Dana, vol. 11(1), pp. 39-64, 1995

# Sampling methods and errors in the Danish North Sea industrial fishery

*Peter Lewy*
Danish Institute for Fisheries Research, Department of Marine Fisheries Biology, Charlottenlund Castle, DK-2920 Charlottenlund, Denmark.

## Abstract

For the Danish North Sea industrial fishery a three-phase sampling procedure involving poststratification on target fisheries and fishing areas has been introduced for calculation of the precision of estimates of catch weight and number at age by species. Three target fisheries have been considered: Sandeel (*Ammodytes lancea marinus*), Norway pout (*Tripsoteris esmarki*) and sprat (*Sprattus sprattus*). Herring (*Clupea harengus*), whiting (*Merlangius merlangus*) and haddock (*Melanogrammus aeglefinus*), the three most important by-catch species, were included in analysis as well. Results for 1993 show that the coefficient of variation of annual catch by weight of the three target species and herring is about 6% or lower and 10-15% for whiting and haddock. The coefficient of variation of annual catch in numbers at age is less than 8% for the same four species and 15-30% for whiting and haddock. Optimum stratified allocation for a specified objective function with fixed costs has been found. The actual sampling allocation is not very different from the optimum implying that the potential gain in precision of estimated catch is limited.

*Keywords*: Industrial fishery, North Sea, sampling error, optimum allocation.

## Introduction

The present work deals with the sampling methods and errors in the Danish North Sea industrial fishery. The objective of the paper is partly to calculate precision of estimates of catch weight by species and catch in numbers by age and partly to optimize the sampling resource allocation to obtain as accurate estimates as possible. Precision of the estimated catch data is important for the quality and reliability of the assessment of a number of fish stocks because they often constitute the main part of input to the assessments.

In most fish stock assessments projections of e.g. catch and biomass are deterministic and do not include stochasticity. This is in correspondence with the assumption that input is known without error.

Several authors have investigated errors in estimated catch and effort data and the implications for projected catch or biomass applying Virtual Population Analysis, VPA, cohort analysis or a tuned VPA (Pope 1972, Pope & Gray 1983, Rivard 1983, Gavaris & Gavaris 1983, Pelletier & Gros 1991). In these cases incorporation of errors in catch data into projections is in conflict with the estimation procedure of fishing mortality and stock size, which assumes that catch data are known without errors.

The so-called integrated methods avoid this contradiction by assuming stochastic errors in input data and utilizing standard statistical methods (e.g. Doubleday 1976, Fournier & Archibald 1982, Deriso *et al.* 1985, Lewy 1988, Sparholt 1990). This implies that the stochastic variation of parameters and for instance projected TAC's and stock biomass can be evaluated by standard techniques. Furthermore, submodels included (e.g. the relation between effort and fishing mortality) and hypotheses concerning the parameters can be tested.

In integrated approaches the variance/covariances of estimated catch at age must be estimated or assumed to be known. In Lewy (1988) a model is described where mean catch at age is assumed to depend on catchability and selection parameters and where the variance of the catch is unknown. These parameters and variance parameters have been estimated simultaneously by Maximum Likelihood. Additional analyses showed that the correlation between estimates of the parameters describing mean catch at age and the variance of the catch was very high, often higher than 90%. This strongly indicates that further information is required. A general conclusion may be that it is not possible to estimate simultaneously both the parameters describing the catches and their variance. A solution to this problem is to get independent estimates of the sampling error by using data obtained from a sampling programme. In principle this must be done for the total international catches. The present analysis is the Danish contribution for some North Sea stocks.

The sampling design is different for the industrial fishery and most human consumption fisheries because only the total catch is known without knowing the species composition in the former. This may also apply to some human consumption fisheries in the tropics where the catch is composed by many species, but of which only the most valuable species are recorded, see Sparre and Venema (1992). In these cases the sampling system has to include estimation of the species composition.

