RIVO-Netherlands Institute for Fisheries Research

P.O. Box 68	P.O. Box 77
NL 1970 AB Ymuiden	NL 4400 AB Yerseke
The Netherlands	The Netherlands
Phone: +31 255 564646	Phone: +31 113 672300
Fax: +31 255 564644	Fax: +31 113 573477
Internet:postmaster@rivo.wag-ur.nl	

RIVO report
Number: C023/04

Evaluation of Stock Assessment Models

Dr. M. Dickey-Collas, dr.ir. R.E. Grift, drs. L.J. Bolle, dr. S.B.M. Kraak

Commissioned by:	Ministerie van Landbouw, Natuur en Voedselkwaliteit dhr. ir. L.R.M. Lomans Postbus 20401 2500 EK DEN HAAG
Project number:	319 12430 02
Contract number:	Ond/2002-1/6c/1
Approved by:	Drs. E. Jagtman Hoofd afd. Biologie en ecologie
Signature:	________________________
Date:	29 March 2004

Number of copies: 35
Number of pages: 84

Table of Contents:

Summary

This project was designed to answer the question: "What is the contribution of the different data sources to the parameter estimates and uncertainties of the models that are the basis of the biological advice?". The report answers this question by reviewing the data used in the assessment models, investigating how the models use of these data and the sensitivities of the models to the data, and to explore the inherent model assumptions. The work looks at both assessment models and the projection of populations into the short term.

The first major section of the report gives a background to stock assessments. It explains the "catch equation" and the problems in estimating fishing mortality. Then the report focuses specifically on the input data. Our understanding of the effect of precision in the catch estimates on assessment quality is progressed by an analysis of the precision of weight and numbers landed in the pelagic fisheries of the Netherlands. It is found that the uncertainty of estimates of the mean weight of the fish in the catch is very low (<5% variability) and that the uncertainty of the estimates of numbers of fish in the Dutch catch is acceptable (<25% variability). Some minor stocks however, have very poor precision in the numbers of fish in the catch, such as North Sea horse mackerel, and this was thought to be due to the low sampling levels of a spatially very variable stock. The advances made in this study, when considered in the context of previous work, suggests that the level of precision found in the Dutch estimates are comparable with those of other nations and have a small impact on the precision of the stock assessments and the quality of biological advice. It is concluded that sampling at current levels and with current methods is adequate in terms of the quality of the assessment for most stocks. It is likely that increases in precision will not increase the quality and reduce the variability of the assessments. It is important to consider both the representative nature and the precision of sampling, hence the biology and ecology of a stock must always be considered prior to the design of a sampling strategy. One drawback of the analysis of precision is that it is based on a so-called non-parametric bootstrap analysis. This analysis is conditional on the market samples that have been taken in the past. If the sampling is not representative of the true catches, the bootstrap analysis may give a precise but biased answer.

The spatial nature of the age profiles and sampling is compared to the commercial landings. Whilst the age structure of the samples is found to be different by area (to be expected considering the variability in the population), the sampling levels are found to be proportionately representative of the catch. This finding indicates that the analysis of the precision of the market-sampling programme, which was discussed above, is likely to be applicable.

The project then concentrates on the assessment models themselves and shows that there are structural and conceptual differences between them. Hence it is not surprising that the choice of model can affect the final stock assessment. However this study also shows that within each model, many of the choice options within the model (that affect the assumptions) can have large influences on the final stock assessment. The basic models used by ICES are described: XSA, ICA, TSA, SAD, AMCI and ISVPA and it is found that their inherent differences are either statistical or conceptual. It is shown that the results of different models on the same stock are fairly consistent if the model assumptions are similar. This predominantly explanatory section then concludes that the model is as good as its input data and the relevance of the model assumptions to that stock or fishery. It also shows that the weighting system within each model favours the catch data, and hence not just the precision but also the accuracy of the catch is very important to the quality and robustness of the assessment.

The sensitivity of the short-term projections to the input data is then tested. It is shown that assumptions about fishing mortality, recruitment and trends in weight can impact on the quality of the projection. Stocks with reduced age profiles are particularly susceptible to assumptions about the numbers of recruits and ages close to recruitment. Strong trends in fishing mortality and the weight of a fish at each age that are not reflected in the projection inputs will impact on the final outcome. It is shown that the catch forecasts of the commercially important species to the Netherlands, have a tendency to over predict, in that stock sizes are generally lower and

fishing mortalities higher when compared *post hoc* to the actual values in the historic estimates.

By successfully achieving its objectives, this project has shown that virtually all the assumptions (uncertainties) in the models can affect the biological advice. Unfortunately there is no generic solution to the problems associated with each stock. Therefore it is crucial that each assessment is robustly scrutinised and its quality assured. Previous work and section 6 shows that accurate catch statistics are crucial to the quality of the assessment. Hence as the precision of the estimates of numbers in the Dutch catch is already high, and thus has little effect on the quality of the assessment, it is crucial that the accuracy of the estimate of international catch also be high. All projections should be carried out carefully within the context of recent trends in the population and it is probably impossible to carry out valid projections in the context of short lived or highly over-fished species that have extremely variable, or non-predictable recruitment. The use of short-term projections as a valid management tool may therefore be questioned in some instances.

Samenvatting

Dit project is opgezet om de volgende vraag te beantwoorden: “Wat is de bijdrage van de verschillende gegevensbronnen aan de parameter-schattingen en onzekerheden van de rekenmodellen die ten grondslag liggen aan de biologische adviezen?”. Om deze vraag te beantwoorden, is in dit rapport een analyse gedaan van de gegevens die in de rekenmodellen gebruikt worden, een onderzoek naar hoe de rekenmodellen deze gegevens gebruiken en naar hoe gevoelig de modellen zijn voor de gegevens en de aannames die gemaakt worden. Deze studie beschouwt zowel de rekenmodellen die de historische toestand van de visbestanden schatten als de rekenmodellen die een korte termijn prognose voor de visbestanden geven.

Het eerste deel van het rapport beschrijft de achtergrond van de rekenmodellen die voor toestandsbeoordelingen van visbestanden gebruikt worden. Het begrip “vangst vergelijking” wordt uitgelegd en de problemen bij het schatten van visserijsterfte worden besproken. Het begrip over het effect van precisie op de kwaliteit van vangstschattingen is toegenomen door een analyse van de precisie van de schattingen van visgewichten en aantallen vissen in de aanlandingen van de Nederlandse pelagische visserijen. De onzekerheid in de schattingen van gemiddelde visgewichten in de Nederlandse aanlandingen is erg laag (<5% variabiliteit) en die van de schattingen van de aantallen vis acceptabel (<25% variabiliteit). Sommige kleine visbestanden, bijvoorbeeld de Noordzee horsmakreel, hebben echter een erg lage precisie voor de aantallen aangelande vis. Het lijkt er op dat dit veroorzaakt wordt door een klein aantal monsters van een ruimtelijk zeer variabel visbestand. De precisie van de Nederlandse schattingen uit deze studie is vergelijkbaar met de uitkomsten van eerdere onderzoeken. Ook blijkt, de andere studies in aanmerking genomen, dat de precisie van de Nederlandse schattingen een minimale invloed heeft op de precisie van de bestandsschattingen en de kwaliteit van het biologisch advies. We concluderen dat het niveau en de gebruikte methodes van de huidige bemonstering adequaat zijn voor wat betreft de kwaliteit van de toestandsbeoordeling van de meeste visbestanden. Waarschijnlijk zal een hogere precisie niet leiden tot toename van de kwaliteit van de toestandsbeoordelingen. Het is belangrijk om zowel de representativiteit als de precisie van de bemonstering in ogenschouw te nemen en dus moet de biologie en de ecologie van een soort worden beschreven voordat de bemonstering wordt opgezet. Een nadeel van de analyse van de precisie is dat deze is uitgevoerd met behulp van een zogeheten non-parametrische bootstrap analyse. Voor marktmonsters die in het verleden verzameld zijn, kan de analyse een verkeerd beeld geven. Wanneer de bemonstering niet representatief is geweest voor de ware vangsten, kan de bootstrap analyse een precies antwoord geven dat een systematische fout bevat.

De ruimtelijke verschillen in leeftijdsopbouw van de marktmonsters zijn vergeleken met de aanlandingsgegevens. Hoewel de leeftijdsopbouw van de marktmonsters per gebied blijkt te verschillen (wat ook te verwachten is gegeven de variatie in de populatie) blijkt dat de bemonstering representatief is voor de aanlandingen. Dit betekent dat de hierboven beschreven analyse van de precisie van de marktbemonstering, aannemelijk is.

Vervolgens wordt ingegaan op de rekenmodellen voor toestandsbeoordelingen en wordt aangetoond dat er structurele en conceptuele verschillen tussen de modellen zijn. Het is niet verrassend dat de keuze van een model de uitkomst van de toestandsbeoordeling kan beïnvloeden. De studie laat echter ook zien dat binnen elk model verschillende opties (die de aannames beïnvloeden) waar uit gekozen kan worden, een grote invloed kunnen hebben op de uiteindelijke toestandsbeoordeling. De binnen ICES gebruikte rekenmodellen worden beschreven: XSA, ICA, TSA, SAD, AMCI en ISVPA en er wordt aangetoond dat de intrinsieke verschillen van statistische of conceptuele aard zijn. De uitkomsten van de verschillende modellen toegepast op dezelfde visbestanden blijken tamelijk consistent te zijn als binnen elk model gelijkwaardige aannames gemaakt worden. De conclusie van dit hoofdstuk is dat een rekenmodel zo goed is als de kwaliteit van de invoergegevens en de relevantie van de modelaannames voor een visbestand of visserij. In dit rapport staat tevens dat de wegingen binnen elk model voorrang geven aan de vangstgegevens en dat daarom niet alleen de

precisie, maar ook de nauwkeurigheid van de vangstschattingen erg belangrijk is voor de kwaliteit en de robuustheid van de toestandsbeoordeling.

De gevoeligheid van de korte termijn prognoses voor de invoergegevens is onderzocht. Aannames betreffende visserijsterfte, rekrutering, en trends in visgewicht kunnen invloed hebben op de kwaliteit van de prognose. Met name visbestanden met voornamelijk jonge vissen zijn gevoelig voor aannames betreffende de aantallen rekruten. Sterke trends in visserijsterfte en visgewichten per leeftijdsgroep, die niet worden meegenomen in het model, beïnvloeden de kwaliteit van de prognose negatief. Vangstvoorspellingen voor de voor Nederland commercieel belangrijke soorten blijken meestal hoger uit te vallen dan de *post hoc* berekeningen omdat de bestanden meestal kleiner blijken te zijn en de visserijsterfte hoger blijkt te zijn dan aangenomen in het model.

Dit project heeft al zijn doelstellingen behaald, in die zin dat is aangetoond dat alle aannames (de onzekerheden) in de rekenmodellen het biologisch advies beïnvloeden. Helaas bestaat er geen generieke oplossing voor de problemen die geassocieerd zijn met elk visbestand. Het is daarom van cruciaal belang dat elke toestandsbeoordeling grondig onderzocht wordt en de kwaliteit ervan verzekerd wordt. Uit eerder onderzoek en uit hoofdstuk 6 blijkt dat nauwkeurige vangststatistieken cruciaal zijn voor de kwaliteit van de toestandsbeoordeling. Aangezien de precisie van de schattingen van de aantallen vis in de Nederlandse vangsten hoog blijkt te zijn, en daarom nauwelijks negatief effect heeft op de kwaliteit van de toestandsbeoordeling, is het van cruciaal belang dat de nauwkeurigheid van de schattingen van de internationale vangsten ook hoog is. Alle prognoses moeten met zorg uitgevoerd worden en rekenschap geven van recente trends in de populatie. Het is waarschijnlijk onmogelijk om geldige prognoses te doen voor kortlevende vissoorten of sterk overbeviste visbestanden met extreem variabele of onvoorspelbare rekrutering. In sommige gevallen moeten daarom vraagtekens gezet worden bij het gebruik van korte termijn prognoses als een beheersgereedschap.

1. Background and project objectives

The project was devised to explain and investigate the influence of model choice on the stock assessment of commercially important fish stocks to the Netherlands. The specific question was:

"What is the contribution of the different data sources to the parameter estimates and uncertainties of the models that are the basis of the biological advice."

The species to be studied are cod, herring, whiting, plaice and sole in the North Sea and north east Atlantic mackerel and western horse mackerel. The project was constructed to complement and not repeat, the F project (which is due to report next year). Hence this project will pay closer attention to the species other than plaice and sole. However to ensure a broader view is maintained plaice and sole will be included wherever it is felt relevant and useful.

The object of this project is to review the stock assessment and forecasting procedures and in particular:

- *The input data*- the precision of the catch data, the impact of precision, the survey data and spatial representation of the sampling relative to the catch (chapter 3).
- *The assessment models*- the models used, assumptions of the models, the relevance of choice of model and weaknesses in the assumptions (chapter 4).
- *The stock projections*- the suitability and sensitivity of stock projections to the assumptions about catch, fish weight and fishing effort in the final and projected years (chapter 5).

2. Introduction

Fish stock assessment has developed greatly since the initial work on virtual population analysis (VPA). Although, many fish stocks around the world are now managed without VPA, using simple biomass indicators and harvest control rules, within the ICES community and the northern EU, more complicated stock assessment models (VPA type) are still commonly used and are becoming even more complicated. Models are only as good as the relevance of their internal assumptions and the data that goes into them. VPA type models are all based on the catch equation and the differences between them are represent different ways of estimating some of the parameters from the catch at age data.

The simplified catch equation:

$$C = \frac{F}{Z}\left(1 - e^{-Z}\right)N_0$$

[1]

which is the basis of a VPA model.

where C is the catch, F is the fishing mortality, Z is the total mortality and N_0 is the number at time 0. This equation was developed and modified by Gulland (1965) and then Pope (1972), along with an equation for survival of a year class (or cohort).

Survival of a cohort

$$N_{i+1} = N_i e^{-(F_i + M)}$$

[2]

Where N_i is the number in a year i, M is natural mortality, and F+M equals Z (the total mortality).

The VPA type models work backwards iterating from the oldest ages back to the youngest ages to find F. The process assumed that the catch at age and natural mortality were known, while the values of F and N for the oldest ages for each cohort needed to be assumed to initiate the analysis. This meant that the results for the most recent years of the analysis were more uncertain than for earlier years (Table 2.1).

If the stock was heavily exploited, the results obtained for the younger ages in the cohort became independent of the artificial input values of the oldest age (called convergence). If the stock was lightly fished then convergence took longer.

Table 2.1. Hypothetical age-year matrix. The letters stand for cohorts and the shaded letters show the oldest ages for each cohort within the matrix. In the last year the oldest available ages for cohorts N to U, are less that the oldest age in the matrix, *e.g.* the oldest age available for cohort T is 2.

Age/Year	'90	'91	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01	'02
1	I	J	K	L	M	N	O	P	Q	R	S	T	U
2	H	I	J	K	L	M	N	O	P	Q	R	S	T
3	G	H	I	J	K	L	M	N	O	P	Q	R	S
4	F	G	H	I	J	K	L	M	N	O	P	Q	R
5	E	F	G	H	I	J	K	L	M	N	O	P	Q
6	D	E	F	G	H	I	J	K	L	M	N	O	P
7	C	D	E	F	G	H	I	J	K	L	M	N	O
8	B	C	D	E	F	G	H	I	J	K	L	M	N
9+	A	B	C	D	E	F	G	H	I	J	K	L	M

Pope and Shepherd (1985) developed the cohort approximation to improve the solution of the catch equation, reducing the iterations required (using tuning data to calibrate the trends in the catch at age matrix). Most other methods, since developed, are techniques for either reducing the uncertainty in recent years (or the oldest age group), or reducing the parameters that need to be estimated.

The model chosen must be appropriate for the problem to be solved. For example, a stock assessment that supports management decisions may require a different model than an exploration of the population dynamics of a species.

As the ICES working group on methods (WGMG) stated: "*Each method will have its own limitations, and if methods are used on stocks where underlying assumptions are violated, the results will potentially be misleading, even though a method may have been approved for general use. It is naïve to expect that any method will be universally applicable and this implies that stock assessment Working Groups will need a wide range of methods and software at their disposal.*" ICES (2003a)

As a result there are many stock assessment models available and used in the advice process of ICES and the EU. Some have been tested and validated by ICES and some have not. Some existing models (*e.g.* XSA) are being redeveloped and there are plenty of new developments on the way (AMCI, Surba, SeaStar etc., ICES 2003a).

Within the context of the Netherlands, this project tests the relevance of the model assumptions within each fish stock assessment and the quality of the input data to the model. It will describe where the uncertainties lie and problems with the current assessments.

3. Input data

The data requirements for a stock assessment using a VPA type model are:

- the numbers of fish in the catch at each age
- the values of natural mortality experienced by fish in the population at each age

To work out the spawning stock biomass (SSB) from the VPA results you also need:

- the weight of an average fish in the population at each age
- the proportion of the population sexually mature at each age
- the proportion of the annual fishing and natural mortality experienced by an average fish before spawning

In addition, to check the calculations of fish numbers and biomass the following are used:

- the weight of the total catch
- the average weight of a fish in the catch at each age

The resolution of the data, in terms of time, varies with model, but most of the assessments carried out by ICES use models that work on an annual resolution. In addition to the catch data described above "tuning data" are used to increase the accuracy and/or precision in the estimation of terminal fishing mortality. This means that the results of the un-converged part of the stock assessment is compared to external indicators of stock trends, such as the catching potential of fish (CPUE) or fisheries independent surveys of the fish population. The tuning process is a statistical minimisation of the differences between the VPA results and the external indicators. Within ICES tuning data come in three main forms:

- CPUE series, where the catch rate of fish at each age is compared to the estimated numbers at those ages in the population (Pope and Shepherd, 1985)
- Age disaggregated survey data, where the catch rates of certain age groups in surveys are compared to the estimated numbers of these age groups in the population
- Biomass survey data, where a biomass or spawning biomass of the population is compared to the estimated biomass. These types of data also include proxies for biomass, such as larval surveys or other population characteristics. Biomass data are used within the objective function of the model (*i.e.* the root assumption and solution of the model, Patterson, & Melvin, 1996).

The source of these data varies. Most catch data are determined from "market sampling" and involve the raising of samples to the entire catch. This is complicated by the international nature of market sampling. The way the catch in numbers at age, the weight of the total catch and the mean weight of a fish at each age in the catch are determined varies in each national laboratory. Other data are assumed and based on historic tradition (*e.g.* natural mortality-M and proportion of mortality before spawning). There is no consistent method for estimating the weight of a fish in the population between stocks; it can be estimated from market samples or from survey results (some other ICES stocks use growth models), likewise the method for determining the proportion mature varies between stocks. The compilation of CPUE series also varies between institutes but the growth in internationally coordinated surveys has increased the quality control of survey data and standardised the compilation of results.

Within the Netherlands, the catch is sampled differently dependent on species. A summary of the methods used is given in table 3.1 and described in greater detail in de Vries *et al.* (2003).

Internationally the demersal stocks are sampled at a much higher rate per tonne landed than the pelagic stocks (Tables 3.2a and b). The actual number of fish measured per stock is very similar but relative to the catch, ten times more cod, sole and whiting are aged than herring or mackerel. This is further complicated by the demersal stocks being in one discrete area (North Sea) while the mackerel and horse mackerel stocks cover a much wider area.

3.1 Precision of catch estimates

This issue of sampling the catch raises questions about the precision of the catch estimates, and how the assessments are affected by variations in the precision. Work on the precision of catch estimates of North Sea herring, cod and plaice was carried out by the EU EMAS project (O'Brien *et al.*, 2001a; Pastoors *et al.*, 2001). These studies were instigated following worries about the error introduced into the assessment by poor precision in the catch at age estimation and by the contrasting methods used by different institutes to raise their sample results to total landings (Simmonds *et al.*, 2001). Further work on sole was carried out in the F project and analysis of the precision of mackerel and horse mackerel was carried out as part of this project (Dickey-Collas & Eltink, 2003). Further developments within the EU, under the fisheries data directive (Commission Regulation, EC 1639/2001) also require that sampling is increasingly scrutinized and the precision of estimation is beginning to be investigated.

It is important to realise that precision is not the same as accuracy. The most precise estimate is the value from one single sampled individual. Thus if a species varies greatly in terms of age and size over regions and yet only a few individuals are sampled of this species from one specific region, the estimate would have a high precision but would not reflect the true picture of the whole catch, *i.e.* it would be inaccurate. So attempts to increase precision should not be made at the loss of accuracy. The assessment of precision in the catch estimates detailed in this report does not reflect the imprecision due to ageing errors, inaccuracy due to misreporting of the catch (*i.e.* it assumes an accurate catch census) or testing whether the sampling is truly representative of the catch, but reflects the variability within the sampled catch and raising procedures.

As mentioned above, each fishery institute estimates and raises the catch in numbers in a different way, so O'Brien *et al.* (2001a) suggested that bootstrap techniques (*i.e.* techniques that simulate re-sampling) be used to generate several virtual realisations of national-catch age compositions and weights at age. By the very nature of this method, the combined international catch will have greater precision than separate national estimates, as highlighted by Pastoors *et al.* (2001). Unfortunately only RIVO has carried out this analysis on sole, mackerel and horse mackerel so the precision of the full international catch estimates cannot be determined. However the precision of the Dutch estimates for sole, mackerel and horse mackerel can be compared with those of cod, herring and plaice.

3.1.1 Method

This study uses bootstrapping. The method is used because it is not possible to resample the population by repeated market sampling within a year. Bootstrapping is a method where the observed data are re-sampled to provide information on the variability of the estimates of the population. The bootstrap creates a new sample with the same number of observations as the original sample by sampling (with replacement) from the original data. The re-sampling was done at the level of the market samples, not at the level of the individual fish within the market samples. In other words, with each iteration, the bootstrap created a new set of market samples, with the same number of market samples as were in the original set, by sampling randomly with replacement from the original set of market samples. The individual fish contained within the samples stayed the same. The new set of market samples was then processed and raised in the same manner as the original market sampling data.

The method used followed that of O'Brien *et al.* (2001a), Pastoors *et al.* (2001) and Simmonds *et al.* (2001). National data were used to provide 100 replicates of market sampling sets of length, weight and catch numbers at age. The market samples were stratified by quarter for horse mackerel and by month for mackerel and the area of stock distribution fished. The process of manually allocating unsampled catches to sampled strata was not taken into account, as this tends to operate at international rather than national level. The algorithm used for the 100 bootstraps was as follows:

Set up-

- read in original length, age and weight data,
- create a list of unique identifiers for sampling units – using sample number
- calculate values that will not change with each bootstrap- commercial weight in tonnes and numbers of samples.

Bootstrap loop-

- set seed for random number generator
- form bootstrap set of length, weight and age samples by re-sampling samples (note that pelagic fish are sampled representatively at RIVO, thus there is no need for age-length-keys).
- calculate length and age distributions
- calculate numbers at age, mean length at age and also mean weight at age from the length distributions and a length weight relationship
- append the estimates from the iteration to the output file.

Summary statistics-

- Once 100 replicate sets of catch numbers at age and weight at age were produced, mean values and coefficients of variation were calculated per age, and raised to larger units (*e.g.* per year) by weighting by the numbers at age.

3.1.2 Results

Horse mackerel

It is clear that there are large interannual differences in the precision of the estimation of numbers at age in horse mackerel in the Dutch catch (Figure 3.1). In Western horse mackerel, numbers in quarters 1 and 3 are the most poorly estimated (Figure 3.1) and numbers at ages 1 and 2 also have poor precision (CV > 30%, Figure 3.2). The numbers of mature fish are estimated well (CV <30%) until ages greater than 11 (Figure 3.2). However this contrasts greatly with the precision of numbers at age for North Sea horse mackerel (Figures 3.1 and 3.3). It was rare in quarters 3 and 4 for the CV to be below 30%, and overall, only in 2001, did the precision of the numbers of mature fish fall below CV= 30%. This imprecision leads to problems with the calculated catch based on sum of products (Figure 3.3); these problems are absent in Western horse mackerel (Figure 3.2).

This imprecision in the estimates of numbers at age is not reflected in the weights at age (Figure 3.2 and 3.3). For Western horse mackerel the CV of mean weight at age is rarely greater than 5% and for North Sea horse mackerel it is less than 10%. So for horse mackerel in the Dutch catch, the greatest imprecision comes from estimates of numbers at age, which is very high for the North Sea stock.

Mackerel

For many years it has been traditional for the Dutch mackerel estimates to be calculated by month and then raised to quarter at a national level (and then raised to year internationally). This was carried out to account for changes in age-size distribution due to migration patterns that occur on a temporal scale less than quarters (Eltink, 1987). Whilst this method reflects the population dynamics well, a comparison of the mean CV of numbers at age in mackerel illustrates that no extra precision is achieved by this method compared to quarterly compiling of data (Figure 3.4) and the estimation of the number of older fish is much worse (when data are raised from monthly estimates than when data are raised from quarterly estimates). To maintain consistency with other studies (Simmonds *et al.*, 2001; Pastoors *et al.*, 2001) and with the analysis of horse mackerel in this study, further analysis will be concerned with quarterly raising only.

There are large interannual differences in the precision of the estimates of numbers at age (Figure 3.5). During the main targeted fishing season (Q1 and some of Q2) in the adult area the precision is high (low CV) for ages 4 and older. The precision of ages 4 and 5 is below CV=30% in all quarters for most years (Figure 3.5) and when the quarterly estimates are combined to generate the annual estimate, most of the mature mackerel in the catch (ages 3 to 9) are estimated with a CV <30% (Figure 3.6). However the estimates of older fish are more erratic, particularly in quarters 3 and 4 (Figure 3.5 & 3.6) when the mackerel are mostly caught as by-catch. Unlike horse mackerel, the precision of annual estimates of numbers of younger fish in the catch is not so poor (Figure 3.6).

The CV of the mean weights at age for mackerel is lower than 5% (Figure 3.6). It should also be noted that the greater imprecision in the catch in numbers from the monthly raising (rather than quarterly) also introduces greater problems in reconciling actual catch with catch estimated by sum of products (Figure 3.6).

3.1.3 Discussion

The precision of the estimates of mean weight was high (CV was low). The raising method of using annual and area length-weight relationships has probably resulted in this high precision. This high precision may not reflect the variability in the field and only reflect the estimation method used here (which is really a smoothing or averaging technique). However with regards to this study the estimation of mean weight at age in mackerel and horse mackerel from the Dutch catch is precise.

It is clear that large variability exists in the precision of numbers at age between years and ages. The estimation of numbers in the catch for younger ages of horse mackerel is poor, whilst this is not really the case for mackerel. This suggests that for mackerel the juveniles seem to be more equally distributed and represented in the samples than for horse mackerel. Combining quarterly data to arrive at annual estimates of precision results in an improvement in the precision (as expected). The CVs of numbers at age of mackerel and horse mackerel behave in a similar way, being relatively higher for the juveniles and the older fish. This pattern is seen in the national and international sampling of North Sea plaice, herring, cod and Irish Sea cod (Maxwell *et al.*, 2001; O'Brien *et al.*, 2001a; b; Simmonds *et al.*, 2001). In both horse mackerel and mackerel the precision of numbers of fish older than 10 is poor. This raises questions about the appropriate plus group to be used in the assessment.

It appears that the present sampling regime for North Sea horse mackerel by the Netherlands is not effective. This should be reviewed. The poor precision on the younger Western horse mackerel should also be considered. However this study only refers to catches by the Netherlands, and before any action is taken based on this study, other countries with major proportions of the catch of horse mackerel and mackerel should investigate the variance in their estimations of numbers at age. Only then can the precision of the international catch be estimated. It has been shown that different institutes have remarkably different levels and patterns over age in the precision of their estimates (O'Brien *et al.*, 2001b; Maxwell *et al.*, 2001).

Currently, in most ICES assessments, the input data is combined from strata (both temporal and spatial) and in most cases these strata are arbitrary and based on management units and not on the variability in fish populations or fisheries. Although this investigation suggests that the choice of monthly raising of mackerel samples to annual results, may be less precise than quarterly, it does not answer on what level the sampling should be stratified. It also does not address the structural variability in the mean length within a population. A good assessment model should not just seek precision but also reflect the true variability within the population. As Eltink (1987) pointed out, in NE Atlantic mackerel there are great differences within quarters in the size of mackerel, and he concluded that month should be used as a temporal stratum.

3.1.4 Comparison with other national catches

Judging by the findings of Pastoors *et al.* (2001) it appears that the precision level reached by national sampling by the Netherlands of the main adult catch of NE Atlantic mackerel and Western horse mackerel is similar to the levels reached by some of the sampling procedures by other institutes. Scotland has less precision in its estimation of herring numbers (Figure 3.7) and the precision of Denmark's estimates for herring numbers is similar to the precision of the Dutch estimates. However numbers in the catch of younger cod and plaice are more precise than Dutch mackerel and horse mackerel (Figures 3.8, 3.9,3.10). The estimation of the older ages groups of mackerel and horse mackerel are much more precise than any estimates of cod in either the North Sea or Irish Sea (Figures 3.8 and 3.9). Only the estimates for plaice appear to be generally more precise across all age groups (Figure 3.10), although again Denmark has low precision in its older fish. As far as RIVO is concerned, its sampling of plaice, herring and younger cod appears to be robust to this precision analysis.

Comparing the Dutch sampling across six species (Figure 3.11) suggests that the precision on ages 2 to 5 of mackerel and horse mackerel is worse than for cod, plaice, sole and herring, but average for the older fish. The large geographic region of the mackerel and horse mackerel stocks contrasts with that of North Sea cod, plaice, sole and herring. The demersal fisheries target a very specific area of the southern North Sea while the pelagic fisheries are spread across the NE Atlantic continental shelf. The precision of the herring might not be so high if catches in ICES areas VI and VII were included in the analysis. As mentioned above, the seasonality of the mackerel fishery and the targeting of specific life stages in the horse mackerel fishery probably make the estimation of numbers in the catch less precise. It is a pity that the current analysis was only carried out on the Dutch catch, so the precision of the total international catch of mackerel and horse mackerel cannot be assessed or the precision of national sampling programmes compared.

3.1.5 Impact on the assessment

This weakness of not having the precision estimate for the total international catch means that the impact on the assessment of the poorer precision in Dutch sampling of the younger age groups of mackerel and horse mackerel cannot be assessed. As shown by the study on herring (Simmonds *et al.*, 2001, Figure 3.7) different institutes can have markedly different precision of the catch in numbers. The stock assessments only use total catch, representing sampling from different fleets, so other laboratories must determine their precision of sampling for these two species. Only then can Monte Carlo techniques, as used in the EMAS project, be used to analyse the impact of sampling precision on the mackerel and horse mackerel stock assessments.

The nature of the analysis is such that the combining of national catches will change the overall precision of the estimate of total catch in numbers (see Figures 3.7, 3.9, 3.10). However we can use a previous worked example, from the EMAS project as a way of commenting on the likely impact of this rate of precision on the stock assessment. The scatter of precision estimates in the mackerel and horse mackerel were not that dissimilar to that found in the four fleets sampled for North Sea herring (Simmonds *et al.*, 2001, Figure 3.7). The North Sea herring are assessed with ICA, the model as used for mackerel. ICA is a model that uses the assumption of separability (this will be explained below) and it already assumes that the catch in number matrix is not absolute and contains errors. Hence it is probable that an imprecision of 20% will not have a great effect on the assessment estimates of F, SSB and recruitment and this was found to be the case. With over 1000 bootstrapped runs, with precision estimates reflecting 75% of the catch (*i.e.* bootstrapping was carried out by countries that account for 75% of the catch) applied to the total catch of herring, Simmonds *et al.* (2001) found that market sampling only contributed a "negligible" proportion of the historic uncertainty of the assessment (between $^1/_8$ to $^1/_{30}$ of the uncertainty came from the catch at age imprecision). SSB and recruitment were less affected than the estimation of F.

