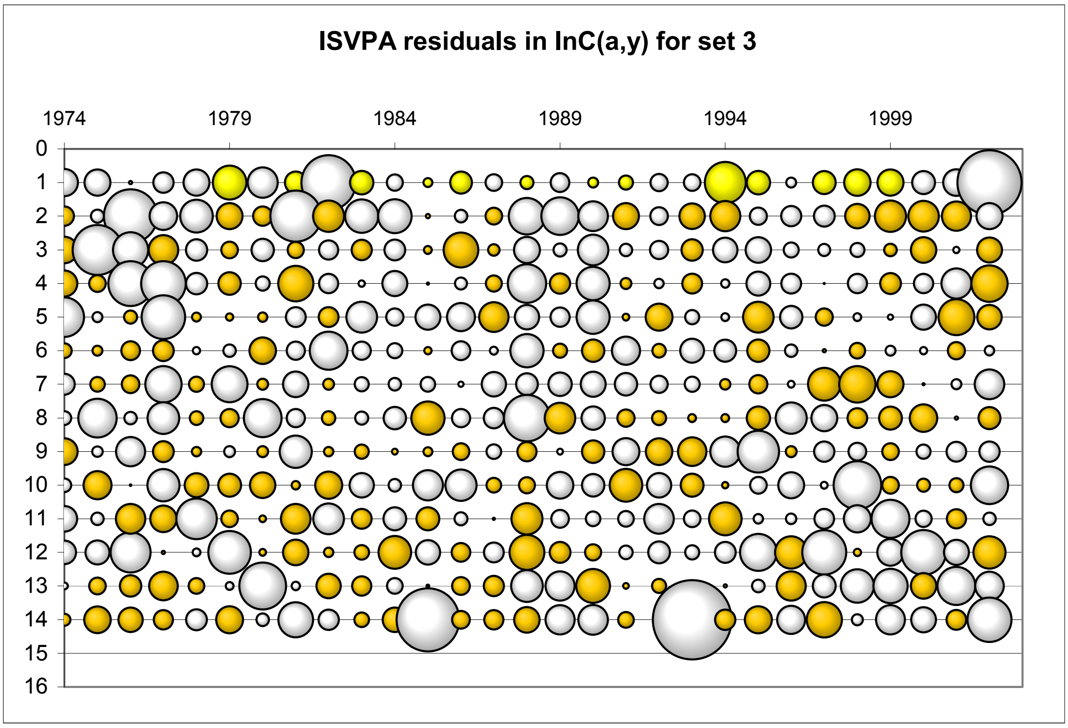


Figure 4.2.2.15. Log catch residuals for ISVPA fit using simulated data set 3.

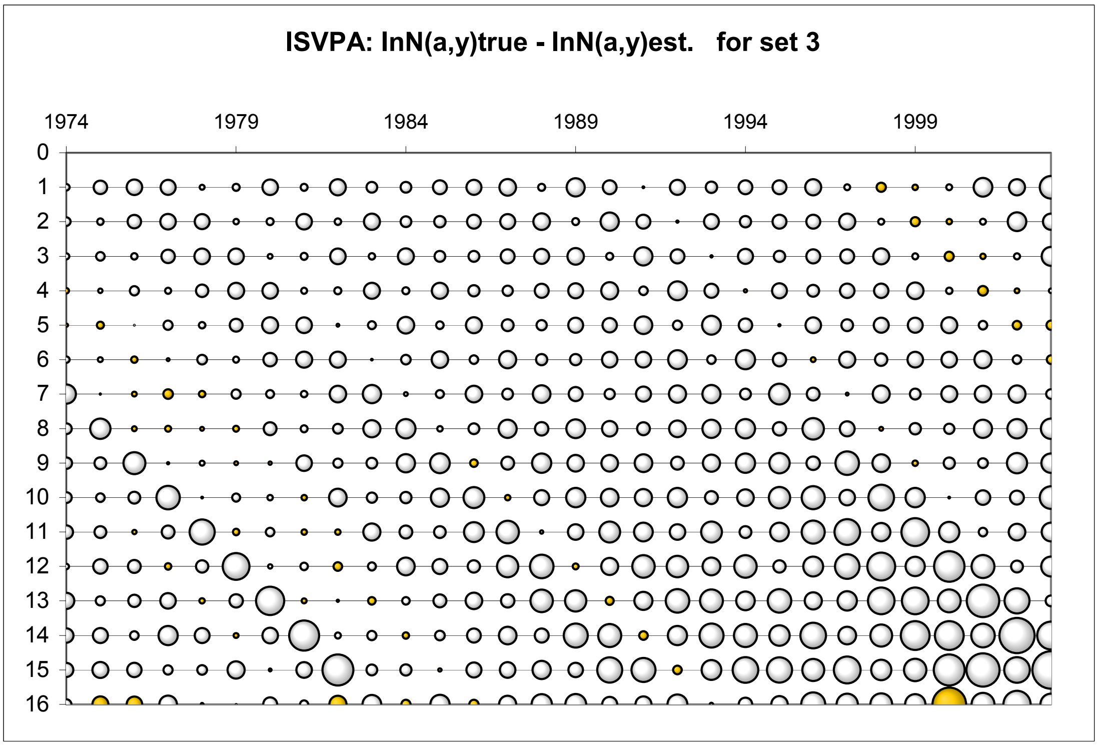


Figure 4.2.2.16. Log population size in numbers (N) residuals by year and age for ISVPA fit to simulated data set 3.

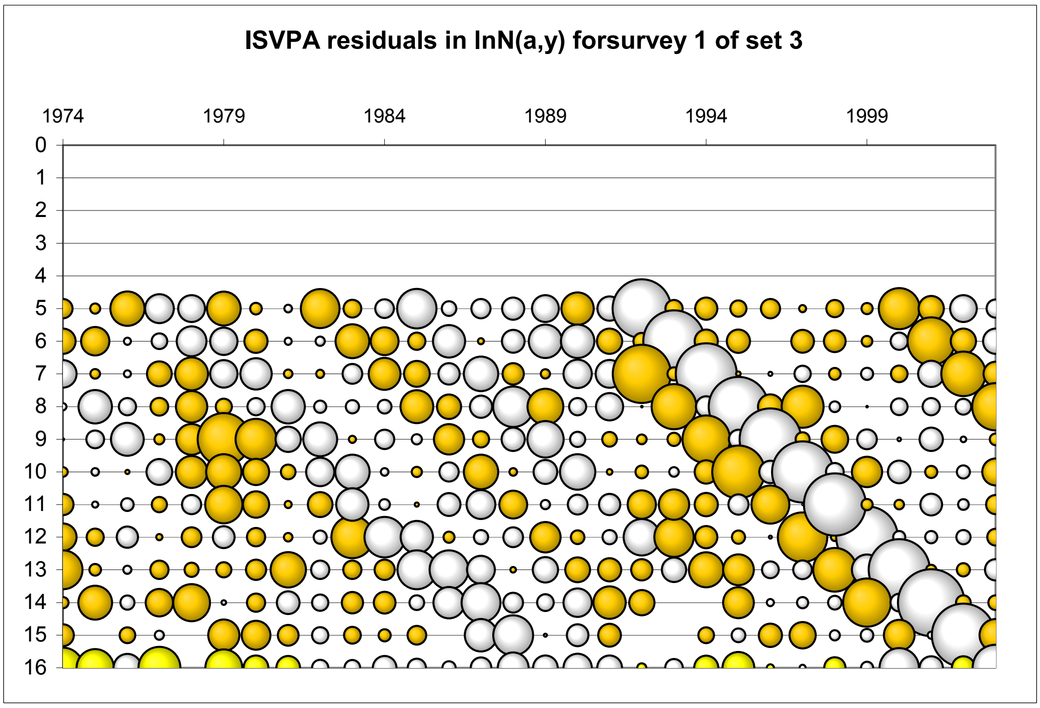


Figure 4.2.2.17. Log survey residuals (survey numbers vs. predicted) for ISVPA fit with data set 3 for survey 1. Residuals are given by age and year.

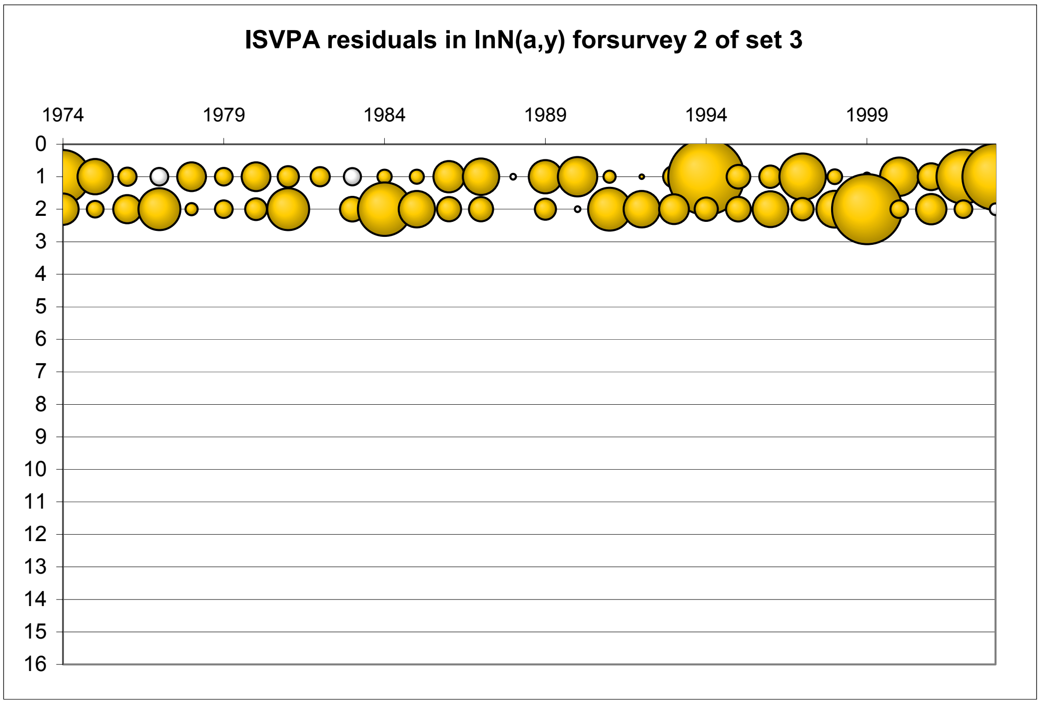
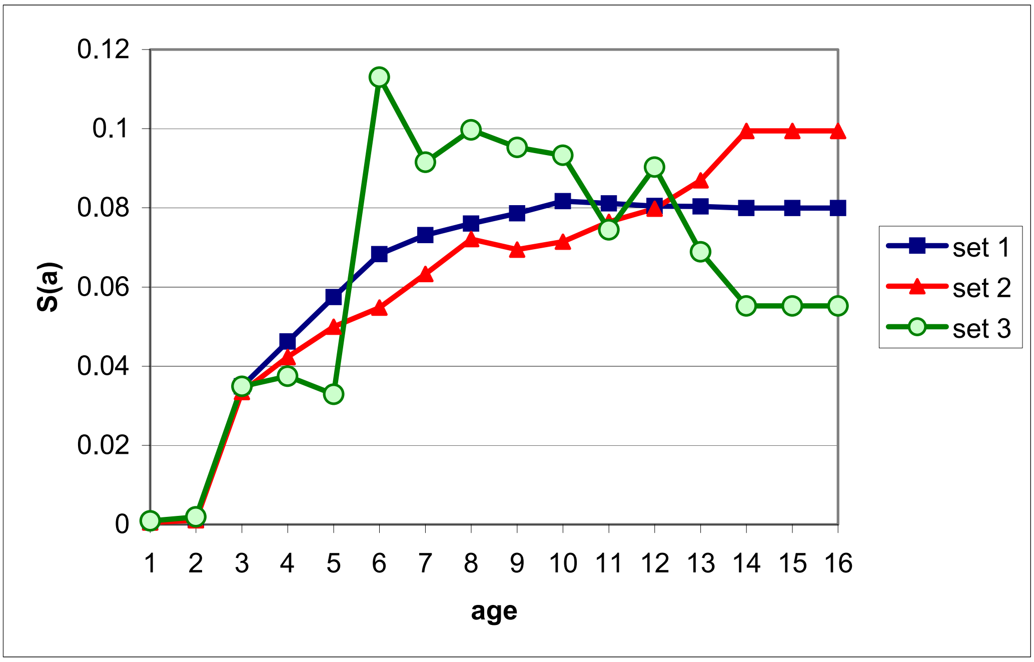
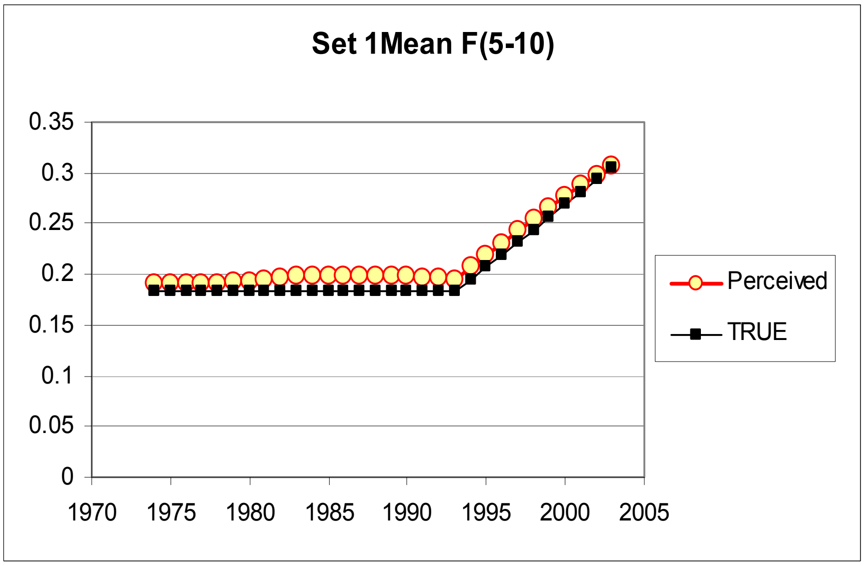


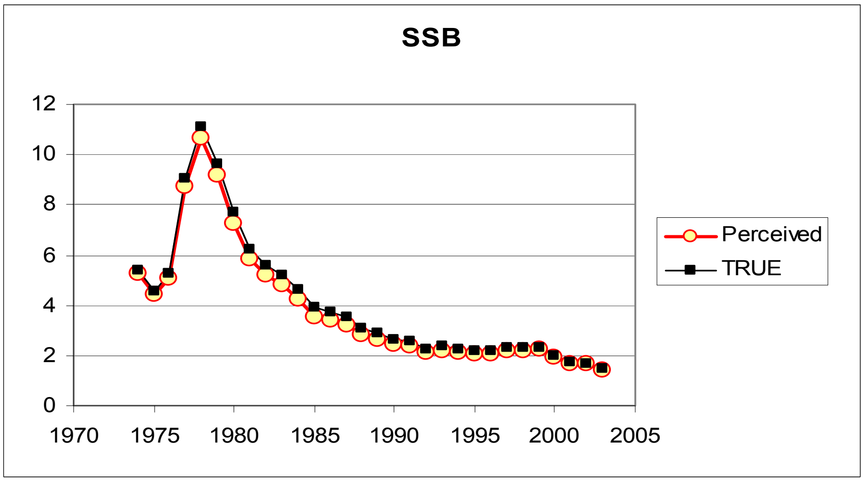
Figure 4.2.2.18. Log survey residuals (survey numbers vs. predicted) for ISVPA fit with data set 3 for survey 2. Residuals are given by age and year.



