CM 2004/K:16

ANALYSES OF ICES STOCK ASSESSMENT RELIABILITY AND PROPOSAL FOR METHODOLOGICAL IMPROVEMENT

By Tristan Rouyer, Didier Gascuel and Emmanuel Chassot.

Abstract

Although stock assessment is a central issue in European fisheries management and it has been used for a long time, a very few studies deal with its reliability. We here analysed the assessment quality for ten fish stocks. We quantified the impacts of errors on the main parameters, and we then used a tuning method based on the best CPUE indices that can be obtained with the tuning fleets available. This approach produced good results. Assessments for 1998 with 2003 data were compared to 1998 assessments, for some stocks said to be badly assessed and for others that did not present any serious problem. We first detailed the haddock (VIA) stock and then we presented general results. TACs for 1999 were calculated for the 1998 and 2003 assessments for comparison purpose. The errors found were decomposed into effects caused by the main factors. Results show variable quality from stock to stock. We found stocks with high differences between the two assessments and with important TAC errors, whereas other ones displayed little differences and low errors on TACs. Decomposing the error into different parameters puts into relief the parameters highly sensitive relatively to the stock assessed. The new method used for tuning the VPA allowed in a general way to reduce the assessment errors by improving fishing mortality and stock number estimates. This underlines that few quality data lead to better estimates than important data set of variable quality. The approach presented appears very useful and promising to improve stock assessment.

Introduction

Fisheries management in Europe mainly lies on single-species approach based on cohort analysis principles developed by Gulland (1965), and leaned on some of Beverton and Holt (1957) hypotheses. This classic population dynamics framework allowed to develop powerful tools, as the Virtual population Analysis (VPA), for the study of fish population dynamics. The VPA method consists in studying abundance through catch data in order to produce estimates of fishing mortality and numbers at age, that explain the stock history and its recent trends. The Extended Survivor Analysis (XSA) developed by Shepherd (1999), is an extension of the Survivors method of Doubleday (1981). It is tuned VPA, a method using abundance indices to calibrate the VPA, that is still a problematic point for fish stocks assessment with several methods developed to solve it. Today XSA is a method widely used for management purpose and particularly by the ICES Working Groups to assess fish stocks. These assessments provide a scientific basis on which the Advisory Committee on Fishery Management (ACFM) produce management advice. This roughly consists in stating a Total Allowable Catch (TAC) that will be used in the discussion process leading to the determination of fishing quotas. Even if the methods used in this process have been applied for a long time, a few studies deal with the quality and the reliability of the results obtained.

This study is part of the European Advisory System Evaluation program (EASE), that aims to "*evaluate the performance of assessments, as well as the usefulness of the corresponding advice for management*". This work is an attempt to use the important available log-time series data in order to tackle these subjects and to consider possible methodological improvements. A selection of stocks was overviewed with the three following objectives : i) Quantitatively evaluate the impact of assessment errors ii) Consider possible correction iii) Decompose these errors in their main components. This study does not aim to solve assessment errors, but to provide a concrete vision of their management impact for different stocks. A range of error sources is proposed and appears very promising to reduce a certain extent of the errors.

Materials and Methods

Data source

Data were extracted from the reports of the Working Groups on 2003 and 1998 assessments (ICES 1999, ICES 2004), corresponding to the selected stocks. We also used reports of the ACFM to obtain details on TAC determination, for the same years and the same stocks.

Notations

In a general way, the following notations are used in the whole study. By convention, parameters will be noted A_j^i, where "i" is the year of the assessment from which "j" the year of the parameter estimate, is considered. For example, F_{1995}^{1998} is the fishing mortality for 1995 estimated in the 1998 assessment.

Stocks selection

The objective was to get a sample of stocks representative of different cases that potentially affect the assessment quality in different ways. We first selected stocks for which the quality of assessment was questionable. Ten stocks were thought to be a sufficient number to examine the different stock states. We took the situations described by Lesueur (2003) as a basis for this selection, and the stocks were chosen according to the classification made in this study. The criteria used were the significance and the variability of the mean bias on the spawning stock biomass and on the fishing mortality. According to these criteria, we also selected stocks without any apparent assessment problem.

General method

The first assumption was to consider that the knowledge on a given stock is improving each year by different means. The first one is the convergence property of the VPA (Jones, 1961; Tomlinson, 1970; Pope, 1972). The error done on the fishing mortality estimates, and in particular the one of the last year, is decreasing at each following year of assessment. In addition, the ICES Working Groups work on the more recent and reliable available data, it is incremented at each assessment with the new year information and eventually reviews are added when

consistent scientific information is available. We assumed that the last assessment gives the best estimates for all parameters, and it was taken as the reference one.
The comparison between this reference and a previous assessment should allow to quantitatively evaluate the errors that occurred between these assessments. We assumed that five years previous to the reference was a sufficient time span for changes, and allow the convergence property to take a noticeable effect. The 1998 assessments were chosen for the comparison as the 2003 assessments were the more recent available data. The idea was to reprocess the 1998 assessments with the 2003 data, in order to compare the results with those from original 1998 assessments. Thus, the time period used for the comparison was the widest in common of both assessments, 1998 and 2003 (Fig. 1).
All the errors are percentages calculated on the reference basis (2003 assessment).