Considering the similarities in sampling method, assessment technique and fisheries behaviour, it is probable that like herring, the imprecision due to market sampling will have virtually no influence on the uncertainty of the stock assessment of NE Atlantic mackerel and western horse mackerel.

As there has been no accepted assessment of North Sea horse mackerel the poor precision of the sampling data has no effect on the management of the stock. However sampling must be maintained to provide catch data for future assessments.

To conclude this section, with regard to precision, the current sampling levels and strategies appear to provide adequate coverage. Now that we have a method for assessing precision of estimates from the catch, more work is required into how to stratify the sampling and give thresholds of acceptable variance/CV in the mean age at length. Only then can optimal sampling criteria (Sukhatme & Sukhatme, 1970) be established and appropriate sampling strata be determined.

3.2 Survey data

Although not within the specific aims of this project, the investigators felt that it was important to mention the recent studies on the quality of survey data within the stock assessment context. These studies were carried out as part of the EU EVARES project aimed at North Sea cod, plaice and herring (EVARES, 2003). The project looked at the precision within the surveys and how the variable precision affected the assessment results, in a similar way to the investigations of precision of the catch described above. As with the catch analsysis the CVs were found to vary with species and age. Kell *et al.* (2003) found that the CV on the estimation of numbers of plaice in the surveys was lower than that for the numbers of especially older cod (Figures 3.12 and 3.13). Both the Beam Trawl Survey (BTS) and Sole Net Survey (SNS) had CVs on the estimated numbers less than 30% for plaice, which was similar to the CVs on the estimated numbers of the younger cod, but the pattern with age was different for plaice between the two surveys (Figure 3.12). All of the three surveys for North Sea cod had similar patterns with age (Figure 3.13).

Once these estimates of precision in surveys were combined with the precision estimates of market sampling within the assessments, they appeared to have similar effects on historic estimates of fishing mortality variability and bias in both plaice and cod (Kell *et al.*, 2003). The survey precision affected some outputs of the assessment, whilst the precision in the catch at age data (commercial data) affected other outputs of the assessment. Fishing mortality was mainly affected by commercial sampling error but survey sampling error caused bias in fishing mortality estimates. For SSB, both commercial and survey sampling error had a similar effect on assessment estimates but surveys sampling caused bias. Future estimates (the projections) were more affected by sampling error in the surveys than by the commercial sampling error.

It appeared that survey errors were more likely to cause bias than market sampling errors. The surveys positively biased the estimates of fishing mortality of plaice and negatively biased those of cod. This was thought to be either due to the way in which the surveys are used to calibrate the VPA procedures and configurations or due to the catch and population model used in the VPA, where the surveys may be attempting to correct these biased impressions of the stock given by the market sampling data. Kell *et al.* (2003) also noted that changing tuning fleets may lead to different biases and to thus different advice about the stock.

Work on North Sea herring showed that the acoustic survey was the most consistent of the age-disaggregated surveys (Simmonds, 2003). Comparisons with the IBTS and the MIK net surveys shows that the acoustic survey was more precise than the IBTS for all ages but less precise than the MIK survey (age 0). The precision of the larval survey was similar to that of the

MIK net and acoustic survey (ages 2-7, Simmonds, 2003). To date the precision estimates of market sampling in North Sea herring have not been combined with those of the surveys, to determine the sources of bias and imprecision, due to sampling error, in the stock assessment.

3.3 Spatial Representation of sampling

This section will discuss the spatial representation of the sampling relative to the catch by analysing the spatial variation in age length keys. Age length keys describe the relationship between the size of fish and their age. An example of an age length key is presented in Table 3.3.1.

Age length keys are derived from market sampling programmes and research surveys. These programmes aim to describe the age distributions in the commercial fisheries and in the surveys. With age-length keys, length-frequency distributions of catches can be transformed to catches per age group. The catch per age group of the fishing fleet is important input to the stock assessment model.

Age length keys are not used by RIVO for the compilation of herring, mackerel and horse mackerel catch at age data. In these three cases representative sampling is used. However the age profiles of the samples was still compared to investigate the spatial variability in sampling.

A significant spatial variance of age length keys would imply that the sampling programme should cope with this variance. Consequently, the programme should representatively collect fish from different areas. The programme is then only representative if it represents the spatial variation in catches of the commercial fleet.

Spatial variation in age length keys may be due to spatial variation in growth rates or to different dispersion of fish with different growth rates. If, for example, fish in the southern North Sea grow faster than in the northern North Sea, they reach a certain length at an earlier age. Consequently, the relationship between length and age, and thus the age length key, is different.

In this section, the spatial variation in age length keys (or age profiles for pelagic species) for seven species is analysed using statistical models. The method employed, tests for spatial variation in the fraction of the population that has a certain age at a given length. Because age lengths keys may also vary over time, spatial effects are corrected for temporal effects of year and quarter.

3.3.1 Statistical analysis

If a typical age length key is plotted, *i.e.* the fraction of a certain age group against length class, it has a parabolic shape. For example, the fraction of fish that are age 3 is small at short lengths, and increases at larger lengths until it reaches a maximum, and then goes down again at even larger lengths. This is because at short lengths most fish are younger than age three and at large lengths most fish are older than age 3. (Figure 3.14).

This curve describes the fraction of the whole population belonging to a certain length class that belongs to a certain age group. It is, however, rather difficult to analyse spatial variation when the age length key is approached in this way. Therefore, a different approach was chosen. This approach analyses spatial variation in the fraction of the population of age x and older, that is age x. So it analyses, for example, which fraction of the fish of age 4 and older is actually age 4. This fraction can be compared among different areas. This comparison allows for the spatial variation in age length keys to be tested. To statistically test for spatial variation in these age length keys, the continuation-ratio logits method was applied (Kvist *et al.*, 2000; 2001; Pauly *et al.*, 2002; Rindorf and Lewy, 2001). The method is explained in more detail in Box 1 and will be summarized here.

Usually, the fraction of fish that have a certain age of the total population at that length class is calculated (Table 3.3.2). From the number of fish per age group in a sample of a certain length class, the fraction in the total sample can be calculated. In the continuation-ratio logits method, the fraction of fish in this age group is not calculated of the total sample, but only of fish that are of that age and older. With increasing length, this fraction will decrease because fish grow when they become older. The relationship between length and this fraction can be statistically modelled (Figure 3.15).

To test for the spatial variation in age length keys, seven areas from the North Sea were defined and samples categorized to area. The difference in age length keys among areas was statistically tested. The definition of areas followed that of the ECOTOETS project (Figure 3.16).

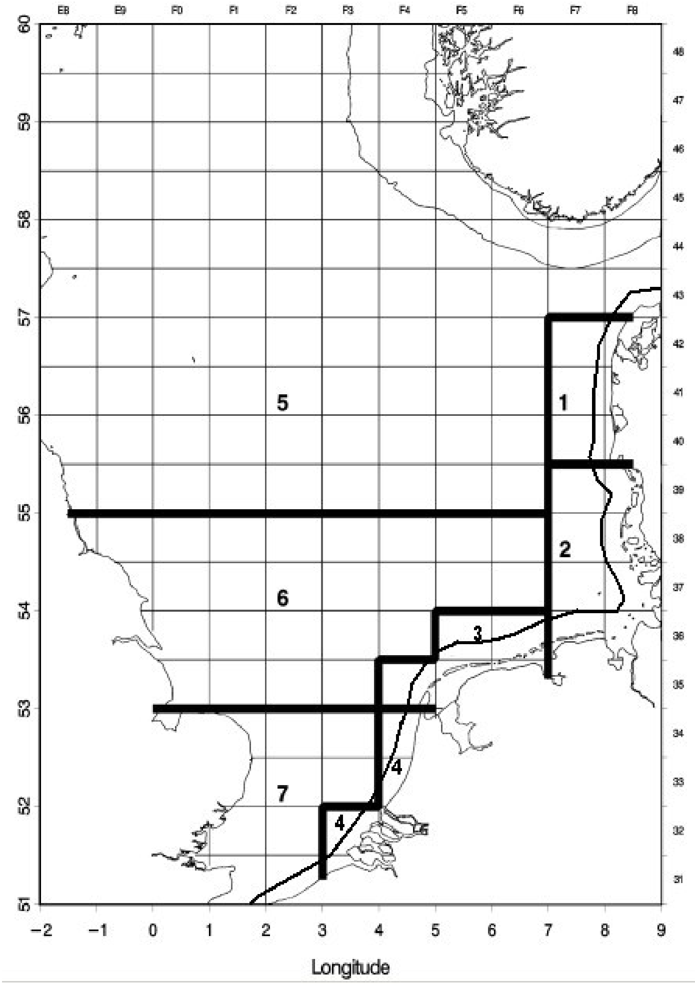


Figure 3.15. Definition of areas following the ECOTOETS project.

The analysis was carried out for ages 0 to 9 of seven species: plaice, sole, cod, whiting, mackerel, horse mackerel and herring. To cope with possible differences in age profiles between sexes, the effect of sex was included in the models. Also the year and quarter of sampling were included in the analysis to correct for changes in age lengths keys over years and among quarters. For each of these species, only the most abundant age groups were analysed and for each species, the best fitting statistical model[1] was chosen. With the parameters of these models, the age lengths keys can be constructed and differences among areas can be illustrated.

[1] The fit of the model was judged by the fraction of the deviance explained, the contribution of terms to the explanation of deviance in relation to the degrees of freedom these terms costed, and by convergence of the model.

Having tested the differences in the age length profiles, the representative nature of the sampling was tested for plaice. Using the ECOTOETS areas (Figure 3.15), the landings per area per quarter were compared with the sampled landings per area per quarter.

Box 1. Continuation-ratio logits

The variation in age length keys was analysed by employing the continuation-ratio logits method (Kvist *et al.*, 2000; 2001; Pauly *et al.*, 2002; Rindorf and Lewy, 2001). This method is appropriate when ordered categories (age) represent a progression of stages, so that individuals must pass through each lower stage before they go on to higher stages (Allison, 1999). We applied this method to model the progression of fish through age classes and to test for the spatial variation in the fraction of fish that have a certain age x. Other than in usual age length keys, not the fraction of the total population is regarded, but the fraction of fish that are age x and older. This fraction (*P*) can be modelled with a logistic model and was defined for each sex and species as:

$$P_{l.a} = N_{l.a} \ / \ (N_{l.\geq a}), \tag{1}$$

where $P_{l.a}$ is the proportion of the sample of length class (*l*) that had age (*a*), $N_{l.a}$ is the number of fish in the sample with length *l* and age *a*; $N_{l.\geq a}$ is the number of fish in the sample with length *l* and age *a* and older. $P_{l.a}$ has a binomial distribution and was logit-transformed: :

$$g = \log (P_{l.a} \ / \ 1\text{-}P_{l.a}). \tag{2}$$

With this transformation, the response variable (the transformed proportion) could be linked to a linear model. This model was defined as a function of length (*l*), year (*y*), quarter (*q*) and area (*a*). Quarter and area were defined as factors:

$$g \sim S + L + Y + Q_i + A_i + (L \text{ x } A_i) + (Y \text{ x } A_i) + (Q_i \text{ x } A_i), \tag{3}$$

where S=Female, male or unknown, L = length, Y = year, Q= 1 to 4 are 4 the quarters,and A=1 to 7 are the 7 areas and (L x A), (Yx A) and (Qx A) indicate the interactions between length, year and quarter respectively with area. These interaction terms allowed for a different effect of area at different lengths, years and quarters.

The age length keys were modelled for each age group using the SAS software (Allison, 1999). The aim of the investigation is to test whether age length key vary by area. If area did not contribute significantly to the explanation of deviance (P>0.05), there was no statistical difference in the age length keys per area.

3.3.2 Results

Cod. In total of 36259 individual cod data on the age and length from the period of 1968-2003 could be analysed (Table 3.3.3). For almost all ages that could be analysed, age length keys varied significantly among areas, although the effect of area decreased with higher ages. At ages 1, 2,3 and 4 there was a highly significant (P<0.01) effect of area on the age length keys whereas at ages 5, 6 and 7 it was significant ($0.01 \geq P < 0.05$) and at age 9 it was not significant ($P \geq 0.10$). The difference in age length keys among areas is illustrated in Figure 3.16. It shows that cod of the same length are generally older in northern regions than in southern regions. Over time, the age length key has also significantly changed (Figure 3.17): on average, fish of the same size have become older because the fraction of age-1 fish has decreased.

Herring. The analysis for herring was executed with data of in total 202532 individual fish from the period of 1960-2003 (Table 3.3.4). At all ages, age length keys varied highly significantly among areas.

Horse mackerel. The age profiles of horse mackerel also varied highly significantly among areas (Table 3.3.5). In total, this analysis was based on 42961 individual horse mackerel from the period of 1982-2002.

Mackerel. Of mackerel observations of in total 81089 fish from 1968-2003 could be analysed (Table 3.3.6). At all ages, the age profiles varied significantly among areas. For all ages, except age 6, the effect of area was highly significant. For all ages, the effect of area depended on the year and quarter (interactions were significant).

Plaice. The analysis for plaice was executed with data of in total 343285 individual fish from the period of 1957-2003 (Table 3.3.7). For all ages analysed, age length keys varied highly significantly among areas. Also, for all ages, there is a significant difference in age length key between the two sexes. All interaction terms between length, year and quarter with area were also significant. The effect of the significant interaction between length and area is illustrated in Figure 3.18. Because the effects of area and length do interact, the slopes of the logistic curves differ among areas, and not only the intercept as is the case for cod (Figure 3.16). For example, the curve is steeper at area 7 than at area 4.

As was the case with cod (Figure 3.17), the age length key of plaice has also significantly changed over time (Figure 3.19). However, the age length key changed in the opposite direction as that of cod: on average, fish of the same size have become younger because the fraction of age-4 at a certain length class increased.

Sole. The analysis for sole was executed with data of in total 229228 individual fish from the period of 1951-2003 (Table 3.3.8). Also for sole, age length keys varied highly significantly among areas for all ages analysed (Figure 3.20). As was the case for plaice, age length keys also varied significantly between sexes and over time but this change was smaller than was shown for cod and plaice (Figure 3.21).

Whiting. The analysis for whiting was executed with data of in total 49466 individual fish from the period of 1968-2003 (Table 3.3.9). Only for ages 0 to 6 a significant difference in age length keys among areas could be detected. The analyses for ages 7 and 8 are, however, based on such low numbers of fish that their results are questionable. For ages 0-6, there was also a significant difference between sexes and over years.

Spatial nature of sampling. It is clear that the sampling of plaice is representative of the catches, in terms of the total landings of the Dutch fleet from an area and the associated sampled catch in the Netherlands from that area (Figure 3.22). While the sampling in quarters 1 and 4 has a stronger relationship (therefore more representative), in all quarters the sampled catch is proportionate to the landed catch from an area.

3.3.3 Discussion

For all species of which long term data were analysed, differences in age length keys among areas were demonstrated. These differences were large as was demonstrated for cod, plaice and sole. For a female plaice of 30 cm, the fraction of 4-group fish in the population of age 4 and older fish, varied from 25 % in the Danish coastal zone to 95 % in the southern North Sea. The spatial variation in age length keys is not surprising because growth rates will also vary in space due to variation in *e.g.* water temperature. It also shows that the sampling covers a range of the population or stock.

The conclusion that there is indeed spatial variation in age length keys of the commercially important species described here, implies that the sampling of the commercial catch should thus be spatially representative. If the sampling of the commercial catch would be spatially representative, the proportion of fish sampled from a certain area would not differ significantly from the proportion of fish caught in that area. In terms of plaice, it is clear that this is the case and the samples do reflect the landings (Figure 3.22), suggesting that the sampling strategy does reflect the variable landings by area and hence is representative.

This analysis shows that spatial variation in age length keys exist for the commercially important species. For plaice it is suggested that the sampling is spatially representative of the catches. These findings combined with the analysis in section 3.1 suggest that the current method and strategy for sampling of landings is appropriate for the collection of market data used in stock assessments. However, there should still be caution as the bootstrap method can result in an under estimation of the variance (Manly, 1991).

Table 3.1. Summary of the market sampling procedure of seven species at RIVO.

Species	Samples collected	Lengths measured	Otoliths removed	Type of sampling	Categories	Adjusted total catch
plaice	market	lab	lab	representative	yes	G
sole	market	lab	lab	representative	yes	G
cod	market	market	market	length stratified	yes	G
whiting	market	market	lab	length stratified	no*	G
mackerel	on board	lab	lab	representative	no	M
horse mackerel	on board	lab	lab	representative	no	M
herring	on board	lab	lab	representative	no	M

lab= laboratory, on board= on fishing vessel
Adjustments to total catch = changes to official catch, G= gutting correction, M= misreporting correction
* officially whiting landed in categories but actually only one size category used at market

Table 3.2a. National and total sampling of four demersal stocks in 2002.

Stock	Country	Landings ('000 tonnes)	No of fish for length ('000s)	No of fish aged ('000s)
Cod-North Sea	Denmark	16.3	6.8	6.6
	UK- (Scot)	15.4	46.2	8.9
	Norway	5.1	3.5	0.2
	Netherlands	4.7	3.3	2.1
	UK (E/W/NI)	3.3	34.7	3.8
	France	3.1	-	-
	Belgium	2.7	-	-
	Sweden	2.2	1.1	0.7
	Germany	2.1	2.1	0.5
	Poland	0.0	-	-
	Total	**54.9**	**97.7**	**22.8**
	Sampling rate per tonne		1.78	0.42
Whiting-North Sea	France	8.5	1.6	2.6
	UK- (Scot)	7.8	59.3	3.9
	Netherlands	2.4	5.9	1.2
	UK (E/W/NI)	1.4	16.9	1.1
	Germany	0.4	2.7	-
	Belgium	0.3	-	-
	Denmark	0.1	-	-
	Sweden	0.0	-	-
	Norway	0.0	-	-
	Total	**20.9**	**86.4**	**8.8**
	Sampling rate per tonne		4.13	0.42
Sole-North Sea	Netherlands	12.1	4.0	4.0
	Belgium	1.4	2.7	0.4
	Germany	0.8	1.0	0.1
	Denmark	0.6	0.2	0.0
	UK (E/W/NI)	0.5	16.1	1.9
	France	0.3	-	-
	UK- (Scotland)	0.2	-	-
	Total	**15.9**	**24.0**	**6.4**
	Sampling rate per tonne		0.69	0.16
Plaice-North Sea	Netherlands	29.1	7.7	7.7
	Denmark	12.6	1.9	1.9
	UK (E/W/NI)	8.6	19.9	1.3
	UK- (Scot)	8.2	5.9	-
	Belgium	4.8	2.9	0.4
	Germany	3.9	9.9	0.1
	Norway	2.0	-	-
	France	0.5	-	-
	Sweden	0.0	-	-
	Total	**69.7**	**48.2**	**11.4**
	Sampling rate per tonne		1.51	0.40

Table 3.2b. National and total sampling of three pelagic stocks in 2002.

Stock	Country	Landings ('000 tonnes)	No of fish for length ('000s)	No of fish aged ('000s)
Herring–North Sea	Norway	75.0	3.5	3.5
	Denmark	70.8	9.0	0.9
	Netherlands	55.3	9.2	1.6
	UK (Scot)	30.9	17.0	4.1
	Germany	27.2	14.9	0.8
	France	25.4	-	-
	UK (E/W/NI)	14.7	-	-
	Sweden	3.4	-	-
	Faroe Islands	1.4	-	-
	Belgium	0.0	-	-
	Total	**304.2**	**53.6**	**10.9**
	Sampling rate per tonne		0.18	0.04
Mackerel-NE Atlantic	Norway	184.3	24.7	3.9
	UK (Scot)	165	27.6	6.1
	Ireland	72.2	7.2	2.0
	Spain	50.5	17.6	3.0
	Russia	45.8	27.7	1.9
	Denmark	34.4	1.4	1.3
	Netherlands	33.5	7.9	2.5
	Germany	26.5	36.7	1.5
	UK (E/W/NI)	26.1	3.8	1.1
	France	21.9	-	-
	Faroes	19.8	0.2	0.2
	Sweden	5.2	-	-
	Portugal	2.9	29.2	2.6
	Belgium	0.0	-	-
	Iceland	0.0	-	-
	Total	**688.1**	**184.0**	**26.1**
	Sampling rate per tonne		0.27	0.04
Horse mackerel-western	Netherlands	57.2	26.3	1.8
	Norway	36.7	2.8	0.9
	Ireland	36.5	4.7	1.2
	Spain	31.5	36.3	1.8
	France	20.2	-	-
	Germany	15.9	27.7	0.4
	Portugal	14.3	0*	1.5
	Denmark	12.5	-	-
	UK (E/W/NI)	8.3	-	-
	UK (Scot)	2.9	-	-
	Faroes	0.7	-	-
	Sweden	0.6	-	-
	Belgium	0.0	-	-
	Russia	0.0	-	-
	Total	**237.3**	**97.8***	**7.6**
	Sampling rate per tonne		0.41	0.03

* sampling of horse mackerel by Portugal uncertain

Table 3.3.1. Theoretical example of an age length key. The table presents the fraction of the population in a given length class that has a certain age. Of the fish of 11 cm, for example, 75 % is age 1 and 25 % is age 2.

Length (cm)	Age 1	Age 2	Age 3	Age 4
10	1.00	0.00	0.00	0.00
11	0.75	0.25	0.00	0.00
12	0.50	0.50	0.00	0.00
13	0.25	0.50	0.25	0.00
14	0.00	0.25	0.50	0.25
15	0.00	0.00	0.50	0.50

Table 3.3.2. Theoretical example of the continuation-ratio logits method. Here, the age of 100 fish of one length class was determined. From the number of fish (B) per age group (A), the fraction in the sample can be calculated (C). In the continuation-ratio logits method, the fraction of fish in this age group is not calculated for the total sample, but only for fish that are of that age and older. In this case, a fraction of 0.1 of all fish that are age 1 and older(100 fish) is actually age 1 (10 fish). A fraction of 0.33 of all fish that are age 2 and older (90 fish) is actually age 2 (30 fish). A fraction of 0.67 of all fish that are age 3 and older (60 fish) are actually age 3 (40 fish). A fraction of 1.0 of all fish that are age 4 and older (20 fish) are actually age 4 (20 fish).

A Age	B Number of fish	C Fraction population	D Fraction of age x and >age x
1	10	0.10	0.10
2	30	0.30	0.33
3	40	0.40	0.67
4	20	0.20	1.00
Total	100		

Table 3.3.3. Results of the logistic regression of age length keys for each age group of cod. The effects of sex (S), length (L), year (Y), quarter (Q), area (A), interaction between length and area (L x A), year and area (Y x A) and quarter and area (Q x A) are presented. Significance of terms is indicated with symbols: n.s.: not significant ($P \geq 0.10$); *: $0.05 \geq P < 0.10$; **: $0.01 \geq P < 0.05$; ***: $P < 0.01$; x: term could not be not included in the model. The probabilities are the likelihood ratio statistics that test whether a term contributes significantly, after other terms have been included in the model. R^2 is the faction of de[1]viance explained and observations is the total number of fish on which the analysis was based (please note that the sum of these observations is larger than the total number of fish in the database; a fish of 4 yrs old for example, is included in the analyses for age 0, 1, 2, 3 and 4).

Age	S	L	Y	Q	A	(L x A)	(Y x A)	(Q x A)	R^2	Observations
0	x	x	x	x	x	x	x	x	-	36214
1	n.s.	***	***	***	***	x	x	x	0.72	34316
2	***	***	***	***	***	x	x	x	0.56	18921
3	**	***	***	***	***	x	x	x	0.54	7638
4	***	***	**	***	***	x	x	x	0.40	3489
5	***	***	**	***	**	x	x	x	0.28	1830
6	n.s.	***	n.s.	***	**	x	x	x	0.20	873
7	n.s.	***	n.s.	n.s.	**	x	x	x	0.17	448
8	x	x	x	x	x	x	x	x	-	250
9	n.s.	***	**	n.s.	n.s.	x	x	x	0.17	108

1 Deviance is the amount of variance in logistic models. In logistic models, R^2 is comparable to the R^2 in linear models. In logistic models, R^2 is also called pseudo-R^2.

Table 3.3.4. Results of the logistic regression of age length keys for each age group of herring. Legend as in Table 3.3.3.

Age	S	L	Y	Q	A	(L x A)	(Y x A)	(Q x A)	R^2	Observations
2	***	***	***	***	***	***	X	X	0.70	201399
3	***	***	***	***	***	***	X	X	0.42	178976
4	***	***	***	***	***	X	X	X	0.28	108261
5	***	***	***	***	***	***	X	X	0.17	58125
6	***	***	***	***	***	***	X	X	0.13	31864
7	***	***	***	***	***	***	X	X	0.06	18278
8	-	**	***	**	***	***	X	X	0.03	11015
9	-	***	***	***	***	***	X	X	0.03	6419

Table 3.3.5. Results of the logistic regression of age length keys for each age group of horse mackerel. Legend as in Table 3.3.3.

Age	S	L	Y	Q	A	(L x A)	(Y x A)	(Q x A)	R^2	Observations
3	-	***	-	***	***	X	X	X	0.65	39514
4	***	***	***	***	***	X	X	X	0.41	35464
5	-	***	***	***	***	X	X	X	0.30	32271
6	**	***	***	***	***	X	X	X	0.26	28489
7	**	***	***	***	***	X	X	X	0.21	25003
8	*	***	***	***	***	X	X	X	0.25	20550
9	-	***	***	***	***	X	X	X	0.18	17133

Table 3.3.6. Results of the logistic regression of age length keys for each age group of mackerel. Legend as in Table 3.3.3.

Age	S	L	Y	Q	A	(L x A)	(Y x A)	(Q x A)	R^2	Observations
2	-	***	***	***	***	***	***	***	0.58	71726
3	***	***	***	***	***	***	***	***	0.38	58838
4	*	***	**	***	***	-	***	***	0.26	46352
5	-	***	-	***	***	-	***	***	0.21	35635
6	***	***	-	***	**	-	***	***	0.16	26567
7	***	***	-	***	***	-	***	***	0.14	19825
8	-	***	-	*	***	-	***	***	0.12	14525

Table 3.3.7. Results of the logistic regression of age length keys for each age group of plaice. Legend as in Table 3.3.3.

Age	S	L	Y	Q	A	(L x A)	(Y x A)	(Q x A)	R^2	Observations
3	***	***	***	***	***	***	***	***	0.40	264362
4	***	***	***	***	***	***	***	***	0.33	202803
5	***	***	***	***	***	***	***	***	0.26	148684
6	***	***	***	***	***	***	***	***	0.20	106956
7	***	***	***	***	***	***	***	***	0.17	76788
8	***	***	***	***	***	**	***	**	0.13	55035
9	***	***	***	***	***	**	***	***	0.11	40951

Table 3.3.8. Results of the logistic regression of age length keys for each age group of sole. Legend as in Table 3.3.3.

Age	S	L	Y	Q	A	(L x A)	(Y x A)	(Q x A)	R^2	Observations
3	***	***	***	***	***	***	***	***	0.36	185836
4	***	***	***	***	***	***	***	***	0.25	131307
5	***	***	***	**	***	***	***	***	0.20	89857
6	***	***	***	***	***	***	***	***	0.16	62220
7	***	***	***	***	***	***	***	***	0.14	43846
8	***	***	***	*	***	***	***	***	0.11	31944
9	***	***	***	-	***	*	***	*	0.09	24444

Table 3.3.9. Results of the logistic regression of age length keys for each age group of whiting. Legend as in Table 3.3.3.

Age	S	L	Y	Q	A	(L x A)	(Y x A)	(Q x A)	R^2	Observations
0	***	***	**	***	***	X	X	X	0.83	49297
1	***	***	***	***	***	X	X	X	0.62	45932
2	***	***	***	***	***	X	X	X	0.30	30368
3	***	***	***	***	***	X	X	X	0.16	16913
4	***	***	***	***	***	X	X	X	0.10	7276
5	***	***	***	-	***	X	X	X	0.06	2427
6	***	***	***	-	*	X	X	X	0.05	768
7	-	***	-	-	-	X	X	X	0.07	231
8	-	-	*	-	-	X	X	X	0.11	67

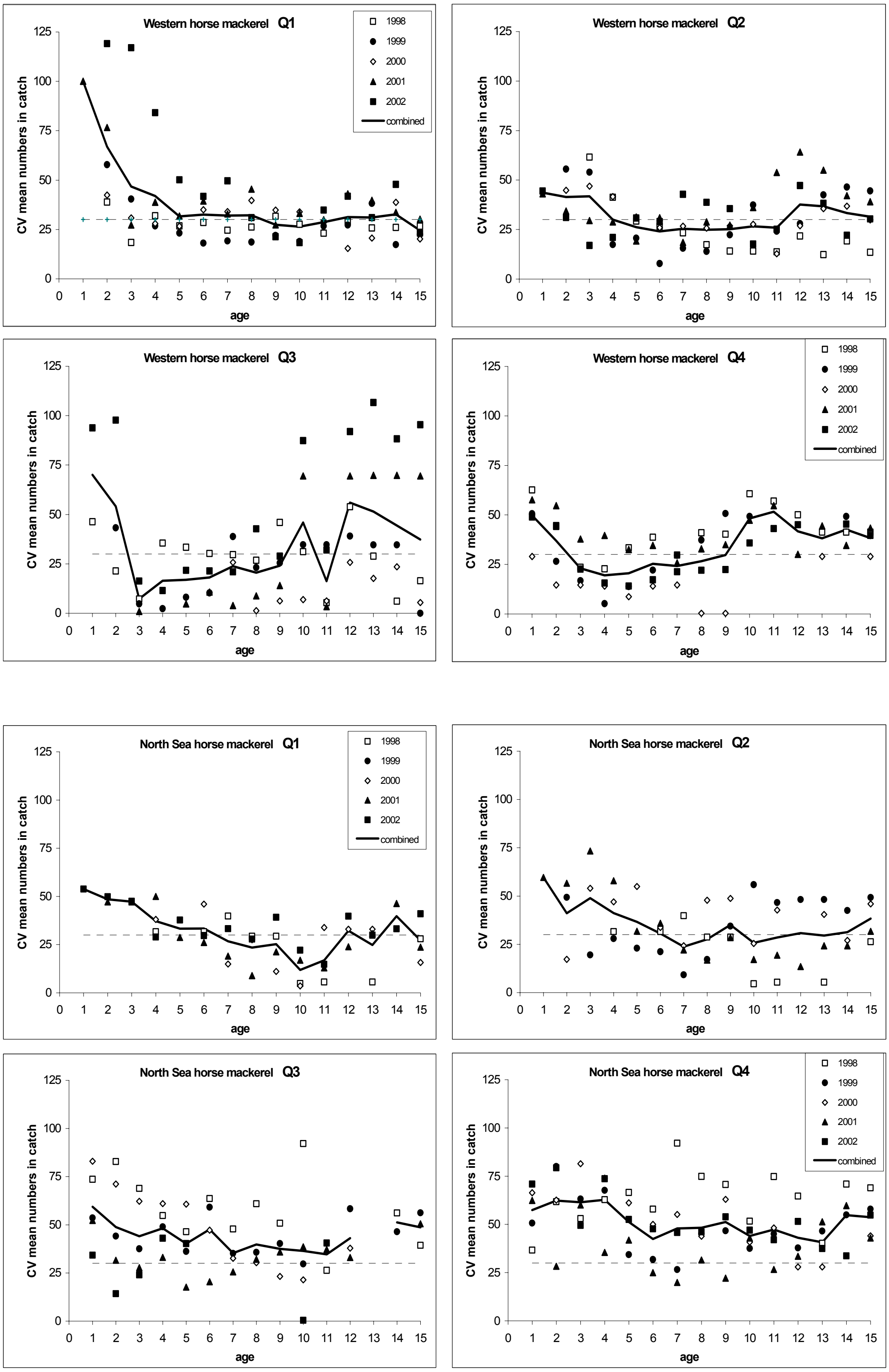


Figure 3.1. Western and North Sea horse mackerel in the Dutch catch. Precision (CV) in the estimation of numbers at age in the catch by quarter. Dotted line denotes CV= 30%.