Only very limited work has been done dealing with sampling in such fisheries (e.g. Lassen *et al.* 1973, Lewy 1979), while sampling aspects involved in the human consumption fisheries including determination of age and other attributes have been studied frequently (Doubleday & Rivard 1983, ICES 1994).

In the Danish industrial sampling system the weight of the catch by species and numbers at age are estimated. The stratified sampling system in the industrial fishery described in the present paper introduces a stratification, which differs from the one previously used. Previously the samples for estimation of catch weight by species were stratified by geographical unit, the ICES international rectangles, so that the total industrial catches were separated by ICES statistical rectangle using log book information. Afterwards, the catch by rectangle was split into species components using biological samples from the rectangles available, whereupon numbers at age could be estimated for the various species. The weakness of this system is that biological samples are often missing for some rectangles from which catches are recorded. In such cases alternative 'neighbour' samples must be selected to provide the species composition, which results in subjective methods.

Among various alternatives to stratify by ICES statistical rectangles *target fisheries* defined as a combination of gear type, mesh size, the time period and the area have been used as strata in the present analysis. The reasons for choosing target fisheries as basis for the sampling programme are:

1. *Management reasons*. Usually management of fish stocks is linked to target fisheries. In case of the industrial fishery it concerns for instance by-catch or mesh size regulations in a specified fishery. As a consequence it is necessary to collect data accordingly in order to obtain the relevant data for analysis of these matters.

2. *Homogeneity*. When estimating the species composition of catches the separation of samples into target fisheries ensures homogeneous groups and thereby reducing the variability. However, if the area covered by a target fishery is inhomogeneous with respect to a particular species the variability of this species may instead be increased.

3. *Precision of estimates*. A target fishery is a much coarser stratification than the ICES rectangle previously used. Therefore, it is possible to obtain a large number of biological samples within each strata. This ensures that the necessary data are available for estimating the precision of the estimates in contrast to when rectangles are used as strata. Furthermore, the subjectivity referred to above can be removed totally.

The concept of target fishery has been described e.g. by Murawski *et al.* (1983), Rogers & Pikitch (1992), and Lewy & Vinther (1994).

## Materials and methods

*The Danish industrial sampling system*

Samples from the Danish North Sea industrial fisheries and catch statistics for 1993 were analysed with the objective to estimate the landings in weight and the numbers caught by age group for six species. The information obtained originated from two databases:

- Fishery database
- Biological database

The Fishery database combines information from log books and catch statistics. For each month and for each individual vessel trip the database contains information on: the weight of the landing, the gear and mesh size used and the ICES statistical rectangles where the fishery took place. The species composition was not known from this source.

The biological database consists of samples taken from the industrial landings collected by the Fishery Inspection during the whole year from the largest ports along the west coast of Jutland. Samples were obtained by lowering a 10-litre pail into the hold and taking out a random sample, weighing about 7 kg, from the cargo. The biological samples were analysed with the purpose of determining the species composition, the length distribution, the mean weight by length group and the age distribution by species. For each sample the vessel trip information on gear, mesh size, time and fishing area was available as well.

Full analyses of the biological samples were carried out in three consecutive different steps:

In the *first step* concerning the estimation of the species composition of a sample, total catch of a sample is split into the various species, which were weighed. The proportion by species is obtained by dividing the weight by species by the total weight of the sample. This step was done for all samples collected.

The next two steps were carried out only for a random selected subsample.

For each species the *second step* consists of 1) to count the number of fish and to measure the length of the fish in order to estimate the length distribution, 2) to weigh the fish by length group in order to estimate the mean weight per fish by length group. The mean weight per fish by length group and species is simply the weight divided the number of fish. Total mean weight per fish by species is obtained as the weighted average of the mean weight by length group. The variance of the mean weight by species and length group was calculated as well.