ISVPA selection patterns for data sets 1-3 (normalised by sum=1)

Figure 4.2.2.19. Calculated selection patterns at age for commercial catches from ISVPA runs Using simulated data sets 1, 2, and 3.





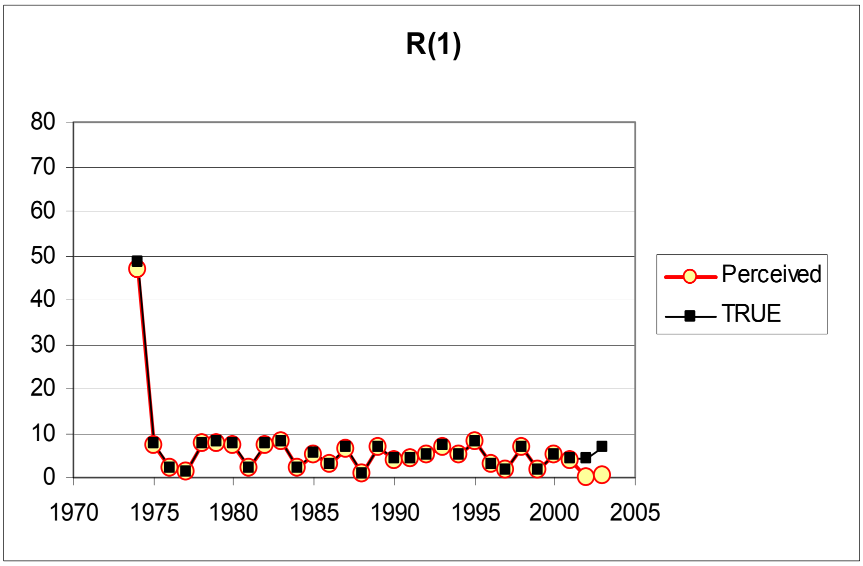


Figure 4.2.3.1. Estimated vs. "true" F (mean, age 5-10), SSB and age 1 recruitment using SeaStar with simulated data set 1. Perceived is the model result, "true" is the actual data.

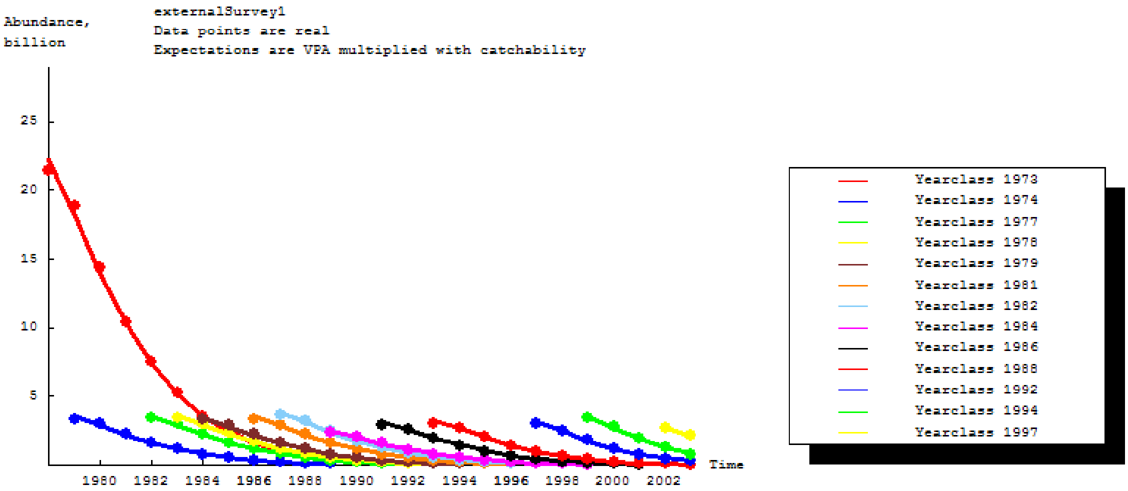


Figure 4.2.3.2 Comparison of historic stock and perceived data from survey 1 along tuned cohorts using SeaStar – data set 1.

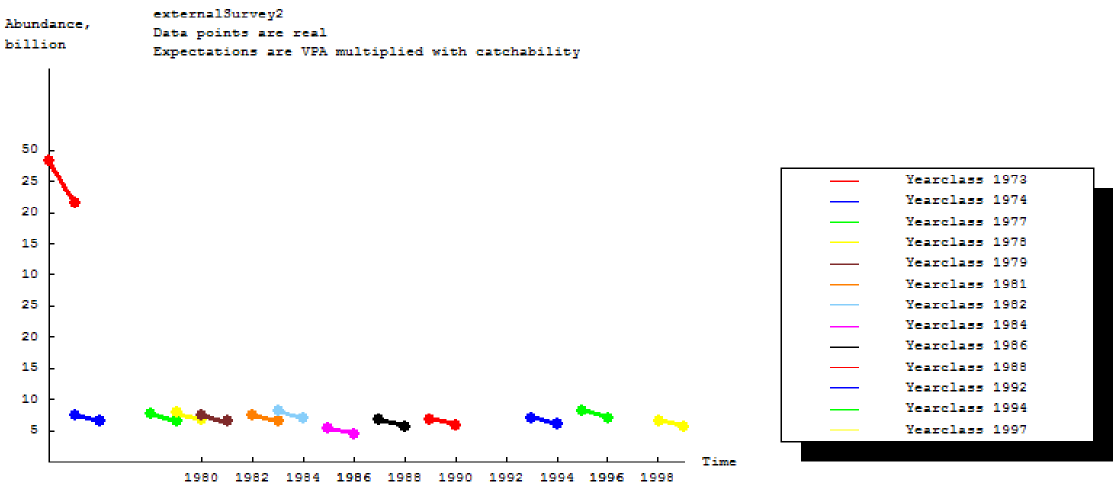


Figure 4.2.3.3 Comparison of historic stock and perceived data from survey 2 along tuned cohorts using SeaStar – data set 1.

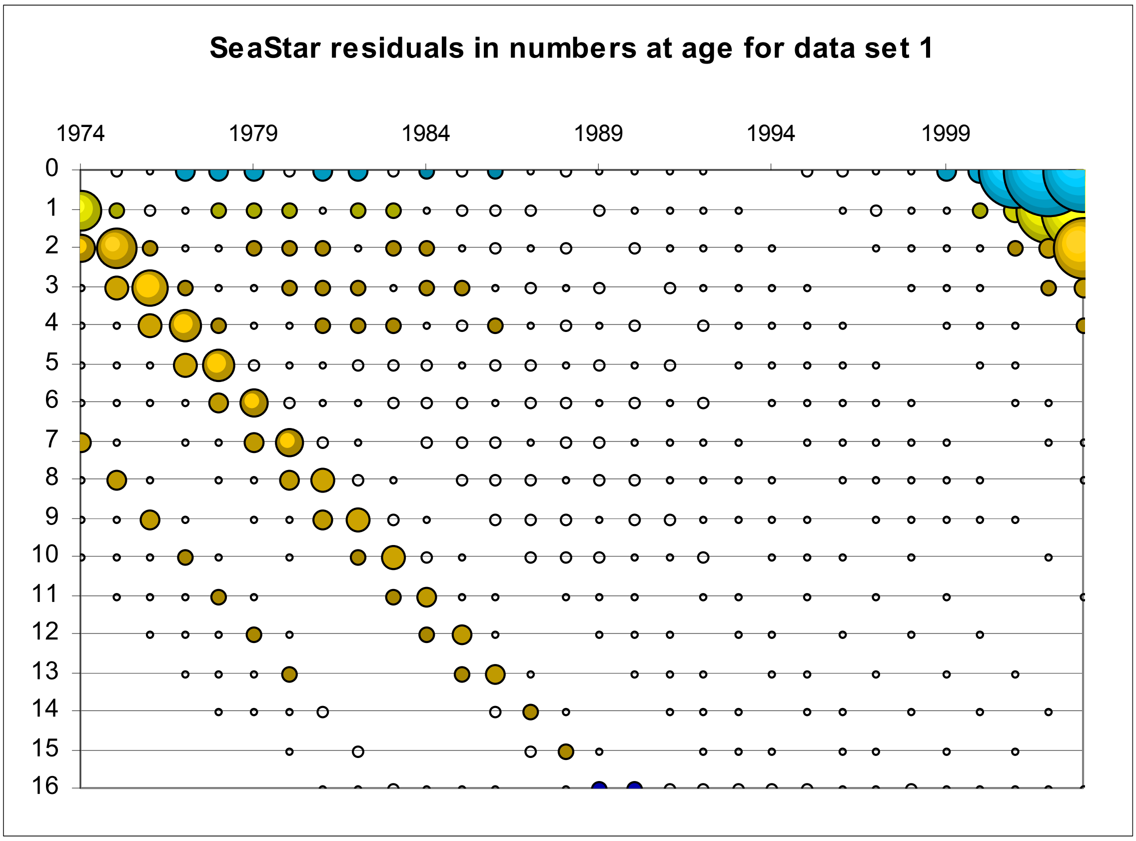


Figure 4.2.3.4 Comparison of perceived and true historic stocks using SeaStar – simulated data set 1.

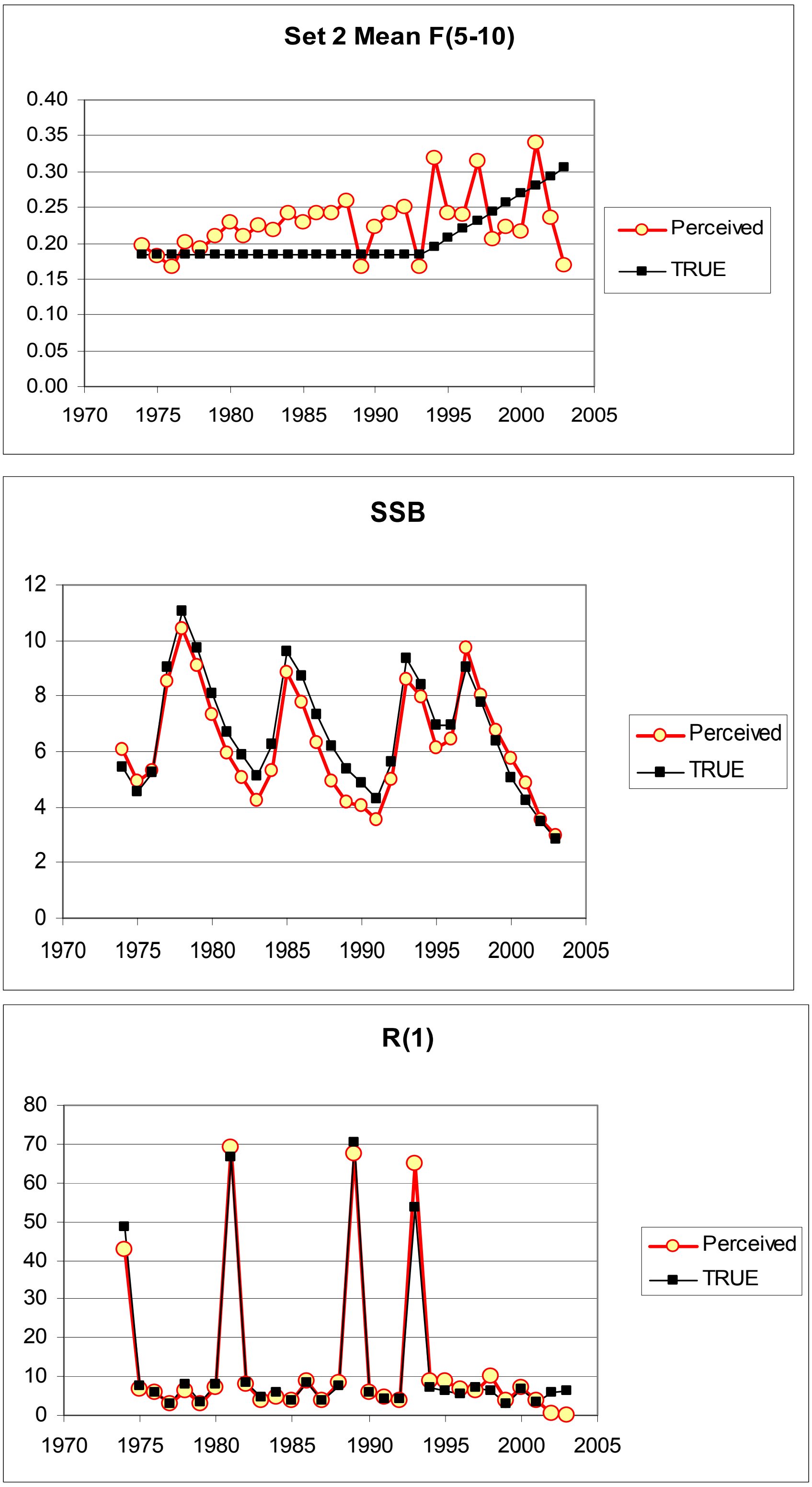


Figure 4.2.3.5. Estimated vs. "true" F (mean, age 5-10), SSB and age 1 recruitment using SeaStar with simulated data set 2. Perceived is the model result, "true" is the actual data.

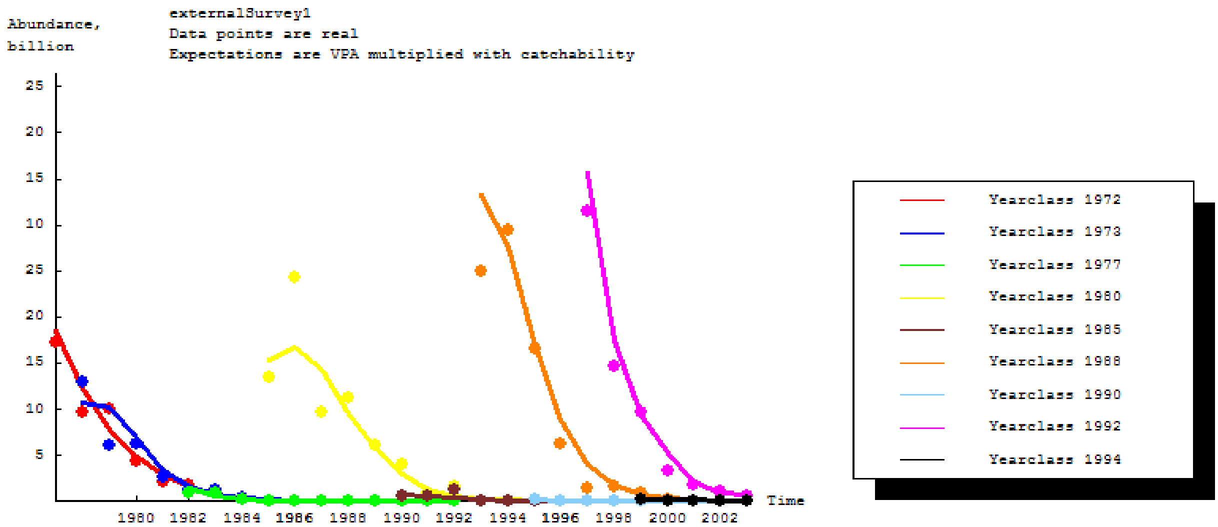


Figure 4.2.3.6 Comparison of historic stock and perceived data from survey 1 along tuned cohorts using SeaStar – data set 2.