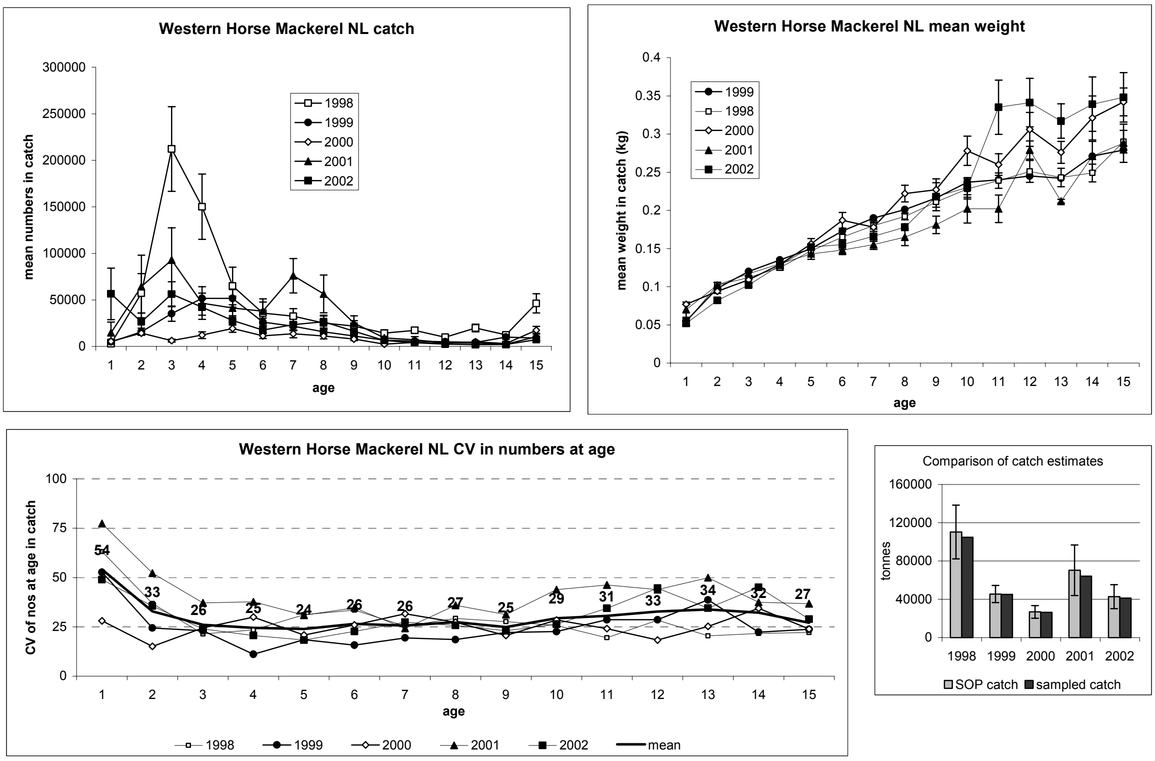


Figure 3.2. Western horse mackerel in the Dutch catch. Mean numbers at age and weights at age in the catch (error bars denote 1 standard error). Precision (CV) in the estimation of numbers at age in the catch by year. Comparison of SOP estimated catch with actual catch.

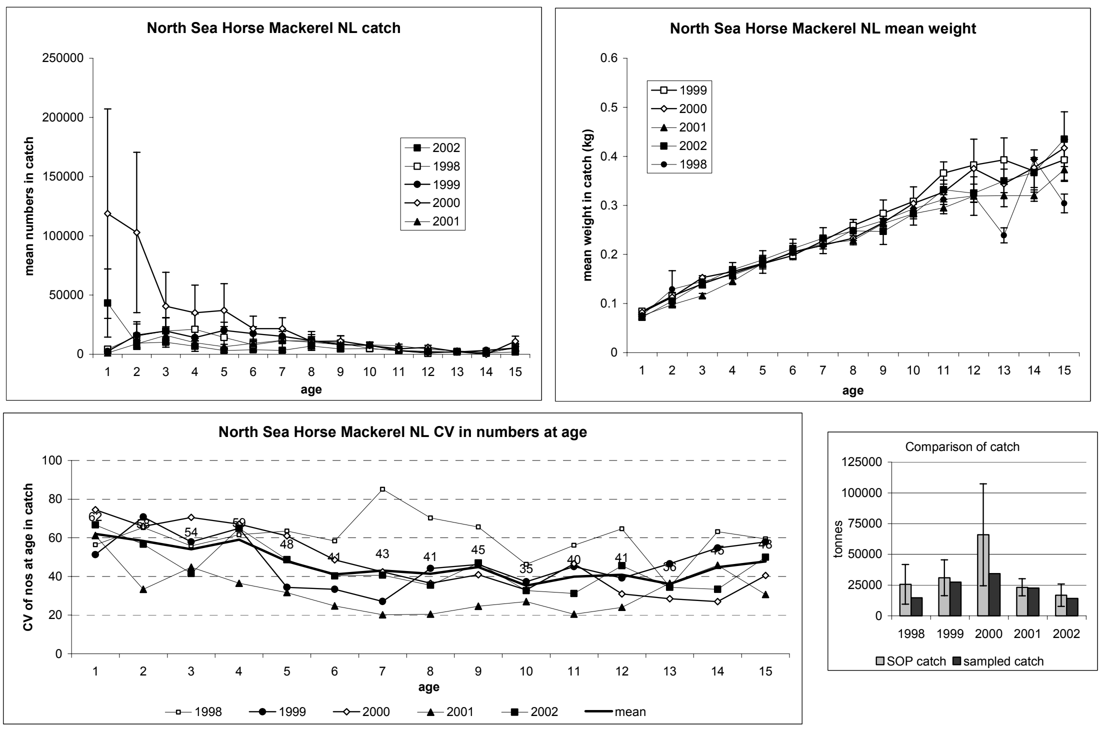


Figure 3.3. North Sea horse mackerel in the Dutch catch. Mean numbers at age and weights at age in the catch (error bars denote 1 standard error). Precision (CV) in the

estimation of numbers at age in the catch by year. Comparison of SOP estimated catch with actual catch.

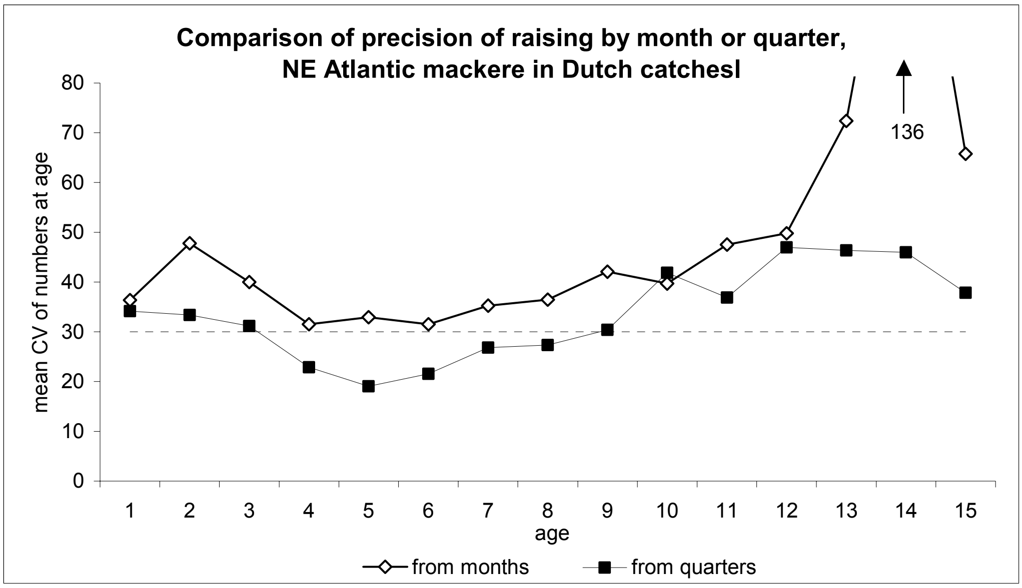


Figure 3.4. NE Atlantic mackerel in the Dutch catch. Comparison of precision (CV) in the estimation of annual numbers at age in the catch raised from months and from quarters. Dotted line denotes CV= 30%.

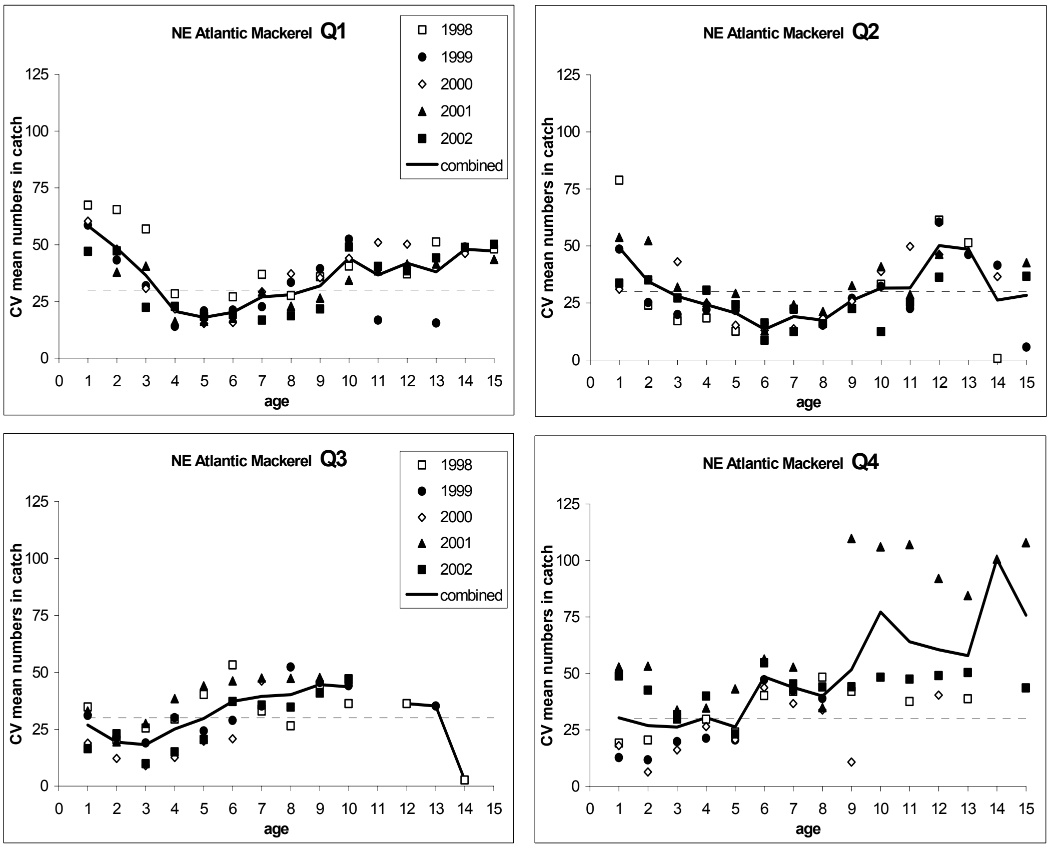


Figure 3.5. NE Atlantic mackerel in the Dutch catch. Precision (CV) in the estimation of numbers at age in the catch by quarter. Dotted line denotes CV= 30%.

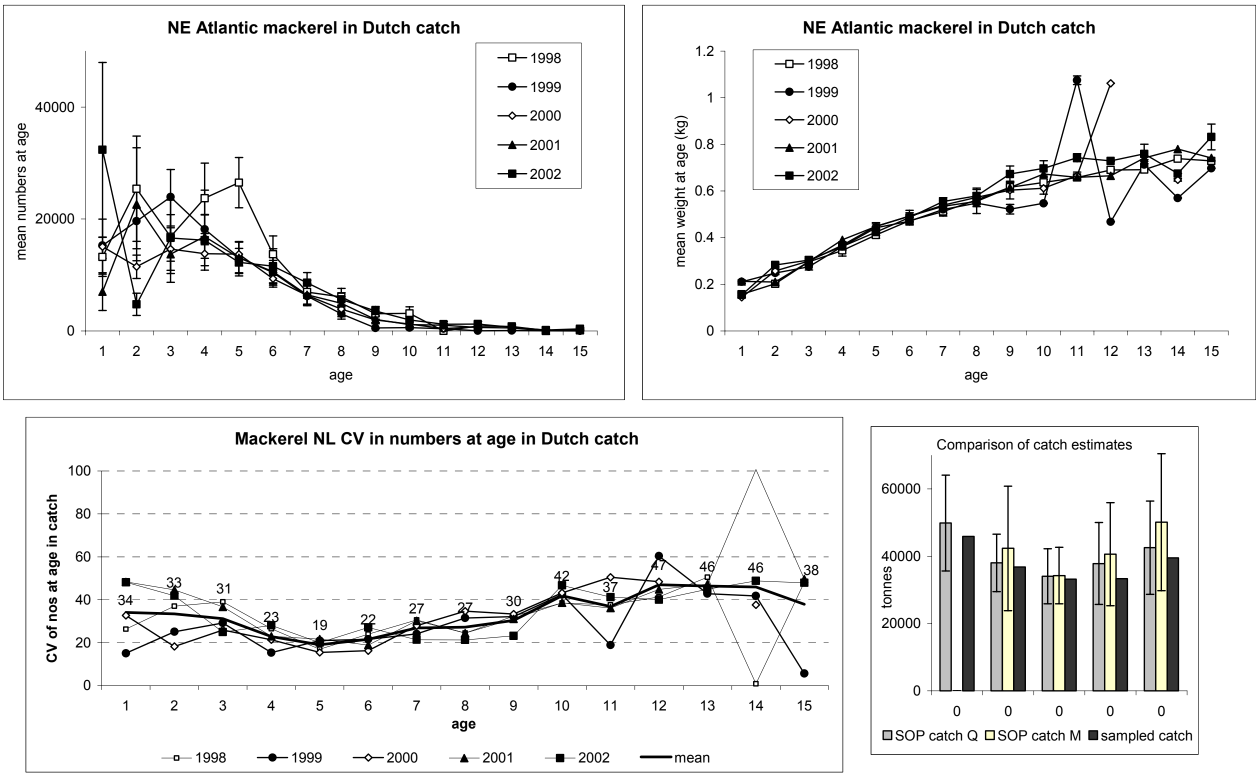


Figure 3.6. NE Atlantic mackerel in the Dutch catch. Mean numbers at age and weights at age in the catch (error bars denote 1 standard error). Precision (CV) in the estimation of numbers at age in the catch by year. Comparison of SOP estimated catch (from both monthly and quarterly raising of samples) with actual catch.

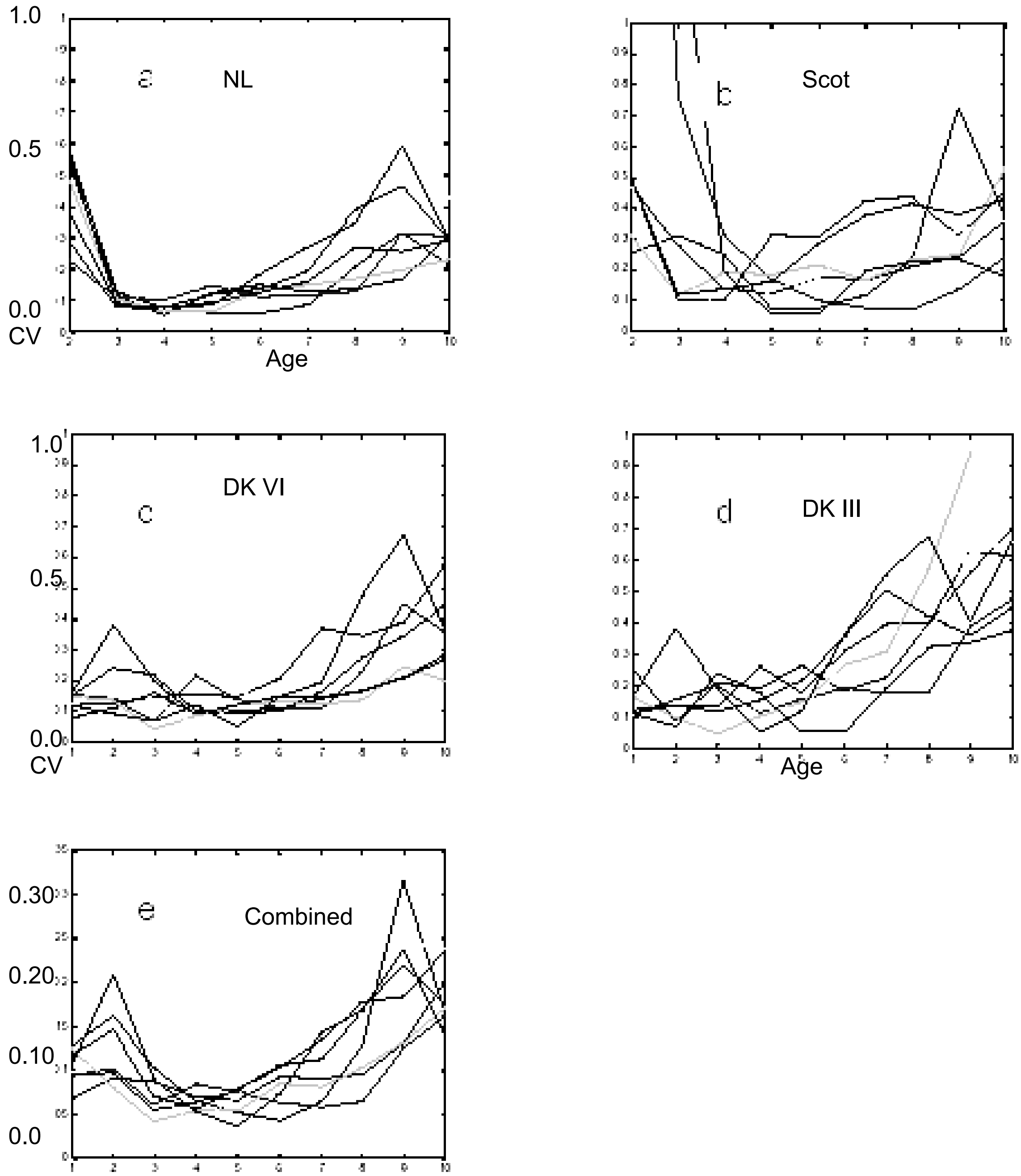


Figure 3.7. National CVs for North Sea herring catch at age (wr+1) for a) Netherlands, b) Scotland, c) Denmark in ICES area VI, d) Denmark in ICES area III. CVs (as proportion of 1) at age for North Sea herring for 75% of combined international catch (e). Taken from the EMAS project and published in Simmonds *et al.* 2001.

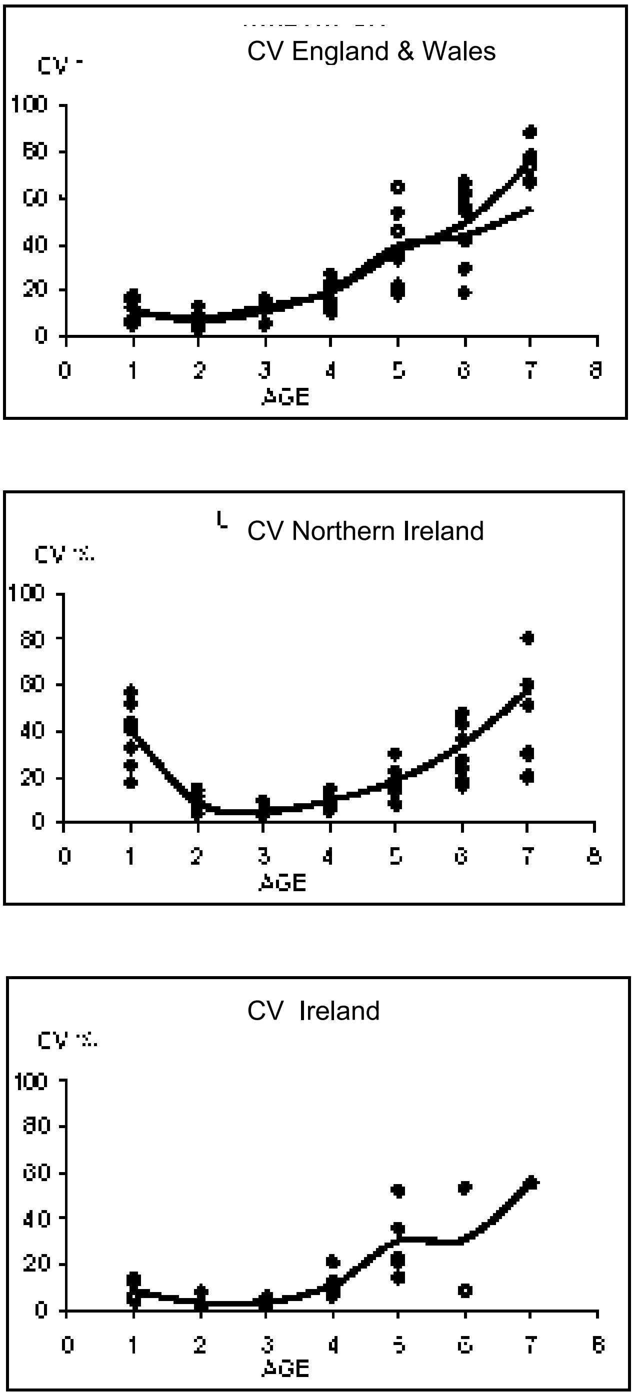


Figure 3.8. Irish Sea cod (ICES VIIa). CV of annual estimates of catch numbers-at-age. Points show estimates by year, lines show average across years. Taken from Maxwell *et al.*, 2001.

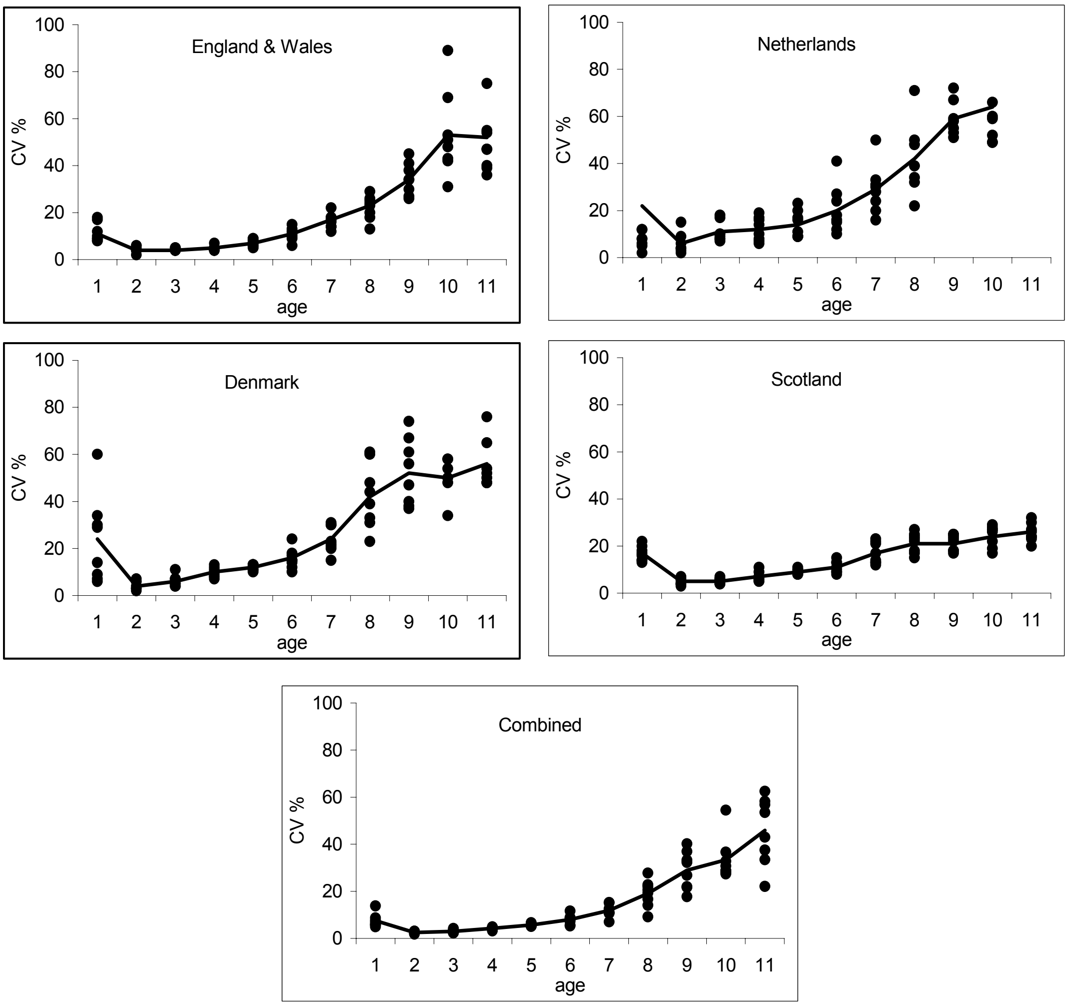


Figure 3.9. North Sea cod (1991-1998). CV of annual estimate of numbers at age in the catch. Points show estimate per year, line shows average. Taken from the EMAS project and published in O'Brien *et al.* 2001b.

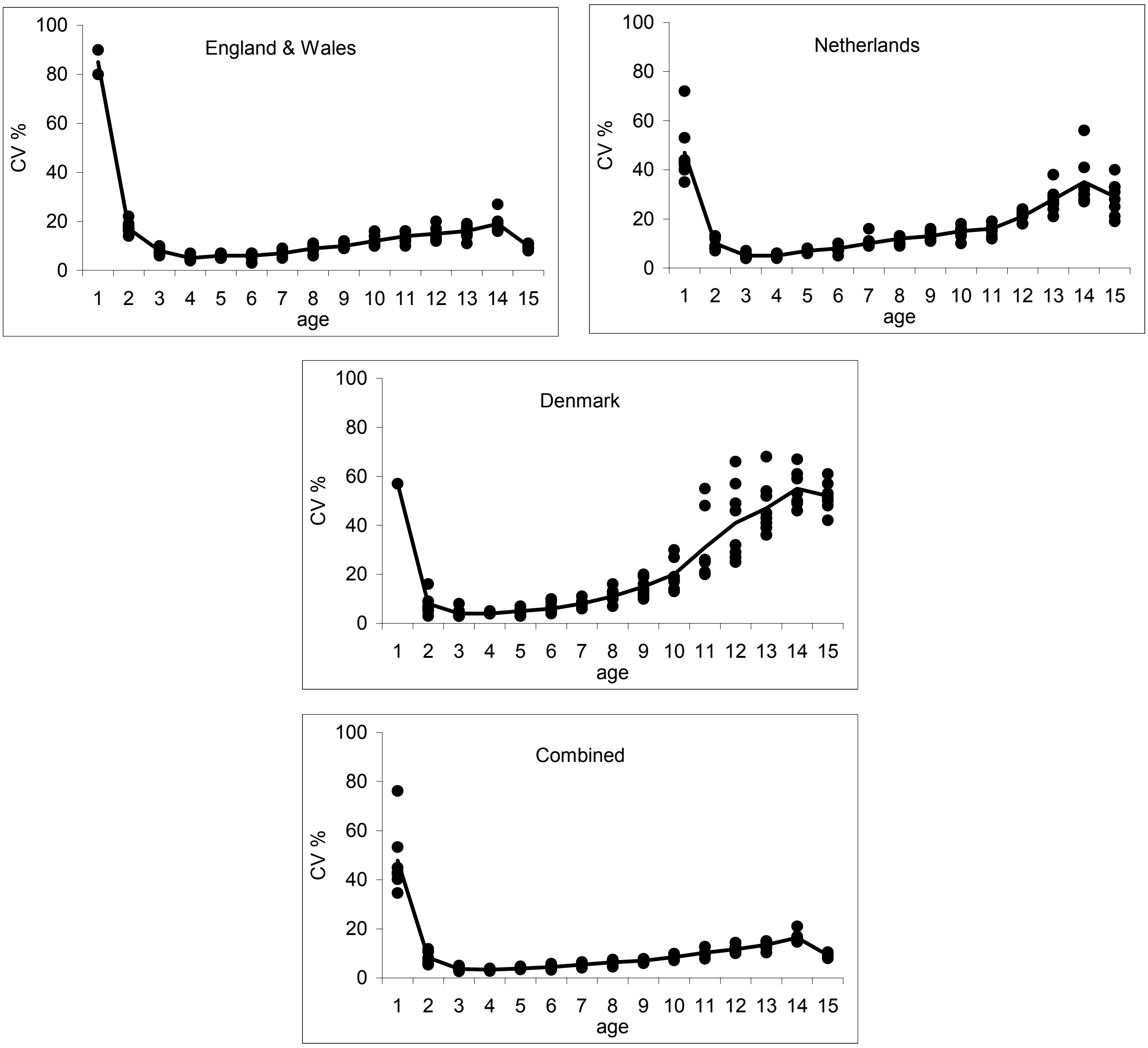


Figure 3.10. North Sea plaice (1991-1998). CV of annual estimate of numbers at age in the catch. Points show estimate per year, line shows average. Taken from the EMAS project and published in O'Brien *et al.* 2001a.