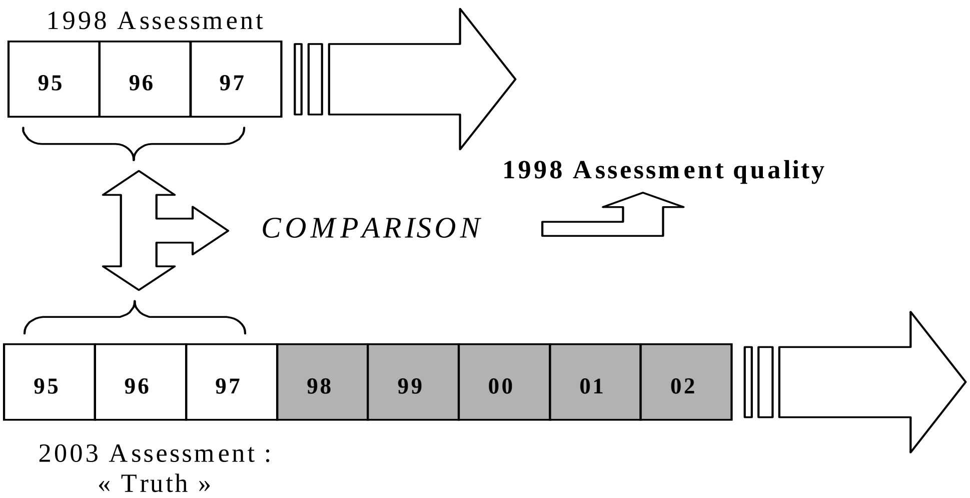

Fig. 1: Illustration of the general method used

Assessment comparison

The stock of haddock (VIA) was first detailed, and then results were given for all the stocks. A yield per recruit was realised with the 2003 assessments data and plotted with the one from the 1998 assessments. Each parameter needed was calculated with the same method as for the 1998 assessment, but with the 2003 assessment data. For example the exploitation pattern used for the 1998 assessment is generally a vector, constituted at each age with the mean of the fishing mortality calculated over the three last years, here 1995, 1996 and 1997. So we used as exploitation pattern a vector, composed at each age with the mean of the fishing mortality calculated over the three same years from the 2003 assessment data. The short-term predictions of the total catch in 1999 were plotted and compared using the same method than the 1998 assessment for the parameters calculation. The curves obtained reflected by their difference the global error appeared between the two assessments.

TAC calculation

The TAC calculation was determined using the same method as the ACFM. Results of the short-term predictions for the total catch in 1999 were used as it is explained. The TAC_{1999} is the total catch in the short-term forecast, corresponding to the target fishing mortality given by the ACFM. For example F_{pa}, was often used as target fishing mortality. The F_{ref} was often the mean of the fishing mortality calculated on the most exploited ages of the exploitation pattern. Then the $F_{multiplier}$ corresponding to the F_{target} was determined with the following relation :

$$F_{target} = F_{ref} \times F_{multiplier}$$

For the TAC_{1999}^{2003} calculation, F_{target} was the one from the 2003 ACFM report. All the parameters used were those from the 2003 assessment : exploitation pattern, catch weights at age, natural mortality, 1998 recruitment, 1998 number at age.

New tuning method

The last years of an assessment are the most important for management objectives, in particular they are often used to calculate the exploitation pattern for the 2 or 3 years to come. However the parameters estimates of these last years are often the worst as they are provided by the very last data, the most unreliable part of the series. The calibration of the VPA improves the survivors estimate, and thus has a great effect on the last years parameters estimate. The method widely used in Working Groups use Catch Per Unit Effort (CPUE) data provided by professional and scientific fleets. The quality of the data provided by these tuning fleets is decisive for the reliability of the estimate produced by this process, and then for management advice. The objective was to improve the tuning process by a quality-oriented approach on tuning fleets. The first step was to find a criterion that indicates the tuning fleets quality. The principle of tuned VPA is to dispose of CPUE, the tuning fleets, that efficiently track an abundance index (e.g., for example the stock number at age). Our point was to produce the best CPUE indices calculated with the tuning fleets available in assessments according to their quality. So we tried to find the most parsimonious selection of fleets and ages that could produce CPUE that track the most efficiently the stock number at age. We retained the coefficient of determination (R^2) as a criterion that indicates the quality of the correlation between the stock number at age and the CPUE series.
As a single time period was needed in order to compare the R^2 among the fleets, it has been calculated for each age over a 10 year period, even if the CPUE time series were longer. This time span was assumed to be long enough to underline the real consistency between the abundance at a given age and the CPUE at this same age given by the different fleets. In addition, a ten years long time period was short enough to focus on the recent trends more than the old one, that could have lead to a good R^2 despite a bad correlation in the most recent years. As well this time span was more simple to manipulate as tuning fleets have very different time series, and because most of them start before 1988. So it allowed us to dispose of an homogeneous and directly comparable data set.
With these criteria we made a selection of ages and fleets. We tried not to check the ages separately, but to consider the whole fleet consistency. Assuming that each fleet targets a specific group of ages, those that seemed not to be concerned by this group (mainly the youngest and the oldest ages) were excluded. However an age with a bad R^2, situated in this group was always selected. Our point was to rely on the fleet consistency more than the ages consistency.

Furthermore, if we assumed that an age was badly tracked by the CPUE in this group, there was no reason to be more confident with the other ages information and in this case, the bad correlation brings an information that should be conserved.
The idea was to find the best abundance indices as possible. A new CPUE was calculated over the 10 year. For each age and each year, it is a weighted mean of the selected fleets.

$$CPUE_{i,j} = \sum_{k=1}^{n} \left(\frac{\alpha_k \times CPUE_{i,j,k}}{\sum_{k=1}^{n} \alpha_k} \right)$$

Where n is the number of fleets, k is the fleet indices, i is the age, and j is year.

This CPUE was calculated in order to give more weight to good quality fleets, it is to say with high R^2. The new CPUE was calculated with the criterion of maximizing the mean R^2 calculated over the ages composed by the weighted CPUE from the different available fleets. The maximisation process used allowed us to find the CPUE weights that maximized R^2. We obtained a new CPUE matrix calculated with these weights, that we used as a tuning fleet. All the tuned VPA done, were computed by the software VPA95, the most used by Working Groups. This allowed to use exactly the same VPA procedure for all processes. Then the new set of CPUE obtained was input in this software as the only tuning fleet (Fig. 2.). The software was run with the same settings than those used by the Working Group for each 1998 stock assessment. The new stock number at age matrix obtained was then computed into the maximisation procedure. Again, the new weighting found produced a new CPUE matrix, used in the VPA software. This chain was reprocessed until the parameters were stabilised, but was stopped when an inconsistency occurred (for example negative CPUE). The Stabilisation of parameters achieved, the fleet the most useless was deleted. It is to say that the fleet corresponding to the smallest α_k was not used by the end of the procedure. All the process was done again until the mean R^2 dropped or until one fleet remains (this never occurred). The objective was also to find the more parsimonious set of tuning fleets, as modelling too many fleets leads to take into account noise and to model it. When the convergence parameters occurred we tested the difference between the new R^2 found and the former. The procedure was stopped when the R^2 significantly dropped, this meant that the last fleet removed contained important information.