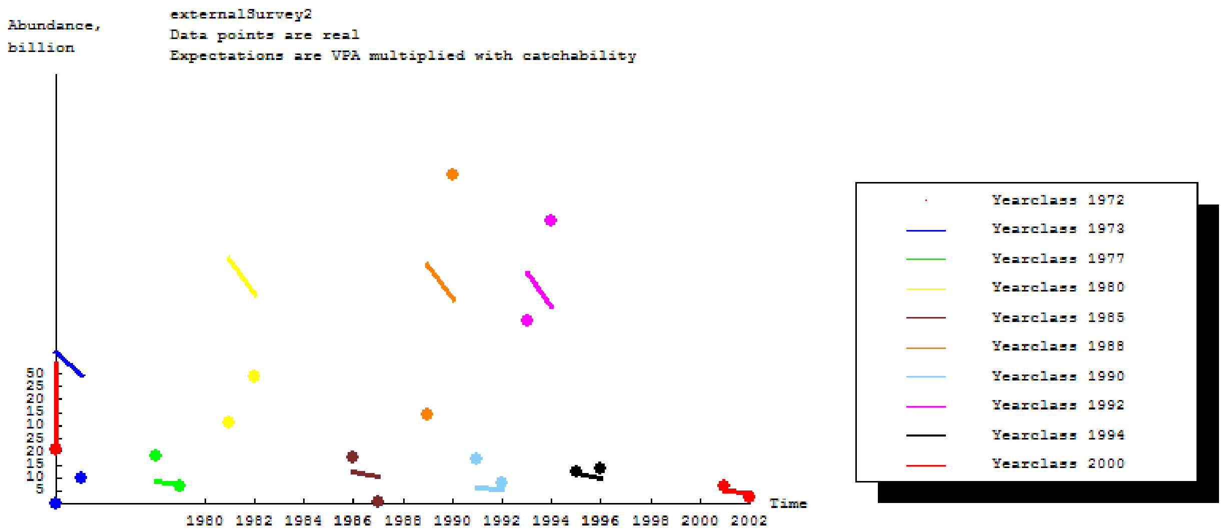


Figure 4.2.3.7. Comparison of historic stock and perceived data from survey 2 along tuned cohorts using SeaStar – data set 2.

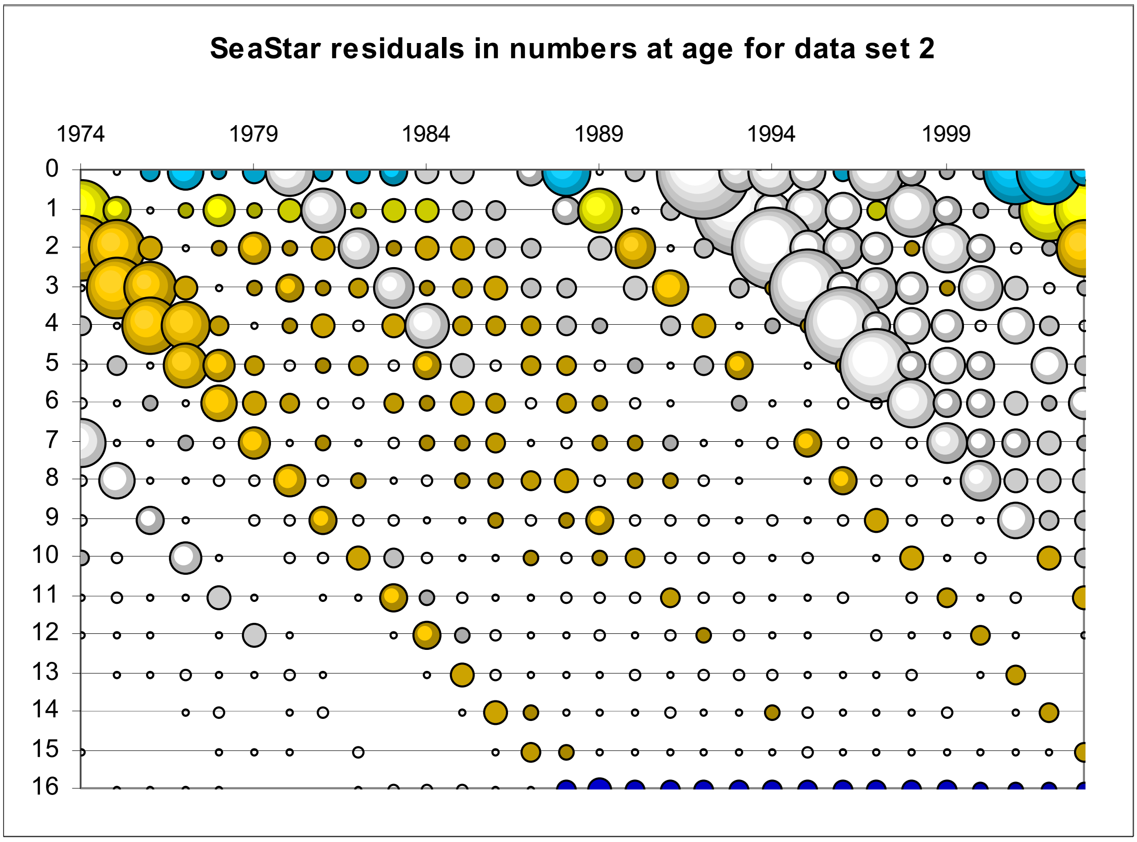


Figure 4.2.3.8. Comparison of perceived and true historic stocks using SeaStar – data set 2.

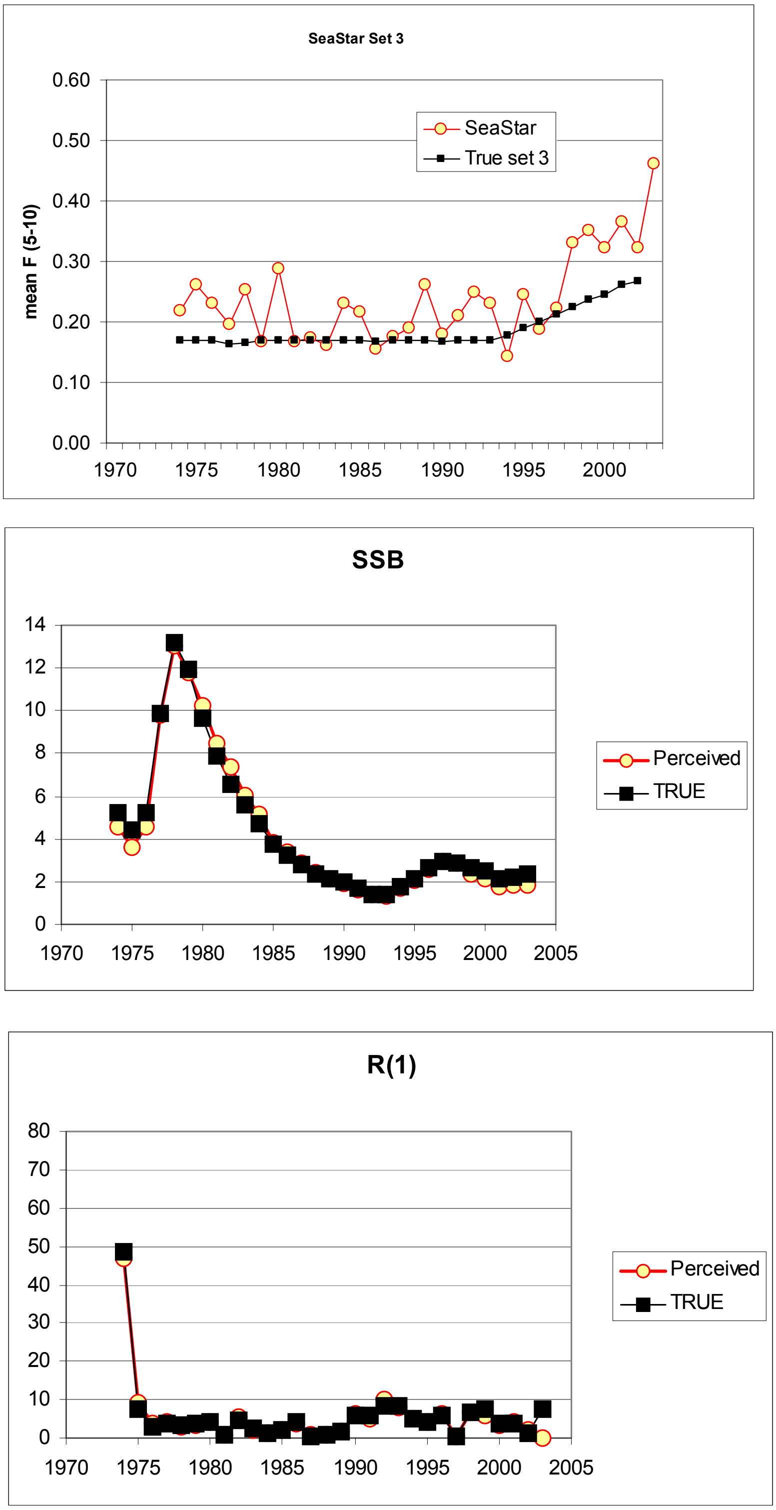


Figure 4.2.3.9. Estimated vs. "true" F (mean, age 5-10), SSB and age 1 recruitment using SeaStar with simulated data set 3. Perceived is the model result, "true" is the actual data.

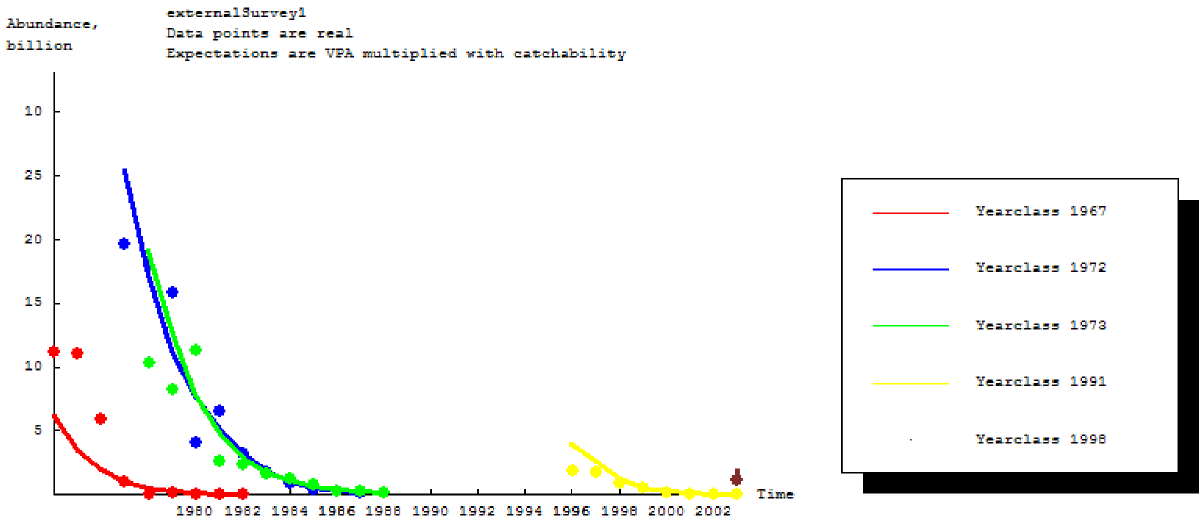


Figure 4.2.3.10. Comparison of historic stock and perceived data from survey 1 along tuned cohorts using SeaStar – data set 3.

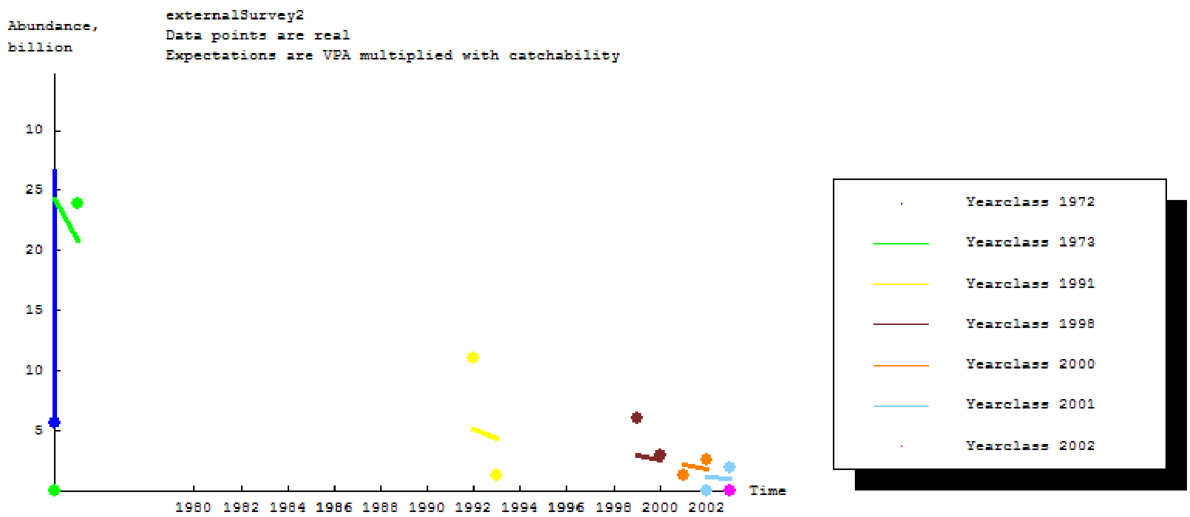


Figure 4.2.3.11. Comparison of historic stock and perceived data from survey 2 along tuned cohorts using SeaStar – data set 3.