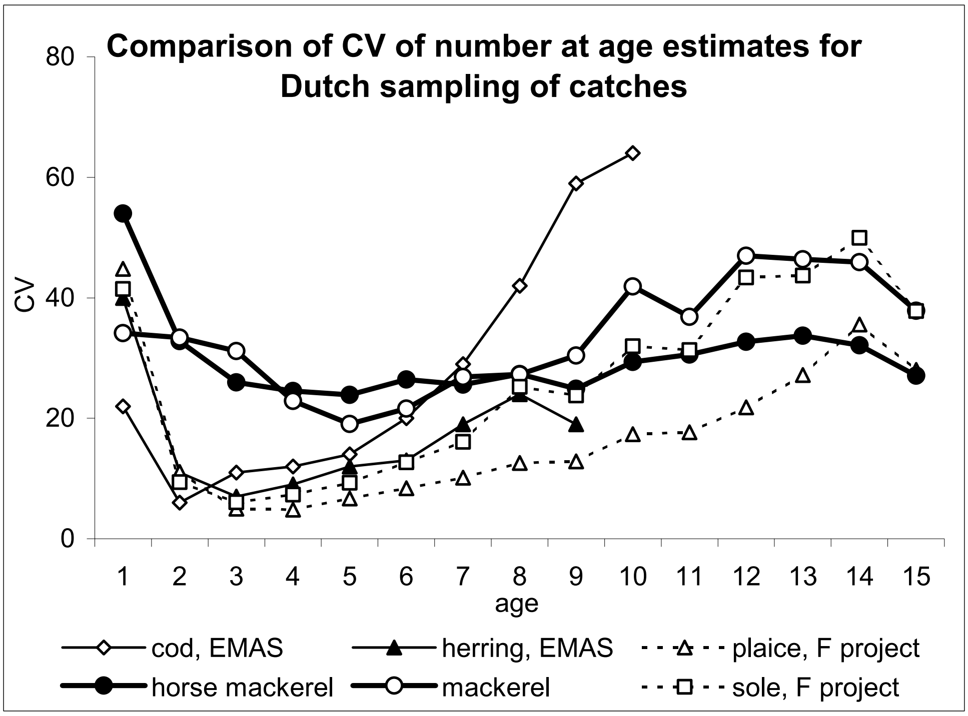


Figure 3.11. CV of annual estimate of numbers at age in the Dutch catch by species. Horse mackerel and mackerel 1998-2002; cod and herring 1991-1998 (EMAS project), plaice and sole 1991-2001 (F project).

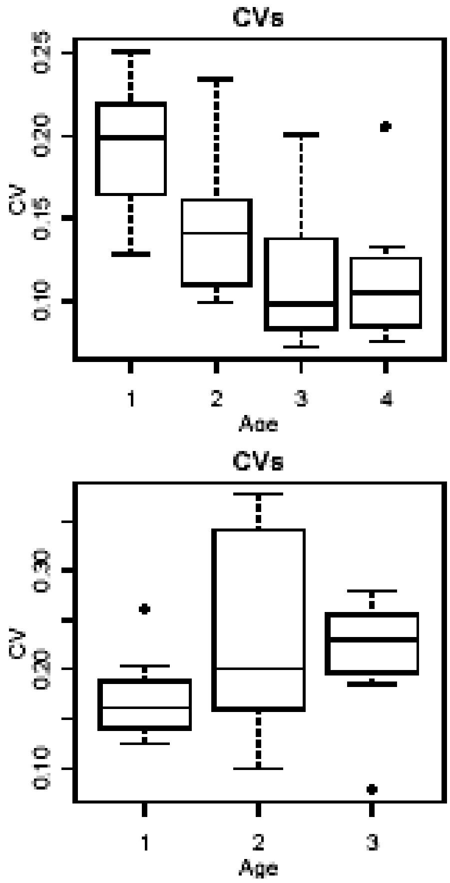


Figure 3.12. Statistical analysis of bootstrapped North Sea plaice research vessel sampling errors. CV= coefficient of variation of the estimated numbers at age. Top plot = BTS and bottom plot = SNS. Box whisker plots show median, quartiles and range. Taken from Kell *et al* (2003), based on EVARES project.

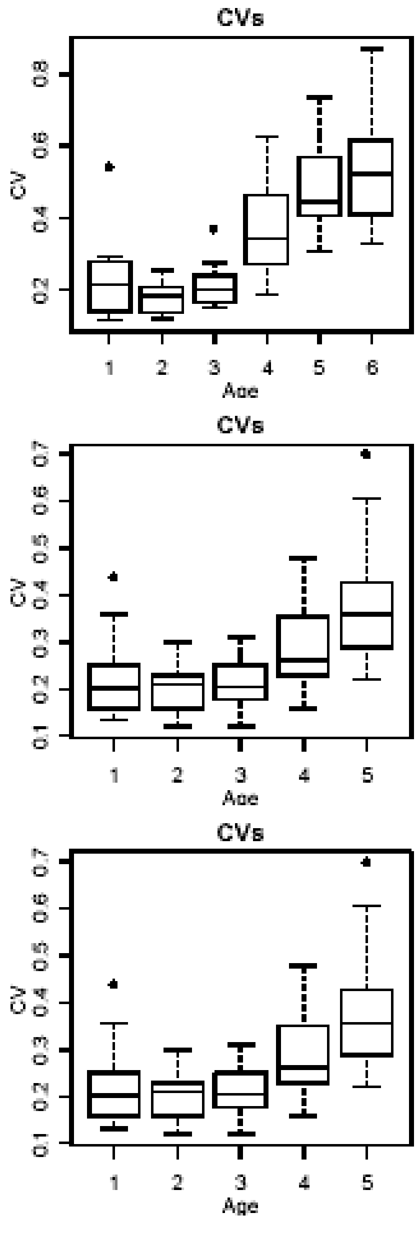


Figure 3.13. Statistical analysis of bootstrapped North Sea cod research vessel sampling errors. CV= coefficient of variation of the estimated numbers at age. Top plot = English GFS, middle plot = IBTS and bottom plot = Scottish GFS. Box whisker plots show median, quartiles and range. Taken from Kell *et al* (2003), based on EVARES project.

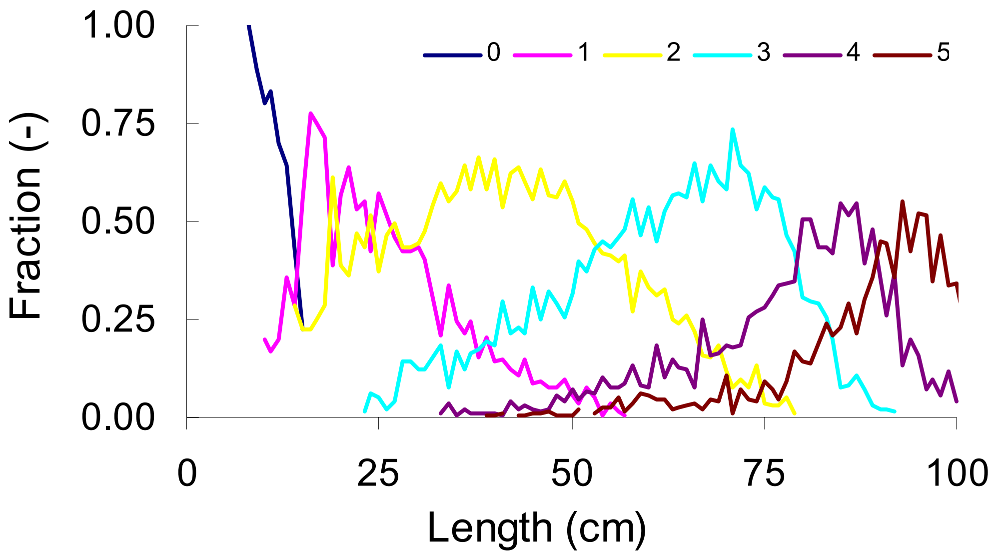


Figure 3.14. Age length key for cod with the length on the x-axis (cm) and the fraction of the population of a certain age group on the y-axis. Data from 1998-2002 for age groups 0-5.

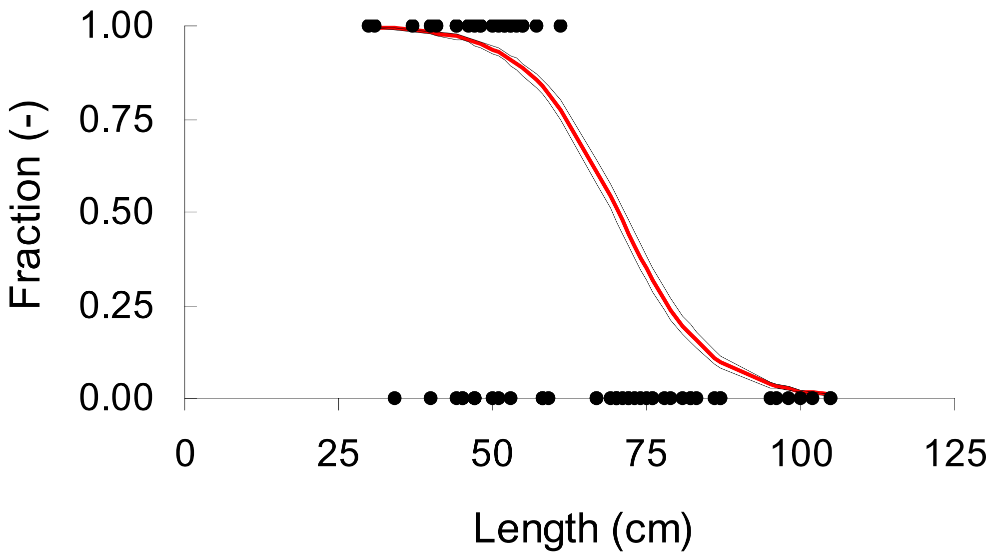


Figure 3.15. Example of the application of the continuation-ratio logits method for cod of age 3 and older. The dots represent observations of cod of a given length (x-axis) that are age 3 (y=1) or age 3 and older (y=0). The lines represent model predictions (red line) and 95 % confidence limits for the fraction of cod of a given length class of all cod of age 3 and older, that are 3 years old.

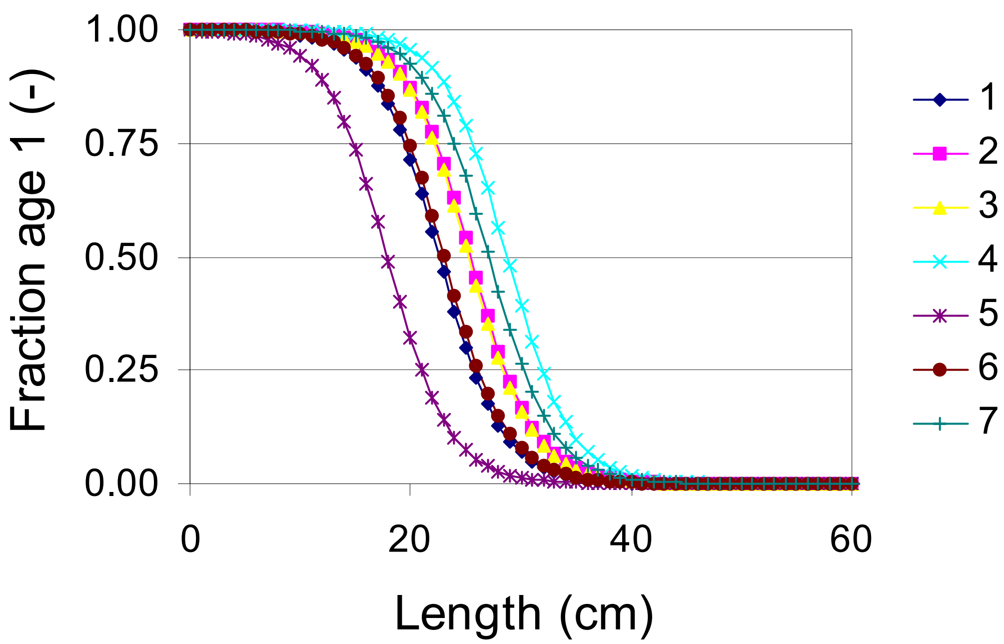


Figure 3.16. Illustration of the difference in age length keys among areas 1 to 7 for age-1 cod sampled in the 1st quarter of 2000. For each area, the fraction of age-1 and older cod that is age 1 is plotted against length.

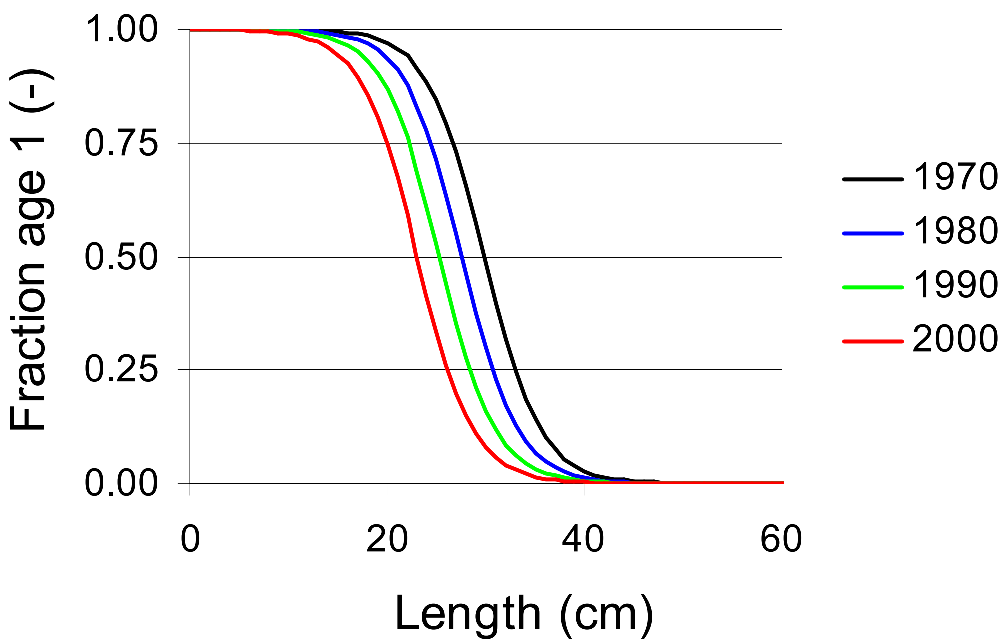


Figure 3.17. Illustration of the difference in age length keys among years for age-1 cod sampled in the 1st quarter of 2000. For each year (1960, 1970, 1980 and 2000), the fraction of age-1 and older cod that is age 1 is plotted against length.

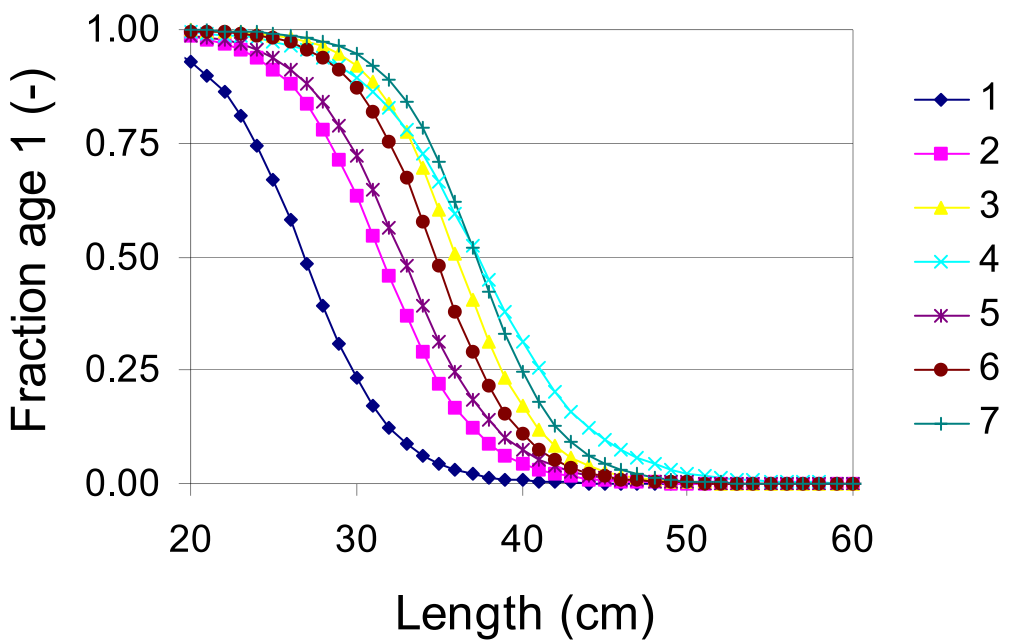


Figure 3.18. Illustration of the difference in age length keys among areas 1 to 7 for age-4 female plaice sampled in the 1st quarter of 2000. For each area, the fraction of age-4 and older female plaice that is age 4 is plotted against length.

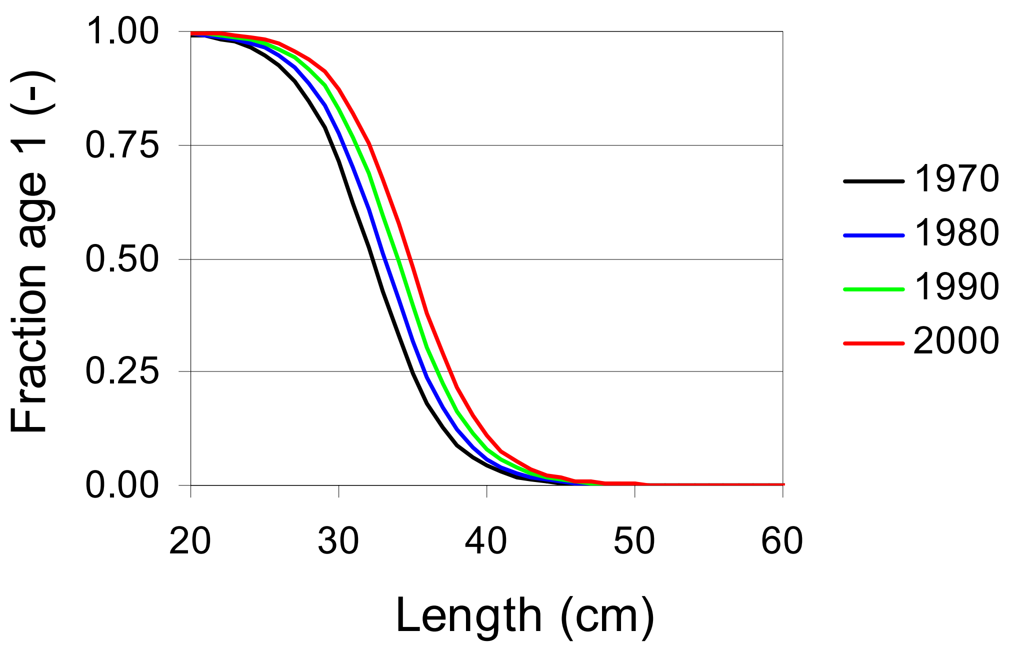


Figure 3.19. Illustration of the difference in age length keys among years for age-4 female plaice sampled in the 1st quarter of 1970, 1980, 1990 and 2000. For each year, the fraction of age-4 and older female plaice that is age 4 is plotted against length.

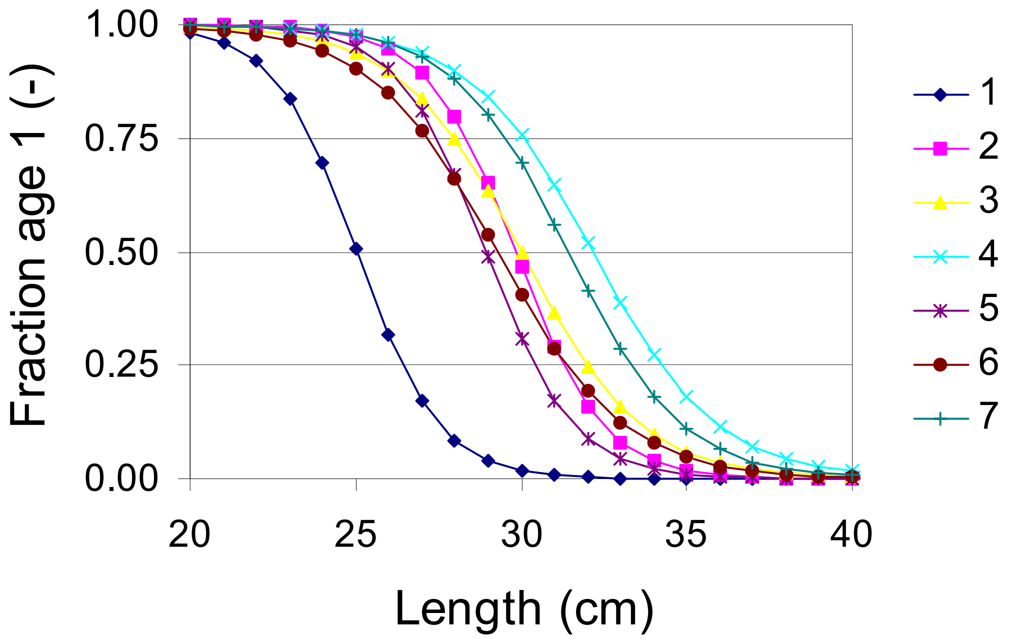


Figure 3.20. Illustration of the difference in age length keys among areas 1 to 7 for age-3 female sole sampled in the 1st quarter of 2000. For each area, the fraction of age-3 and older female plaice that is age 3 is plotted against length.

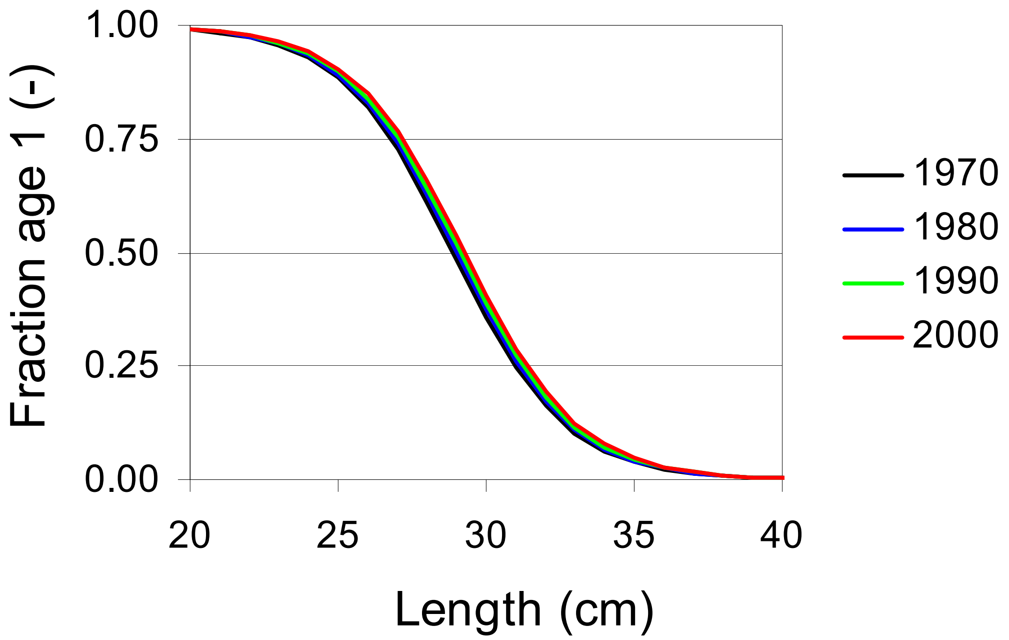


Figure 3.21. Illustration of the difference in age length keys among years for age-3 female ole sampled in the 1st quarter of 1970, 1980, 1990 and 2000. For each year, the fraction of age-3 and older female sole that is age 3 is plotted against length.

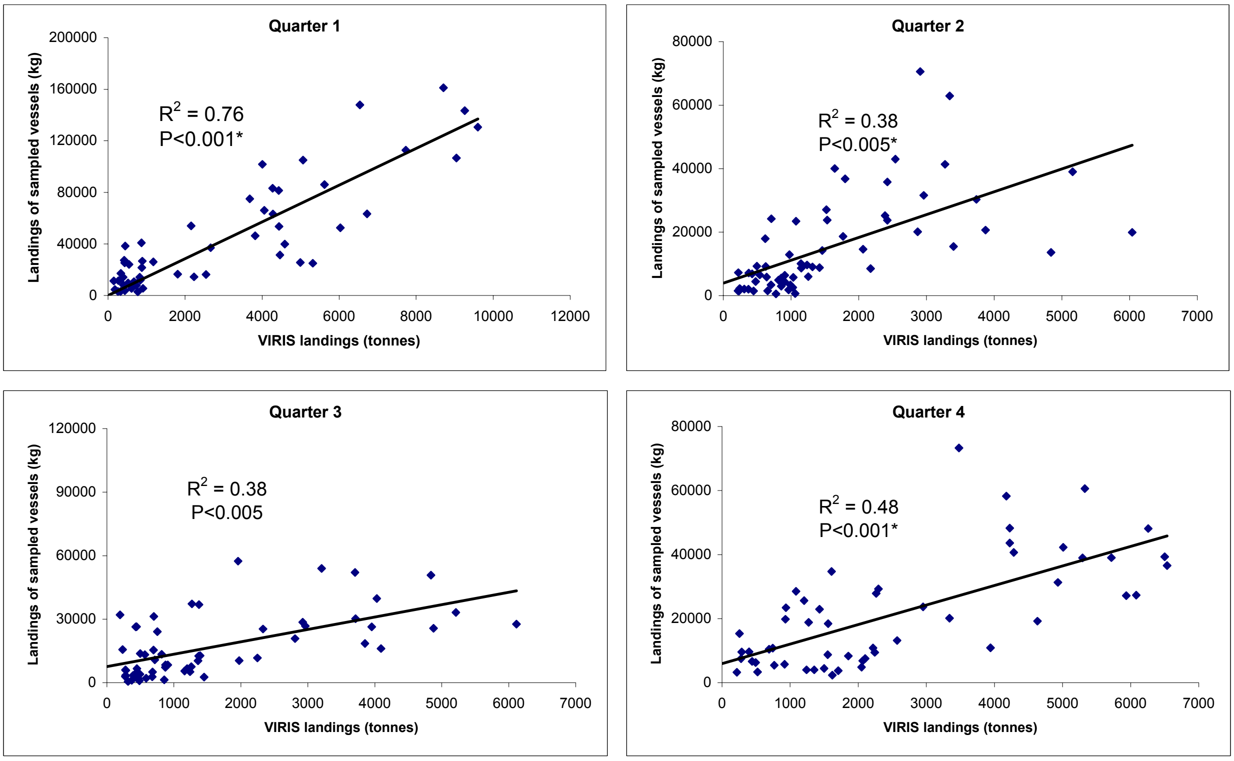


Figure 3.22 Comparison of the landed catch of plaice per ECOTOETS area of the southern North Sea to sampled weight of the area by the Netherlands Institute of Fisheries Research (1990- 2001) by quarter. Asterisk denotes significant correlation.

4. Model assumptions

The purpose of this section is to review whether any of the current assessment models are inappropriate or if any of the assumptions in the models are broken by the stocks and the fisheries. This section will also compare and contrast, where possible the results of different models on the same stocks. However, as will become apparent, the outputs of some of the models are very sensitive to the assumptions within the model (*e.g.* the relationship between surveys and catch in mackerel, section 4.3.5) and some models cannot be used on all stocks, hence only models relevant to a stock assessment will be compared.

4.1 Assumptions of models

The basic theory of fisheries assessment models was introduced in section 2. In this section the models XSA, ICA, TSA, SAD, AMCI and ISVPA will be briefly described and their assumptions noted in the context of the assessments of important species to the Netherlands. The use of these models on stocks of interest to the Netherlands are shown in the text table below:

Species	Stock	XSA	ICA	TSA	SAD	AMCI	ISVPA
Plaice	North Sea	**X**	z	#	#	z	z
Sole	North Sea	**X**	z	#	#	z	z
Cod	North Sea	**X**	z	#	#	z	z
Whiting	North Sea	**X**$		**X**$	#		z
Mackerel	NE Atlantic	#	**X**	#	#	x	x
Horse Mackerel	Western	#	#	#	**X**	#	#
Herring	North Sea	x	**X**	#	#	z	z

X- main model used for assessment x- method used as comparison in the WG
z- can be carried out for comparison #- cannot use as methods or assumptions are invalid
$- in 2002 TSA was used for whiting and this changed to XSA in 2003.
Note: XSA cannot be used on mackerel & horse mackerel, as the only tuning fleet is biomass. The SAD model was specifically created to cope with the strong 1982 year class in horse mackerel. The current version of TSA cannot use more than 1 fleet or survey or biomass indicator.

4.1.1 Assumption of separability

An important concept to introduce at this point is "*separability*". Most traditional VPA type models assume that the landings data are exact. No allowance is made for the fact that catches are measured with error. With this in mind the models compare the difference between each year and each cohort, e.g in table 2.1 there are 13 years of data for 9 age groups. This means that the model needs to estimate F between years and ages 12 x 8 = 96 times.

However if you assume that the data are approximations, you can model the processes. This will reduce the number of parameters that need to be estimated. You can assume that the selection pattern by the fishery, *i.e.* relative F at age, has remained unchanged, and that only the relative strength of F has changed between years. This is called the *separable* assumption. For example in table 2.1, it means that you only need to estimate 20 parameters: 8 for the relative F at age (the selection pattern) across the 9 age groups and 12 for the relative strength of F over the 13 years. This simplifies the statistical processes and also reduces the parameters to be estimated (20 compared to 96).

Models such as ICA, AMCI, SAD and ISVPA use various types of separability assumptions, and these models should only be used if it is likely that the selection pattern has not changed over the separable period.

4.1.2 Other assumptions within the models.

Assumptions about catch data: In most models, it is assumed that reported landings are representative of actual catches (although some TSA or Bayesian methods may not make this assumption), but in general the effects of discarding and misreporting are not explicitly considered. For many stocks the catch data are thought to be highly inaccurate and may become worse with collapsing stocks and stringent management measures. Poor catch data means a poor assessment. Also problems arise within the assessments when inaccurate catch data are used with accurate survey data. It is difficult to reconcile divergent signals from the two data sources, and it is likely that these contradictory signals are the cause of much of the retrospective bias[1] in the assessment time series. In effect, the assessment over years is moving from an estimate derived from false catch data and true survey trends to an estimate only based on false catch data within the converged part of the assessment (See figure 4.1.1, and section 2). Often it is thought that the retrospective bias may be a problem of the most recent years (the non-converged part) but it may be the converged part, based on poor catch data, that is the “dodgy” part of the assessment. Potential alternatives to catch-based assessment methods are briefly discussed in Section 4.4.

How the plus group is treated: Assumptions about the plus group[2] vary with virtually each model. The assumptions may be important in cases where considerable amounts of fish reach that stage, but they may also conceal unrealistic large or small numbers surviving to the oldest true age. Due to the way VPA uses the plus group, slight differences in the plus group in recent years can affect the entire time series during the period of convergence (see section 2), thus scaling the biomass estimates up or down.

Assumptions about catchability in surveys and CPUE: Most assessments are “tuned” or calibrated with surveys or commercial CPUE series. Hence the catchability (or selectivity) of the surveys must not show a trend over the time series[3]. Most models used at present cannot cope with trends in the catchability of surveys. Changes in catchability have also been suggested as a cause of retrospective bias. In the same way, CPUE series can show trends in their catchability. Restrictive management measures and effort restrictions can change the catchability of CPUE series, as can misreporting the catches or wide scale discarding of undersized fish. Incorrect assumptions about the dynamics of CPUE series led to the poor assessments that allowed the North Sea herring to collapse in the 1970s (Burd, 1985, Cushing, 1992). CPUE series must always be used carefully as many species remain catchable at low densities due to their shoaling behaviour, or attraction to certain prime habitats.

[1] Retrospective bias is the phenomenon that, starting from the most recent assessment, and then leaving out the last year’s data, and then the one-before-last year’s data etc., will lead to assessment results where the successive historical estimates deviate from each other in a systematic way, *e.g.* SSB tends to be overestimated compared to each successive assessment with one year’s data added.

[2] The plus group is the group of all fish of a certain age and older. For instance the 10+ group contains all fish of 10 years old and older. The oldest measured age may be much higher, *e.g.* 25 years old, but there may be very low numbers of fish that have that age.

[3] The assumption of no trend implies that over the time series a constant number of fish by age is caught with a given effort at a given abundance.