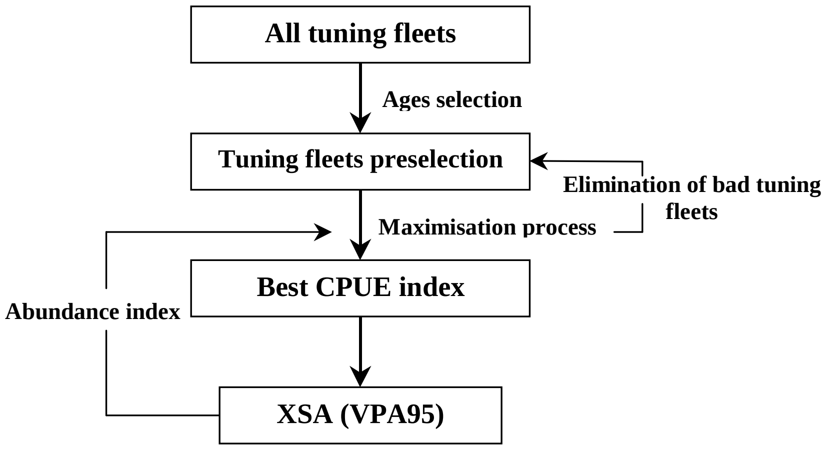

Fig. 2. General outline of the tuning method

As the stock number at age from the 2003 assessment was still considered as the best, we used it for validation purpose. The fleet selection that resulted from the procedure described was compared with another fleet selection obtained using the 2003 stock number at age matrix. It was made on the assumption that if the fleets selected during the procedure really tracked efficiently the abundance, it still must be the case with the 2003 assessment data. The ages' selection should be roughly the same when using each of the stock number at age matrices. This also should allow to check when the process did not manage to find any consistent result, whether it was possible to find a CPUE weighting with the 1998 tuning fleets that was well correlated with the 2003 stock number at age data.

Error decomposition

The TAC errors found were decomposed in the main parameters used for the calculation. The decomposition is realised with the correction made on the parameter estimates. The recruitment was corrected by the use of the 2003 recruitment estimate, R_{1998}^{2003} in spite of R_{1998}^{1998}, the exploitation pattern and the stock number at age by the use of the results found with the new tuning method.
As the reference points for each stock were considered to be related to the stock biology, improvements of these parameters were taken into account for the improvement of the TAC determination. The residual error is assumed to be caused by the VPA convergence and the reviews made by the working groups, between 1998 and 2003. This last point is mainly composed of changes in the catch at age matrix and in the number of age groups.

Results

Selected stocks

The ten stocks selected are demersal stocks mainly located in the north of Europe. According to Lesueur (2003), excepted for Megrim, Plaice and Saithe, all the stocks showed significant mean bias on fishing mortality and on SSB (excepted for Monkfish). The selection was thus composed of stocks whose parameters were said to be badly estimated, and with other stocks without any apparent bad estimate (Table 1).

Table 1: Selected stocks

Stock	Scientific Name	Area	Working Group
Haddock	*Melanogrammus aeglefinus*	VIA	WGNSDS
Sole	*Solea solea*	VIIE	WGNSDS
Southern Hake	*Meluccius merluccius*	VIIIC, IXA	WGHMM (03) / WGSSDS (98)
Megrim	*Lepidorhombus whiffiagonis*	VIIB,C,E-K, VIIIA,B,D	WGHMM (03) / WGSSDS (98)
Anglerfish	*Lophius budegassa*	VIIB-K, VIIIA,B	WGHMM (03) / WGSSDS (98)
North Sea Cod	*Gadus morhua*	IIIA, IV, VIID	WGNSSK
Haddock	*Melanogrammus aeglefinus*	IV, IIIA	WGNSSK
North Sea Saithe	*Pollachius virens*	IV, IIIA	WGNSSK
North Sea Sole	*Solea solea*	IV	WGNSSK
North Sea Plaice	*Pleuronectes platessa*	IV	WGNSSK

Comparison of the stock state

For the haddock VIA, a global trend of F underestimation was shown. This underestimation increased in the last years of the assessment. A general trend to overestimate the SSB was also noticed, in relation to the fishing mortality behaviour. The recruitment was also widely overestimated especially in the three last years. The stock perception is generally too optimistic in the 1998 assessment than in the 2003 assessment.

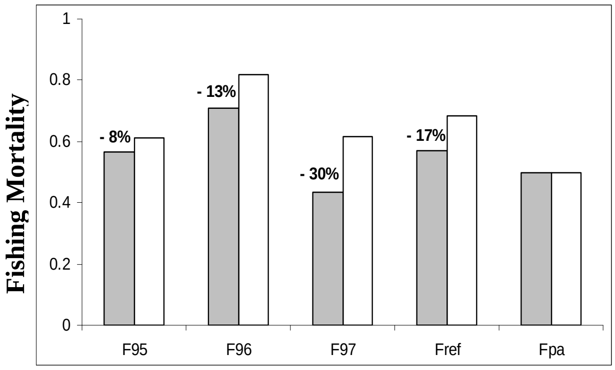

Fig. 3.a. Haddock VIA, error in fishing mortality. 1998 assessment in grey, 2003 assessment in white.