The *third step* is to estimate the age distribution of the fish by species. This was done by age reading of the otoliths. In most cases all otoliths were read. In cases with many biological samples covering a specific fishery and month, a number of samples were randomly selected for which no otoliths were read. In the cases where age determination was carried out and when the otoliths showed a clear separation between the length groups with respect to age most of the fish were aged using the length only. The mean weight by age was estimated assuming that the weight of a fish only depends on the length and not the age.

Using target fisheries as basis for stratification requires of course that both total catch and biological samples can be divided into categories defined to ensure that biological samples are representative for the catches. A necessary condition for this is that the criteria used for definition of target fisheries are identical for both catches and biological samples. As the fishery database does not contain information on the species composition, total catch by month was classified into the target fishery categories applying the following procedure: each vessel trip in the North Sea was classified according to target fisheries using the information on gear, mesh size, period and fishing area included in the fishery database. Combining the catch by trip and month within each target fishery total catch by target fishery and month can be obtained.

For the North Sea industrial fishery in 1993 three target fisheries were considered: a sandeel fishery, a Norway pout fishery and a sprat fishery. The actual definition of the classifications used was based on analyses of the biological samples such that e.g. sandeel should be the dominating species in weight in the defined sandeel fishery etc. In the sprat fishery either sprat or the by-catch of herring should be dominating. This kind of definition is in correspondence with Lewy & Vinther (1994) in which target fisheries also were defined by similarity of the species composition in the catch. The actual used definition was: firstly, only trips were considered for which logbook information was available. Sprat fishery was characterized by the use of pair trawl as gear or that the fishery took place in the first or fourth quarter in the area limited by 2° to 8° east and 54° and 57°. Considering the remaining vessel trips the sandeel fishery was defined by an upper limit of the mesh size used of 10 mm or that the fishery took place in the months March to July in the North Sea area between 56° and 58°. Norway pout fishery was defined as the trips for which the fishing area was north of 57° 30'. Finally, the few remaining trips including trips, where no log book information was available, were considered. In these cases the trips were classified as sandeel fishery if the trip took place in the second quarter. Otherwise they were classified as sprat fishery.

Besides the three target species, sandeel, Norway pout and sprat, the three most important by-catch species were included in the analyses as well. These are herring in the sprat fishery, whiting in the Norway pout and sprat fisheries, and haddock in the Norway pout fishery. No by-catch was considered for the sandeel fishery as the sandeel on average constituted about 95% of the total. The six species accounted for about 90% of the total North Sea industrial catch of 877 000 tonnes in 1993.

Furthermore, catches and biological samples were stratified by fishing area as well. The reason is that data from the industrial fishery and research vessels indicate geographic differences in age distribution and growth. The fishing areas considered are shown in Figure 1. Seven sandeel areas, two Norway pout areas (the fishing areas 4 and 5, Figure 1B) and three sprat areas (the fishing area 1-3, Figure 1B) were considered.

For whiting and haddock, only one area, total North Sea, was considered. These areas are the same as previously used for the estimation procedure for industrial catches by the Danish Institute for Fisheries Research.

A total of 589 biological samples was collected by the Danish Fishery Inspection. For 319 of the samples only the species composition was determined by the Fishery Inspection. The remaining 270 samples were fully analysed by the Danish Institute for Fisheries Research with respect to species composition and for the six species considered the length distribution, weight for each length group and age distribution.