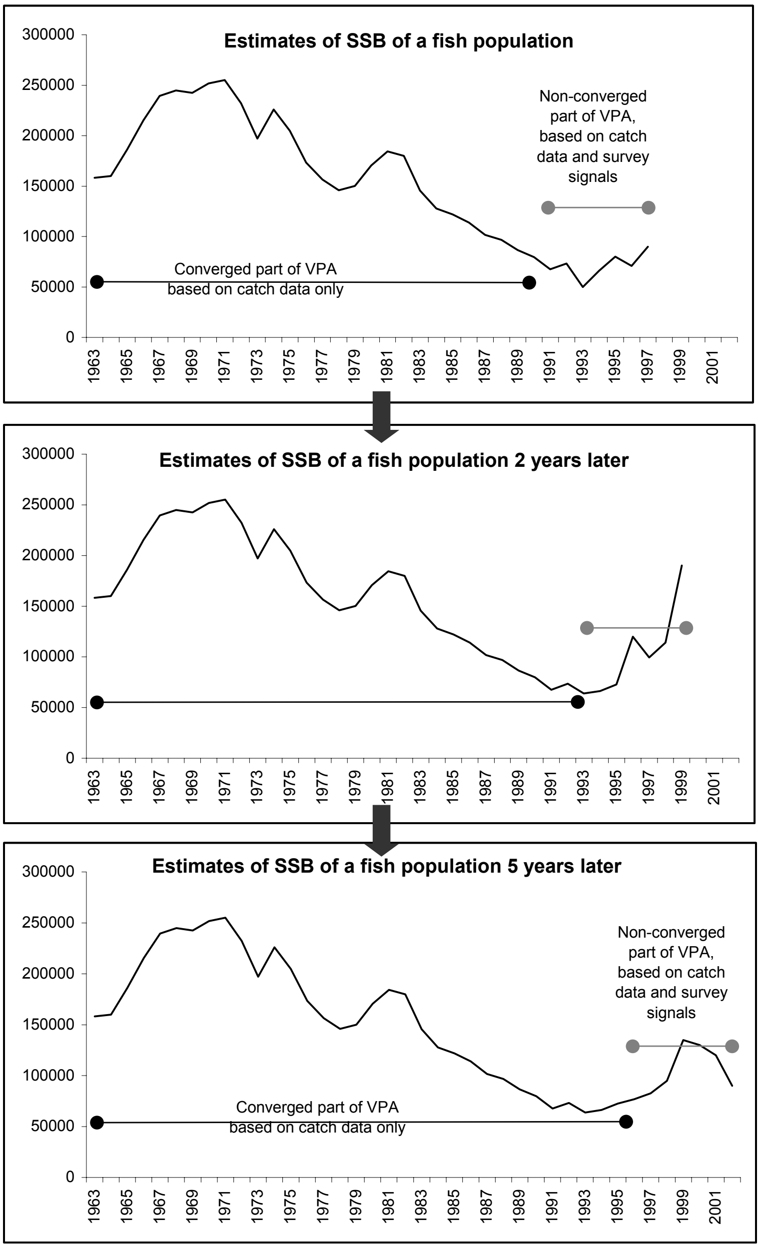


Figure 4.1.1. A hypothetical estimate of an SSB based on data from incorrect catch data and true survey data, over a 5 years period. Converged VPAs, generally use only catch data.

In addition, assumptions are often made about the underlying relationship between the abundance of fish and the tuning indices of CPUE or from surveys. Linearity is the most common assumption (*i.e.* that the survey estimates vary in a linear way with the assessment estimates of abundance), although many models now allow for non-linear relationships to be chosen. When biomass indicators are used, an assumption must be made as to whether the relationship to estimated biomass is either relative or actual (*i.e.* 1 to 1). In other words, after an egg or acoustic survey gives an SSB of 50,000 tonnes, should the assessment model also produce an estimate of 50,000 tonnes or just use the trend in the time series of surveys. For these reasons, as with catch data, surveys or CPUE series must be rigorously investigated and tested for bias or trends, and the underlying processes between survey results and the population must be understood.

Assumptions about natural mortality (M): There are many assumptions related to natural mortality rates. The most common is that natural mortality is static over years, and does not vary with population density or the environment. Very large fluctuations will have dramatic effects on the assessments of short-lived species and long-term trends in M will affect the quality of any stock assessment. Natural mortality can in general not be estimated reliably together with the fishing mortality from catch and survey data because the data can be explained almost equally well by a range of each of the mortalities, provided the other is adjusted accordingly.

All these assumptions listed above can affect the quality of the assessment. Many of them (*e.g.* natural mortality) also impact on the quality of projections into the future. In addition there are other assumptions that are now becoming less common but often occurred in earlier assessments (or early in existing time series) such as growth or maturity being constant between years, or stock weights being equal to catch weights. Most of these latter assumptions will affect the estimation of biomass rather than fishing mortality (unless biomass is within the objective function[1] of the model, *e.g.* in ICA). When a population or ecosystem is stable and productivity remains the same, these assumptions will not have a large impact. However if the productivity of the system changes then these assumptions will cause problems, particularly in the estimates of spawning biomass.

4.2 Description of the models

In this section the main models used on stocks of interest to the Netherlands will be described. Note that whilst the actual minimisation process of each model will be given, it will not be explained and is purely there for statistical completeness.

4.2.1 Extended Survivors Analysis (XSA)

This is the model used on North Sea cod, sole and plaice. It is based on the basic VPA methods and involves tuning of the recent estimates of fishing mortality (F) with estimates from surveys or commercial CPUE data, as described in section 2 (Darby & Flatman, 1994; Shepherd, 1999). It was developed from the method of Doubleday (1981). XSA differs from other VPA approaches by using estimates of survivors in the final year to be used in the minimisation process. Most other VPA models just used the numbers of fish on the 1 January of the final year within the analysis. Statistically it uses an iterative re-weighted least squares method.

The basic process is:

1. the abundance of survivors for each cohort are treated as the principle variables to be estimated.

2. estimate the population abundance for all other ages and years by VPA (Pope, 1972) using the estimated survivors as terminal populations.

3. calibrate the tuning indices (CPUE or surveys) using the abundance estimates of each tuning fleet and simple catchability models.

4. use the calibrated abundance indices, for all ages in each cohort as the basis for estimating survivors.

XSA assumes the catch data to be exact, and hence it does not make any assumptions about separability, thus allowing the selectivity pattern of the catch to change by year and age. Shepherd (1999) says that this assumption may be inappropriate where "*catch data are poorly sampled or otherwise defective, but where one or more sets of reliable survey data were nevertheless available*." It does treat the survey or CPUE fleets as separable and hence

[1] The objective function of an estimation model usually represents a quantity that has to be minimised, *e.g.* the sum of the squares of the differences between the observed and the estimated values (called least square minimisation), in order to derive at the values of parameters that fit the resulting model best to the data.

determines a selectivity curve for each fleet. It also allows the operator to shrink[1] the most recent estimate of F to a mean value of a chosen number of most recent years. This implies an assumption that the fishery has not changed greatly over the recent time period and gives greater stability to the assessment. However, the use of shrinking F to the recent mean may cause conflict with strong signals in the surveys, particularly if the catch data are inaccurate and thus a bias towards the mean is introduced (Shepherd, 1999).

The current version of XSA cannot deal with biomass indictors as a tuning index for the assessment. Hence egg or larvae surveys and biomass only acoustic surveys are of no use in an XSA assessment. This means that the ICES mackerel, horse mackerel and herring assessments cannot be fully replicated in XSA.

XSA calculates the variances of the survey fleets in relation to the catch data and uses them as weighting factors as an integral part of the process and thus it prevents weighting factors being applied manually during the assessment, if so required.

XSA has been heavily scrutinised by ICES and the International community (ICES 2003a). However the sensitivities of the model to extreme changes in catch or biomass have yet to be assessed. Simulation studies suggest that XSA behaves well, if the assumptions used to create the simulated population are similar to the assumptions underlying XSA (Kell pers comm). Simulation studies have yet to be carried out on the behaviour of XSA under severe catch limitations.

4.2.2 Integrated Catch at age (ICA)

This is the basic model used on all the herring stocks and NE Atlantic mackerel. In its most basic form it is a separable model for the most recent part of the time series attached to a traditional VPA for the more distant past in the time series. As a separable model, it accepts that the measurements of the catch are not exact and the reduction in the number of parameters estimated improves the precision (ICES 2003a). Unlike XSA it also gives outputs of the uncertainty in F, due to the fit of the model. Its major weakness is of course the separability assumption, *i.e.* assuming that the exploitation pattern (relative F at age) of the fishery on the fish has not changed during the period of separability[2] (Patterson, 1998). It does allow for two periods of separable constraint to exist, with either an abrupt or gradual change between them.

The minimisation process also requires that selection (relative F) on two ages within the catch at age matrix are fixed and predetermined. One age (usually the first at full exploitation and called the reference age) is set at an exploitation of 1 and the selection at the final true age is then set relative to that value (Patterson, 1998). This results in part of the exploitation pattern being assumed prior to the assessment and hence it is very important when using ICA, that the model fit is widely explored. The selection of the plus group is assumed to be equal to that of the last true age.

---

[2] Shrinkage implies that the F in the most recent year is constrained not to deviate far from the average over the past several years. This, therefore, represents the assumption that F did not change much compared to previous years. The operator can determine the strength of the shrinkage.
[2] The operator can choose the length of the separable period, *i.e.* the period for which separability is assumed.

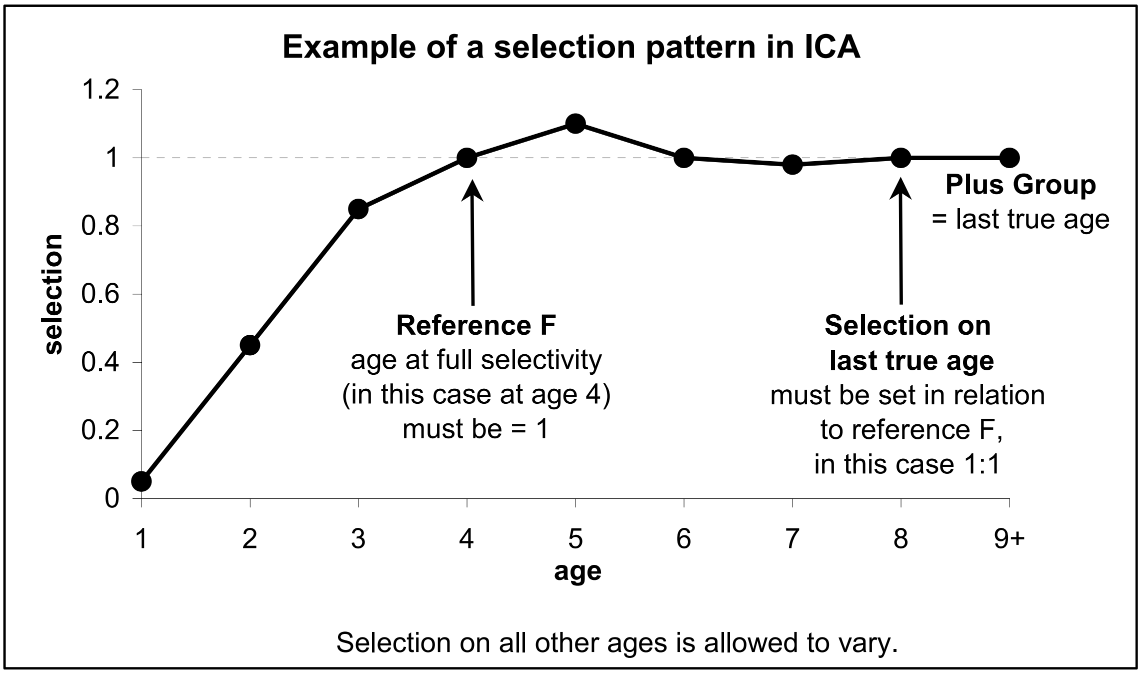


Figure 4.2.1. Hypothetical selection pattern in ICA, showing that the selection on the reference age and the last true age in the catch data are set within the assumptions of the model. The ratio of the two can be varied at initiation of the model to give smooth or dome shaped exploitations etc.

The model is based on the work of Deriso *et al.* (1985) and Patterson and Melvin (1996), with additional work by Kimura (1986) and Gudmundsson (1986). The model is fitted in two stages. Firstly a simple iterative procedure (non-linear least squares minimisation) is used to find a conventional separable VPA that fits best to the available survey data. Subsequently a more complicated multidimensional minimisation can be carried out that involves re-weighting the parameters and output functions. Most stock assessments do not use this second stage.

Importantly, Patterson (1998) comments that the underlying assumption of uncorrelated errors in the catch is unrealistic (a common assumption in almost all models), but suggests that there is no alternative at present. Most catch at age matrices will have correlative errors due to age effects, strong cohort effects or other effects.

ICA allows for a choice of relationship between the surveys and the catch at age data. The relationship can be either absolute, linear or a power model (see section 4.1). ICA differs from XSA by the surveys being within the objective function of the model. The entire time series of surveys is used to estimate their relationship to the VPA numbers (this relationship is called the catchability coefficient, *q*) and then this catchability coefficient is used in the separable part of the model to determine terminal Fs.

Apart from offering the re-weighting of the survey and catch data within the model (the second stage described above and not often used), ICA offers the operator the opportunity to weight the importance of the input data manually. This is usually used, and has been very useful in the assessment of North Sea herring, where variances of the surveys were investigated in an independent project (see section 3.2 on EVARES). It is important to remember that although the catches are not considered to be exact, the default weighting procedure results in a heavy weighting of the catch over the survey data. This is due to each age in the catch at age matrix getting a weighting of 1, compared to each survey getting a weighting of 1. So a catch matrix of 9 ages over 25 years would have a weight of 225, an acoustic survey with ages of fish would have a weight of 1 and a biomass survey (*e.g.* of larvae) would also have a weight of 1. These weightings can be changed but are rarely adjusted to any large degree, hence the catch at age data still dominate the assessment.

Like XSA, ICA also gives the option of shrinking the recent estimates of F to the mean of a selection of years, but unlike XSA this is an extra function and not inherent within the model. Applying a weak level of shrinkage to the F in XSA is advised to maintain the stability of the model, but this is not necessary in ICA.

There is a worry with many separable models including ICA (ICES 2003a), which is due to the separable pattern being projected backward through the converged part of the assessment via the plus group and older fish. This is one of the properties of these types of model.

ICA has been widely tested and assessed by ICES and the international community (ICES 2003a). As with XSA, if the assumptions are not violated, it can be a strong and useful tool for the assessment. However, simulation studies have yet to be carried on the behaviour of ICA under severe catch limitations, such as halving the quota or closing areas.

4.2.3 Time Series Analysis (TSA)

Very few investigators in the ICES community can work with TSA, and most are based in Aberdeen, Scotland. It is used on North Sea whiting and west of Scotland haddock and cod. It is a method based on the science of studying the flight of rockets (rocket science) and uses techniques that follow trajectories (Gudmundsson, 1994). The main statistic technique is called the Kalman filter (Fryer *et al* 1998). It works on the idea that the trajectory in the near past will probably reflect the trajectory of the near future. The filter acts as a time-series smoother, and is highly conservative. Unlike most other models, the residuals from a TSA model fit should not be random, but should show trends, reflecting the fluctuations away from the trajectories. It is so conservative that recently it failed to show the cessation of fishing on cod in the NAFO 3Ps area, despite the closure of the fishery (Needle, 2001). However this conservative nature may be very useful, if stringent management measures are not properly implemented or accepted by the industry, *e.g.* whilst official landings for a year may reflect an unpopular cut in quota of 40%, this may not be the true picture as fleets fail to fully implement the drastic change, and TSA will also not show the drastic change.

TSA is a powerful tool for those who understand it and the programming of its code. It is highly flexible and can be adjusted to many situations (Fryer, 2002), *e.g.* it can be run with and without survey information. The weakness of TSA is that the operators are very few in number and hence there is little transparency in its use. At present it is poorly described and documented, and although already used for some assessments (Needle and Fryer, 2002), few people can replicate the work. With these weaknesses it is difficult to comment further on the assumptions of the model and its utility, although it can replicate the results of XSA if similar assumptions are made (ICES 2003b).

4.2.4 Combined Separable VPA /ADAPT (SAD)

Throughout the last ten years, the mackerel and horse mackerel working group has had problems finding an assessment model that is suitable for the unusual characteristics of the western horse mackerel stock (ICES 2001a, ICES 2002a). Any assessment model constructed for the western horse mackerel should take into account the special characteristics of the catch at age data set. The stock and catch was dominated by a series of strong cohorts, *i.e.* the extremely strong 1982 year class and the also strong but much less abundant 1987 year class. In addition, the selection pattern changed greatly over the last decade (Eltink pers comm.). The only fishery independent information available for calibration of the population model is a time-series of egg survey estimates of spawning biomass (hence XSA and TSA cannot be used). As no age-disaggregated information was available, an assumption is required that selection at age is constant between years, for years to which a separable model is fitted. This assumption is valid for recent years in which there are no dominant cohorts but not for those years when the 1982 year class dominated the catch.

This need for different structural models for recent and historic periods has been met by the development of the SAD assessment model. The SAD model applies a separable model to recent data and then links to a VPA transformation of historic catch, as was described for the ICA model. It is developed from the ICA model developed by Patterson and Melvin. (1996), see section 4.2.2, and the ADAPT model which is primarily used in North America (and not described in detail in this document). The separable model is fitted to the shortest period possible (3 years), which is the most recent part of the time series. The separable model estimates from the earliest year of the separable period to initiate a historic VPA for the cohorts exploited in that year.

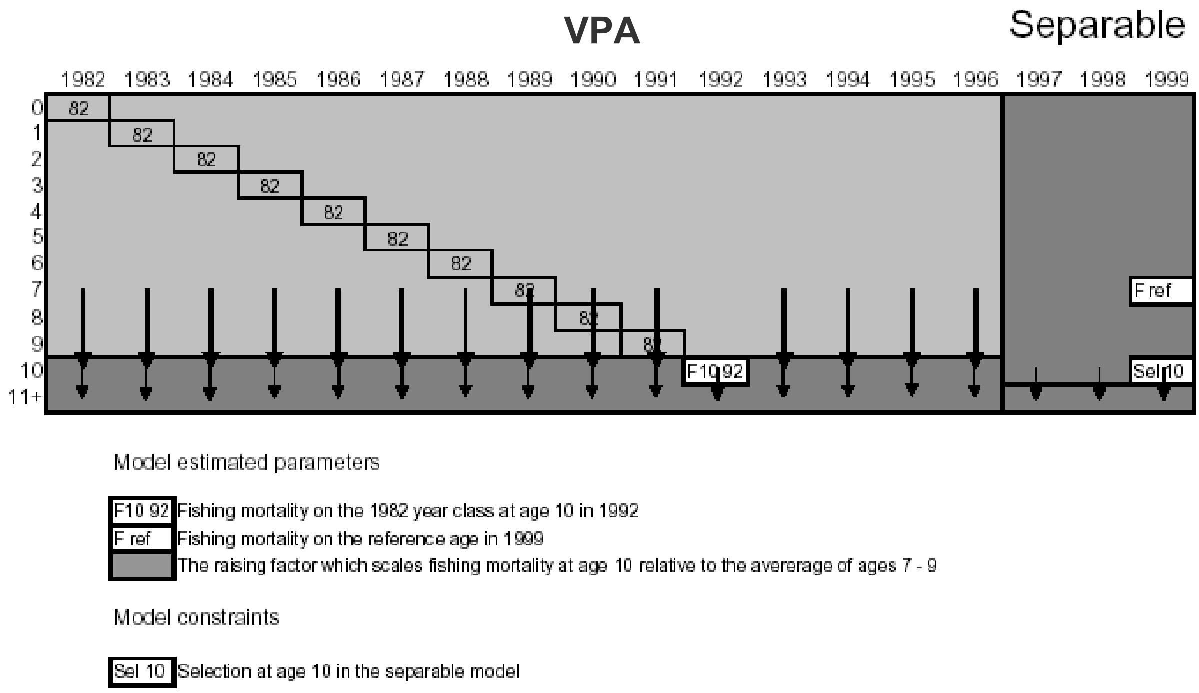


Figure 4.2.2. A conceptual version of the SAD model, designed specifically for the assessment western horse mackerel (ICES 2001a).

Another effect of the very strong year class is the impact on the plus group. Generally, fishing mortality on the plus group is taken to be equal to that on the oldest true age. However, this is incorrect for years after 1992, as the fishery was still directed at the 1982 year class, which was in the plus group. In order to allow for the directed fishing of the dominant 1982 year class, fishing mortality on this year class at age 10 in 1992 was estimated as a parameter within the model. Estimation is again by non-linear minimisation of the sum of squares, which will not be explained further.

The parameters estimated are:

1) Fishing mortality on the reference age for the separable model (age 7), see figure 4.2.1.
2) The scaling of the fishing mortality for age 10 and the plus group relative to the average of ages 7 - 9
3) Fishing mortality on the 1982 year class at age 10 in 1992 and the corresponding plus group in later years

Maturity at age and stock weights at age and the proportions of F and M before spawning, are assumed to be known precisely, as are the catch data in the VPA part of the analysis. The egg survey SSB estimates are considered to be absolute measures of stock abundance. The development of SAD did not deal with the depressed growth caused by the 1982 year class.

Despite all this extra development, the SAD model is still not widely accepted for the assessment of horse mackerel. It has problems reaching a solution and may be over parameterised (ICES 2003a). ACFM rejected the most recent SAD assessment of horse mackerel because of the sensitivity of the historic biomass estimate to the length of the separable period; a change in separable period from 4 to 5 years gave large differences in the historic VPA converged part of the assessment. This was due to the very different selectivities that were produced by the 4 or 5 year separable period (Eltink pers comm.).

The SAD model is a good example of a model that was developed specifically for one stock and its associated problems. This approach is along the lines of AMCI (see below) and the North American ADAPT methods. It clearly has strengths, but does require an even greater degree of quality assurance when utilising the model. ICA was developed with specific problems in mind and now appears to be quite a robust tool, but new methods must always be tested and are often not suited for any of purpose other than that for which they were developed.

4.2.5 Assessment model combining information from various sources (AMCI)

AMCI was designed to offer flexibility and cope with data coming from different sources and of different quality. The programme was designed to allow for exploration of the data and exploration of combining techniques, as well as stock assessments. AMCI is a separable assessment model applied to the whole period, but the selectivity pattern can be allowed to slowly change. In the extreme, this means that AMCI can be made to respond as a non-separable VPA type model. The specification of the selectivity is highly flexible and uses a technique of “gain” similar to the Kalman filter of TSA. The data exploration and assessment ability of AMCI is further enhanced by its ability to project forward in time, like a traditional short term prediction. The catchabilities of the surveys can also be allowed to change over time, in a similar way to the selectivity of the catch at age data. Unlike XSA and ICA, the flexibility of the model allows the parameters to be weighted in a manner that the catch data do not dominate.

The developer of AMCI (Dankert Skagen pers comm, Marine Institute, Bergen) describes AMCI as less of a “statistical model” than the other catch at age based methods . The models that appear in AMCI are basically functional relationships between model components and between model and observations. This approach was used in recognition that the statistical properties of the data are generally complex and poorly known, and to avoid making the final outcome of the assessment conditional on strong statistical hypothesis.

The main building blocks in AMCI are:

1. A population model which projects the population forwards in time according to specified parameters.

2. Observation models that generate modelled counterparts to observed data.

3. Objective functions that measure the fit of the modelled data to the observations. There is a choice of functions, but the most common is the weighted sum of squared log residuals.

4. Optimisation routines that find the parameter sets that give the best fit of the model to the data

5. Supporting routines for model specification, data handling, diagnostics, organising bootstrap, retrospective runs, printout etc.

Additional properties are that fish tagging data can be used, and that there are area and fleet components to the assessments. Weighting between variables is decided by the user although there is some internal implicit weighting, which implies that the weights in AMCI are not directly comparable with those in ICA.

AMCI is not widely used at present although comparative runs for mackerel are produced by the assessment working group. AMCI is used to assess Atlanto-Scandian herring and is being used to explore the population dynamics of blue whiting (ICES 2002b). It is relatively new, and as such any major draw backs of the method have yet to be documented and its assumptions fully tested.

4.2.6 Instantaneous Separable VPA (ISVPA)

This model has been, and is, being developed in Russia (Vasilyev, 2001). It is a separable model that covers the entire assessment period. One separable model can be applied to the whole period, or two different separable models can be applied to two periods, as determined by the user. It is totally different than AMCI, in that it is based on statistical assumptions and approaches, and it searches for model solutions by changing the underlying assumptions. It is not widely used as yet, and has been proposed for the assessment of blue whiting and Atlanto Scandian herring (ICES 2002b). It has proved a useful tool whilst examining the population abundances of herring off Ireland (ICES 2003c). No stock assessment by ISVPA has been accepted by ACFM as yet.

As with ICA, certain assumptions need to be made about the selection patterns. The selection at the oldest age is equal to that of the previous age; selections are normalized in such a way that their sum across all ages adds up to 1. It does not, however require a reference age, or require assumptions about the shape of the selectivity pattern. As with other separable models the catchabilities of the survey fleets or biomass estimates are estimated within the minimisation of the objective function of the catch data. The plus group is not modelled, but the abundance is derived from the catch assuming the same mortality as for the oldest true age.

The difference between ISVPA and the other models discussed so far is the wide choice of objective functions that can be used to carry out the analysis. There are 4 possible statistical methods to find a model solution for the catch data and survey data. The method to find a model solution for the SSB surveys is more simple. The different functions will not be explained here.

ISVPA offers variance estimates of F and biomass from within the model and also provides the opportunity to bootstrap the populations to obtain ranges on the modelled population, using the determined confidence intervals (See section 2 for a description of the bootstrapping technique).

ISVPA offers three different ways to deal with errors in the catch data, one is to be a pure separable model where the catch data are not exact (the effort-controlled version), one acts as a virtual VPA where no errors are assumed (the catch-controlled version) and one offers a mix of the two (the mixed version). The catch controlled version still uses one selectivity pattern and a separable model to arrive at the terminal fishing mortalities. As mentioned above the approach of ISVPA is highly statistical, and with that in mind the model also offers methods and choices for constraining residuals based on catch, mortalities or selection (the explanation of these is not relevant to this context).

This package is still in development and recent experience within RIVO has shown that the minimisation process is still unstable. However the latest version which only arrived at RIVO in late February 2004, may improve this as it allows surveys to be incorporated. In addition some of the stocks for which ISVPA has been proposed as an assessment tool appear to have highly variable selectivity patterns that break some of the assumptions about separability. In a recent study it was shown that ISPVA has problems when selectivity patterns change, even when two periods of separable constraint are used (Eltink 2004). There might also be a small bias in the estimated selectivity patterns.

4.3 Application of Models in current assessments

The effects of the use of certain models on the perception of the stocks of importance will be reviewed in this section. Many of the models cannot be applied to all the stocks (see the text table at the start of section 4.1, *e.g.* the use of biomass indices rules out the use of XSA for most pelagic stocks and SAD was specifically designed for western horse mackerel). Other assessment packages are still being developed (AMCI), or are still poorly understood (TSA). The version of ISVPA that is globally available only uses catch data and is highly unstable and RIVO only received the new version that allows for survey fleets to be incorporated into the assessment in late February. So this section will not comprehensively apply each model to each set of data, but review whether the assessment model of choice (Table 4.3.1) for each stock is relevant and suitable for the assessment.

4.3.1 North Sea Cod

North Sea cod is usually assessed by XSA (Table 4.3.1). The current reductions in effort and management measures mean that it is highly likely that the selectivity pattern has changed greatly in recent years. That, and the decline in abundance of older fish, means that separable models are of limited use for this stock. It is possible that ISVPA and AMCI could be used with their settings adjusted to reduce the separability assumption (reduce the gain in AMCI to allow selection to vary greatly, or use the catch controlled setting for errors in ISVPA, although the assumption that the catch data is exact may not hold for North Sea cod). TSA is used for other cod stocks (*e.g.* west of Scotland), and a preliminary assessment by TSA was carried out for North Sea cod in 2003 (ICES, 2003b; Figure 4.3.1). This shows a very similar trend to the XSA assessment and gives a final SSB similar to the XSA value (Figure 4.3.2). Even if the catch data from the last two years are removed and the TSA is survey driven for those years, the estimates remain fairly stable (Figure 4.3.1).

It is clear that many of the surveys in the assessment give similar signals (Figure 4.3.2) as most of the estimates of SSB and F are well within the confidence limits of an average stock assessment. However the cod assessment has a range of problems; the two main ones being

1. the change in the perception of the state of the stock after new data are added to the time series each year. This has been occurring throughout the 1990s. In fisheries science terms, this is called "showing a strong retrospective bias" and is described section 4.1.2. There are many possible reasons for this, including changes in the catchability of the surveys, but a clear culprit is the divergent signals coming from the survey trends and catch data trends (see section 4.1.2) as the survey-tuned part of the time series tries to reconcile itself with the converged part of the VPA time series.

2. any fish stock that is undergoing new or extreme management measures or changes in the behaviour of its fisheries is very difficult to assess. As explained above, fishing mortalities in the most recent years are the most difficult to estimate. Changes in the selection of the fishery during this period will make the "shrinkage option" in XSA (that adds stability) an inappropriate choice within the model. Add to this the uncertainty in the catch, due to strong suspicions of misreporting, in the most recent years and the actual estimate of SSB is very uncertain (ACFM, 2003). These problems are made worse at low population numbers and reduced age profiles (*i.e.*fewer older fish than

previous, reducing the average age). This results in both the population and the assessment of that population becoming less robust.

The choice of XSA for North Sea cod is probably the most appropriate at the moment. However AMCI and ISVPA should be considered as supplementary tools in the future. They were not used under this project because the models are still being developed and quality tested.

4.3.2 North Sea Sole

Sole is assessed with XSA. There are many characteristics of sole that make the stock easier to assess than plaice or cod. Sole has a wider age range in the population and there is less discarding compared to plaice. The fishery catching most of the sole actively targets this species. Moreover, selectivity has not changed much in recent years. It is one of the few stocks that is still assessed using CPUE data and XSA was specifically designed with CPUE data in mind. The strong year classes that are seen in the catch are also seen in the survey data (ICES 2003b). With all these considerations, XSA is the most appropriate assessment model to use on North Sea sole.

4.3.3 North Sea Plaice

Anybody up to date with fisheries assessments in the Netherlands should be aware of the problems in the recent plaice assessment. The problems have been investigated in detail by the assessment WG (ICES 2003b) and in the recent quick scan report for LNV (Dickey-Collas *et al.*, 2003) and will be further examined in the F project. Plaice in the southern North Sea suffers a very high mortality rate on the younger ages due to discarding, and the traditional stock assessments do not take discarding into account (ACFM, 2003). Constant discarding rates are not that problematic, but large variation or a strong trend in rates can cause huge problems for the quality of the associated assessment (Keeken, *et al*, 2003; 2004; Kraak *et al.* 2002).

The assessment is considered to be uncertain (ACFM, 2003). Sensitivity analysis demonstrated that the perception of stock status is sensitive to the inclusion of simulated discards, notably when there is an increase in discard rates (Kraak *et al.* 2002). This has caused problems with the estimation of year class strengths. Year classes appear in the surveys that then appear less abundant in the catch. This causes large problems with the retrospective patterns in the assessments and the between-year constancy in the perception of the state of the stock. The main assessment tool for North Sea plaice is XSA. This responds well to changes in the selectivity, but as with cod, changes in management measures make the assessment more uncertain and make the assumption of using shrinkage in F (pulling F to the recent mean) less valid. XSA is very sensitive to the number of ages used for F shrinkage (ICES 2003a).