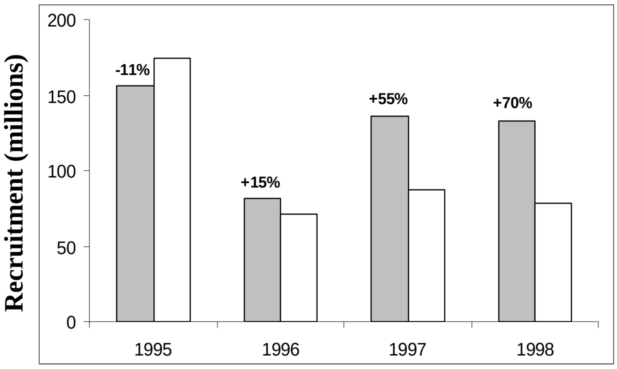

Fig. 3.b. Haddock VIA, error in recruitment. 1998 assessment in grey, 2003 assessment in white.

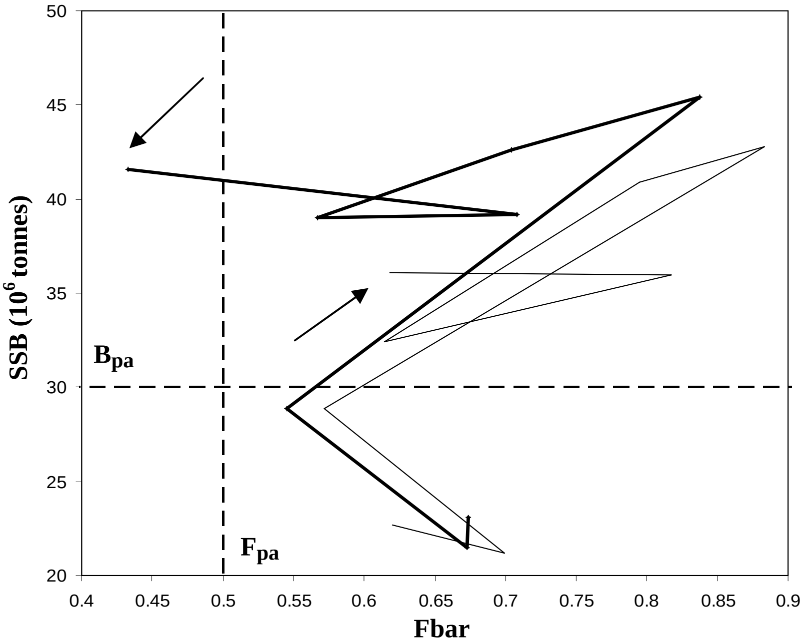

Fig. 4. SSB/F graphical display from 1990 to 1997. 1998 assessment in bold line, 2003 assessment in thin line. Arrows points year 1997.

The error on the main parameters estimate used for the yield per recruit, the short-term projections and for the TAC determination are detailed in Fig. 3.a. The exploitation pattern, a mean of the three last years fishing mortalities, was underestimated by 17%. The F_{1995}^{1998} , F_{1996}^{1998} and F_{1997}^{1998} estimates showed an increasing underestimation from 8% to 30%, that explain the underestimation of the exploitation pattern . The fishing mortality for the precautionary approach did not change, it was not reviewed between the two assessments. The Fig. 3.b shows a growing overestimation of the recruitment over the years considered, this as well in relation with the underestimation of the fishing mortality as an increase in F leads to decrease the stock number estimates. The stock number at age vector used for the short-term prediction and for the TAC determination in the 1998 assessment, was composed with the 1998 XSA estimate. For this stock, the recruitment was calculated by a software that lead to an important overestimation for R_{1998}^{1998}.

The stock perception greatly changed between the two assessment (Fig. 4). The 2003 assessment showed the same main pattern as for the 1998 assessment, but an effective slide towards greater fishing mortalities and lower SSB was noticed. This slide was greater in the last years of the assessment as it was noticed previously. This figure also shows a change in the relative position with the reference points used to define safe biological limits. The VIA Haddock was above B_{pa} in the two assessments, but the 2003 assessment showed a decrease in SSB. On this figure, F_{1997}^{1998} was below F_{pa} whereas F_{1997}^{2003} was above. These two parameters followed the trend described previously, but the fishing mortality crossed the boundary of F_{pa} towards overexploitation. This clearly showed an optimistic point of view induced by the 1998 assessment. The stock was totally outside safe biological limits according to the 2003 assessment. This conclusion could not be the same for the 1998 assessment, where the 1997 plot was inside the safe area.

As it is illustrated in Fig. 5.a, the yield per recruit was not very affected by the underestimation of the fishing mortality. F_{sq}^{2003} was higher than F_{sq}^{1998}, nevertheless it did not change the yield. It was just like a slight spreading of the curve on the right. The stock exploitation was not changed through the yield per recruit diagnosis, as the yield obtained with the 2003 assessment data was quite the same as the ones obtained with the 1998 assessment data, despite an increase in fishing mortality.
On the contrary, the total catch predicted for 1999 was largely different when using the 2003 assessment data than these from the 1998 one (Fig. 5.b). The curves plotted showed a great decrease of the yield. The fishing mortality at status quo increased, whereas the predicted total catch decreased.

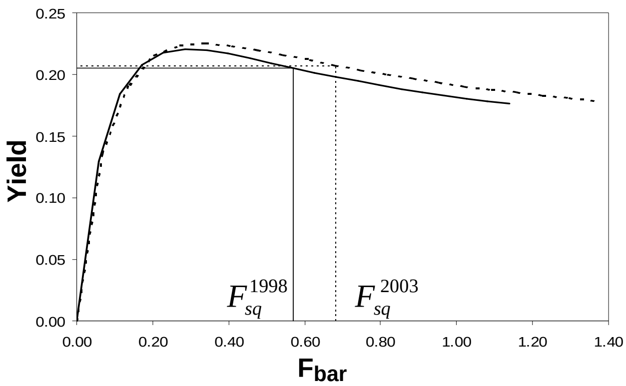

Fig. 5.a. Haddock VIa, yield per recruit. 1998 assessment in plain line, 2003 assessment in dotted line.