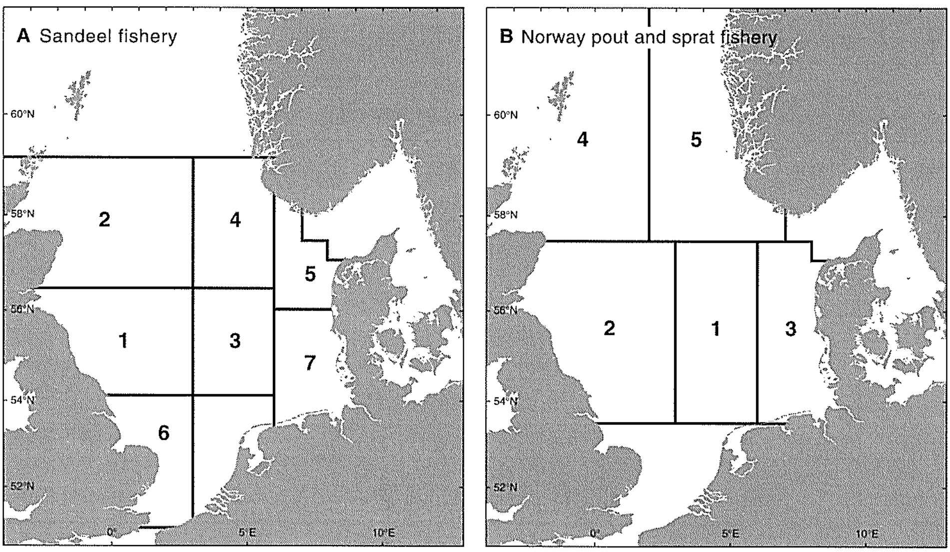



Figure 1. Definition of North Sea fishing areas used as strata.

The number of fish length and weight measured differs from the number of fish aged for sandeel and sprat only. The annual number in 1993 of sandeel and sprat measured are 20 220 and 6194, respectively, compared with 13 660 and 5704 aged fish within the same year.

A summary of catch statistics and sample sizes is given in the Tables 1-3.

For the sake of convenience target fishery is mentioned 'fishery' in the following.

Table 1. Catch in tonnes by quarter and fishery in the Danish North Sea industrial fishery in 1993.

| Quarter | Sandeel fishery | Norway pout fishery | Sprat fishery | Total |
|---|---|---|---|---|
| 1 | 14963 | 36880 | 38563 | 90406 |
| 2 | 372944 | 3345 | 12354 | 388643 |
| 3 | 71857 | 45904 | 120856 | 238617 |
| 4 | 13127 | 64959 | 81059 | 159145 |
| Total | 472891 | 151088 | 252832 | 876811 |

Table 2. Number of biological samples by quarter and fishery in 1993.

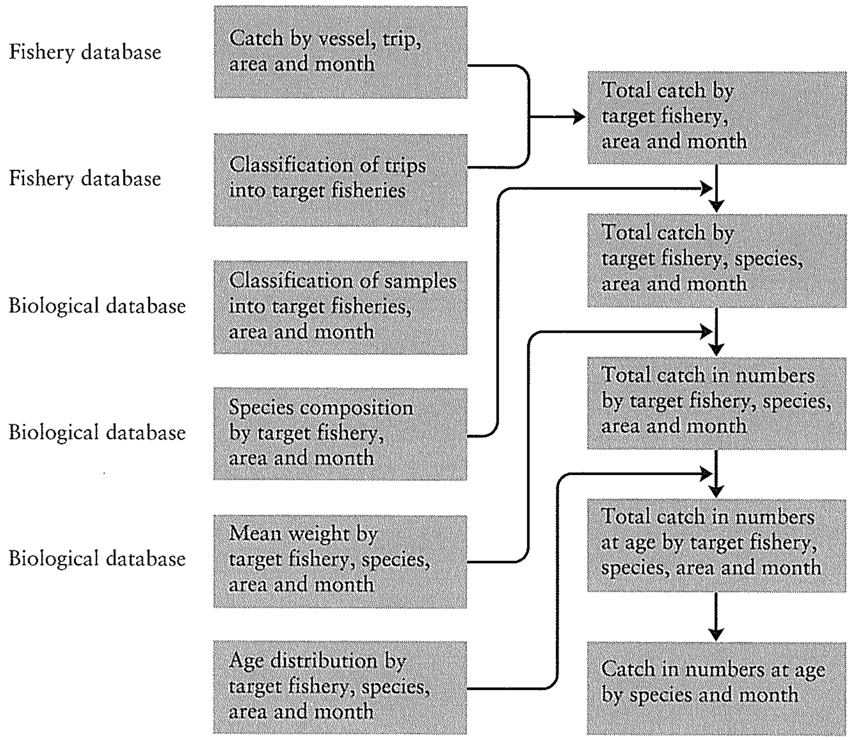
| Quarter | Sandeel fishery | Norway pout fishery | Sprat fishery | Total |
|---|---|---|---|---|
| 1 | 24 | 38 | 32 | 94 |
| 2 | 138 | 0 | 0 | 138 |
| 3 | 50 | 50 | 113 | 213 |
| 4 | 22 | 43 | 79 | 144 |
| Total | 234 | 131 | 224 | 589 |