With regard to model choice a recent comparison of XSA and ICA on North Sea plaice showed that both models gave similar results (Figure 4.3.3). The perception of the stock, without catch data for discards, is very similar with both methods. Issues about the assumption of separable constraint or reference ages cannot be fully investigated until a robust time series of total catch is available (landings + discards).

As a virtual bycatch fishery, the assessment of plaice must be considered within the context of sole. The selectivity patterns on the two species are likely to be highly linked. This would not be the case if discards are ignored, however the selectivity of the total catch of plaice should be related to the selectivity of sole (Dickey-Collas *et al.*, 2003). This will be further investigated within the fourth project of bestek 6c, “alternatieve beleidsmaatregelen”.

4.3.4 Herring

ICA has been used for at least the last eight years for the assessment of North Sea herring (Table 4.3.1). Being tuned by three surveys that target different ages (MIK net survey on 0 groups, the trawl surveys on younger age groups and acoustic surveys on older age groups) and a biomass index (the larvae survey); it would be best to use ICA, AMCI or ISVPA as the assessment tool. The selection pattern is assumed to be stable. The EVARES project showed that the larvae survey was important in the assessment and estimation of SSB (EVARES, 2003; Simmonds, 2003). No TSA has been developed yet that uses a biomass indicator. Independent scientists recently reviewed the assessment of North Sea herring and none of the basic assumptions within the model were questioned (North Sea Commission, 2003).

The recent working group (ICES 2003c) compared the performance if ICA with XSA. To make the models comparable, both ICA and XSA were run without the larvae survey data. The approach used was to choose XSA settings that reflect as many of the assumptions of the ICA model of North Sea herring. The shrinkage of F in XSA was set very low and also quite high. XSA gave very similar results to ICA. However, the retrospective bias in XSA was slightly smaller. The higher shrinkage of F resulted in a slightly lower estimate of SSB (Figure 4.3.4).

4.3.5 North east Atlantic mackerel

Unlike all the stocks above, there are no age based survey data for mackerel. This means that only ICA, ISVPA and AMCI can be used to assess the stock. The egg surveys occur every three years and provide information on both mackerel and horse mackerel egg abundances.

The main and obvious assumption in the assessment is that the egg surveys give an absolute estimate of SSB. Normally only the trend in an ichthyoplankton survey is used to calibrate the SSB trend from the catch (using the catchability coefficient *q*[1]), but the use of a biomass index as absolute is fairly radical (see section 4.1.2). This means that the assessment models must "pull" the signal from the catch data up to the level of the egg survey estimates. The use of an absolute relationship was proposed as a solution to the instability in q caused by the very few data points in the time series of surveys. In the model, the assumption of an absolute relationship for a survey makes $q=1$. In the mackerel assessment, the use of an absolute relationship has an impact on the assessment and increases the estimates of SSB substantially (Figure 4.3.5). The estimate for SSB in 2002 using a relative egg index is 50% of the estimate with the survey as absolute.

However the choice of model has little impact on the assessment. If both AMCI and ISVPA assume an absolute relationship between the modelled SSB and the egg surveys, they give very similar estimates in SSB to the ICA model (Figure 4.3.6). It is very difficult to decide which of the ISVPA statistical approaches is correct; therefore, a range of them is given in figure 4.3.6. Hence, the choice of model and the assumptions about selectivity are not important issues in this assessment compared to the choice of the survey index being absolute or relative.

4.3.6 Western Horse mackerel

As stated above, western Horse mackerel has been difficult to assess (section 4.2.4). In the last ten years many different models have been used (from ADAPT, the North American package, to Bayesian methods and SAD, section 4.2.4). This year ACFM did not accept the SAD assessment, due to worries about the selectivity affecting the converged part of the time series (ACFM 2003) (see section 4.2.2.).

Like mackerel, there are no age-based surveys to calibrate the assessment. The only index available is the horse mackerel egg survey. However, poor understanding of the fecundity of horse mackerel simplifies the issue of the index being a relative or an absolute estimate of SSB (Abaunza *et al*, 2003), as the egg survey cannot be turned into an absolute estimate of SSB so it must be used as a relative index of egg abundance alone.

---

1 q is the catchability coefficient, which is the ratio between survey estimate and VPA estimate of numbers at age or SSB.

The assumptions about the selectivity of the horse mackerel have complicated the assessment. However, different assessment models can give similar perceptions of the state of the stock (Figure 4.3.7). The ISVPA and SAD model broadly agree, although the ADAPT suggests a much higher biomass. Recent work on the assumptions of ISVPA about selectivity in relation to horse mackerel suggest that ISVPA cannot cope well with changes in a fishery as seen on horse mackerel recently (Eltink, 2004) but this has yet to be rigorously tested with the new version of ISVPA.

4.4 Discussion

The models used by ICES for the assessment of fish stocks of commercial importance to the Netherlands are all based on cohort analysis, with either survivors analysis or an assumption of separable constraint. These are all driven by the catch data, with different ways to incorporate survey or fleet information to estimate terminal values of F. The newer techniques, ISVPA and AMCI, must be critically reviewed and rigorously tested prior to their use in an assessment context. Simulated populations are an appropriate way to test these models.

The models used to assess most of the stocks are justifiable and few make assumptions that cannot be supported. Horse mackerel is an obvious problem, but work is ongoing to rectify the need for a novel and yet robust method. As stated above, a model is only as good as the appropriateness of the assumptions and quality of data that it uses. These catch-based methods are susceptible to criticism due to unacknowledged misreporting of the international catch, either by area or by biomass. In the same way unaccounted discarding also causes problems. As yet assessment models that only use survey data, such as SURBA, which is under development (Needle pers comm., Aberdeen, Scotland), are not used in the assessment process. These survey-only methods give quite robust estimates of F but at present cannot give a “realistic” estimate of biomass. SURBA can only estimate biomass as a relative biomass series.

It is clear that assumptions about shrinkage of F, the selection pattern, the length of a separable period and the underlying relationship of survey estimates to catch estimates can impact on stock assessments. This is why the assessment working groups carry out exploratory analyses prior to the main assessment and scrutinise the diagnostics of the assessment. No model can be used “off the shelf” and at present there is no generic model that can address all the problems in all the stocks. The sensitivity of some of the models to their assumptions does raise the need for a high standard of review and clear transparency in the assessment process. However as shown by this report, explaining many of the methods and techniques can be a long and complex procedure and should be carried out with care.

There must be a distinction made between good stock assessments and good management. Just because a stock has a robust assessment, does not necessarily mean that it will be exploited in a sustainable way (Rice & Cooper, 2003). There are many factors that can “muddy the water” between assessment and management, one of which is the quality of the projections.

Name	Code	Species	Stock	ICES Area	Assessment Method	Surveys	CPUE	Ages	Assessment Years	Management Agreement
Plaice	PLE	*Pleuronectes platessa*	North Sea	IV	XSA	2	0	1-10+	1957 to date	EU & Norway, 1999
Sole	SOL	*Solea solea*	North Sea	IV	XSA	2	2	1-10+$	1957 to date	-
Cod	COD	*Gadus morhua*	North Sea, Eastern Channel & Skagerrak	IV, VIId, IIIa	XSA	3	0	1-7+$	1963 to date	EU & Norway, 1999
Whiting	WHG	*Merlangius merlangus*	North Sea & Eastern Channel	IV, VIId	TSA	0	0	1-8+	1980 to date	-
Mackerel	MAC	*Scomber scombrus*	North East Atlantic	NE Atlantic	ICA	1	0	0-15+	1972 to date	EU, Faroes & Norway, 1999
Horse Mackerel	HOM	*Trachurus trachurus*	Western	IIa, IVa, Vb, VIa, VII, VIII	SAD	1	0	0-15+	1982 to date	-
Herring	HER	*Clupea harengus*	Autumn Spawners North Sea	IV, VIId, IIIa	ICA	4	0	0-9+	1960 to date	EU & Norway, 2001

* The horse mackerel assessment was rejected by ACFM in 2003.

$ In 2003 the plus group of sole and cod were reduced from 15 and 11 to 10 and 7 respectively.

Table 4.3.1. Comparison of stock asses and data input for seven stock of i Netherlands.

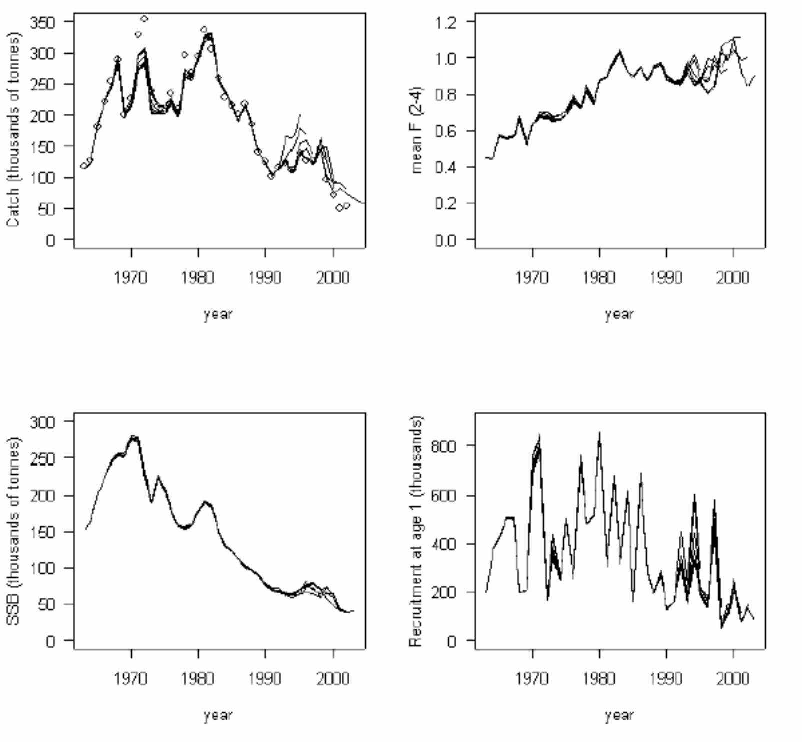


Figure 4.3.1. North Sea Cod: TSA Retrospectives with two years of catch data removed. The longest lines in each plot are the results from a TSA run using survey data to 2003 but catch data only to 2000, the next longest line is results from a TSA run using survey data to 2002 but catch data only to 1999, etc.

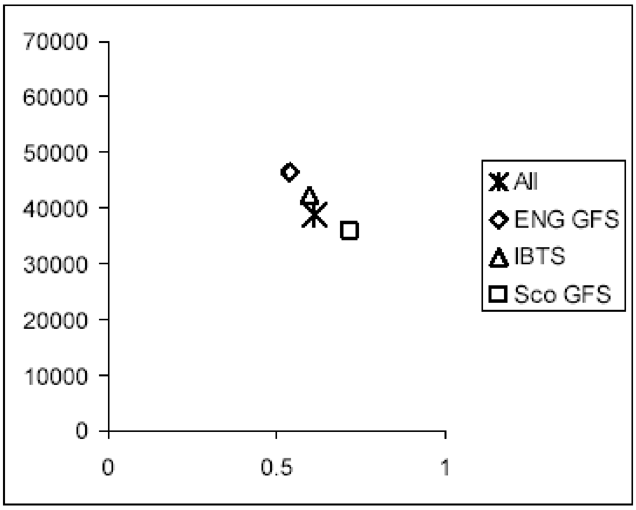


Figure 4.3.2. North Sea Cod: Spawning stock biomass and average fishing mortality at ages 2 – 4 estimated from XSA models fitted to each of the individual survey series with low shrinkage weight and the final combined assessment model.

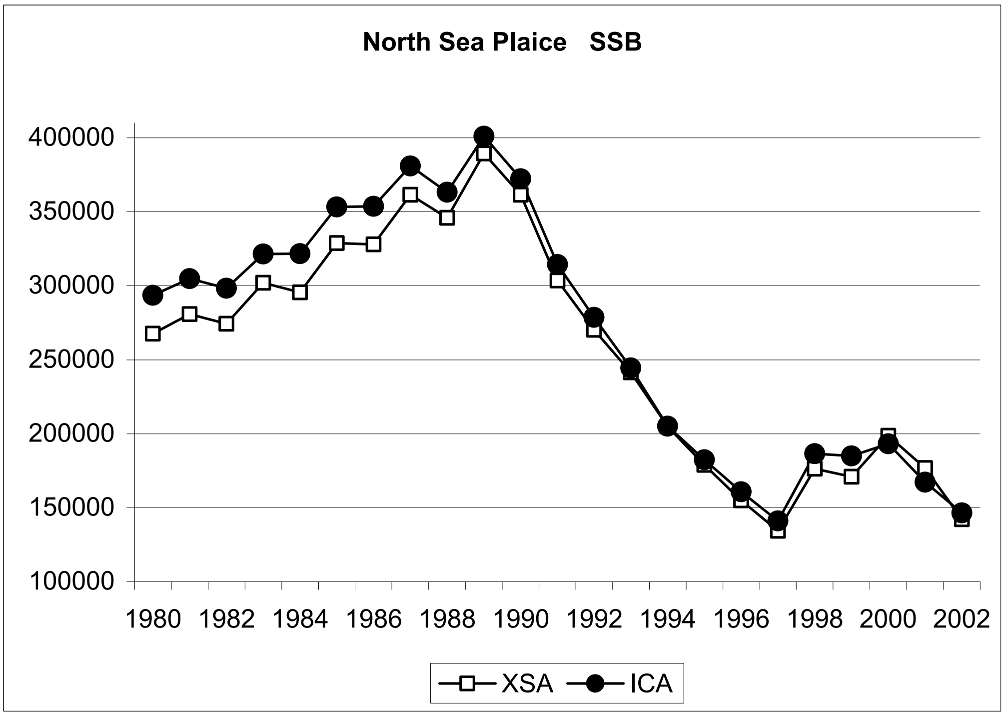


Figure 4.3.3. North Sea plaice: Comparison of XSA and ICA assessment of North Sea plaice (1980-2003 time series). Discard data not included.

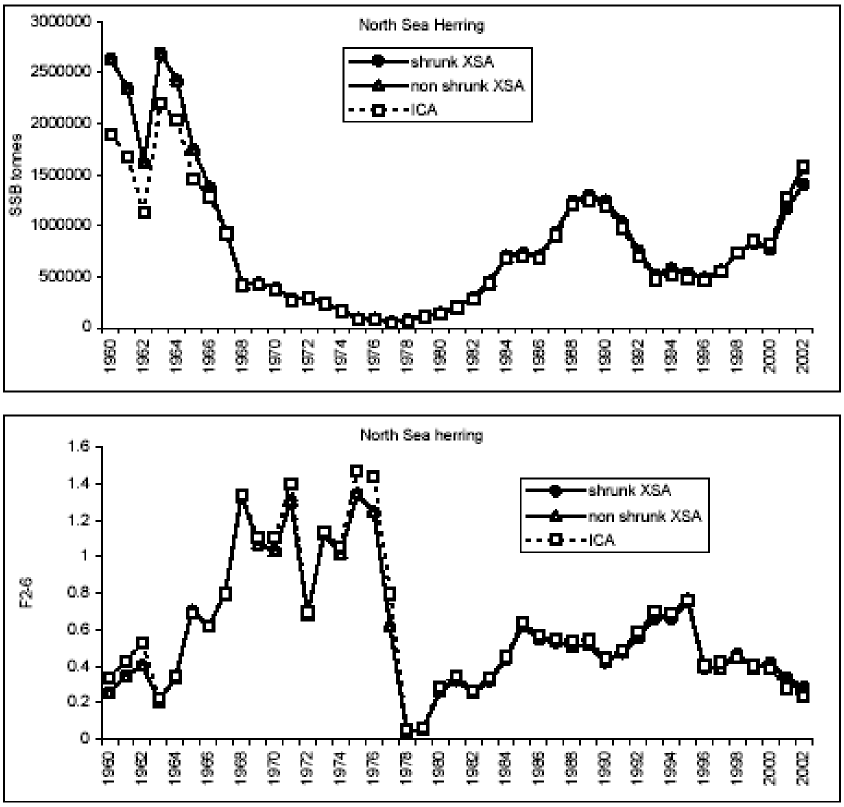


Figure 4.3.4. North Sea herring. Comparison of XSA and ICA assessments. XSA settings as similar as possible to the ICES ICA assessment (ICES 2003c).

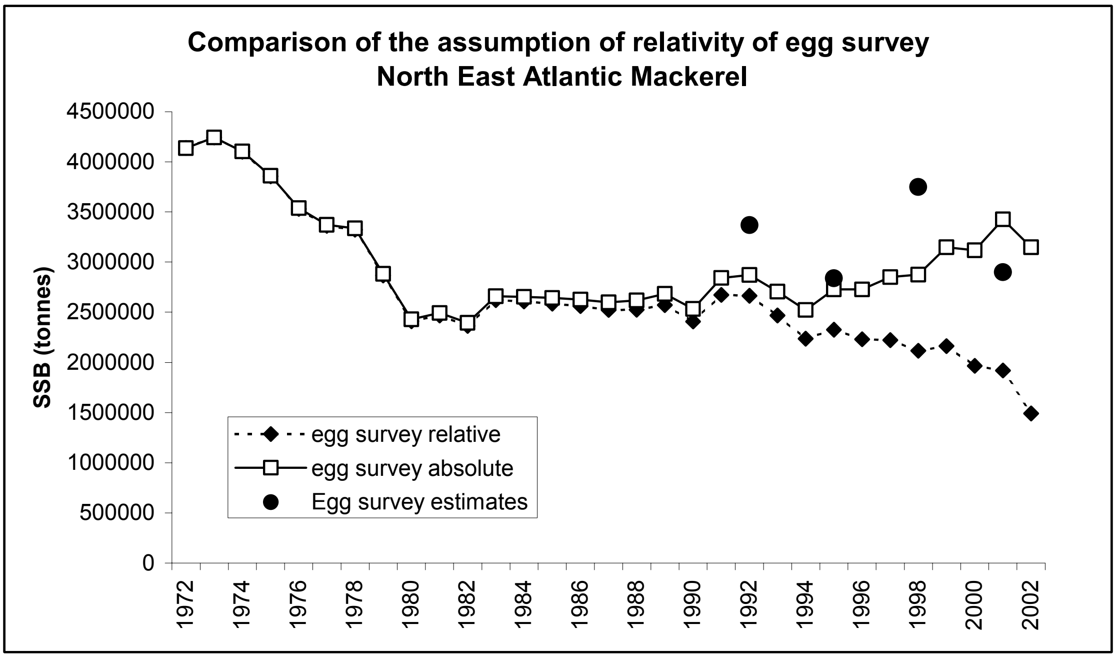
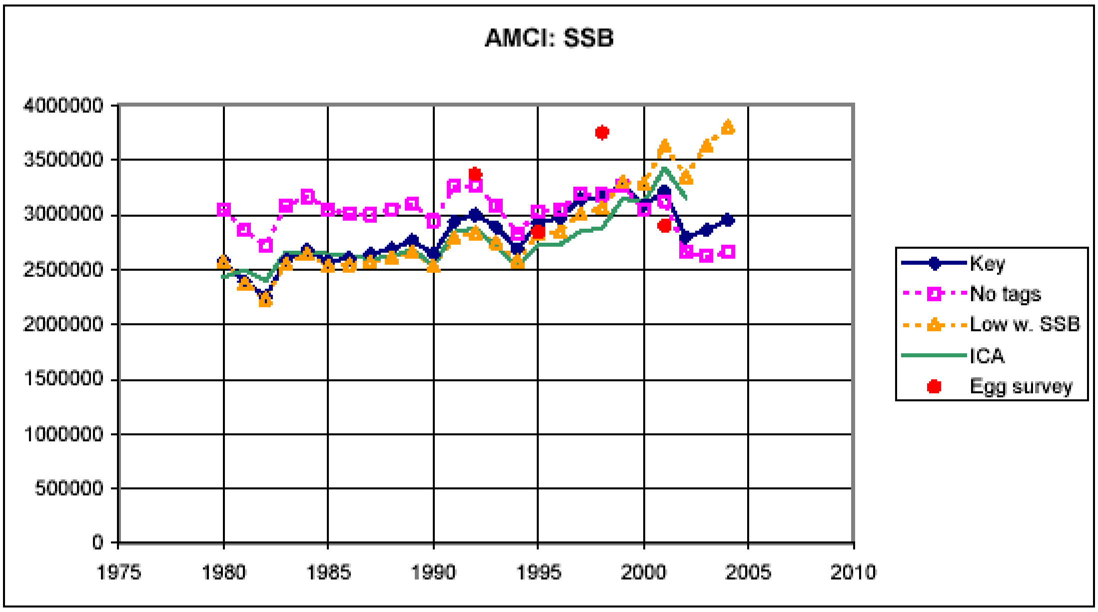


Figure 4.3.5. North Sea Atlantic Mackerel. Comparison of the assumption about the nature of the egg survey on the assessment of North East Atlantic mackerel. Assessment carried out using ICA with the same assumptions other than the relative or absolute nature of the egg surveys.



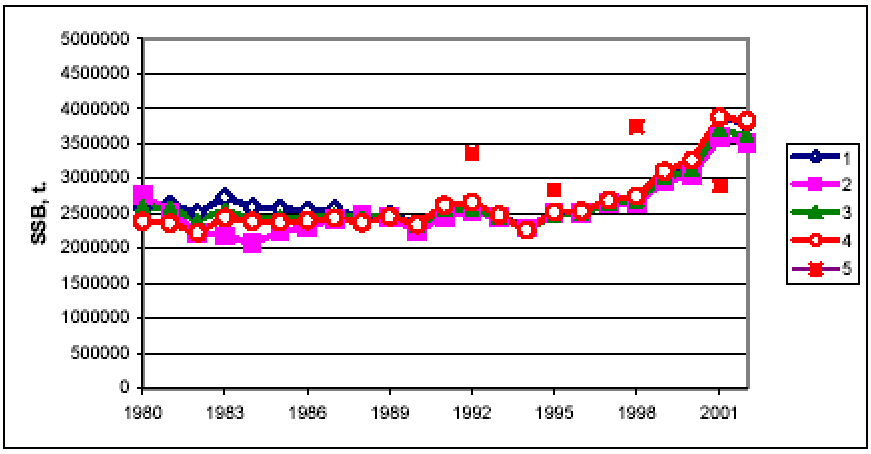


Figure 4.3.6. North East Atlantic mackerel. Comparison of ICA, AMCI and ISVPA. a) AMCI and ICA- the terms key, no tags and low w. SSB refer to different AMCI settings, ICA shows the final ICA assessment and the egg survey biomass estimates are also given. Note AMCI projects forward beyond the final year of the data. b) ISVPA, the 5 options show the range of SSB estimates using the different statistical techniques available in ISVPA. The isolated boxes also show the egg survey biomass estimates.

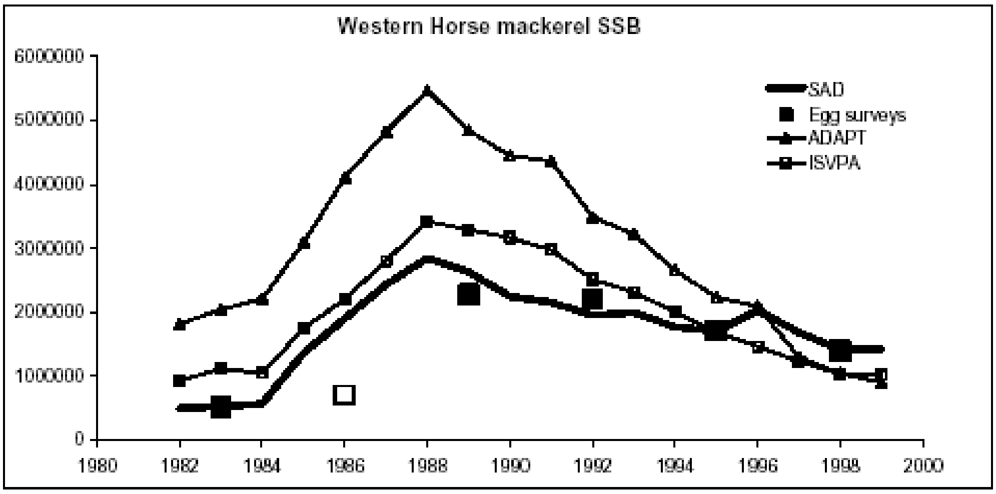


Figure 4.3.7. A comparison of the Western horse mackerel SAD model estimates of SSB with those derived from the ADAPT VPA and the separable ISVPA and the biomass estimates from the egg survey (ICES 2001a).

5. Projections

In the Northeast Atlantic, fisheries management is primarily based on Total Allowable Catch (TAC) regulations and the International Council for the Exploration of the Sea (ICES) annually provides advice on how to manage fisheries in a sustainable way. The stock assessment models described in the previous section are used to assess the historic and current state of the stock. Generally, the data available to ICES working groups do not include data for the on-going year and the stock can only be assessed up to the year prior to that of the working group. Since ICES is requested to advise on management for the year following that of the working group (with TACs – total allowable catches), catch forecasts and stock projections are required[1].

In the short-term projection, the stock numbers surviving the last data year are calculated forward and the predicted stock and catch numbers are converted into catch weights and spawning stock biomass (SSB). This deterministic projection is generally extended from the current year to the following 1 or 2 years. It assumes that the stock numbers estimated in the assessment model are correct. Furthermore, as catch and biological data are not available for the current and TAC years, assumptions need to be made on:

- Overall fishing mortality (F)
- exploitation pattern (fishing mortality at age)[2]
- recruitment
- weights at age
- maturity ogive (maturity at age)

The basis on which these assumptions are made for each of these variables is explained below.

The overall level of F used for management and reference points is quantified by mean F, which is the mean fishing mortality over a certain number of age groups (F_{bar}). An assumption is made on F_{bar} in the current year and a management option table with different catch options related to different levels of F_{bar} is calculated for the TAC year.

Projections are generally made by assuming that F_{bar} remains constant in the current year compared to previous years, this is the F *status quo* assumption. Alternatively, a TAC constrained catch can be assumed, i.e. this is regardless of F, and is called the TAC constraint assumption. Several studies have addressed the issue which of the two assumptions should be used (ICES, 2003a, Kraak *et al.* 2003, Jakobsen & Sparholt 2002); these studies will be briefly discussed in section 5.3. In the present analyses the F *status quo* assumption will be used.

F *status quo* (F_{sq}) is assumed to be equal to either F_{bar} in the last data year or to the average F_{bar} over the last 3 years. Clearly the exploitation pattern (or selection pattern, Figure 4.2.1) in the last three years of a stock assessed with a separable model will not vary, whereas the exploitation pattern of an XSA assessed stock will vary with year. Hence for fish stocks that are assessed using XSA, the exploitation pattern is averaged over the last 3 years. If F_{sq} is assumed to equal F_{bar} in the last data year, then this means that the 3-year averages of F at age are re-scaled to F_{bar} in the last data year. If F_{sq} is assumed to be equal to the average F_{bar} over the last three years, no re-scaling is necessary. Due to the assumption of separability in ICA and SAD, the averaging of the exploitation pattern is not required. In these cases the

[1] Some terminology may help avoid confusion. The year in which the assessment working group carries out the assessment and the projections is called the "current year". The last year from which data are available, which is the year before the current year, is called the "previous year" or the "last data year". The year for which the advice is given, which is the year following the current year, is called the "TAC year".
[2] The term "exploitation pattern" is generally used in the context of projections but refers to the same as the term "selection pattern" used in the previous section.

exploitation pattern of the last data year is used and this pattern may or may not be scaled to the average F_{bar} of the last 3 years, according to which assumption is chosen for F_{sq}.

The numbers of the recruiting age groups are often assumed to be equal to a long-term or short-term geometric mean. Alternatively, year-class strength indices from pre-recruit surveys can be correlated with VPA stock numbers of the recruiting ages (by a method called "RCT3"); from this relationship the number of recruits in the last data year can then be estimated from the survey data. Recruitment in the following years is then generally assumed to be the long term or short term geometric mean.

The standard procedure for short-term projections is to use the average stock and catch weights over the last 3 years for which data are available. However, if trends in growth rates occur this may lead to a bias in the predicted catch weights and SSB.

For many of the fish stocks under consideration a fixed maturity ogive is used in the assessment and predictions. In some cases (herring, horse mackerel and mackerel) annually varying maturity ogives are used. For these stocks it is assumed that the maturity ogives in the current and TAC years are equal to the average of the last 3 years for which data are available.

In the present analyses the sensitivity of short-term projections is examined to assumptions about

- weights
- fishing mortality (F_{bar})
- recruitment

5.1 Methods

The short-term projections of six fish stocks (plaice, sole, cod, whiting, mackerel and horse mackerel) were examined. Herring were not examined because of the complications of the non-directed industrial fishery on the juveniles. For each stock, the weights, recruitment, yield, spawning stock biomass (SSB) and fishing mortality (F) over the last 10 years (1993-2002) were presented to provide insight in the observed trends and variability. Also plotted were the short-term predictions as carried out by the working groups in 2003.

The WG2003 predictions, the WG2000 predictions, and for plaice the WG1997 predictions as well, were re-calculated using alternative assumptions.

For F the alternatives were:

- F_{sq} equals F_{bar} in the last data year
- F_{sq} equals the average of F_{bar} of the last 3 years for which data are available.
- F in the current year is the 'actual' F for that year that was estimated *post hoc* from the most recent assessment[1]; note that these values were not known at the time in the current year.

For weight at age the alternatives were:

- Weights in the current and TAC years equal weights in the last data year

[1] In principle, the F-at-age estimates from the most recent assessment (WG2003, ICES 2003a,b) were used as a proxy for actual fishing mortality, but due to the truncation of the age range in the plaice, sole, cod and whiting assessments, the WG 2002 estimates (ICES 2002) were used for these stocks.

- Weights in the current and TAC years equal the averages over the last 3 years for which data are available

- Weights in the current and TAC years are the actual weights for those years that are only known *post hoc* when the data of those years have become available, but not at the time in the current year.

In these analyses the stock numbers and the numbers of recruits used in the original prediction were maintained. The relative differences in yield (or catch) and SSB predicted for the TAC year are plotted.

The WG2003, WG2000 and plaice WG1997 predictions were also re-calculated assuming different levels of recruitment within the range observed in the last 10 years. In these calculations all other assumptions were maintained as in the original prediction (Table 5.1). The differences in predicted yield (or catch) in the TAC year and SSB surviving the TAC year are plotted relative to the predictions when geometric mean recruitment was used.

5.2 Results

5.2.1 North Sea plaice (Fig. 5.1)

Assuming the weights at age were equal to those of the last data year would have resulted in lower SSB and yield predictions for 2004 and 2001 and higher predictions for 1998 than in the original projections, where the weights at age were assumed to be equal to the average over the last three years (Fig. 5.1e,g,i). Comparison with projections where actual weights were used shows that the potential bias is actually even larger (9-14%, Fig. 5.1g,i). These patterns are related to the clear trends observed in weights at age; weights at age have decreased since 1998 whereas they increased before 1998 (Fig. 5.1a).