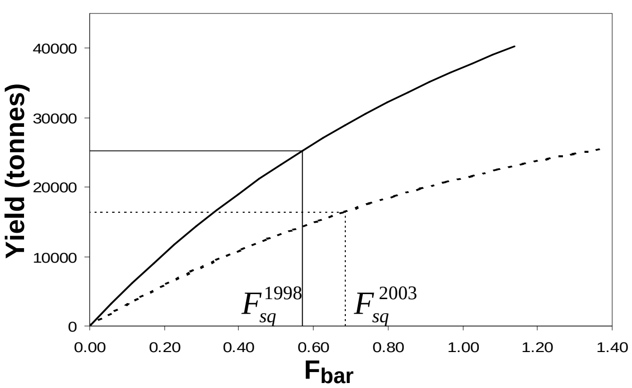

Fig. 5.b. Haddock VIA, short-term prediction, total catch in 1999. 1998 assessment in plain line, 2003 assessment in dotted line

An attempt to improve the TAC estimate

The 1998 assessment data contained three available fleets for the VPA tuning. We used all the fleets as there was no reason to reject any of them. The SCOGFS fleet was selected with all its ages. Age 1 was deleted for the two other fleets as their R^2 was very low, and age 8 was deleted (Table 2). The maximisation process resulted in rejecting the IREGFS fleet after three iterations, as it did not provide any important information. The SCOGFS fleet finally obtained a weight of 3.5 and the SCOLTR fleet a weight of 1. This weighting seemed logical when considering individually the R^2 calculated at each age.

Table 2: Tuning fleets available and the selections operated in the 1998 assessment and in our tuning method

Available tuning fleets	Available ages	1998 assessment	New tuning preselection	New Tuning final fleet selection
IREGFS	Age 1 – Age 5	Age 1 – Age 5	Age 2 – Age 5	
SCOGFS	Age 1 – Age 6	Age 1 – Age 6	Age 1 – Age 6	Age 1 – Age 6
SCOLTR	Age 1 – Age 8	Age 1 – Age 7	Age 2 – Age 7	Age 2 – Age 7

The results obtained for the reference fishing mortality estimate are presented in table 3. The estimate obtained with the new tuning method showed an increase in value. Globally this new estimate constituted an improvement as it was closer to F_{ref}^{2003} than F_{ref}^{1998}.

Table 3: Fishing mortality vectors used as exploitation pattern and their error, obtained by the 1998 and the 2003 assessments, and by the new tuning method.

	F^{2003}	F^{1998}	Error (%)	F new tuning method	Error (%)
Age 1	0.2861	0.3159	10.4	0.3125	9.2
Age 2	0.5202	0.5122	-1.5	0.504	-3.1
Age 3	0.6618	0.6281	-5.1	0.6538	-1.2
Age 4	0.7444	0.5839	-21.6	0.6458	-13.2
Age 5	0.7444	0.5585	-25.0	0.6165	-17.2
Age 6	0.7444	0.5637	-24.3	0.6356	-14.6
Age 7	0.7444	0.7604	2.1	0.8937	20.1
Age 8 (+)	0.7444	0.7604	2.1	0.8937	20.1
F_{ref} = F_{bar} (2-6)	0.68	0.57	-16.2	0.61	-10.3

In order to investigate further, the age structures of the fishing mortality vectors used for the 2003 and the 1998 assessment were plotted (Table 3; Fig. 6). The exploitation pattern for the TAC determination obtained with the new tuning method, showed a slight decrease at ages 1 and 2 in comparison with the 1998 assessment estimate. But all the following ages showed an increase of their value. Comparing to the 2003 assessment estimate, the new tuning method estimate was greater at ages 7 and 8 (+ group), but below for the previous ages.

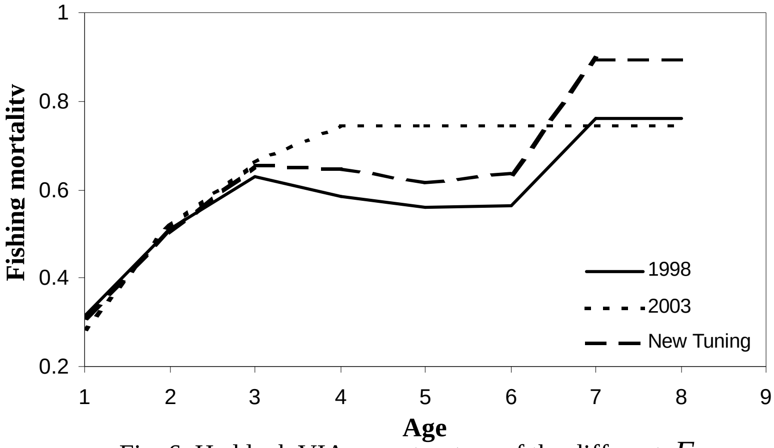

Fig. 6. Haddock VIA, age structure of the different F_{ref}

The curves showed the same general pattern for the 1998 assessment and the new tuning method. The plateau observed for F_{ref}^{2003} might be caused by the TSA software, that should give such a profile to fishing mortality. In addition, this figure showed that the most important underestimation of F_{ref}^{1998} was at ages 4 to 6. Our tuning method produced an improvement for those ages and resulted in the same estimates as for the 1998 assessment at ages 1 to 3. There, it appeared clearly that the exploitation pattern underestimation occurred at ages 4 to 6. However, the reference F used to get the level of effort, is calculated from age 2 to 6. As a consequence, this level of effort was greatly influenced by the underestimation at ages 4 to 6, even if their impact in term of catch weight was reduced compared to age 2 and age 3. The calculation of the level of effort gave the same importance to all ages, despite their different relative impact on the exploitation. For the VIA Haddock, this kind of calculation should have contributed to the underestimation of the level of effort and thus to the TAC underestimation.

The Tuning effect in the error decomposition took into account the change of the exploitation pattern and the change of the stock number at age. The exploitation pattern correction taken individually lead to a weaker reduction in the error. But as the fishing mortality was here directly related to the stock number at age (only the tuning fleets changed, all the options and other input data were the same) there was no sense to consider each one separately.

The other source of the error were not investigated as it would have been necessary to rebuild the historic evolution of reviews. For VIA Haddock, catch number at age and weight at age data was not available.