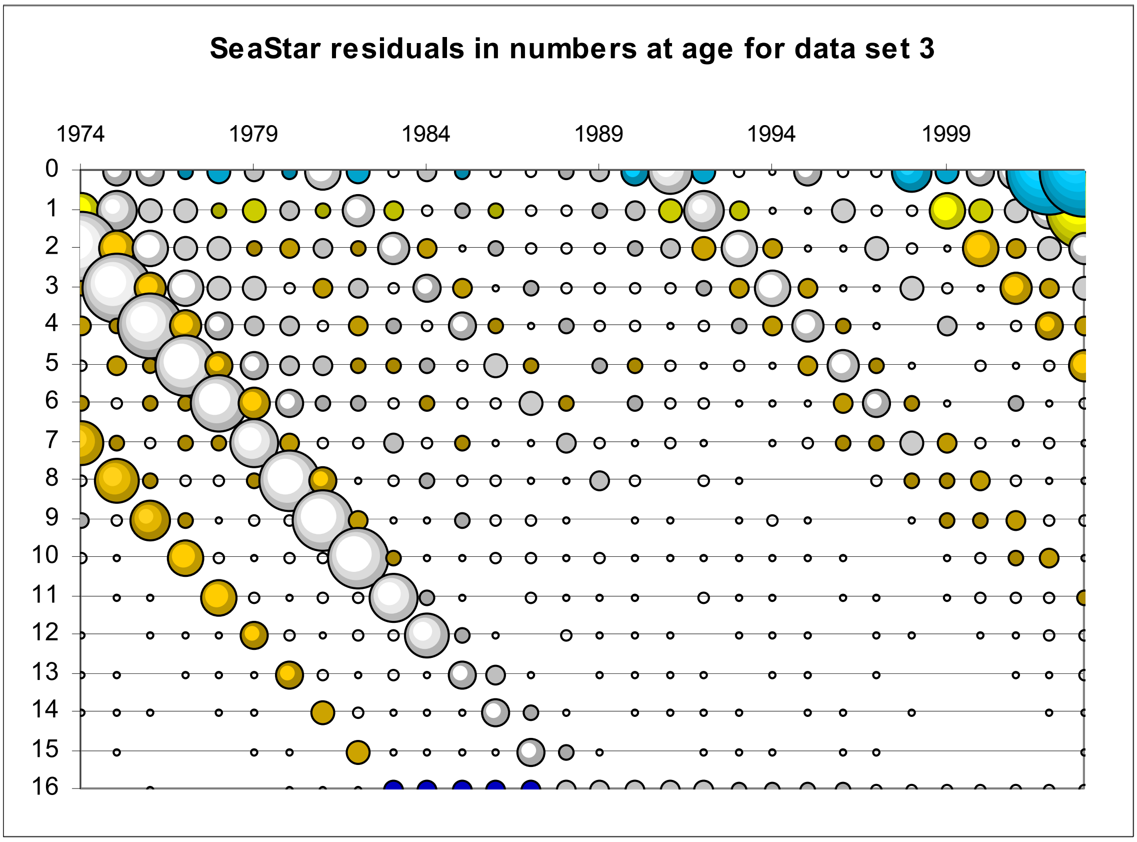


Figure 4.2.3.12. Comparison of perceived and true historic stocks using SeaStar – data set 3.

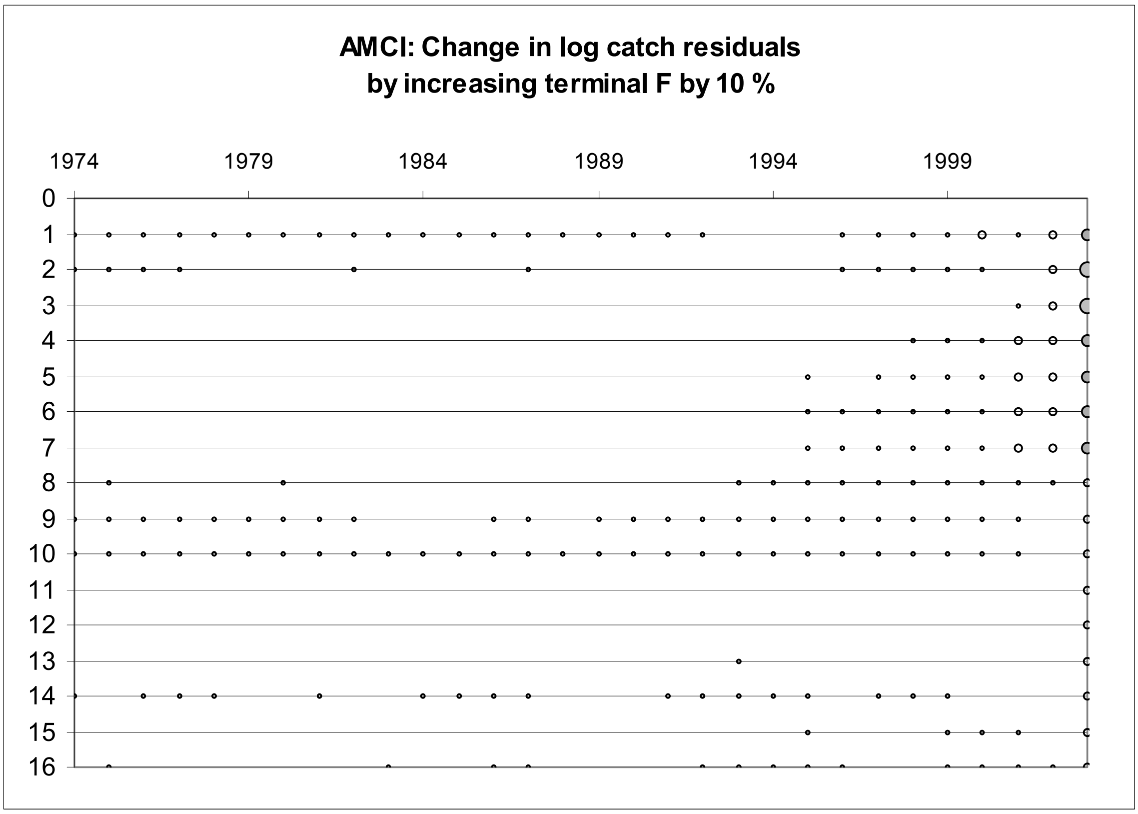


Figure 4.3.1. Perturbation test for AMCI using simulated data set 2 and increasing the terminal F by 10%. Results are log catch residuals by age and year.

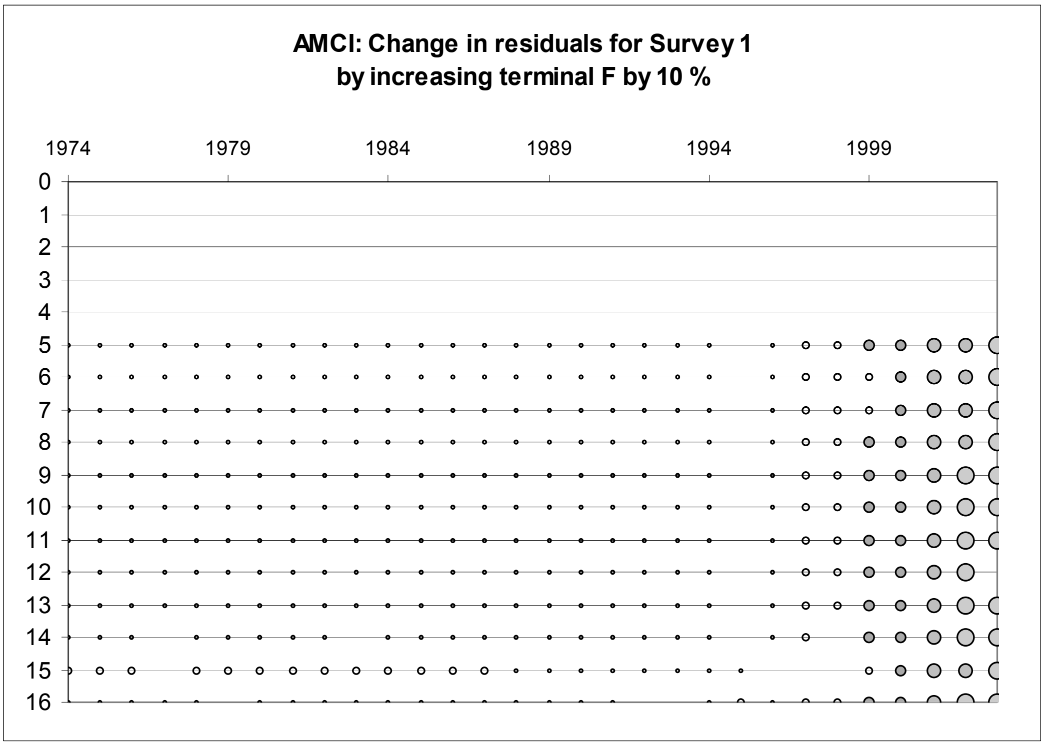


Figure 4.3.2. Perturbation test for AMCI using simulated data set 2 and increasing the terminal F by 10%. Results are log residuals by age and year for survey catches for survey 1.

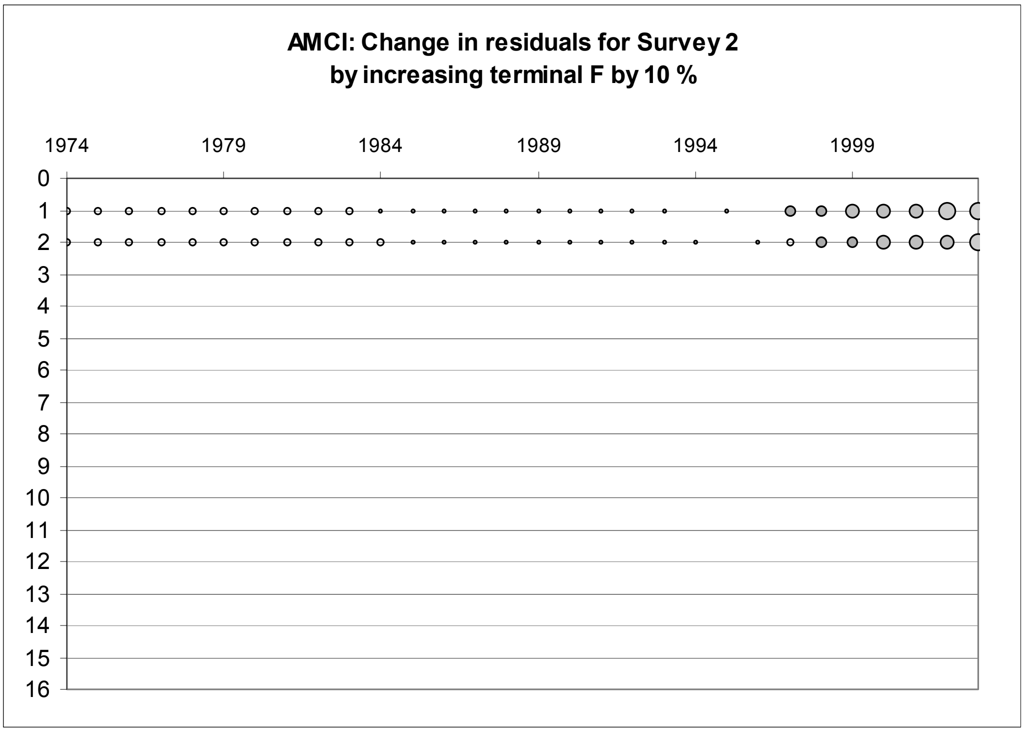


Figure 4.3.3. Perturbation test for AMCI using simulated data set 2 and increasing the terminal F by 10%. Results are log residuals by age and year for survey catches for survey 2.

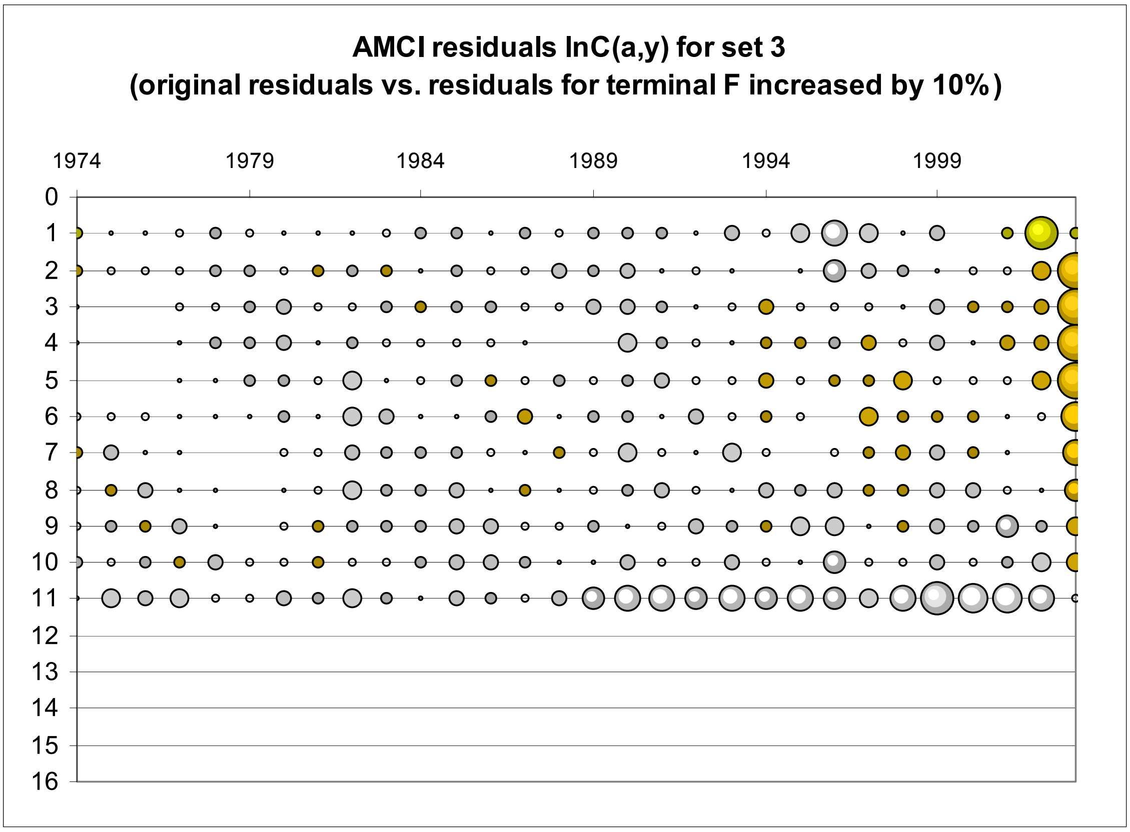


Figure 4.3.4. Perturbation test for AMCI using simulated data set 3 and increasing the terminal F by 10%. Results are log catch residuals by age and year.

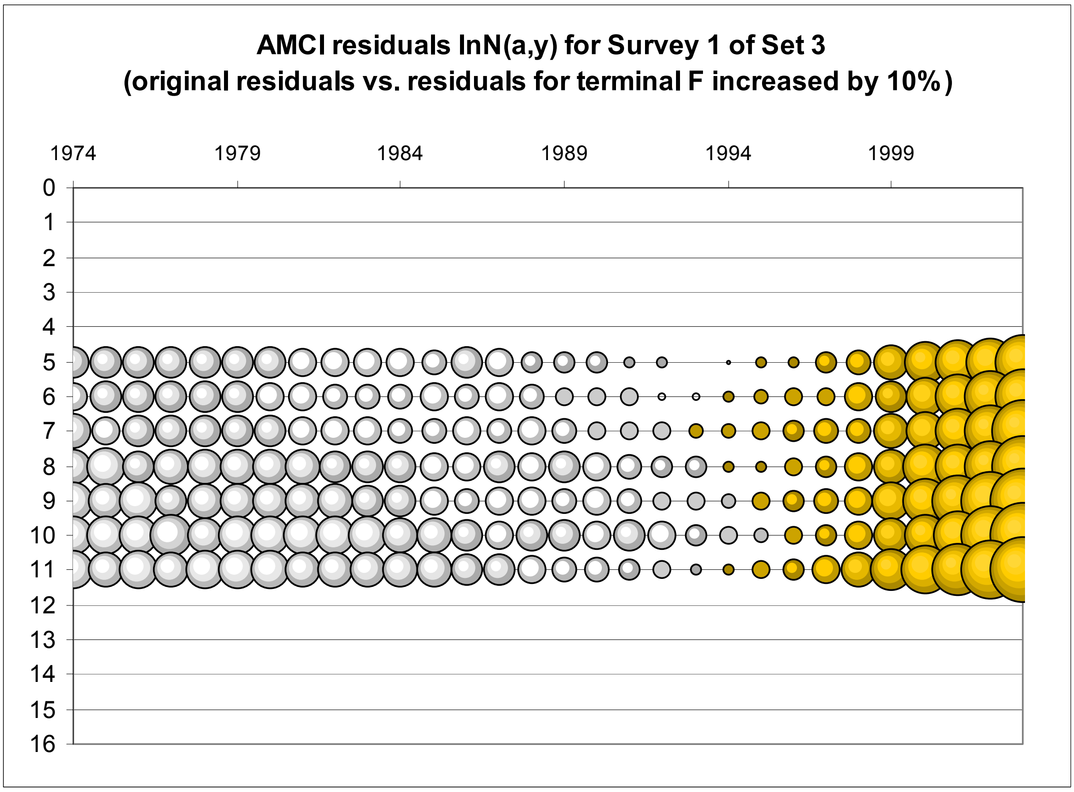


Figure 4.3.5. Perturbation test for AMCI using simulated data set 3 and increasing the terminal F by 10%. Results are log residuals by age and year for survey catches for survey 1.