Whether or not the exploitation pattern is rescaled to F_{bar} in the last data year as opposed to using the average over the last three years has little effect on the predictions for 1998. The effect of rescaling appears to be reversed in 2004 compared to 2001: in 2001 higher yields and lower SSB is predicted if the 3-year average F is used, whereas in 2004 lower yields and higher SSB is predicted (Fig. 5.1e,g). This is possibly related to the upward trend in F prior to WG2000 and the downward trend prior to WG2003 (Fig. 5.1d).

The projection in WG2000 was carried out assuming geometric mean recruitment (417 million). The actual recruitment in 2000 and 2001 was lower (251 and 143 million respectively) which means that the recruitment assumption caused an 8-10% overestimation of in SSB in 2002 and a 4-5% overestimation of yield in 2001 (Fig 5.1h). The WG1998 assumed recruitment in the current year to be 801 million (from the RCT3 method) and 422 million (geometric mean) in the TAC year. Actual recruitment was lower once again resulting in an overestimation of yield and SSB.

5.2.2 North Sea sole (Fig 5.2)

The predictions for sole appear to be less sensitive to weight at age assumptions than those of plaice. If the average weights at are substituted by the actual weights at age, the yield predicted for the TAC year differs by 5% for sole (Fig. 5.2g) and by 11-13% for plaice (Fig. 5.1g,i). This corresponds with the fact that no clear trends on short time scales are observed for weights at age of sole (Fig. 5.2a).

The variation in F is larger in the period 1997-99 than in the period 2000-02 (Fig 5.2b). Consequently the effect of rescaling the exploitation to F_{bar} in the last data year has more effect on the predictions for 2001 (7%, Fig 5.2g) than for 2004 (2%, Fig 5.2e).

The variation in predicted yield and SSB in relation to recruitment (within the recruitment range observed in the last 10 years) is noticeably larger in sole than in plaice. If actual recruitment in 2003 is as high as it was in 1996 than the sole yield will be 30% higher than predicted (Fig 5.2f) and the plaice yield 15% (Fig 5.1f).

5.2.3 Cod in sub-area IV, divisions IIIA and VIId (Fig 5.3)

In 2003 no projection was carried out for cod, due to the uncertainty of the (input data for the) assessment. Fig. 5.3e shows that a projection would have been sensitive to whether or not the mean F-at-age is scaled to F_{bar} in the last data year, which can cause a 10% and a 20% difference respectively in the predicted yield and SSB in 2003. The projection done by WG2000 appears to be relatively insensitive to whether or not the mean F is scaled, but comparison with the prediction where the 'actual' F has been used indicates that SSB and yield were overestimated by more than 20%.

The alternative assumptions on weights-at-age, i.e. mean weight over the last three years or weight of the last data year, appear to have little effect on the predictions carried out in 2000 and 2003 (Fig 5.3e,g). However, comparison with the prediction where the actual weights have been used indicates that the predicted yield in 2001 is overestimated by 11% due to changes in weights-at-age.

5.2.4 Whiting (Fig 5.4)

The whiting projections are carried out for separate catch components: human consumption, industrial by-catch and discards. The predicted catches of the 3 components were combined in Fig. 5.4e-h.

The whiting predictions are relatively insensitive for the use of alternative weight assumptions or the use of the actual weight (2-4% difference in the predicted catches, Fig 5.4e,g), despite trends in weight-at-age (Fig. 5.4a). This is probably because the mayor proportion of the catches consists of fish younger than age 4 and the trends in weight for these age groups are minor (Fig. 5.4a).

The slope of the linear relationship between assumed recruitment and predicted yield (or SSB) differs strongly between Fig 5.4f (current year=2003) and Fig. 5.4h (current year=2000). This difference between assessment years is smaller in the other stocks under consideration.

5.2.5 Western horse mackerel (Fig. 5.5)

The WG2003 was uncertain about the year-class strength of the 2001 year-class. Therefore two projections were performed, one using the assessment estimate of stock numbers and the other using the geometric mean recruitment brought forward 2 years. These two projections give distinctly different perceptions of the stock development in the short term (Fig 5.5d). The sensitivity of the prediction to recruitment in the current and following years is low compared to the previous stocks (Fig. 5.5f,h). This is caused by the fact that the age of recruitment is set at age 0 in the assessment and projection (age at recruitment is 1 for the previous stocks). Only a very small proportion of the recruits are caught in the TAC year, and none of these recruits mature in the following year.

The predictions are relatively sensitive to whether or not F-at-age is scaled to the 3-year average of F_{bar}, especially the WG2003 predictions; predicted yield in the TAC year is 9-16% higher if average F is used (Fig. 5.5e,g). This is probably caused by the variable pattern in F (Fig. 5.5d).

The predictions appear to be less sensitive for weight assumptions than F assumptions if only the predictions using 3-year average and last year weights are compared (5-6% difference in yield, Fig. 5.5e,g), but the prediction using actual weights showed that yield and SSB were underestimated by 8 and 11% due to incorrect weight assumptions (Fig. 5.5g).

5.2.6 Northeast Atlantic mackerel (Fig 5.6)

Like in horse mackerel, the age of recruitment in the mackerel assessment and prediction is set at age 0. The sensitivity of prediction to recruitment, especially the yield prediction, is even lower in mackerel than in horse mackerel. The predicted yield barely increases with recruitment (Fig. 5.6h), because only a very small proportion of the recruitment in the current year (0-group) contributes to the yield in the TAC year.

Neither the F assumption nor the weight assumption affects the predicted yield and SSB in 2004 (WG2003, Fig. 5.6e). This corresponds with the fact that neither F nor weights-at-age varied much in the period 1997-2000 (Fig. 5.6a). The WG2000 used the 3-year average F in the short-term projection. If F_{bar} in the last data year had been used then the predicted yield would have been 8% lower. But if the 'actual' exploitation is used in the projection the predicted yield is 14% (22-8%) higher (Fig, 5.6g).

5.3 Discussion

Several authors have evaluated the quality of short-term projections by comparing the predicted catches, SSB's or F's to the "actual" values as estimated by stock assessment models in more recent years. Sparholt and Bertelsen (2002) examined 33 stocks and showed that on average SSB is overestimated by 14% and in some cases is predicted to be 5-8 times higher than the actual SSB. Van Beek and Pastoors (1999) compared predicted and actual fishing mortality and found no relationship, implying that the predictions of fishing mortality are not accurate. Although these studies clearly indicate that a high level of uncertainty is associated with the short-term projections they do not assess which assumptions are the possible sources of this inaccuracy.

Evidently a mayor source of inaccuracy is the uncertainty of the historic stock numbers, which are calculated forward in the short-term predication. In an evaluation of predictions for plaice and sole in the period 1983-1999, this was found to be the far more important source of uncertainty than recruitment, weights at age and relative exploitation pattern (Kraak *et al.*). Nevertheless, the current study shows that trends (in plaice and cod) and variability (in horse mackerel) in weights at age can have a considerable effect (>10%) on the accuracy of predictions. This conclusion is corroborated by Darby (2002a,b) who showed that trends in weight at age may cause a bias in the short-term forecast if three-year average weights are used. Similarly, if trends in F_{bar} occur, the short-term forecast is sensitive to whether F_{sq} is assumed to be equal to F_{bar} in the last data year or equal to the average F_{bar} over the last three years. This is most evident in cod.

The controversy between the F *status quo* assumption versus the TAC constraint assumption is potentially important for the quality of the prediction. In 2003, the WGMHSA (Working Group on the assessment of Mackerel, Horse mackerel, Sardine and Anchovy) examined both assumptions in the projections for mackerel (Fig. 5.7) by comparing the predicted catches and F's to "actual" catches and F's as estimated 1 year later. They concluded that if large changes in TAC's occur it would be preferable to use the TAC constraint assumption. Kraak *et al.* (2003) carried out a similar study for plaice and sole predictions showing that although in some years the TAC constraint assumption leads to a more accurate prediction of SSB, overall the F *status quo* assumption is more reliable (Fig. 5.8). Jakobsen and Sparholt (2002) estimated the prediction error as a function of the error in the assessment and showed the prediction error to be larger under the TAC constraint assumption than under the F *status quo* assumption.

The short-term projections of plaice, sole, cod, and whiting are sensitive to the recruitment estimates for the current and following years. This sensitivity appears to be less evident for mackerel and horse mackerel, stocks in which the age of recruitment is set at 0. This conclusion may however be misleading because the stock numbers at ages 1 (to 2) are essentially also recruit estimates which cannot be reliably estimated by the stock assessment. This is clearly illustrated by the dependency of the forecasts for 2004-2005 on the assumption of the year-class strength of the 2001 year-class. The observed variation between species and years tentatively suggests that the sensitivity for recruitment may be related to the level of SSB. Cook (1991) shows that in heavily exploited fish populations the forecast is sensitive to recruitment. In this context, Brander (2003) argues that short life expectancy limits the predictability based on the current fishable stock, which causes the projections to be more dependent on estimates or assumptions about future recruitment rates. However, reduced exploitation does not necessarily improve the projection because of increased sensitivity to other quantities such as fishing mortality, which are more difficult to estimate in lightly exploited stocks.

Table 5.1 The Fsq, weights-at-age and recruitment estimates used in the original short term projection as carried out by the working groups (ICES, 1998; ICES, 2001a,b; ICES, 2004a,b).

	Fsq	weights	recruitment			stock numbers		
	=F_{bar} for		age	current year	next years	N1	N2	N3+
ple-2003	last year	avg 3 yrs	1	GM(57-00)	GM(57-00)	<=	RCT3	XSA
2000	last year	avg 3 yrs	1	GM(57-97)	GM(57-97)	<=	XSA	XSA
1997	last year	avg 2 yrs	1	RCT3	GM(57-94)	<=	XSA	XSA
sol-2003	last year	avg 3 yrs	1	GM(57-00)	GM(57-00)	<=	RCT3	XSA
2000	last year	avg 3 yrs	1	GM(57-97)	GM(57-97)	<=	RCT3	XSA
cod-2003	-	-	1	-	-	<=	XSA	XSA
2000	avg 3 yrs	avg 3 yrs	1	GM(89-98)	GM(89-98)	<=	XSA	XSA
whi-2003	avg 3 yrs	avg 3 yrs	1	GM(93-02)	GM(93-02)	<=	XSA	XSA
2000	last year	avg 3 yrs	1	RCT3	GM(80-98)	<=	XSA	XSA
mac-2003	avg 3 yrs	avg 3 yrs	0	GM(72-99)	GM(72-99)	GM*	75%GM**	ICA
2000	avg 3 yrs	avg 3 yrs	0	GM(72-96)	GM(72-96)	GM*	ICA	ICA
hom-2003	avg 3 yrs	avg 3 yrs	0	GM(83-00)	GM(83-00)	GM*	GM* & SAD	SAD
2000	avg 3 yrs	avg 3 yrs	0	GM(83-98)	GM(83-98)	GM*	SAD	SAD

* GM at age 0 brought forward by 1 or 2 years

** 75^{th} percentile of GM at age 0 brought forward by 1 or 2 years

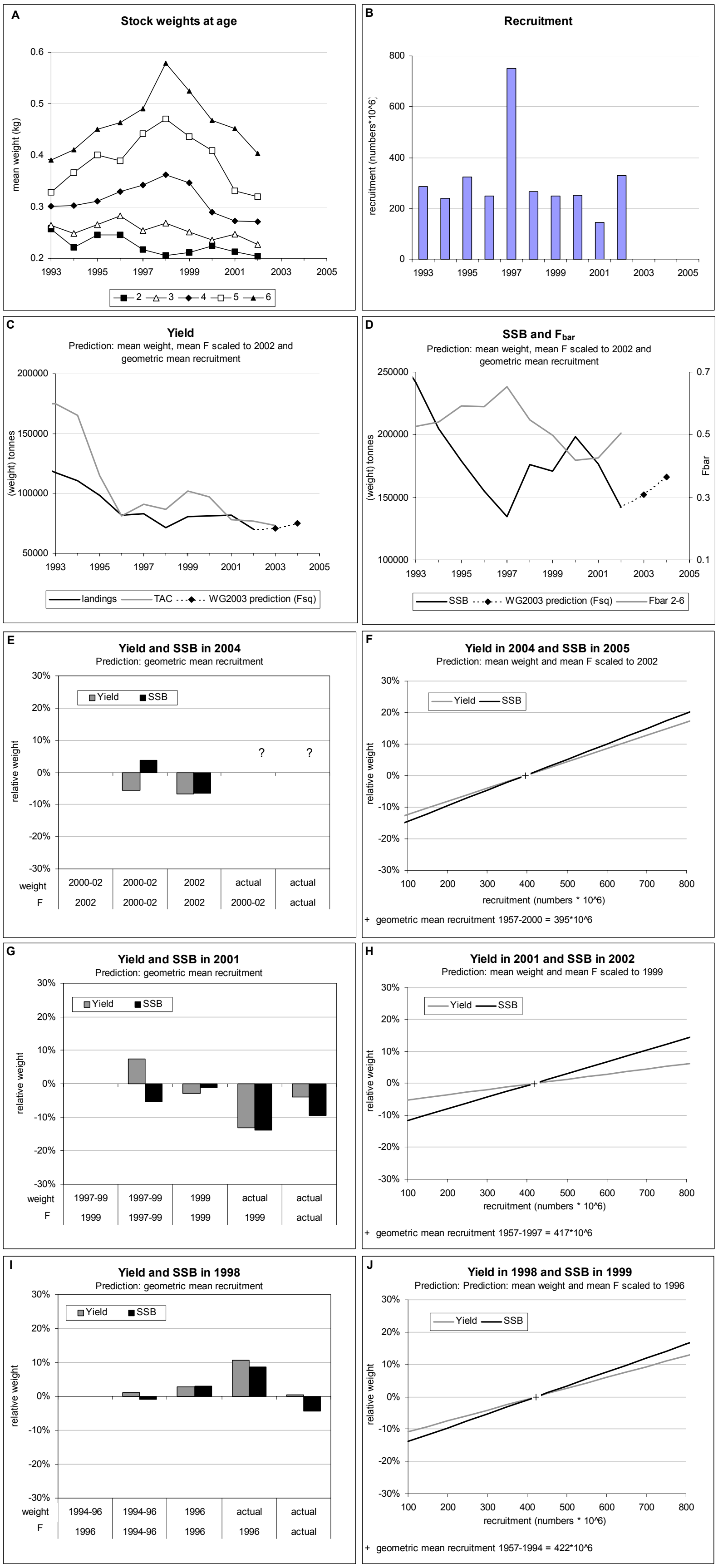


Figure 5.1 North Sea plaice: Weights-at-age (A), recruitment (B), yield (C), SSB and F (D) for the period 1993-2005 (source: ICES, 2004b); predicted yield and SSB based on alternative F assumptions (E,G,I); and predicted yield and SSB assuming different levels of recruitment (F,H,J).

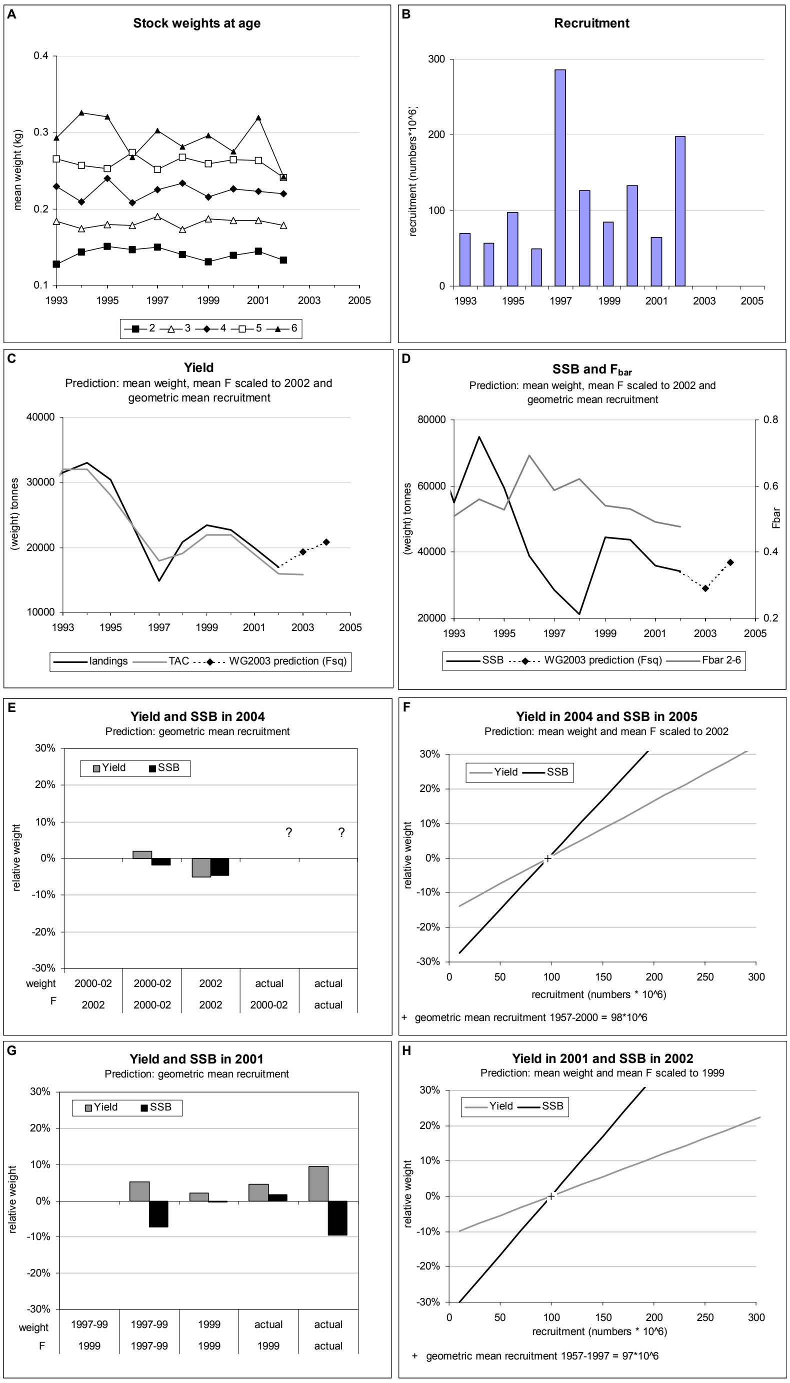


Figure 5.2 North Sea sole: Weights-at-age (A), recruitment (B), yield (C), SSB and F (D) for the period 1993-2005 (source: ICES, 2004b); predicted yield and SSB based on alternative F assumptions (E,G); and predicted yield and SSB assuming different levels of recruitment (F,H).

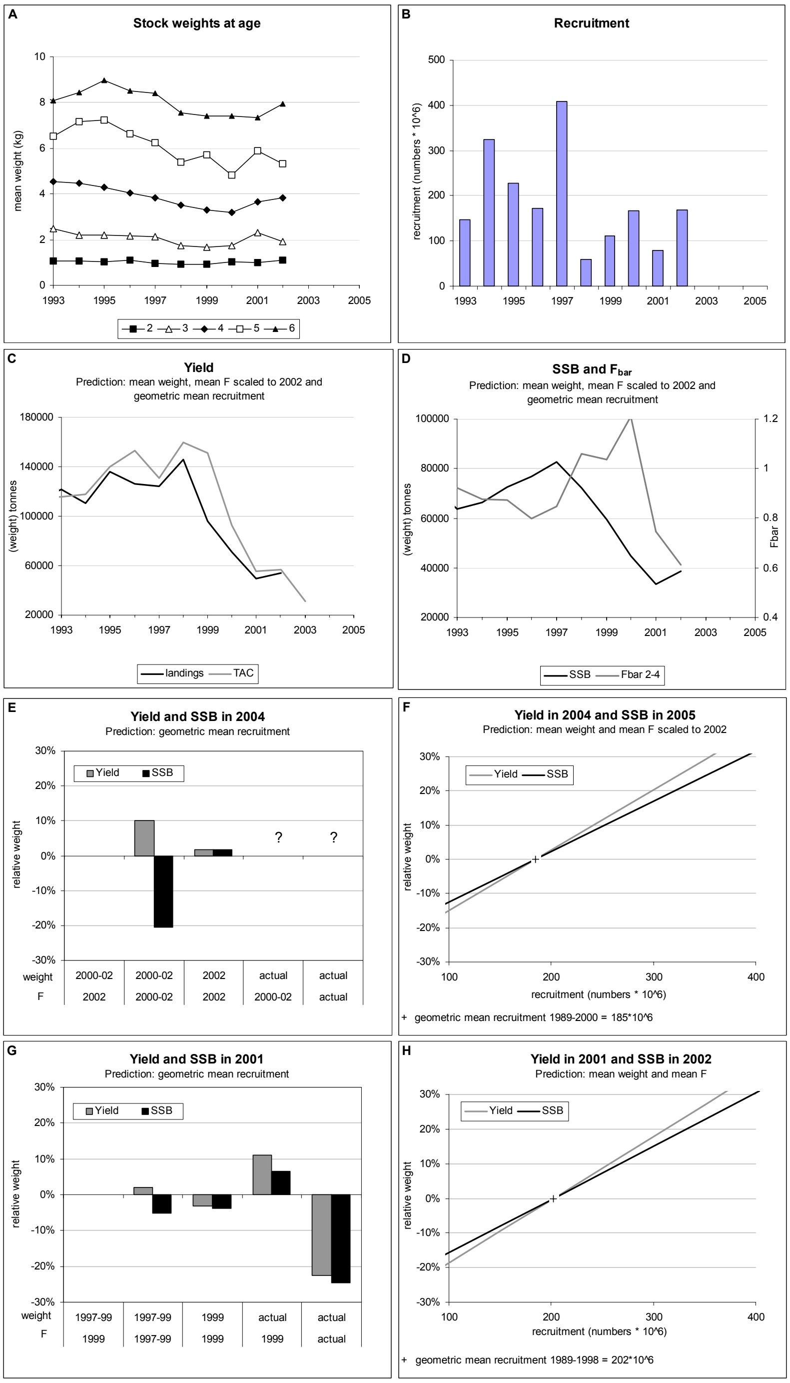


Figure 5.3 Cod in sub-area IV and divisions IIIa and VIId: Weights-at-age, recruitment, yield, SSB and F for the period 1993-2005 (ICES, 2004b); predicted yield and SSB based on alternative F assumptions; and predicted yield and SSB assuming different levels of recruitment.

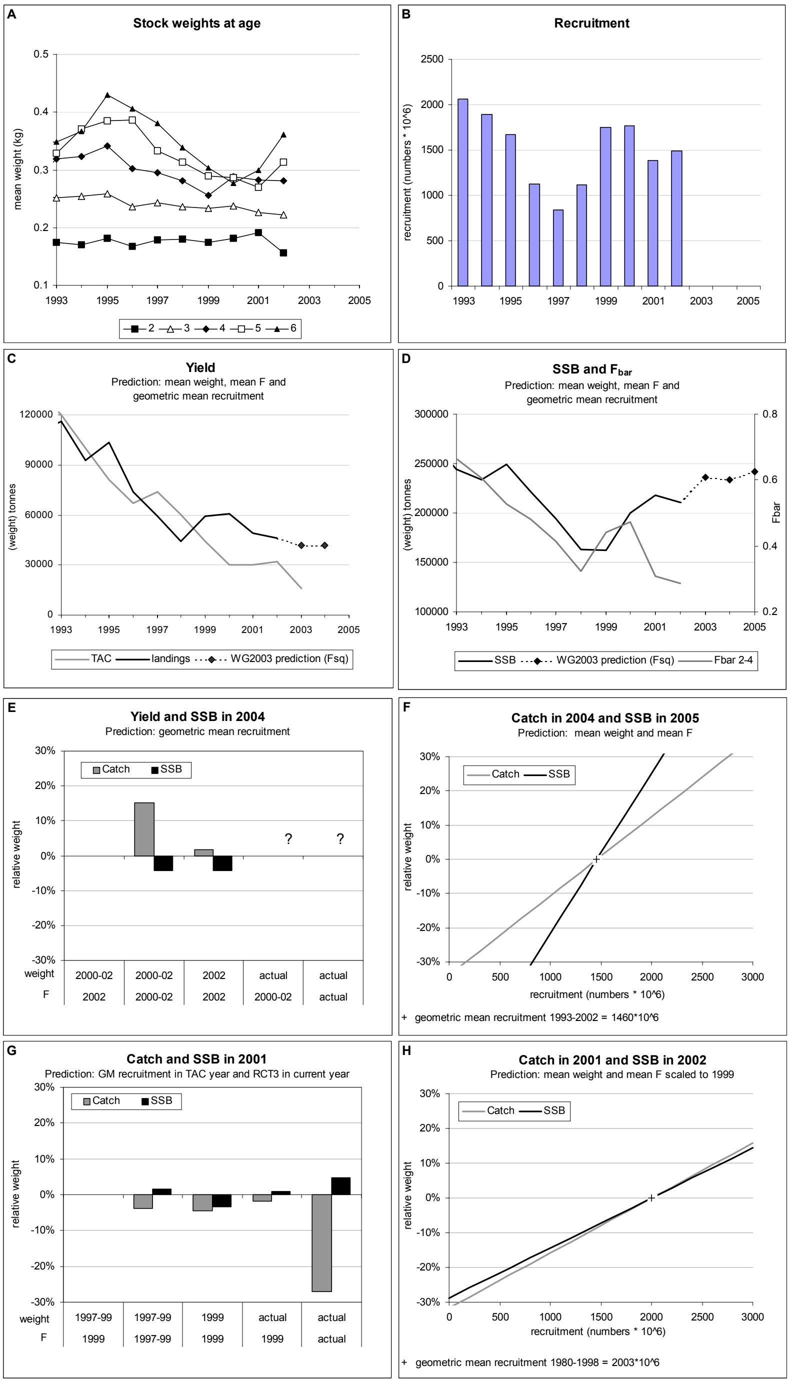


Figure 5.4 Whiting in sub-area IV and division and VIId: Weights-at-age, recruitment, yield, SSB and F for the period 1993-2005 (ICES, 2004b); predicted yield and SSB based on alternative F assumptions; and predicted yield and SSB assuming different levels of recruitment.

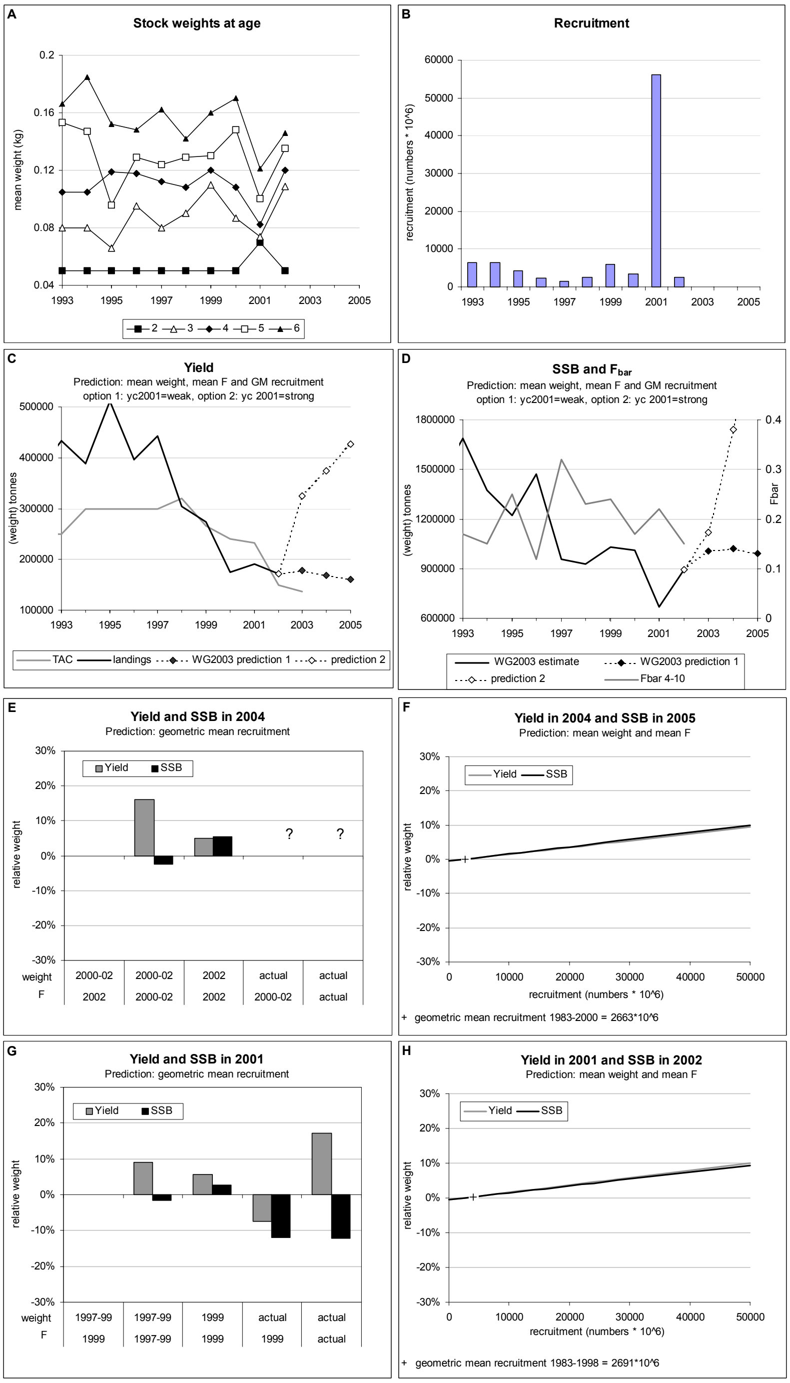


Figure 5.5 Western horse mackerel: Weights-at-age, recruitment, yield, SSB and F for the period 1993-2005 (ICES, 2004a); predicted yield and SSB based on alternative F assumptions; and predicted yield and SSB assuming different levels of recruitment.