Table 4: 1998 Stock number at age vectors and their error, obtained by the 1998 and the 2003 assessments, and by the new tuning method.

	N_{1998}^{2003}	N_{1998}^{1998}	Error (%)	N_{1998} new tuning method	Error (%)
Age 2	54997	89691	63.1	72723	32.2
Age 3	20703	20586	-0.6	24438	18.0
Age 4	22347	26021	16.4	27178	21.6
Age 5	2520	3391	34.6	2656	5.4
Age 6	2401	4893	103.8	4236	76.4
Age 7	875	1770	102.3	1325	51.4
Age 8 (+)	274	294	7.3	221	-19.3
Total	104117	146646	40.8	132777	27.5

N_{1998}^{2003} showed the lowest total population (Table 4). Age 2 and age 6 seemed here to be the most overestimated ages in the 1998 assessment. The new tuning method results presented improvements for the age 2, age 5, age 6 and age 7 estimates. The use of the 2003 assessment data to select the tuning fleets and the ages, showed the same results than the 1998 assessment data. The maximisation process gave the SCOGFS fleet a weight of 4 and a weight of 1 to the SCOLTR fleet, it is to say roughly the same weighting.
Fig. 7 shows the resulted curves of the short-term projection for the total catch in 1999, with the correction of different main parameters, allowing a decomposition of the error. A first sight on the curves show that each parameter brought noticeable correction.

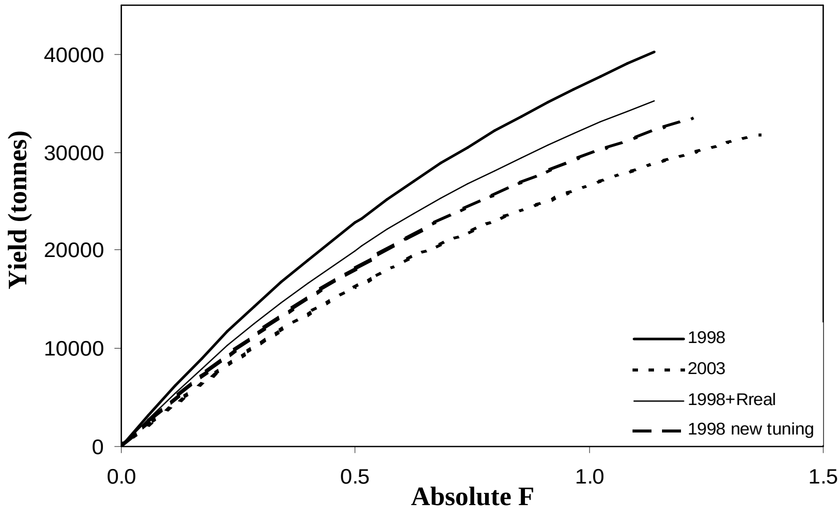

Fig. 7. Haddock VIA, error decomposition of the short term prediction for the total catch in 1999.

The decomposition error showed a great effect of the recruitment overestimation in the 1998 assessment (Fig. 8). 42% of the error on the TAC calculation was caused by this estimate, it is to say about 3000 tonnes. The new tuning method decreased the error by 27%, that corresponded to a 1000 tonnes decrease on the TAC estimate. Finally it still remained a 2000 tonnes error that may be caused by the different parameters whose corrections were not investigated.

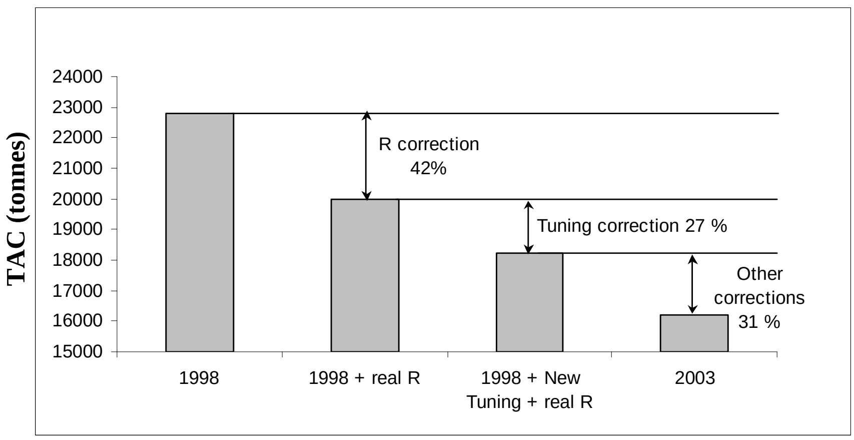

Fig. 8. Haddock VIA, error on TAC and effects of corrections in % of the error.

Results for all stocks

The results of the TAC_{1999}^{1998} and TAC_{1999}^{2003} determination are given in Table 5. The error between the two TAC had a high variability regarding the stock considered. Table 5 shows different cases, some with very important TAC errors and others that were not that affected. Some very important errors were noticed for the TAC determination, they ranged from 171% (southern hake) to –37% (north sea saithe). A large group was composed with misestimate in TAC stocks. The error ranged from 171% to 40% for six stocks, from 21% to –37% for the other four stocks remaining.

Table 5: TAC_{1999}^{1998}, TAC_{1999}^{2003} and the error for the selected stocks.