Table 3. The number of fish aged by fishery, species and quarter in 1993.

| | Fishery | | | Fishery | | | Fishery | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Sandeel | N. pout | Herring | Sandeel | N. pout | Herring | Sandeel | N. pout | Herring | Total |
| Quarter | Species: Sandeel | | | Norway pout | | | Herring | | | |
| 1 | 1568 | 0 | 0 | 0 | 1845 | 0 | 0 | 0 | 1575 | 4988 |
| 2 | 4428 | 0 | 0 | 0 | 441 | 0 | 0 | 0 | 0 | 4869 |
| 3 | 6810 | 0 | 0 | 0 | 1307 | 0 | 0 | 0 | 2762 | 10879 |
| 4 | 854 | 0 | 0 | 0 | 1344 | 0 | 0 | 0 | 1367 | 3565 |
| Total | 13660 | 0 | 0 | 0 | 4937 | 0 | 0 | 0 | 5704 | 24301 |
| Quarter | Species: Sprat | | | Whiting | | | Haddock | | | |
| 1 | 0 | 0 | 790 | 0 | 78 | 59 | 0 | 274 | 0 | 1201 |
| 2 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 4 | 0 | 8 |
| 3 | 0 | 0 | 527 | 0 | 26 | 354 | 0 | 143 | 0 | 1050 |
| 4 | 0 | 0 | 1332 | 0 | 73 | 348 | 0 | 68 | 0 | 1821 |
| Total | 0 | 0 | 2649 | 0 | 181 | 761 | 0 | 489 | 0 | 4080 |
| All | 13660 | 0 | 2649 | 0 | 5118 | 761 | 0 | 489 | 5704 | 28381 |

*Estimation of catch in weight and number at age by species*

The notation used is described in Appendix I. The estimation procedure is illustrated in the flow diagram:




When estimating the proportion of each of the species in the catches it is assumed that the mean proportion depends on the quarter, the fishery and the species only, which means that within a quarter the species composition is assumed independent of the fishing area and the time. This assumption was not supported by analyses, but needs further investigations.

This implies that the catch by month, fishery, fishing area and species is estimated by

$$\hat{L}_{tfgs} = L_{tfg}\,\hat{E}_{q_t fs} \qquad (1)$$

where $\hat{L}_{tfgs}$ is the estimated catch by weight in month $t$ in fishery $f$ taken in fishing area $g$ of species $s$, $L_{tfg}$ is the corresponding (known) catch by month, fishery and fishing area and where $\hat{E}_{q_t fs}$ is the mean of the proportion of the species by quarter, $q_t$, and fishery $f$ obtained from $m_{qf}$ biological samples ($q_t$ is the quarter which corresponds to month $t$). Note that the landings, $\hat{L}_{tfgs}$, are estimated by month and fishing area (and fishery and species) while the mean species proportion, $\hat{E}_{q_t fs}$, is estimated by quarter (and fishery). $\hat{L}_{tfgs}$ is estimated by month and fishing area in order to incorporate spatial and temporal differences in the age composition of the species caught. $\hat{E}_{q_t fs}$ is estimated by quarter and not by month because there were too few samples by month to estimate the variation in the species composition.