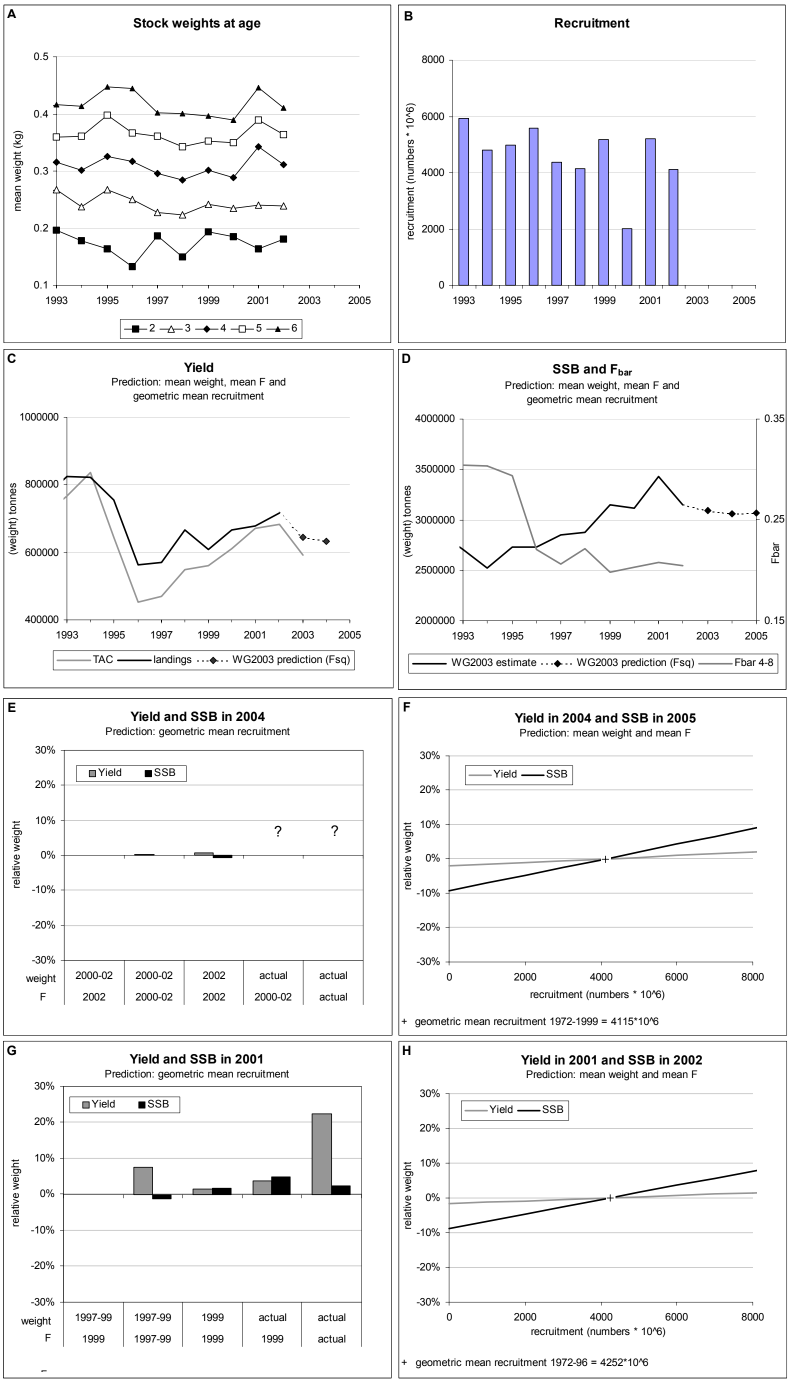


Figure 5.6 Northeast Atlantic mackerel: Weights-at-age, recruitment, yield, SSB and F for the period 1993-2005 (ICES, 2004a); predicted yield and SSB based on alternative F assumptions; and predicted yield and SSB assuming different levels of recruitment.

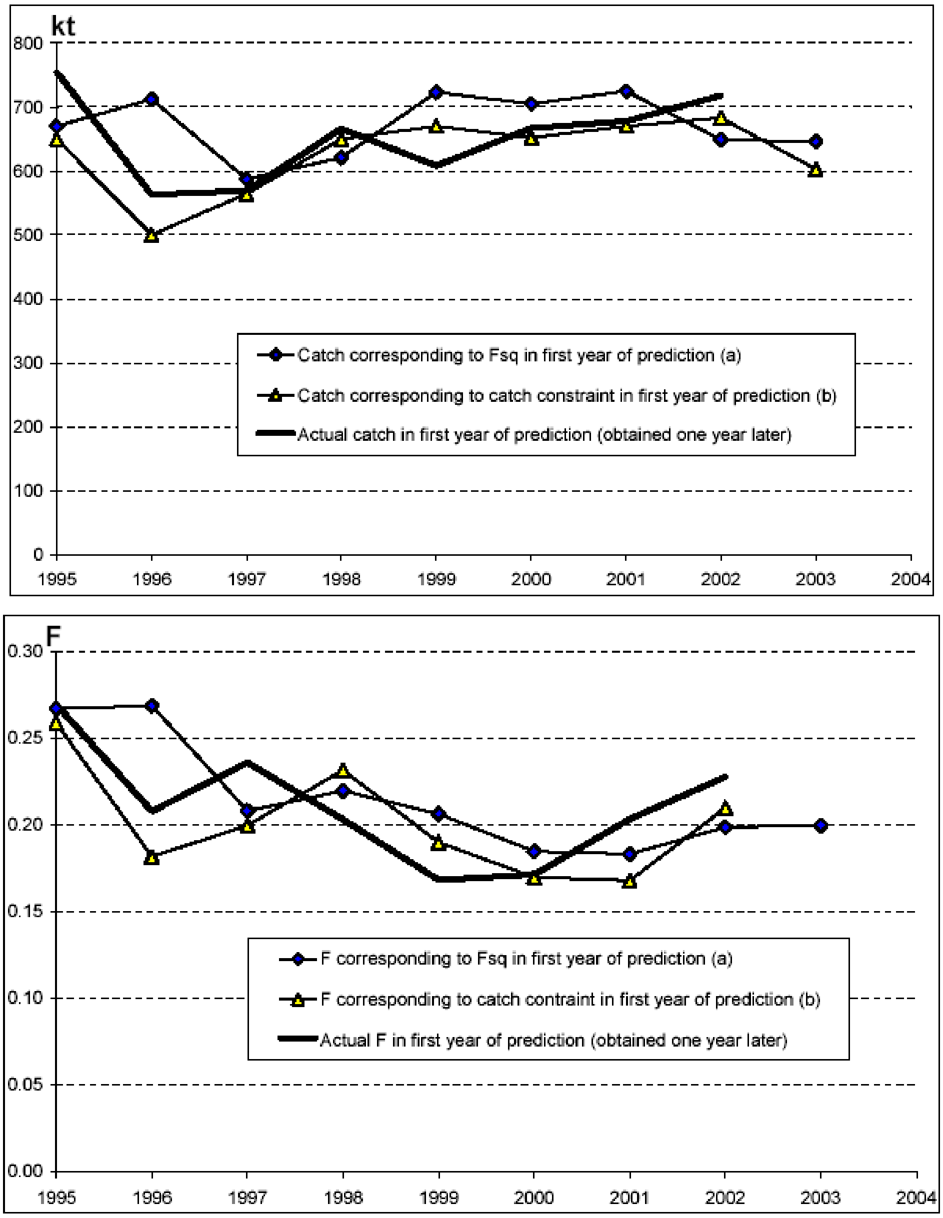


Figure 5.7 Northeast Atlantic Mackerel catch predictions carried out for two options: 1) a catch corresponding to Fsq and 2) a catch constraint. The actual catch and F obtained one year after the predictions are compared to the predicted catches and F's (ICES, 2004a).

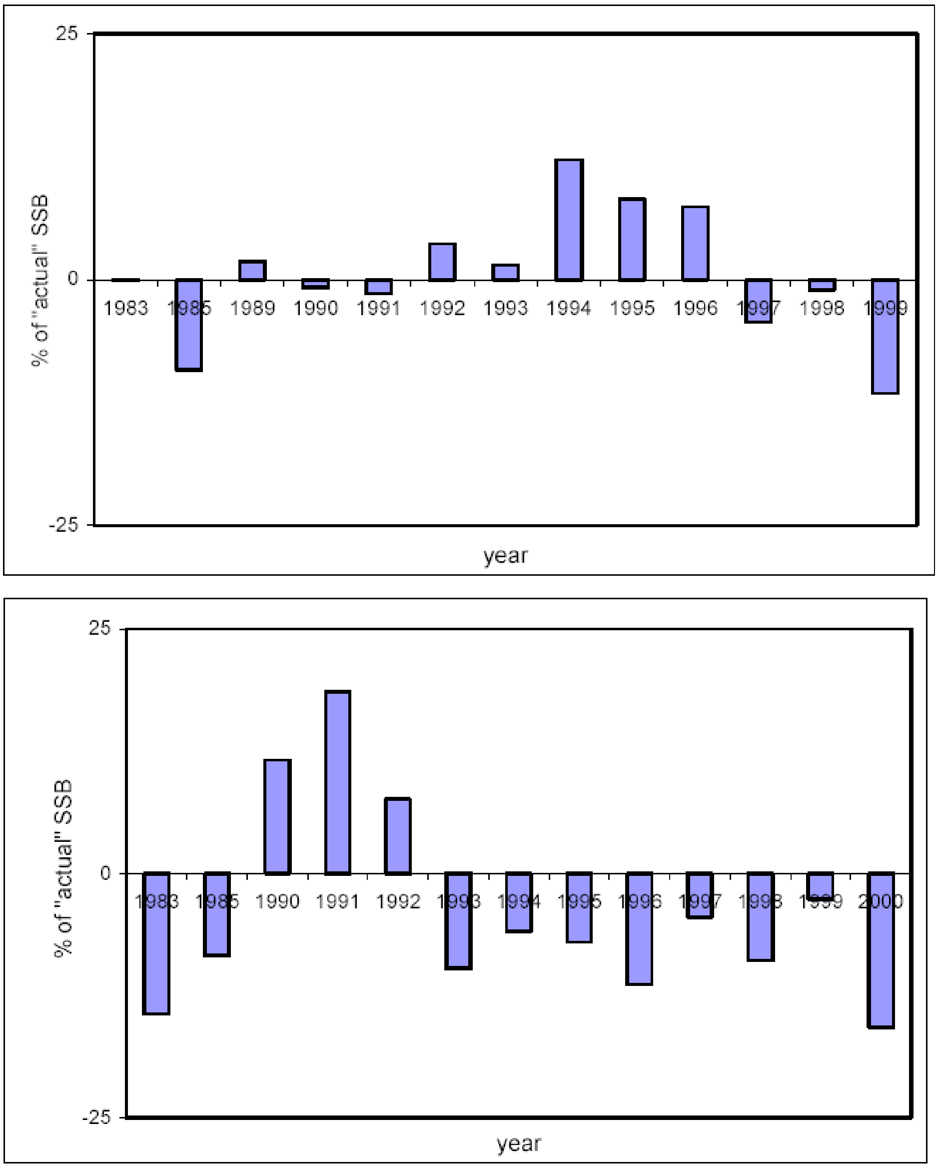


Figure 5.8 Gain in accuracy of the SSB prediction under the TAC assumption versus the *status quo* F assumption for North Sea plaice (top panel) and North Sea sole (bottom panel).

6. Impact of different sources of error data on the quality of the assessment. A quick response to a general question.

This section addresses a specific request by Drs. Schoute made on 25 March 2004. The request was "What has the greatest impact on the quality of the assessment?". This question cannot be addressed in a robust manner in a hurry, but the work below illustrates the sensitivities of the assessment to poor quality data, beyond the precision of the catch (see section 3). The different impact of imprecision and inaccuracies in the input data (including misreported catch) on the quality and reliability of the assessment is investigated. This was done by using a spreadsheet separable model (Xcam_model), attached to a data generator (XGenerator) which was developed by Einar Hjörleifsson (Iceland). As a separable model, it is not dissimilar from AMCI. The data produced by this package are relatively unrealistic, in that the package does not allow year effects to be introduced into the surveys or catch, and hence the data its generates is not as noisy as real assessment data.

Basically a known population was created with described characteristics, and then this population was fished with various errors introduced into the data collection part of the model. This allowed the true population to be compared with the stock assessment results under different scenarios of inaccurate or imprecise catch or survey data. The time series was 25 years long (1975-1999) and had a dip in fishing mortality between 1985 and 1993 (See Figure 6.1). Selection varied with age to a plateau, but was kept constant across the time series. The assessment was tuned with 2 surveys, one aimed at older fish and the other at younger fish. Like most real assessments the surveys carried a lower weight in the assessment than the catch. Each of them carried 10% of the importance of the catch in the solving of the model.

The six scenarios were:

1. **Best estimate** by the separable model without changes to the input data.

2. **Ageing**- a random error of 20% was introduced into the catch matrix

3. **Misreport**- underreporting began in 1992 at 5% of the catch (spread equally across all age groups) and increased gradually to 50% by the final year of the time series.

4. **Discard**- discarding began on fish aged 0-2 in 1992 at 5% of the catch increased gradually to 50% by the final year of the time series.

5. **1 survey trend**- the survey on the adult became better at catching the fish in 1992 and increased its efficiency by 5% per year, to be 40% more efficient by 1999.

6. **2 survey trend**- both of the surveys became better at catching the fish in 1992 and increased its efficiency by 5% per year, to be 40% more efficient by 1999.

It should be pointed out from the start that the best estimate, does not give a perfect assessment of the dynamics of the stock (Figure 6.1). It is 10% out in the estimation of F and SSB in the final year, but correct in the estimation of recruits. It also clear that the imprecision in aging (20%) also has a very small impact on the assessment results (although the residuals to the catch are larger). This good fit is probably due to the lack of trend in the imprecise aging (which varies randomly).

One trend in a survey had an impact, but less than both surveys increasing in catching efficiency. If two surveys had different trends (one more efficient one less efficient), then they would probably cancel each other out. It is clear that discarding and misreporting of the catch have the largest impact (Figure 6.1). The worst estimates of mean $F_{(4\text{-}8)}$ come from the misreported catch (a 46% underestimate), which would cause huge problems in the short-term projections. Discarding does not effect mean $F_{(4\text{-}8)}$, because this only covers the mature fish, whilst in a projection it would impact as projections use F at each age. As expected, the estimates of SSB are heavily underestimated by changes in misreporting or discarding, with a 30% underestimate in SSB in the final year if the discarding trend is not accounted for. The effect on recruitment estimation by a discarding trend is also dramatic (reduction of 37%). This would have important implications on the setting of stock-productivity based reference points, as used by ICES.

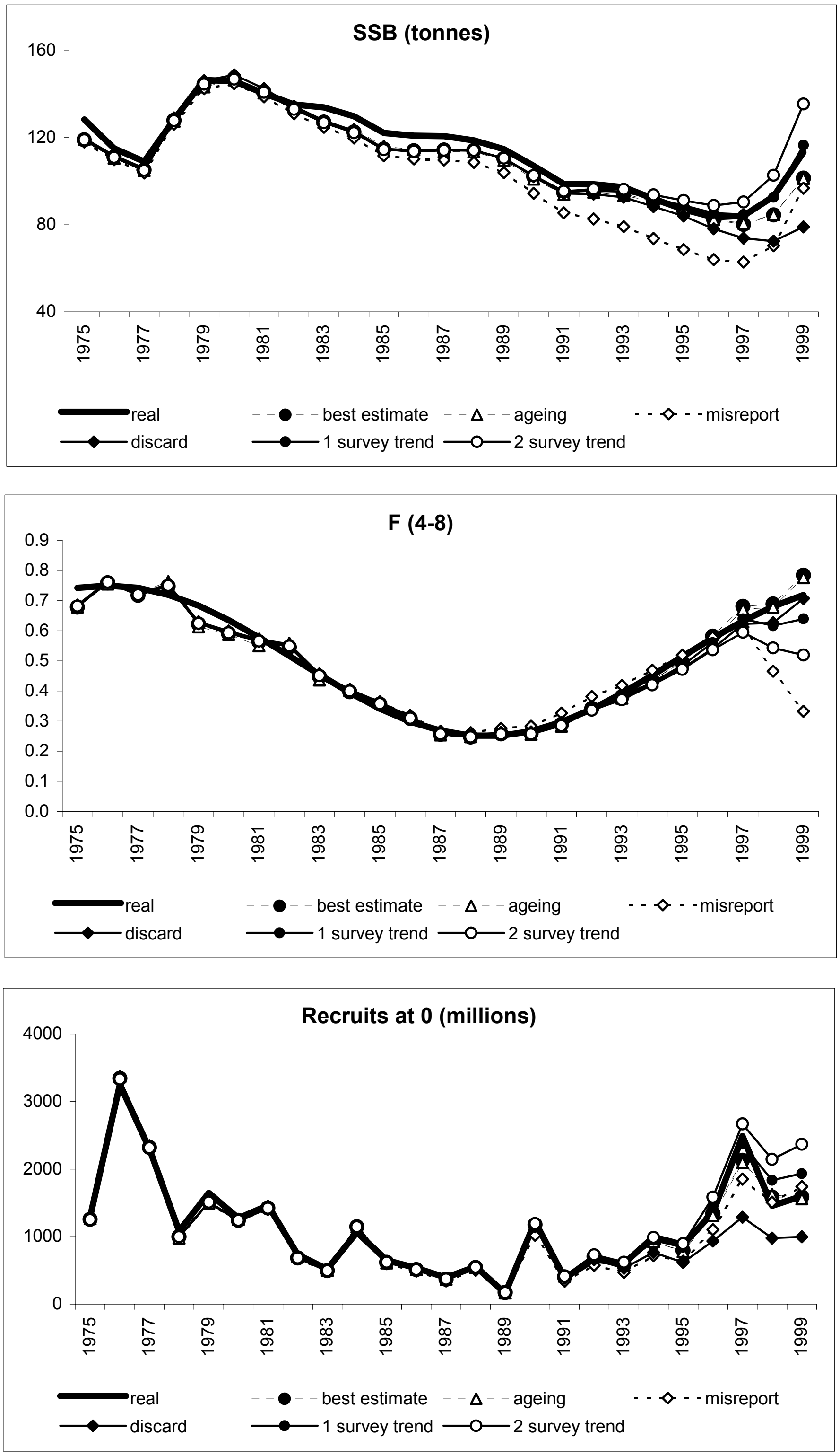


Figure 6.1. Results of simulation study of data inaccuracies and imprecision impacting on the quality of the assessment.

Simulation studies can be used to prove counteracting arguments, however in this case, realistic trends in discarding, misreporting and catchabilities of surveys were fed into the data generator to produce the catch and survey statistics. These show that the imprecision in the estimates of cohorts (ageing error with no trend) has a much smaller impact than the high impact of underreporting or discarding trends. Large scale trends in the surveys will cause problems, but the large differences in catching efficiency (40% increase over 8 years) shown by this study are thought to be much larger than that found in the field. As most assessment models are based on the catch equation (section 2), it is not surprising that the catch has the greatest impact on the quality of the assessment.

7. Conclusions

This study has shown that the precision of the Dutch estimates of weight and numbers at age is good and is unlikely to impact on the quality of the biological advice (section 3.1). It has also shown that the choice of assessment model, or choice of internal assumption, can effect the assessment of a stock (section 4.3). It is clear however that most of the stocks are already assessed with a model that conforms to the requirements of that stock and few assumptions (or uncertainties) are inappropriate (section 4.2). Issues such as changes in separable period, the effect of shrinking F and the relationship between survey and catch estimates can impact on the assessment. Surveys that show similar trends should give similar modelled results (section 3.2).

Assumptions about fishing mortality, recruitment and trends in weight can impact on the quality of projections. Stocks with reduced age profiles are particularly susceptible to assumptions about the numbers of recruits and also ages close to recruitment. Strong trends in fishing mortality and the weight of a fish at each age that are not reflected in the projection inputs will impact on the final outcome (section 5). Projections are normally wrong, once compared to the actual values in the historic estimates and it is the menu of input assumptions that result in these errors. This raises the question of the relevance of short-term projections as a management tool.

The current project did not assess the impact of inaccurate catch data. This is different from the precision (see section 3.1) and is best described as the catch data being wrong, not just poorly estimated. Many studies have shown that incorrect catch data will have large impact on the quality of biological advice and this is particularly relevant in environments like the North Sea where the catch is international and caught under different jurisdictions. Incorrect catch data has been implicated as the cause of retrospective bias in assessment time series (section 4.1.2) and the problem of disparate signals between catches and surveys. As shown in section 4.2, many models weight the catch data extremely highly compared to the survey data and some, like XSA, assume that not only are the catch data accurate but also precise. Bayesian methods (too complicated for the author to understand or describe) or survey-only methods of stock assessment (*e.g.* SURBA) have been and are being developed to address this issue but they are not fully available at present and are yet to be scrutinised by the scientific community. TSA methods can also cope with years of missing catch and are used in this way by ICES. However as stated above they too are poorly documented and are not widely used.

By successfully achieving its objectives, this project has shown that the many uncertainties can affect the quality of the biological advice. Unfortunately there is no generic solution to the problems associated with each stock. It appears that the current levels of precision of the input data do not adversely affect the quality of the biological advice, whereas the accuracy of the catch estimates is crucial.

Acknowledgements

Martin Pastoors and Ingeborg de Boois are thanked for their help and advice in the compilation and execution of this report.

References

Abaunza, P., L. Gordo, C. Karlou-Riga, A. Murta, A.T.G.W. Eltink, M.T. Garcıa Santamarıa, C. Zimmermann, C. Hammer, P. Lucio, S.A. Iversen, J. Molloy & E. Gallo (2003). Growth and reproduction of horse mackerel, Trachurus trachurus (carangidae). Reviews in Fish Biology and Fisheries 13: 27–61

Allison, P. D. 1999. Logistic regression using the SAS system; theory and application. SAS institute inc.. Cary, NC. 304 pp.

Beek, F.A. van and Pastoors, M. A. 1999. Evaluating ICES catch forecasts: the relationships between implied and realised fishing mortality. ICES C.M. 1999/R:04

Brander, K. 2003. What kinds of stock predictions do we need and what kinds of information will help us to make better predictions? Scientia Marina, 67(Suppl.1): 21-33.

Burd, AC (1985) Recent changes in the central and southern North Sea herring stocks. Can. J. Fish. Aquatic Sci., 42 (Suppl 1): 192-206

Cook, R.M.1991. Changes in the precision of short-term forecasts under different levels of exploitation. ICES CM 1991/D:22

Cushing, DH (1992). A short history of the Downs stock of herring. ICES J. mar. Sci., 49: 437-443.

Darby, C. 2002a. Over-estimation bias in the North Sea cod short-term forecasts. Appendix B – Working Document WA2. WGMG. ICES C. M. 2002/D:01

Darby, C. 2002b. Estimation bias in the North Sea short-term forecasts. Appendix C – Working Document WA3. WGMG. ICES C. M. 2002/D:01

Darby, C.D. & Flatman, S. (1994). Virtual Population analysis: Version 3.1 (Windows/DIOS) user guide. MAFF Directorate of Fisgheries Research IT Report 1. 85pp.

Deriso, R.B., Quinn, T.J. and Neal, P.R. 1985. Catch-age analysis with auxiliary information. Canadian Journal of Fisheries and Aquatic Sciences, 42, 815–824.

Dickey-Collas, M. & Eltink, ATGW (2003). The precision of numbers at age and mean weight estimation of mackerel and horse mackerel from Dutch market sampling from 1998 to 2002. Working Document to ICES WHMHSA, 12pp.

Dickey-Collas, M., Pastoors, M., Kraak, S., Keeken, O. van, Oostenbrugge, H van, Smit, J Wilde, JW de (2003). Quick-scan on the different scenarios relating to the ACFM 2003 advice: biological and economic consequwences. CVO report nr. 03.019. 151pp.

Doubledat, W.G. (1981). A method for estimating the abundance of survivirs of an exploited fish population using commercial catch-at-age and research vessel abundance indices. Canadian Special Publication Fisheries and Aquatic Science, 58: 164-178.

Eltink, A.T.G.W. (1987). Changes in age-size distribution and sex ratio during spawning and migration of Western mackerel (*Scomber scombrus* L.). J. Cons. int. Explor. Mer, 44: 10-22. 1987

Eltink, ATGW (2004). Preliminary evaluation of assessments with ISVPA and ICA in the case of shifts in the exploitation pattern. Working document for the ICES Methods Working Group, 5pp.

EVARES (2003). Evaluation of research surveys in relation to management advice. Final Report to the European Commission Contract EVARES - FISH/2001/02 - Lot 1. 326pp.

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.

Fryer, R.J., Needle, C., & Reeves, S.A. 1998. Kalman filter assessments of cod, haddock, and whiting in VIa. Working document for the Working Group on the Assessment of Northern Shelf Demersal Stocks 1998.

Gudmundsson, G. 1986. Statistical considerations in the analysis of catch-at-age observations. Journal du Conseil International pour I'Exploration de la Mer, 43, 83–90.

Gudmundsson, G. 1994. Time series analysis of catch-at-age observations. Appl. Statist. 43: 117-126.

Gulland, JA (1965). Estimation of mortality rates. Annex to Arctic Fisheries Wiorking Group Report. ICES CM 1965/3.

ICES 1998. Report of the Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak (Copenhagen, 5-14 October 1998). ICES C.M. 1999/ACFM:8.

ICES (2001a). Report of the working group on the assessment of mackerel, horse mackerel, sardine and anchovy, Copenhagen 2000. ICES CM2001/ACFM:06

ICES (2001b). Report of the Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak (Copenhagen, 3-12 October 2000). ICES C.M. 2001/ACFM:7

ICES (2002a). Report of the working group on the assessment of mackerel, horse mackerel, sardine and anchovy, Copenhagen 2001. ICES CM2002/ACFM:06

ICES (2002b). Report of the Northern pelagic and Blue Whiting Fisheries Working Group. ICES CM 2002/ACFM19. 289pp.

ICES (2003a). Report of the Working Group on Methods on Fish Stock Assessments. ICES CM 2003/D:03. 171pp.

ICES (2003b). Working Group on the Assessment of Northern Shelf Demersal Stocks. ICES CM 2003/ACFM:04. 625pp.

ICES (2003c). Report of the Herring Assessment Working Group for the Area South of 62°N. ICES CM 2003/ACFM:17. 438pp

ICES 2004a. Report of the Working Group on the assessment of mackerel, horse mackerel, sardine and anchovy (Copenhagen, 9-18 September). ICES CM 2001/ACFM:06.

ICES 2004b. Report of the Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak (Boulougne-sur-Mer, 9-18 September 2003).

Jakobsen, T. and Sparholt, H. 2002. Definition and use of status quo F. Working paper for SGWP and SGPA. ACFM May 2002. Kraak, S.B.M., Pastoors, M., Rijnsdorp, A.D. 2003. Analysis of the ICES short-term forecasts of North Sea plaice and sole: dealing with the “current” year assumption.

Keeken, O.A. van, Dickey-Collas, M. Kraak, S.B.M. Poos, J.J. Pastoors M.A. (2003). The use of simulations of discarding to investigate the potential impact of bias, due to growth, on the stock assessment of North Sea plaice (Pleuronectes platessa). ICES CM 2003/X:17, 23pp

Keeken, O van, Quirins, F, Pastoors, M (2004). Analysis of discarding in the Dutch beamtrawl fleet. RIVO concept report to LNV. Contract number 02.036. 66pp.

Kell, LT, Cotter, AJR, van Keeken, O, Pastoors, M, O'Brien, CM, Piet, GJ, Rackham BD (2003). The influence on stock assessment advice of sampling error in research survey and international market sampling data for North Sea cod (*Gadus morhua* L.) and plaice (*Pleuronectes platessa* L.). ICES CM 2003/X:05.

Kimura, D.K. 1986. Variability, tuning and simulation for the Doubleday-Deriso Catch-at-Age model. Canadian Journal of Fisheries and Aquatic Sciences, 46, 941–949.

Kraak, S. M.B., M. A. Pastoors & A.D. Rijnsdorp. 2002. Effecten van discarding en high-grading op de toestandsbeoordeling van schol: een quick-scan. IJmuiden, RIVO: 13p

Kvist, T., Gislason, H. and Thyregod, P. 2000. Using continuation-ratio logits to analyze the variation of the age composition of fish catches. Journal of applied statistics, 27: 303-319.

Kvist, T., Gislason, H. and Thyregod, P. 2001. Sources of variation in the age composition of sandeel landings. ICES Journal of Marine Science, 58: 842-851.

Manly, F.J. (1991) Randomization, Bootstrap and Monte Carlo Methods in Biology. Chapman and Hall, London 399pp.

Maxwell, D.L, B.D.Rackham, S.Warnes, M.Armstrong, A.W.Newton, R.Officer. (2001). A study of market sampling data, precision of age compositions and the effect of mis-ageing: ICES Division VIIa cod. ICES 2001/P:11

Needle, C.L. (2001). Some alternative assessments for 3Ps cod. Working Paper to the Newfoundland Regional Groundfish Assessment Meeting, St. John's, October 2001.

Needle, C.L. and Fryer, R.J. (2002). A modified TSA for cod in Division VIa: separate landings and discards. Working document to the ICES Advisory Committee on Fisheries Management, October 2002.

O'Brien, CM, Darby, CD, Maxwell, DL, Rackham, BD, Degel, H, Flatman, S, Pastoors, MA, Simmonds, EJ & Vinther, M (2001a). The precision of international market sampling for North Sea plaice (Pleuronectes platessa L.). ICES CM2001/P:13.

O'Brien, C.M. Darby C.D, Rackham, B.D, .Maxwell, D.L,. Degel, H, Flatman S, Mathewson M, Pastoors M.A., Simmonds E.J .and Vinther M (2001b). The precision of international market sampling for North Sea cod (Gadus morhuaL.) and its influence on stock assessment. ICES 2001/P:14

Pastoors, MA, O'Brien, CM, Flatman, S, Darby, CD, Maxwell, DL, Simmonds, EJ, Degel, H, & Vinther, M, Sparre, P & Vanderperren, E. (2001). Evaluation of market sampling strategies for a number of commercially exploited stocks in the North Sea and development of procedures for consistent data storage and retrieval. EMAS). Project Final Report, European Commissions CFP Study Project 98/075. 324pp.

Patterson, K.R. 1998. Integrated Catch at Age Analysis Version 1.4. Scottish Fisheries Research Report. No. 38.

Patterson, KR & Melvin, GD (1996). Integrated Catch at Age Analysis version 1.2. Scottish Fisheries Research Report, SOAEFD, 58, 60pp.

Patterson, KR, Cook, R, Darby, C, Gavaris, S, Kell, L, Lewy, P, Mesnil, B, Punt, A, Restrepo, V, Skagen, DW, Stefansson, G (2001). Estimating uncertainty in fish stock assessment and forecasting. Fish and Fisheries, 2: 125-157.

Pauly, D., Christensen, V., Guénette, S., Pitcher, T. J., Rashid Sumaila, U., Walters, C. J., Watson, R. and Zeller, D. 2002. Towards sustainability in world fisheries. Nature, 418: 689-695.

Pope, JG (1972). An investigation of the accuracy of virtual population analysis using cohort analysis. Res Bull Int Comm Northwest Atl Fish 9: 65-74.

Pope, JG & Shepherd, JG (1985). A comparison of the performance of various methods for tuning VPAs using effort data. J. Cons. Int. Explor. Mer, 42: 129-151

Rice, J & Cooper JA (2003). Management of flatfish fisheries- what factors matter? Journal of Sea Research 50: 227-243.

Rindorf, A. and Lewy, P. 2001. Analyses of length and age distributions using continuation-ratio logits. Canadian Journal of Fisheries and Aquatic Sciences, 58: 1141-1152.

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 Research 56: 584-591.

Simmonds, EJ (2003). Weighting of acoustic and trawl survey indices for the assessment of North Sea herring. ICES Journal of Marine Science, 60: 463-471.

Simmonds, E.J., Needle, C.L., Degel, H., Flatman,S., O 'Brien, C. Pastoors, M.A., Robb, A.P., and Vintner, M. (2001). The precision of international market sampling for North Sea herring and its in□uence on assessment. ICES CM 2001/P:21.

Sparholt, H. and Bertelsen, M. 2002. Quality of assessments. Working paper for SGPA and WGMG. ACFM May 2002.

Sukhatme TV & Sukhatme BV (1970). Sampling theory of surveys with applications. 2nd Edition. Iowa State University Press, 452pp.

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

de Vries, M., Rink, G.J., Groeneveld, K, Beintema, JJ, Poos, JJ, Eltink, ATGW (2003). Market sampling of landings of commercial fish species in the Netherlands in 2002. Internal RIVO report 03.012, of project 406-75502. 317pp.

Electronic sources

Commission Regulation, EC 1639/2001. Establishing the minimum and extended Community programmes for the collection of data in the fisheries sector and laying down detailed rules for the application of Council Regulation (EC) No 1543/2000. http://europe.eu.int/comm/fisheries/doc_et_publ/factsheets/legal_texts/regl_en.htm