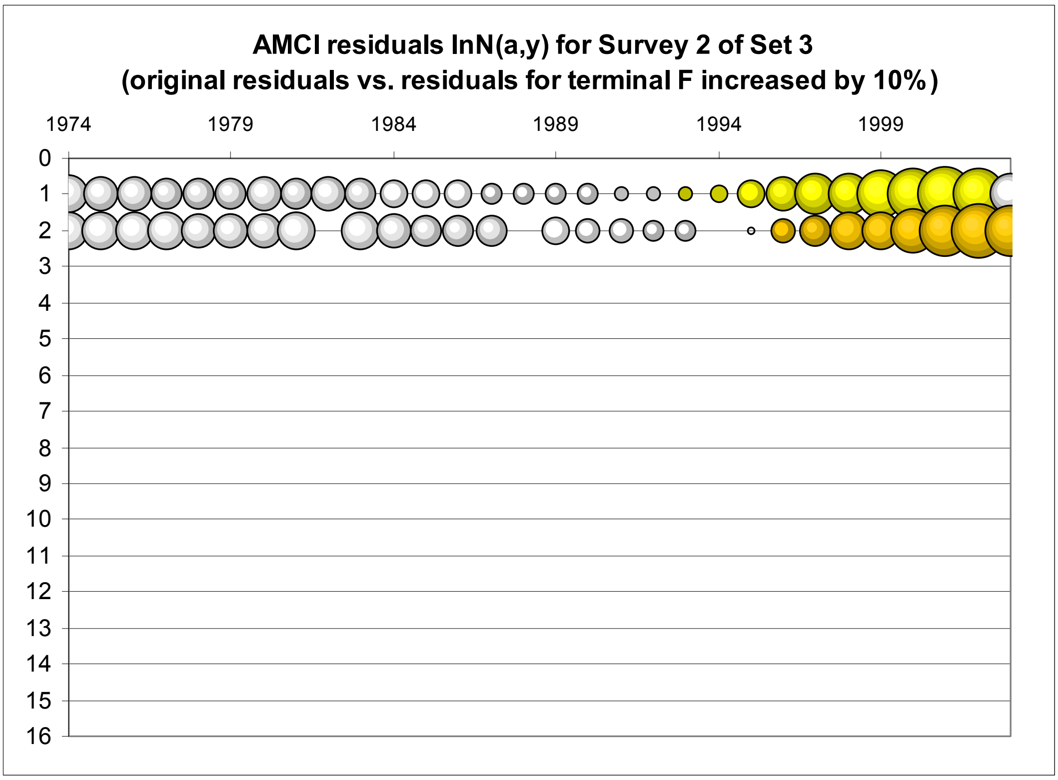


Figure 4.3.6. Perturbation test for AMCI using simulated data set 3 and increasing the terminal F by 10%. Results are log residuals by age and year for survey catches for survey 2.

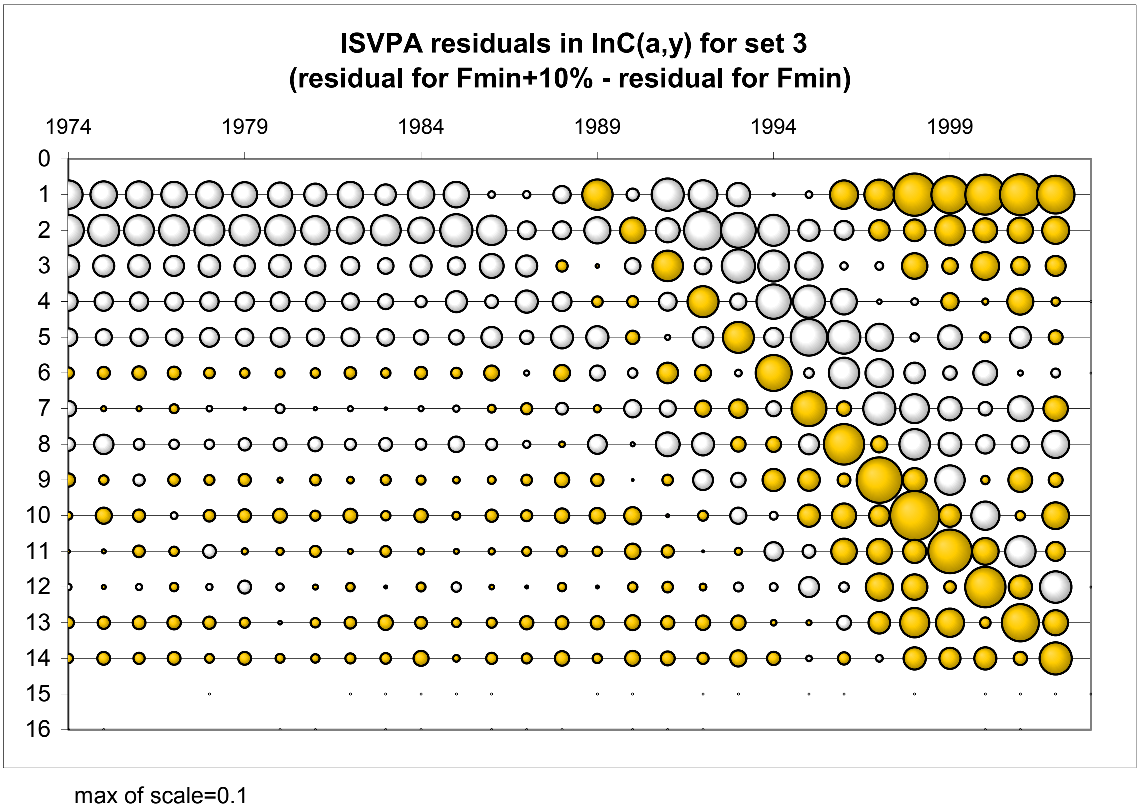
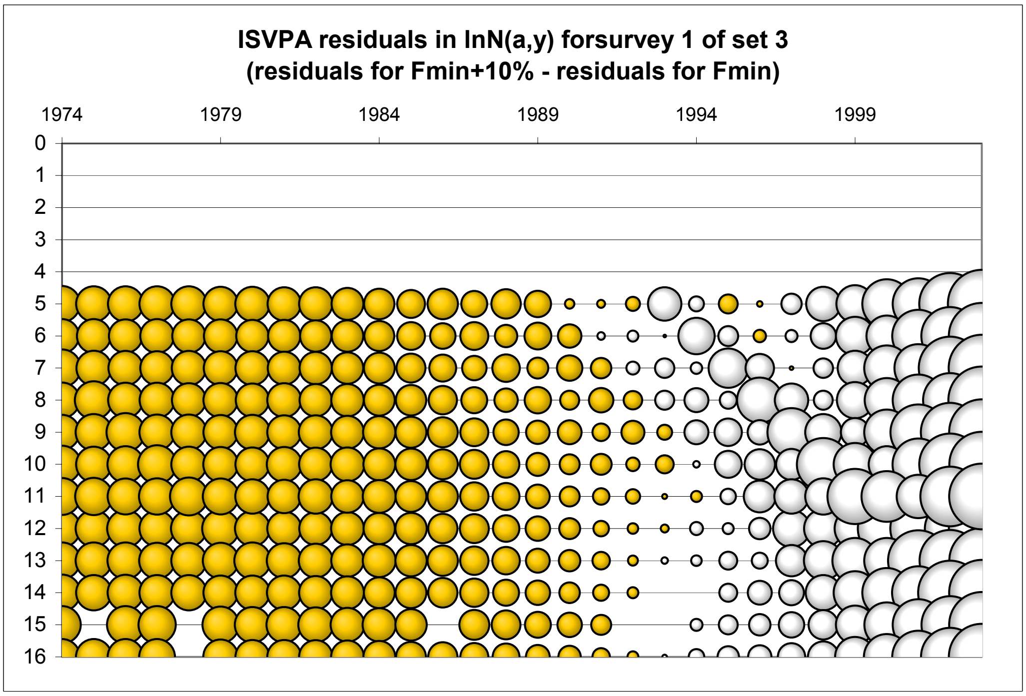
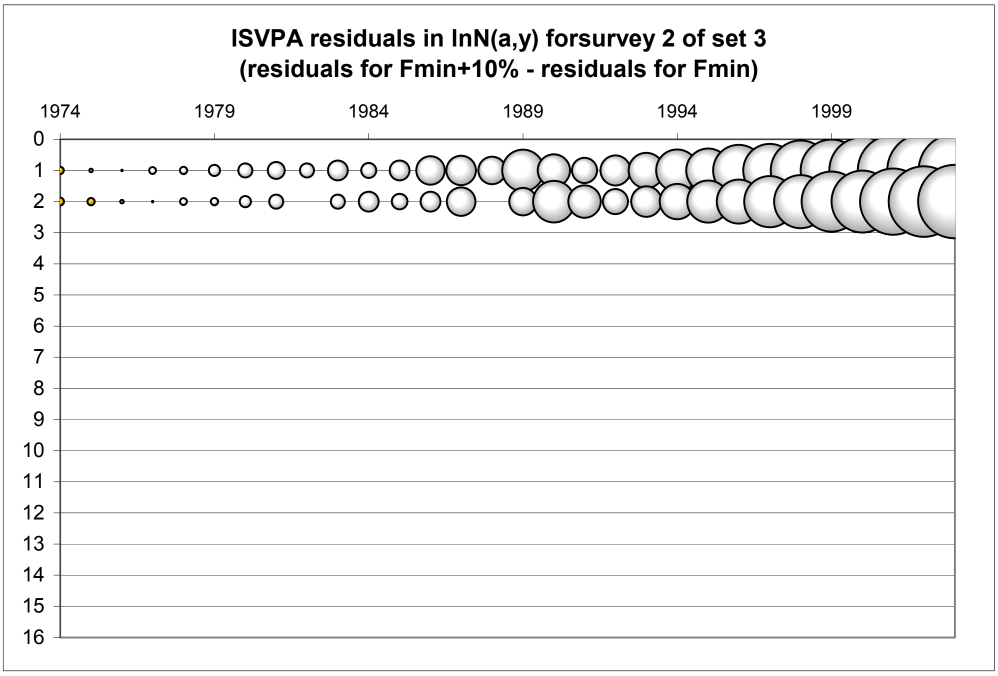


Figure 4.3.7. Perturbation test for ISVPA using simulated data set 3 and increasing the terminal F by 10%. Results are log catch residuals by age and year.



max of scale=0.1

Figure 4.3.8. Perturbation test for ISVPA using simulated data set 3 and increasing the terminal F by 10%. Results are log residuals by age and year for survey catches for survey 1.



max of scale=0.1

Figure 4.3.9. Perturbation test for ISVPA using simulated data set 3 and increasing the terminal F by 10%. Results are log residuals by age and year for survey catches for survey 2.

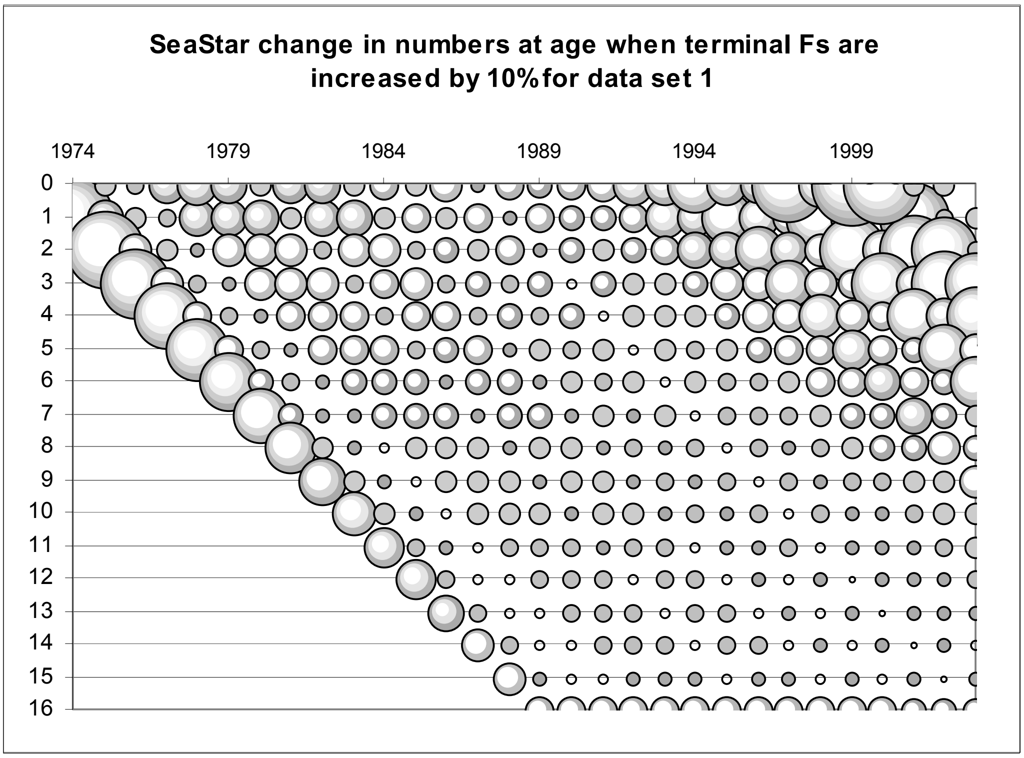


Figure 4.3.10. Perturbation test for SeaStar using simulated data set 1 and increasing the terminal F by 10%. Results are population numbers residuals by age and year.