Stock	TAC_{1999}^{1998} (tons)	TAC_{1999}^{2003} (tons)	Error (%)
Southern Hake (VIIIC, IXA)	9 500	3 500	171
Monkfish (VIIB-K, VIIIA,B)	9 200	5 200	77
North Sea Plaice (IV)	106 000	63 500	67
North Sea Cod (IIIA, IV, VIID)	147 000	98 600	49
Sole (VIIE)	670	470	43
Haddock (VIA)	22 800	16 300	40
Haddock (IV, IIIA)	114 000	94 000	21
North Sea Sole (IV)	20 300	17 000	19
Megrim (VIIB,C,E-K, VIIIA,B,D)	17 400	20 500	-15
North Sea Saithe (IV, IIIA)	104 000	164 000	-37

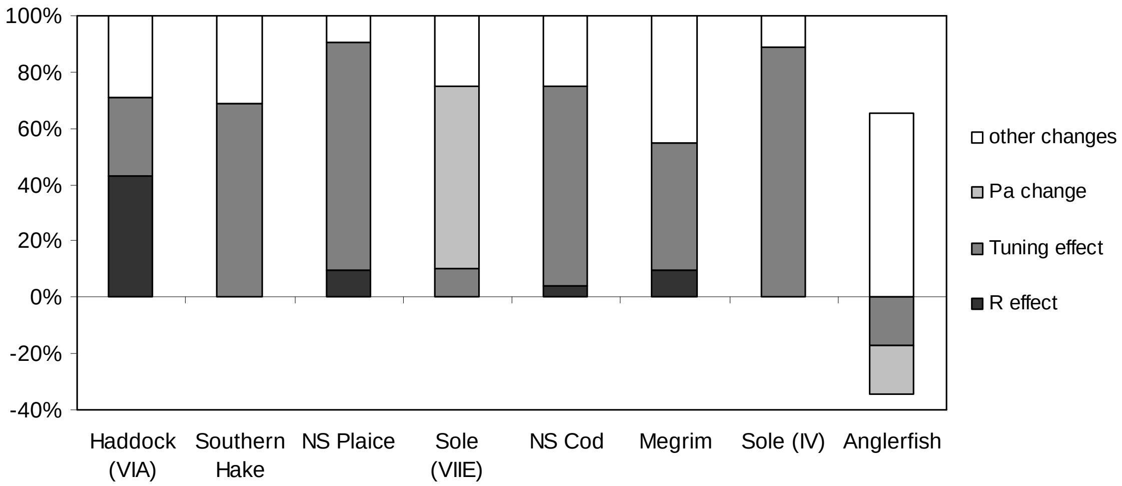

Fig. 9. Percentage effect of different parameters in the TAC error of 8 stocks

Table 5 allows to divide the selected stocks in different categories. Here the North Sea Saithe and the Megrim stocks had their TAC underestimated. The Megrim, the Haddock (IV, IIIA) and the North Sea Sole stocks showed an absolute error below or around 20 %. The stocks of Haddock (VIA), Sole (VIIE), North Sea Cod and North Sea Saithe showed a greater absolute

error, below 50 %. Finally, the stocks of Southern Hake, Monkfish and North Sea Plaice showed a high error.
The extremely high error showed by the southern hake case could be explained by the important change of the stock perception between the two assessments. The 2003 catch at age matrix is not the same as the 1998 one, in the 2003 assessment no discards data were available and catches had a different age distribution. The fishing mortality was reviewed upward with the main difference at the ages1 and 2, the most represented into catches. In addition, the 2003 ACFM report did not determined any Fpa, and we used the 1998 one that could be involved in that important error.
Results of the error decomposition are presented in figure 9. The noticeable effect of the R estimate for the stock of Haddock (VIA) is shown. The correction of this parameter had a rather important impact on the TAC determination of Plaice and Megrim stocks. In many cases the new tuning method used brought important improvements on the TAC determination, the correction impact ranged from 10 % (Sole VIIE) to 85% (Sole IV). The tuning correction was particularly effective for the stocks of Southern Hake, North Sea Plaice, North Sea Cod and North Sea Sole. The impact of other and no investigated changes remained important for all the stocks, with an impact from 10 % (Plaice) to 50% (Megrim).

Discussion

The results showed that errors in the different parameters could be very high and that they are very variable considering the different stocks. In terms of management, these errors could have more or less impact. For example, the 1998 recruitment estimate is a very important point for the haddock (VIA) TAC determination, and thus should be studied with caution. The particular case of VIA Haddock stock clearly showed that each stock is a particular case. Haddock is characterized by an important impact of the young ages on the catch weight that emphasizes the importance of the recruitment estimate for the short-term predictions. Commonly, the estimation of this parameter is a central problem in fish assessment. Here, in a deterministic model, an error in the estimate directly affects the results. The 1998 recruitment estimate represented the age 2 catches for the short-term prediction of total catch in 1999. It was the most exploited age, with 30% of the total catch in weight. And the fishing mortality at the first ages were rather high compared to those of other stocks. Thus this recruitment estimate has an important impact on the TAC calculation. R_{1998}^{1998} was provided by a software that use survey data to provide year class recruitment estimates. Otherwise, the geometric mean is often calculated on the previous year recruits, over a given period. Here if this geometric mean had been used, the error would have been reduced by about 3000 tonnes, as this recruit estimate was just slightly superior to the value from the 2003 assessment. This same parameter estimate had absolutely no impact in the TAC determination for the southern hake. However, as for haddock (VIA) some accurate methods for recruiting estimate produce less reliable results than the geometric mean over an appropriate historic period. The yield per recruit did not change a lot with the errors on the different parameters, it seemed to be relatively robust to the fishing mortality changes. But here the most important diagnosis for the assessment and for management purpose is the 1999 short-term yield forecast. However this diagnosis is more sensitive to the different parameters and errors could greatly affect the results. It emphasizes the need to take into account the impact of the potential errors in the management advice. For example different scenarios could be considered, taking into account the variability and the maximum values reached for the most sensitive parameters.

The tuning method used is an adding to the classic one. It provides an improved CPUE matrix and reduce the number of tuning fleets. The XSA principles were not changed, we only produced a new tuning matrix with the original one, and we input it on the software. The difference with the only use of the XSA software was that fleets were selected before their input in the software, that allowed a more objective choice. XSA does not reject any fleet (excepted on the user demand) and it does not weight them prior to their computation. Here we selected the best set of fleets which were consistent among them on the criteria of the most efficient abundance tracking. The XSA principles are still used, but we added to this the use of the obtainable CPUE instead of several fleets with different quality. This method surely needs further investigations to validate it more securely and to test it in different situations. But if the results presented here about are preliminary and have to be taken with caution, the fact is that this method produced improvements. The error between TAC_{1999}^{1998} and TAC_{1999}^{2003} decreased whatever the stock studied (excepted for anglerfish). Furthermore, other criteria that could prove the improvement of the assessment were quickly checked for VIA Haddock and Southern Hake. The error for the fishing mortality at ages and the stock number at ages matrixes were calculated for each method. Our method gave the lowest error in each case. The method was not as efficient for each stock, and it results provided very variable improvements.