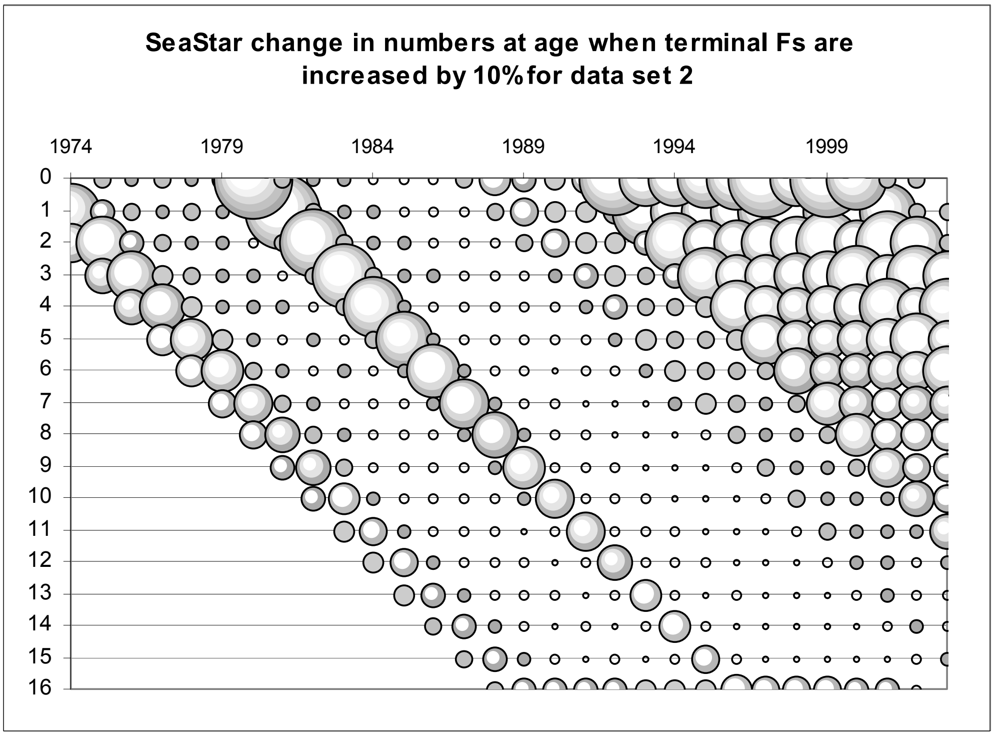


Figure 4.3.11. Perturbation test for SeaStar using simulated data set 2 and increasing the terminal F by 10%. Results are population numbers residuals by age and year.

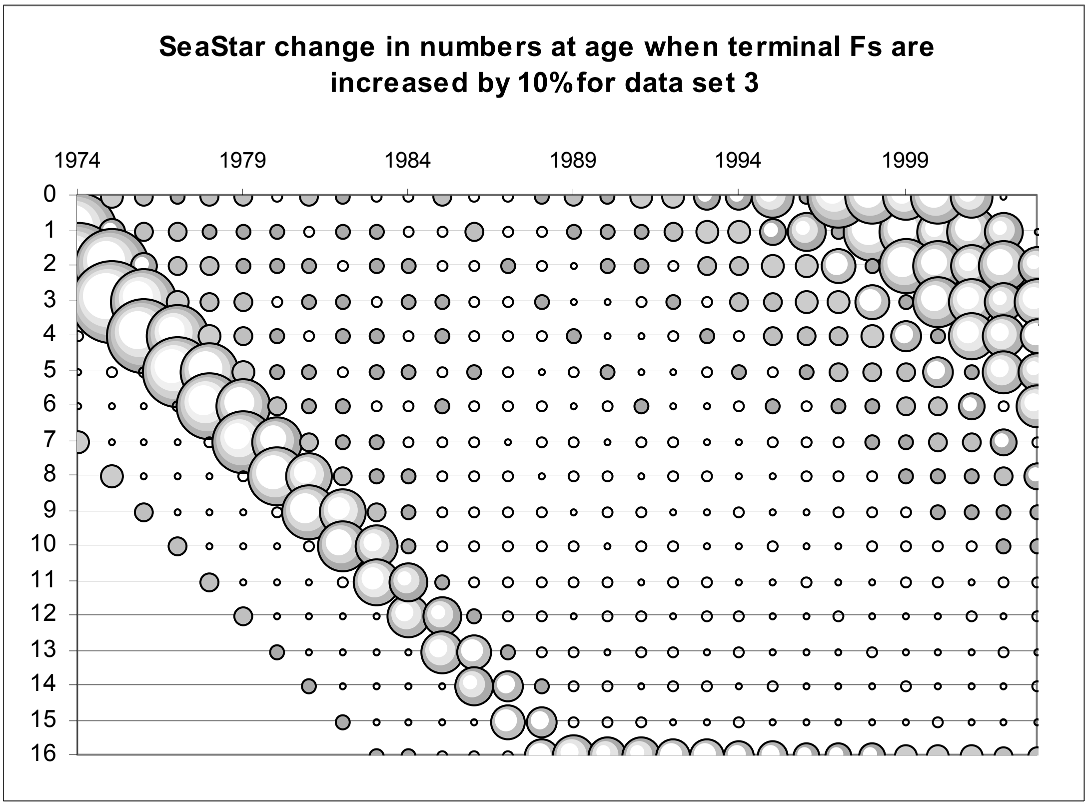


Figure 4.3.12. Perturbation test for SeaStar using simulated data set 3 and increasing the terminal F by 10%. Results are population numbers residuals by age and year.

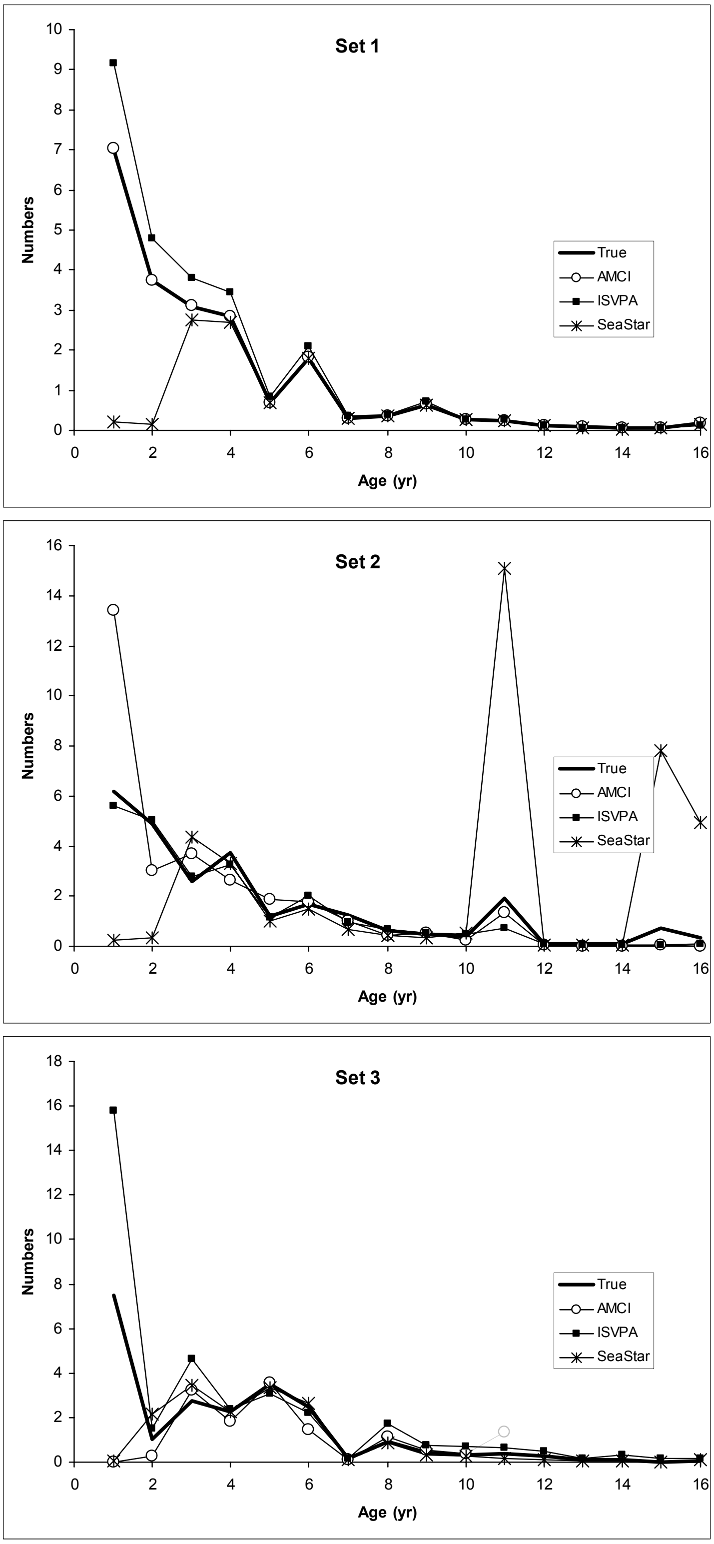


Figure 4.4.1. Population abundance at age in the terminal year for the three test data sets as estimated with AMCI, ISVPA and SeaStar.

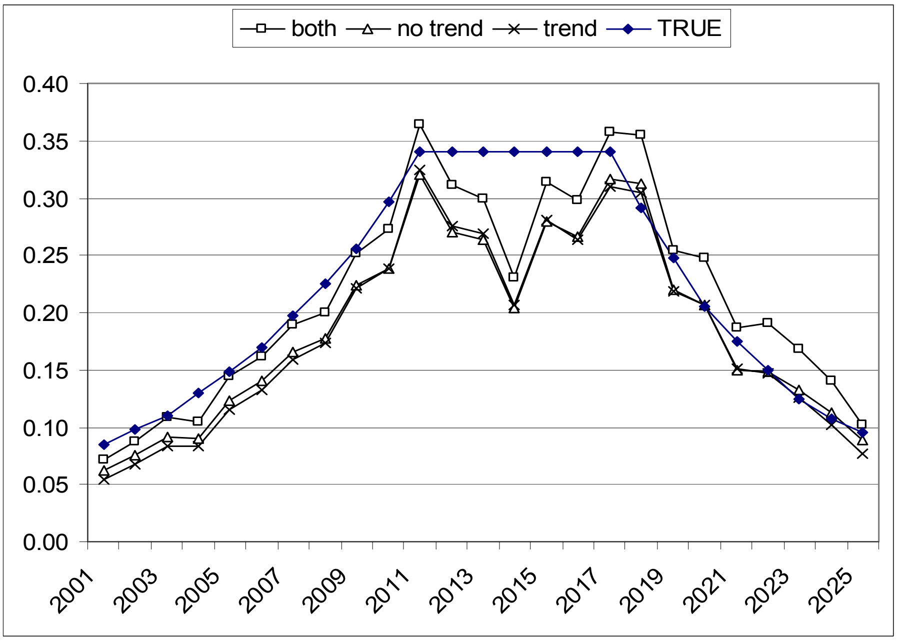


Figure 4.5.1. Tests of SeaStar using simulated data set from the 2003 ICES Methods WG Report. The "true" data are the simulated results, and three options were fit: (1) both tuning fleets Combined, (2) fleets without trend, and (3) fleet with q trend. Comparisons are for mean Fishing mortality rates for ages 8-12, by year.

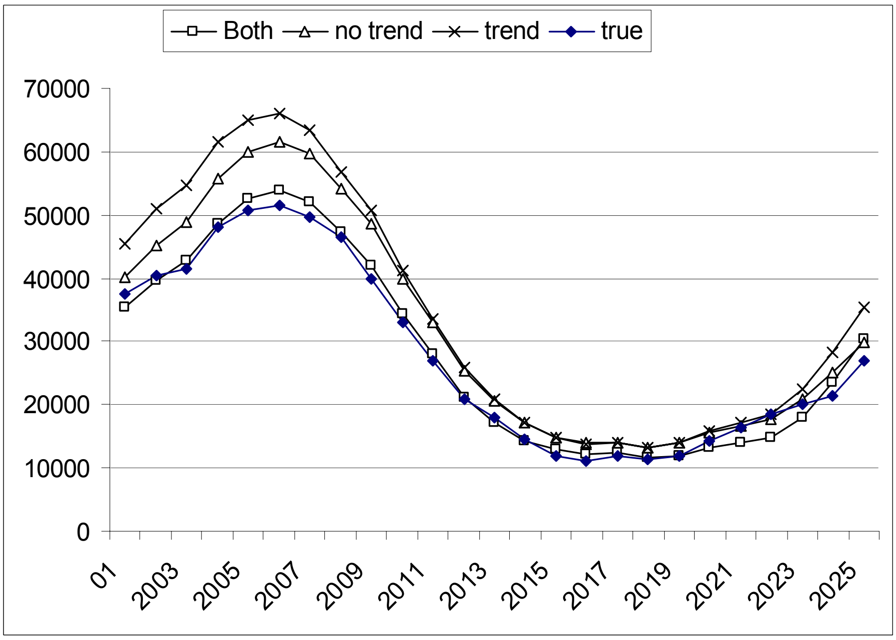


Figure 4.5.2. Tests of SeaStar using simulated data set from the 2003 ICES Methods WG Report. The "true" data are the simulated results, and three options were fit: (1) both tuning fleets Combined, (2) fleets without trend, and (3) fleet with q trend. Comparisons are for SSB.

5 CONSIDERATIONS IN FUTURE ASSESSMENT APPROACHES

In its 2003 report, the Methods Working Group considered the application of AMCI and ISVPA to the blue whiting assessment (ICES 2003b). They had a number of recommendations relative to the further testing of various assessment approaches, and uses of specific data sets. Specifically, The Methods Working Group recommended:

1) The choice of appropriate model to assess the stock is not clear-cut and the approach used (at the meeting of the Methods WG) of exploring a number of competing models is to be commended as an aid to disentangle the apparent conflicting sources of data
2) That one particular CPUE series and historical acoustic data series be excluded or handled differently
3) It was found that the survey data were very noisy and often contradictory. It was suggested that if sampling for age-disaggregated indices was poor, that indices be combined into SSB indices (estimates)

With regard to recommendation (1), the Study Group has continued the exploration of the two assessment approaches applied to each stock in order to isolate the interactions between models and data that result in diverging perceptions regarding stock status.

In situations where different sources of data provide conflicting signals about the underlying dynamics of a fish stock (e.g., the current issues surrounding the blue whiting and Norwegian spring-spawning herring stock assessments), the choice of an appropriate model may have to be made from a range of competing but equally plausible models. This may be because conflict in the data is handled differently by the various models (e.g. models based on ISVPA, SeaStar or AMCI). As long as the assumptions underlying these models are justified by the data, it may be desirable to combine results from these models in some way to reflect model-structure uncertainty, instead of selecting a "best" model and thus ignoring model-structure uncertainty. Accounting for model-structure uncertainty is important for developing robust management advice (Punt and Hilborn 1997, Richards and Maguire 1998, Butterworth and Punt 1999). A possible approach is one based on Bayesian Decision Analysis, which attempts to provide empirically based weightings for stock-assessment results from structurally different models (McAllister and Kirchner 2002). While not necessarily advocating a formal model-based approach for dealing with model (structural) uncertainty, the Study Group recognized that this is an issue that is confronted by many working groups, in ICES and elsewhere.

The issue of model choice is one that is confronted by most management institutions world-wide. There are several schools of thought on advising management, particularly when the outcomes from the application of multiple models differ. First, the generation of multiple model approaches to the same assessment is generally considered a positive step, given uncertainty in fundamental ecological processes, and incomplete and potentially noisy data (e.g., NRC 1998). Given these types data problems (e.g., blue whiting) one may expect assessment results to differ. How should management consider such multiple outcomes? One approach is to apply Bayesian weighting, as explored above. The danger with this approach that it may average "right" and "wrong" outcomes to advise something that is less wrong, but still wrong. The issue then becomes the consequences to the stock and fishery of being moderately wrong. A second approach would be to choose the more precautionary of the scientific results, given the uncertainties that underlie the various models. This assumes that the truth lies in one of the models (or perhaps between their outcomes), and that when conflicting realities require a choice, that deference is given to protecting the resource. The consequences of such a procedure are that the appropriate management scheme is followed if the more conservative model outcome is in fact reflecting the "truth". However, if the less conservative model outcome is more reflective of the truth, then there will be foregone economic benefits, which, depending on the circumstances, may be considerable. In the latter situation, catches may be increased if the management program was initially too conservative, but in the reverse situation, there may be a big economic penalty to pay from a wrong model choice. These are clearly issues for ACFM and management authorities to ponder. As far as the Northern pelagic Working Group is concerned, the fundamental conundrum is that the truth is essentially unknown (especially for blue whiting), and the data sources for fishery independent abundance trends are not particularly informative. Better data would clearly inform choices on particular model outcomes.

It is clear that none of the models evaluated in simulation studies consistently out performed the others. However, all three approaches have important deficiencies related to how much the developers of the method had to be involved in order to undertake routine and special case assessments. A more flexible and transparent assessment engine is considered a priority for the group.

6 CONCLUSIONS

- Objective methods for screening data series for their likely ability to inform stock assessment models have been developed by the Study Group and applied to data sets used in Norwegian spring spawning herring and blue whiting assessments. In several cases, the implications of these analyses are that time series of low precision or short time interval should be dropped from assessment model runs. The implications of these proposed adjustments to the data series have not been preformed by the Study Group – these are recommended to the Northern Pelagic and Blue Whiting Working Group for further investigation.
- The timing of the annual assessment cycle, especially for Norwegian spring-spawning herring (e.g., late April) is such that annual recruitment indices from surveys become available immediately after the Northern Pelagic and Blue Whiting Working Group concludes its work (May surveys in the Barents and Norwegian Seas). While the Study Group members understand that this will always be a problem since data collection is a continuous process, in this case, ACFM might consider shifting the timing of the Working Group slightly so that these recruitment indices could be accommodated in near real time assessments. This issue is considered particularly important since estimates of recruitment tend to be the most divergent signals coming from the models (Figure 1.4.1)
- The Study Group recommends some censoring of the input data sources to eliminate information that is essentially noise and that may conflict with data sources from which consistent signals seem to be emerging. These recommendations are provided in Chapter 2.
- The study group recommends that ACFM and the Working Group consider the development of a new synoptic, and unambiguous fishery independent survey time series for Blue Whiting.
- Extensive simulation testing undertaken by the Study Group did not detect consistent biases in estimated F, SSB or recruitment derived from AMCI, ISVPA or SeaStar models on the order of some differences among assessment outcomes as revealed by the Northern Pelagic Working Group. There may be a slight tendency of SeaStar to underestimate terminal recruitment, but this is a tentative result that requires further work
- Additional simulation testing would be necessary to isolate conflict in the data sets that is the likely source of discrepancies in model results. These simulations would entail more sophisticated and realistic tuning data series with noise and correlation. Such simulations are not trivial and would require considerable forethought
- The Study Group recommends consideration of the development of a more consistent assessment methodology for these stocks that incorporates a flexible, transparent and well-supported software and computational engine.

7 LITERATURE CITED

Anon. 2003. Evaluation of research surveys in relation to management advice (EVARES - FISH/2001/02 - Lot 1) Final Report - February 2003 European Commission Director - General Fisheries

Butterworth, D.S. and A.E. Punt 1999 - Experiences in the evaluation and implementation of management procedures. ICES J. Mar. Sci. 56: 985-998.

Collie, J.S., and M.P. Sissenwine. 1983. Estimating population size from relative abundance data measured with error. Can. J. Fish. Aquat. Sci. 40:1871-1879.

Fournier, D., and Archibald, C. P. (1982). A general theory for analyzing catch at age data. Canadian Journal of Fisheries and Aquatic Sciences, 39: 1195-1207.

Fryer, R.J. (2002). TSA: is it the way? In Appendix D of the Report of the Working Group on Methods on Fish Stock Assessments. ICES CM 2002/D:01.

Gavaris, S. (1988) An adaptive framework for the estimation of population size. Canadian Atlantic Fisheries Scientific Advisory Committee Research Document No. 88/29, 12 pp.

Gudmundsson, G. (1994). Time-series analysis of catch-at-age observations. Applied Statistics 43:117-126.

Methot, R.D. (2000). Technical description of the stock synthesis assessment program. U.S. Dept. Commer., NOAA Tech. Memo. NMFS-NWFSC-43, 46 p.

ICES 2002. Report of the Northern Pelagic and Blue Whiting Fisheries Working Group. ICES CM 2002/ACFM:19

ICES. 2003a. Report of the Northern Pelagic and Blue Whiting Fisheries Working Group. ICES C.M. 2003/ACFM:23

ICES 2003b. Report of the Working Group on Methods of Fish Stock Assessment. ICES C.M. 2003/D:03.

Kizner, Z.I., and D.A. Vasilyev. 1997. Instantaneous separable VPA (ISVPA). ICES Journal of Marine Science 54: 399-411.

McAllister, M. and C. Kirschner. 2002. Accounting for structural uncertainty to facilitate precautionary fishery management: illustration with Namibian orange roughy. Bulletin of Marine Science 70(2): 499-540.

Mesnil, B. (2004) A crash test of Catch-Survey Analysis (CSA) using the NRC simulated data. Working Document for the ICES Working Group on Methods of Fish Stock Assessments, Lisbon, February 2004

NRC. 1998. Improving fish stock assessments. National Research Council. National Academy Press. Washington, DC. 177 pp.

Pope, J.G., and J.G. Shepherd. 1982. A simple method for the consistent interpretation of catch-at-age data. Journal du Conseil International pour I'Exploration de la Mer 40: 176-184.

Punt, A.E and R. Hilborn 1997 - Fisheries stock assessment and decision analysis: the Bayesian approach. Rev. Fish. Biol. Fish. 7: 35-63.

Richards, L.J. and J.J. Maguire 1998 - Recent international agreements and the precautionary approach: new directions for fisheries management science. Can. J. Fish. Aquat. Sci. 55: 1545-1552.

Shepherd, J.G. 1999. Extended survivors analysis: an improved method for the analysis of catch-at-age data and abundance indices. ICES Journal of Marine Science, 56:584-591.

Shepherd, J.G., and M. Nicholson. 1991. Multiplicative modelling of catch-at-age data, and its application to catch forecasts. J. CONS. CIEM., vol. 47, no. 3, pp. 284-294.

Vasilyev, D.A. 2001. Cohort models and analysis of commercial bioresources at information supply deficit. VNIRO Publishing: Moscow.

Vasilyev D. 2003. Is it possible to diminish the impact of unaccounted time trends in age-structured surveys' catchability on the results of stock assessment by means of separable cohort models? ICES C.M. 2003/X:03. 13 pp.

Working Documents

Björnsson, H. WD. Input Data.

Björnsson, H. WD. Some basic things.

Eltink, G. WD. Preliminary evaluation of assessments with ISVPA and ICA in the case of shifts in the exploitation pattern.

Gudmunsdottir, A. WD. Some tables and plots of data used in assessment for NSSH

Tjelmeland, S. WD. What about Separability?

Tjelmeland, S. WD. Elements of the assessment of Norwegian spring spawning herring using SeaStar.

Software Documentation

Skagen, D. 2003. AMCI Version 2.3. Assessment model combining information from various sources. Model description: Instructions for installation and running File formats.

Tjelmeland, S. 2003. SeaStar: Assessment model for Norwegian spring spawning herring. Web site: www.assessment.imr.no

Vasilyev, D. 2003. Documentation contained in the latest version of the ISVPA software

8 APPENDIX

An Exploratory Catch-at-Age Model in AD Model Builder

An exploratory assessment of the blue whiting stock was done using a home-made catch at age model written in AD-model builder. The goal was not to make the assessment but rather to see how much model results could be affected by changing some of the premises of the model. In addition to assessment a short-term prognosis was done using a TAC constraint.

The first step in the assessment was to look at the catch and survey data. A Shepherd-Nicholson model was applied to the catch data giving a CV of 0.56 if ages 0 to 9 were included but 0.37 if age 0 was excluded. The residuals from the Shepherd Nicholson shown in the table below were in some runs used as CV on the age groups though multiplied by a common number that was estimated.

Age	0	1	2	3	4	5	6	7	8	9	10
Residuals	1.17	0.15	0.15	0.11	0.11	0.11	0.13	0.13	0.08	0.24	0.38

Similar analyses were done on the two most important surveys the Norwegian Acoustic survey and the survey in the Norwegian Sea. Both surveys are considered to have changed in 1990 so most of the analysis only used survey data from 1991 and later.

A Shepherd-Nicholson model for the Norwegian Sea survey gives a CV of 0.79 and if the year term is dropped the CV is 1.2. Use of the survey in tuning seems questionable except possibly for the youngest age groups where it might be of some value.

A Shepherd-Nicholson model for the Norwegian Acoustic survey (ages 2 –8) gives a CV of 0.49 and 0.7 if the year term is dropped. The distribution on different age groups is shown in the table below and used as candidate for weighting of different age groups as done with catch in numbers.

Age	2	3	4	5	6	7	8
Residuals	0.30	0.10	0.10	0.09	0.15	0.16	0.17

Figure 9.1 shows the biomass from the two surveys (based on catch weights as weight at age in the surveys was not available). Both the surveys show relatively high values in recent years but they show quite different development over time, possibly due to surveys covering different parts of the stock.

In the analysis described below only the Norwegian Acoustic survey (Figure 8.1) was used but as mentioned before the youngest age groups from the other survey might be used. A number of different alternative model configuration were tested. The settings in alternative 1 which might be called the base case were.

- Recruitment was lognormally distributed around a fixed mean with the CV of the distribution estimated. (P shrinkage).
- Separable model
- Catchablitity of all age groups estimated but independent of stock size.
- CV of residuals in the survey and in the catch at age was assumed to follow 2nd order polynomials with the parameters estimated. CV on age 0 in the catch was set considerably higher than the polynomial indicated.
- Autocorrelation of residuals in catch at age data and survey estimated.
- Standard deviation of catch in tonnes was assumed to be 0.05 i.e catch in tonnes was followed closely. This setup might be questioned.
- Autocorrelation of residuals from the stock-recruitment function (constant) estimated.

The alternatives differed from the base in the following way.

- Alternative 2. No autocorrelation in residuals from the stock recruitment relationship.
- Alternative 3. CV of the residuals in survey and catch data described by the residuals from the Shephard – Nicholson models multiplied by an estimated number, 1 for the survey and 1 for the catches.
- Alternative 4. Same as alternative 3 without autocorrelation in the recruitment residuals.
- Alternative 5. CV of age 0 in the catches follows the same 2nd order polynomial as the other age groups. (Puts more weight on age 0).

- Alternative 6. The weight of the stock-recruitment term in the likelihood function reduced by a factor of 10.
- Alternative 7. Weight of the catch at age data in the likelihood function reduced by a factor of 20.
- Alternative 8. The selection pattern was allowed to change quite freely approaching VPA model.

Results of the runs are shown in Figure 8.2 indicating an estimated $F_{3\text{-}7}$ in the range 0.44-0.6 in 2002 and spawning stock between 3 and 5 million tonnes in the beginning of 2003. The run where little weight was put on the catch at age data (7) gives the most optimistic view of the stock but the run where the selection is allowed to vary freely (8) shows similar results. The last figure in Figure 8.3 also demonstrates that these results (7 and 8) follow the survey biomass most closely. This is to be expected as in alternative 7 relatively more weight is put on the survey data and in alternative way the model is allowed to adjust the F values quite freely to adjust to the catch at age data. It can be said that in both of those alternatives the model can follow the survey data as well as the internal consistency of the survey allows.

Looking at Figure 8.2, the main difference between different runs is in the view of the recruiting year classes. It is difficult to say to what extent recruitment models (P-shrinkage) are appropriate, as it seems highly likely that recruitment in recent years is well above average. Including autocorrelation of residuals can take care of this problem but can questionable if the recruitment will reduce.

Figure 8.3 shows catch and survey residuals from alternatives 1 and 7 showing how alternative 1 follows the catch more closely and alternative 7 the survey.

Figure 8.4 shows the result of running the model for 2 years using a TAC constraint of 1500 kT for alternative 7 and 1000 and 1500 kT for alternative 1. The figure indicates that if alternative 7 (the surveys) is giving the "right" view of the stock an annual catch of 1500 kT is not going to cause major risk to the stock while if alternative 1 gives the right picture of the stock even an annual catch of 1 million tonnes is going to lead to much reduction in the spawning stock.

Conclusions

The alternative models shown, other than the VPA model (alternative 7), have a difficult time in following survey and catch data at the same time. All limitations that are put on the selection pattern in the catches lead to the model not following the most recent surveys.

So is it possible to say what is the right picture of the stock. It must be investigated if recent surveys are covering a larger part of the stock than earlier surveys either due to increased survey effort or change in behaviour, which could be caused by higher maturity at age. The fishery also needs to be mapped to see if there is much probability of change in selection pattern in recent years.

So how to provide advice for this stock? Today the spawning stock is reasonably strong so the advice could be based on the catch where the most likely value of the spawning stock is going to be stable, based on some intermediate run like the base run. This will probably lead to annual catches close to 800 thousand tonnes.

Looking at the age composition of the catches in recent years 33 and 22% of the catches by weight in 2001 and 2002 were age 2 and younger. It might be argued that this is not a serious problem if the catch is only a small proportion of the incoming year classes. The main problem is that nothing is known about the size of incoming year classes and how big proportion the fleet is able to remove is still unknown. Reducing the effort towards the youngest age groups is therefore a priority as is trying to get some measure of recruitment.