This method seemed to have greater effect for tuning fleets with bad R^2, and in the opposite stocks whose fleets had a good R^2 were not that improved. This correlation was of the same kind for the TAC error, the fleets quality determined the error done. The greater the error, the worst the fleets. This last point was confirmed by the Anglerfish stock. The error on the TAC and the fleets R^2 were maybe the worst observed, so that we did not perform to get any improvement for this stock. But here we obtained totally different exploitation pattern that seemed to be more consistent with the 2003 assessment, than the one obtained in the 1998 assessment. In some way this method gave a good sight of the fleets quality that was related to the quality of the assessment. Some stocks had a very good R^2 and main of their tuning fleets, for example for Sole (IV) the R^2 reached 0.98 for many ages and fleets, the TAC error was small. For the Anglerfish, the R^2 hardly reached 0.5 for some ages in the better fleets and the TAC error was big. This underlines the data problem, some were so bad that there were no accuracy expectable and the results remained unreliable whatever the method used.

The data quality is underlined and whatever efficient XSA is, bad data lead to unreliable results. This method gave to the user a sight on the fleets quality and the choice of the coefficient of determination as a criterion was convenient and easy to interpret. But in the case of short time tuning series (less than ten years), comparison between the fleets became impossible. So the fleet choice were very subjective, the main idea was to consider whether a fleet brought information or not. In addition, some fleets could be acceptable with this method, and only relevant on statistical arguments that required good knowledge on fleets and fisheries would allow to make a good choice. Here it was necessary to add to the theoretical approach a concrete knowledge. And if a good knowledge was required, the quality and the abundance of information about stock remained a critical point. The need to have of the maximum of relevant information was here again underlined.

The consistency of this method was confirmed by the choices made by the working group that studied the stock considered. We often selected the same ages and roughly the fleets rejected by the working group were the same excepted when the justification was based on empirical knowledge. It appeared in some cases that we selected very good fleets that were rejected by the working group on relevant non-statistical criteria. The use of the 2003 assessment abundance matrix with the method allowed us to validate our results. We obtained the same selection in many cases, only the ages set was sometimes slightly different.

However this method gave promising results, some points needed to be clarified. We here assumed that the settings used by the 1998 working group were the best, and we did not changed

them. However using this method questions this assumption as the options could be no longer relevant with it. In particular the case where the settings are the less effective could be examined, in order to let the data speak. The convergence of weighting parameters achieved would allow to find the best settings to use.
In addition, as this was expected not to noticeably change the results, we decided not to take into account the alpha and beta coefficients for the scientific fleets (surveys), which represent the beginning of the fishing period and its end respectively. We put alpha at 0 and beta at 1, like a professional fleet on the assumption that the mix of fleets and weighting might reduced the potential effect of these parameters.

Conclusion

This study showed that stock assessment errors do not systematically lead to an important impact on management. Here the degree of impact on management is linked to the stock and the parameters concerned by the errors. Taking into consideration sensitive parameters and their possible range of variation could allow to consider extreme situations that could be reached. Then a wiser management taking into account different possible scenarios could be settled, following a precautionary approach.
Besides, this analysis also deals with the issue of data quality and its effect on the reliability of assessments. The quality of data used in assessments seemed not to be sufficient to expect reliable results. Using the tuning method proposed could improve results and help the selection of tuning fleet made by working groups. A software using this method could be developed and added to VPA95, in order to simultaneously use both processes and make them easier to use.

References

Beverton, R.J.H. and Holt S. J. 1957. On the dynamics of exploited fish populations. UK, London, 533p.

Doubleday, W.G. 1981. A method for estimating the abundance of survivors of an exploited fish population using commercial catch-at-age and research vessel abundance indices. Canadian Special Publication Fisheries and Aquatic Sciences **58**: 164-178.

Gulland, J.A. 1971. The Fish Resources of the Oceans. Fishing News Books, West Byfleet, UK .

ICES 1999. Report of the ICES Advisory Committee on Fishery Management, 1998, **229** .

ICES 2004. Report of the ICES Advisory Committee on Fishery Management, 2003, **261** .

ICES 1999. Report of the working group on the assessment of demersal stocks in the north sea and skagerrak. ICES CM 1999/ACFM:08.

ICES 2004. Report of the working group on the assessment of demersal stocks in the north sea and skagerrak. ICES CM 2004/ACFM:07.

ICES 2004. Report of the working group on the assessment of hake, megrim and monkfish. ICES CM 2004/ACFM:02.

ICES 1999. Report of the working group on the assessment of northern shelf demersal stocks.

ICES CM 1999/ACFM:1.

ICES 2004. Report of the working group on the assessment of northern shelf demersal stocks. ICES CM 2004/ACFM:1.

ICES 1999. Report of the working group on the assessment of southern shelf demersal stocks. ICES CM 1999/ACFM:04.

ICES 2004. Report of the working group on the assessment of southern shelf demersal stocks. ICES CM 2004/ACFM:04.

Jones, R. 1961. The assessment of the long–term effects of changes in gear selectivity and fishing effort. Mar. Res. Scotl. **2**: 19p.

Lesueur, M., Gascuel, D., and Rouyer, T. 2003. Control of assessment for demersal fish stocks in ICES area : analysis for 36 stocks and investigation of some potential bias sources.

Pope, J.G. 1972. An investigation of the accuracy of virtual population analysis using cohorta analysis. ICNAF Research Bulletin **9**: 65-74.

Shepherd, J. G. 1999. Extended survivor analysis: an improved method for the analysis of catch-at-age data and abundance indices. ICES Journal of Marine Science **56**: 584-59.

Tomlinson, P. K. 1970. A generalisation of the Murphy catch equation. Journal of the Fisheries Research Board of Canada **27**: 821-825.