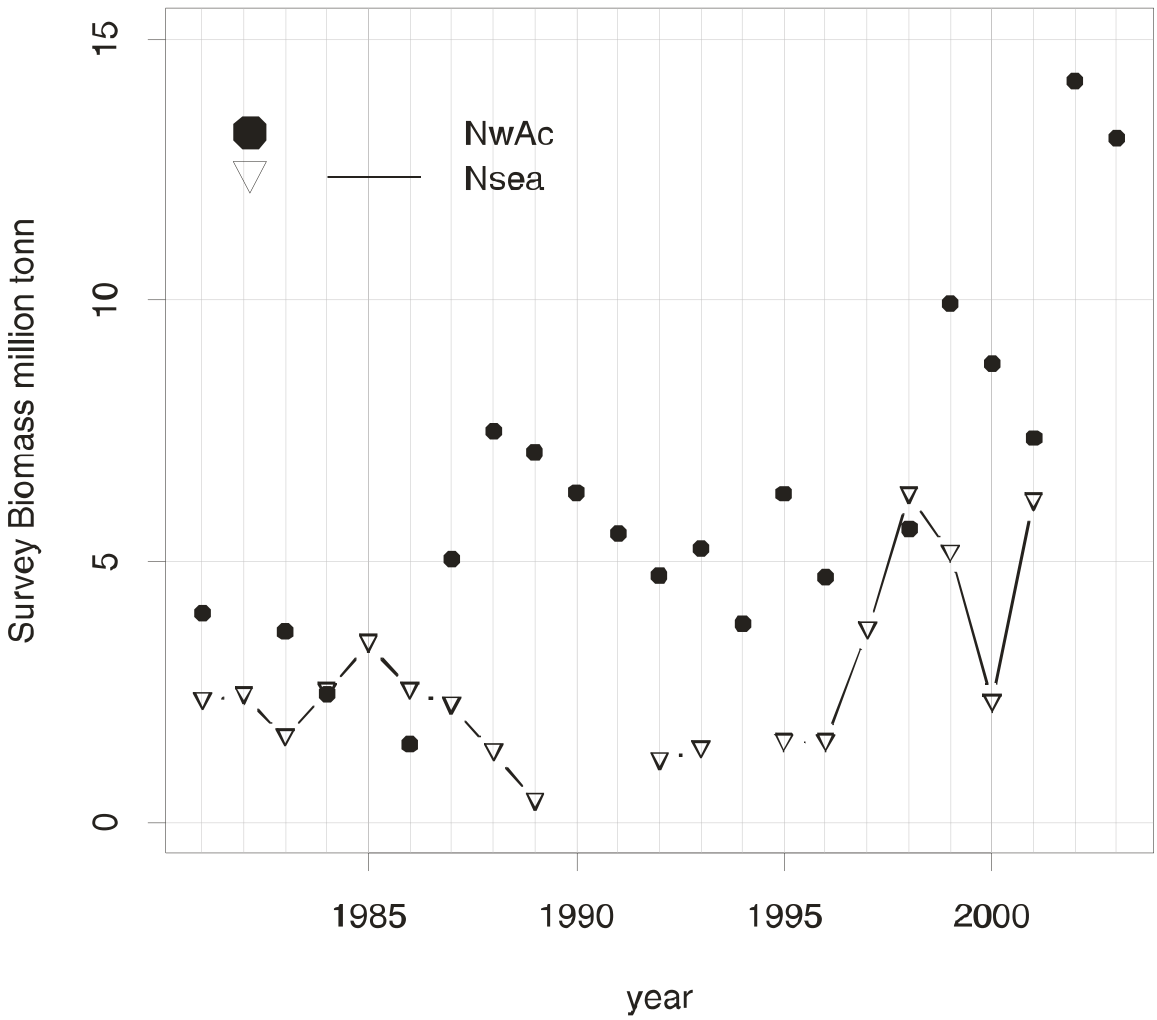


Figure 8.1. Biomass from the Norwegian acoustic survey and the survey in the Norwegian Sea.

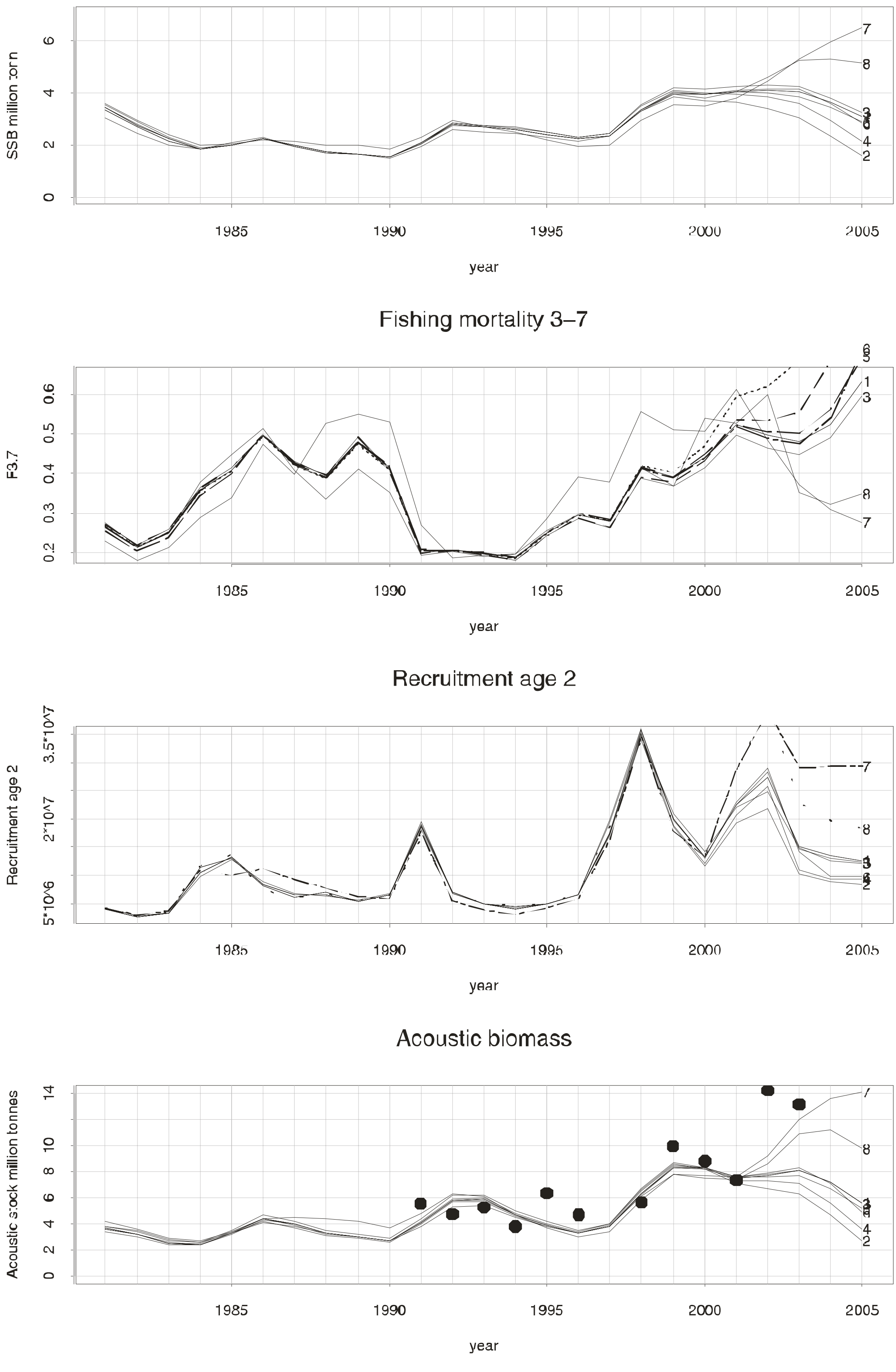


Figure 8.2. Summary of results from the model.

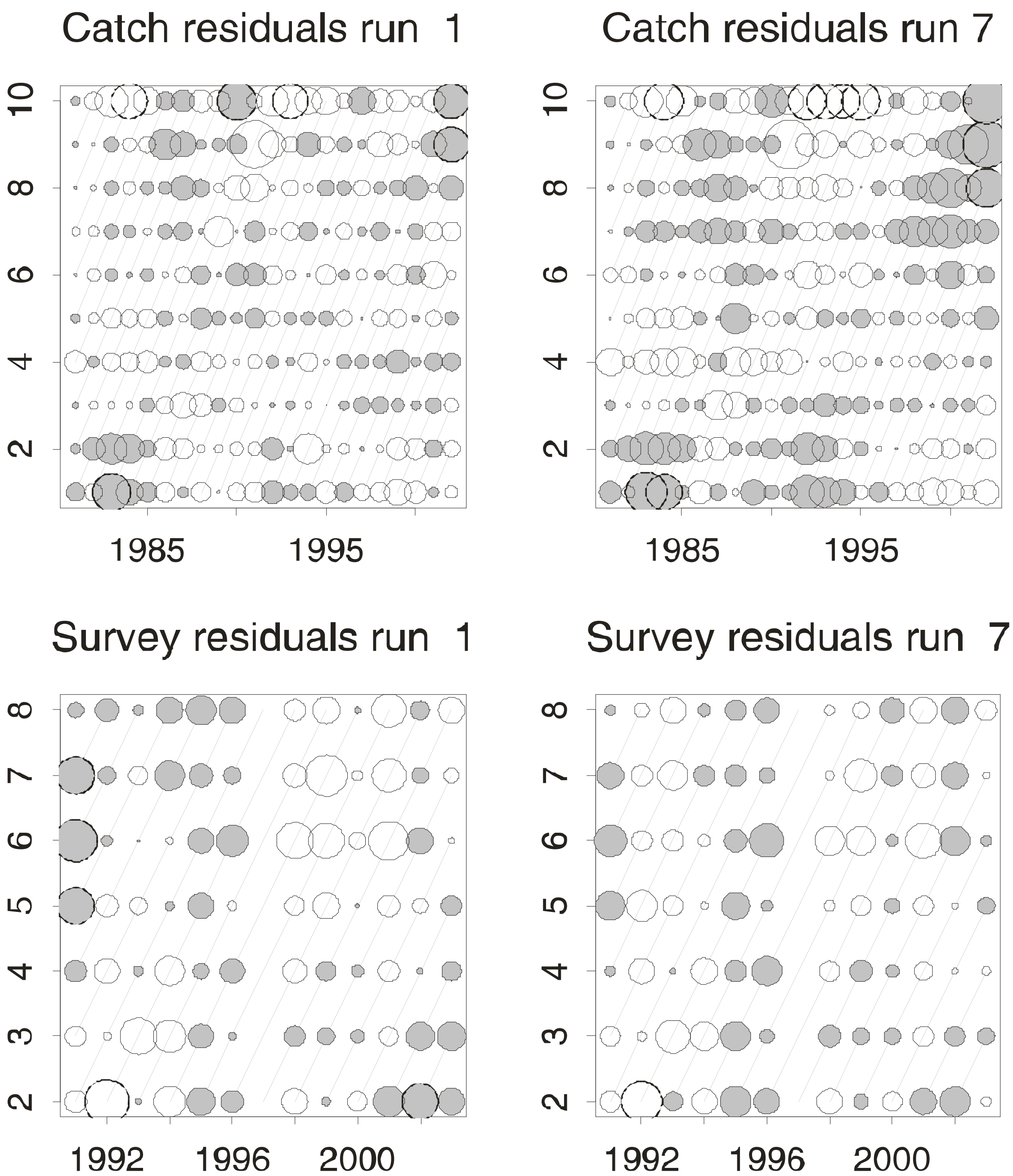


Figure 8.3. Residuals from the model. Shaded circles show positive residuals.

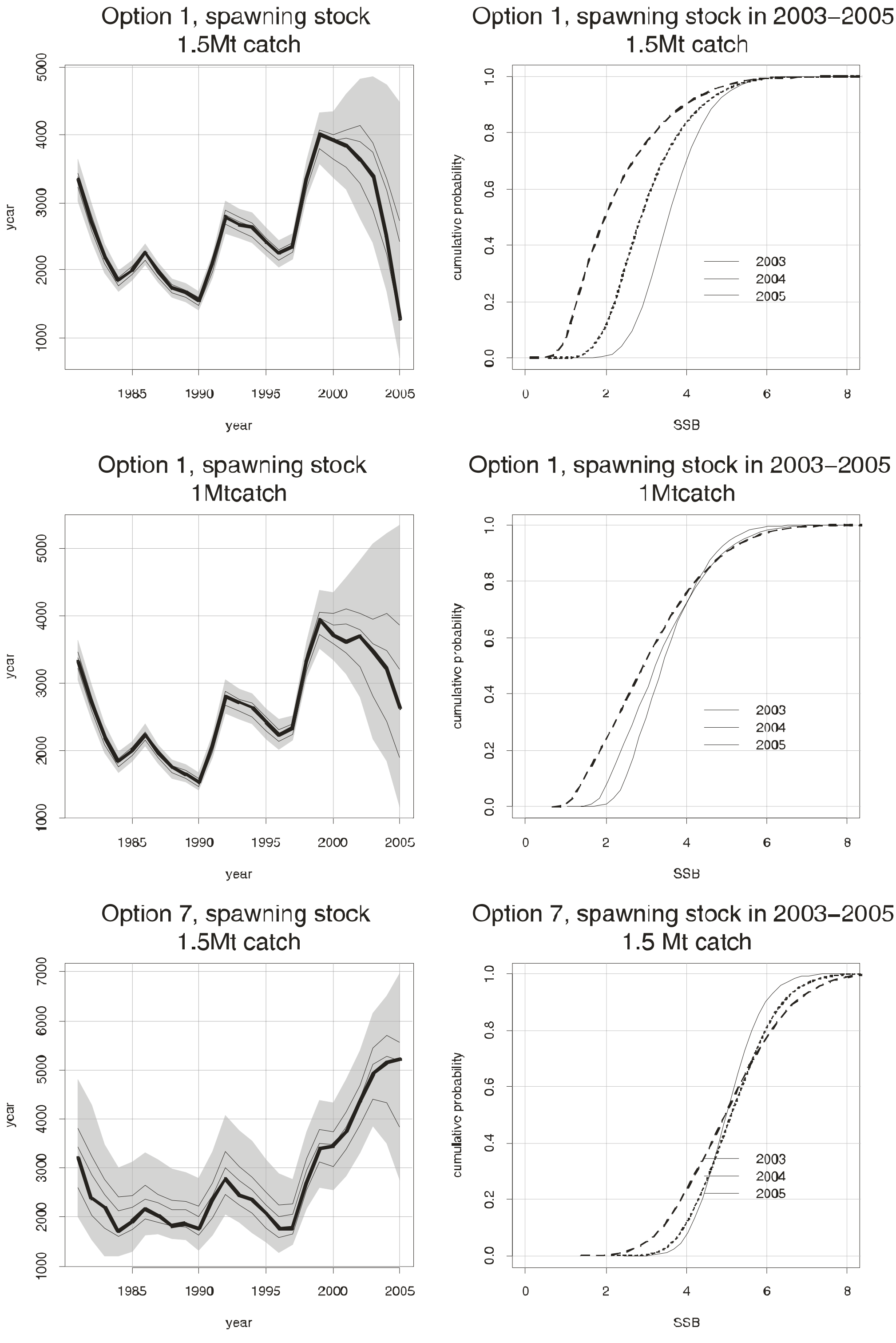


Figure 8.4. Results from running the model until 2005 using a TAC of 1 or 1.5 million tonnes after 2002. The shaded areas show 90% probability, the dashed lines 60% probability, the wide lines the median and the thin line the mean.

9 ABBREVIATIONS

ACFM	Advisory Committee for Fisheries Management
AMCI	Assessment Model Combining Information from various sources
BEI	Bergen Echo Integrator
CADAPT	Cohort Adapt
ICA	Integrated Catch Analyses
ICES	International Council for the Exploration of the Sea
IPIMAR	Instituto de Investigação das Pescas e do Mar
ISVPA	Instantaneous Separable VPA
MARECO	
NSSH	Norwegian spring spawning herring
NSS herring	Norwegian spring spawning herring
NRC	
PGSPFN	Planning Group on Surveys on Pelagic Fish in the Norwegian Sea
SeaStar	Stock Estimation with Adjustable Survey observation model and Tag Return data
SGAMHBW	Study Group on Assessment Methods Applicable to Assessment of Norwegian Spring-Spawning Herring and Blue Whiting Stocks
TAS	
XSA	Extended Survivor Analyses
WGNPBW	Northern Pelagic and Blue Whiting Working Group