$\hat{E}_{qfs}$ was estimated using the Delta-Dirichlet (D-D) distribution, introduced by Lewy (1995). The D-D distribution is a multidimensional distribution describing the simultaneous variation of variables larger than 0 and less than 1, which are correlated as their sum is 1. As far as the author knows, the Dirichlet distribution is the only known distribution which include these characteristics for sets of variables summing to 1. Furthermore, the D-D distribution allows for the possibility that any of the variables may be 0 or 1 with positive probabilities. This distribution may be specifically useful when considering the variation of proportions of several species. Furthermore, the species composition within a specified fishery in industrial biological samples has frequently the characteristic that some species are not present or that one species constitutes the total catch. This situation is taken care of by allowing the variables to be singular in 0 and 1. Furthermore, like the Beta distribution, the D-D distribution is very flexible allowing for several types of shape of the probability density function. It is therefore likely that this distribution fits better to the very skew marginal distribution of the species compositions considered than for instance the log normal distribution. Another reason for applying the distribution is that if the variables are distributed according to the D-D distribution Lewy (1995) has shown that the estimates of the mean and the variance/covariances based on maximum likelihood estimates of the parameters are more efficient than the corresponding empirical moments (the average, the standard deviation etc.).

The fit of the D-D distribution has been tested by quarter and fishery. The results of the tests are given by Lewy (1995). For the Norway pout fishery it was accepted that the species composition follows the D-D distribution, while the test of fit was rejected for the sandeel and the sprat fisheries. However, in the present paper estimates of mean and variance/covariances of the species proportions were based on the new distribution for the sandeel and sprat fisheries as well. The reason is – as mentioned above – that there is no reasonable alternative to the D-D distribution.

Total number of fish by month, fishery, fishing area and species was estimated by dividing the corresponding catch in weight by the mean weight per individual. The corresponding catch in number at age was then estimated by multiplying estimated total catch in number and the age proportion, $\hat{p}_{tfgsa}$, obtained from the aged fish in the biological samples:

$$\hat{N}_{tfgsa} = \frac{\hat{L}_{tfgs}}{\bar{w}_{tfgs}} \hat{p}_{tfgsa} \; . \tag{2}$$

*Optimum sampling allocation*

In order to define optimum sampling allocation an objective function, O, was defined. Such an function should specify the priorities of the managers. The objective function considered is the variance of the weighted sum of the number of fish:

$$O((m_{qf}),(l_{tfgs}),(n_{tfgs})) = VAR(\sum_{sa} \nu_{sa}\hat{N}_{sa}) \tag{3}$$

where

$\hat{N}_{sa} = \sum_{tfg} \hat{N}_{tfgsa}$ and $v_{sa}$ are priority weights given by the managers.

As indicated the objective function depends on the number of biological samples by quarter and fishery, $m_{qf}$, the number of fish measured, $l_{tfgs}$, and the number of fish aged, $n_{tfgs}$, by month, fishery, fishing area and species.

The priority weights used in the analysis is the weighted average weight per fish by species and age (weighted with respect to the numbers caught by month and fishing area). In that case the objective function becomes the sum of the variances for each species of the the so-called SOP, the sum of product of estimated mean weight and catch in numbers at age. Lewy & Lassen (1995) has shown that for each species the mean value of the SOP approximately is equal to the landing, $\hat{L}_s = \sum_{tfgs} \hat{L}_{tfgs}$. Assuming that the price per kg is the same for all industrial catches, the weight per fish is proportional to the value of the individual fish and and the objective function may consequently be interpreted as the sum of variances of the value of the species.

The optimum sampling allocation considered is the sampling scheme which minimizes the objective function under the restriction that the total sampling costs are constant. The optimum resource allocation specifies the number of biological samples by fishery and quarter, the number of measured and aged fish by month, fishery, fishing area and species.

The sampling costs are calculated as follows:

$$c_{\text{tot}} = (c_{\text{fix}} + c_{\text{s}}) \sum_{qf} m_{qf} + \sum_{f} (c_{\text{m}f} \sum_{tgs} l_{tfgs}) + \sum_{f} (c_{\text{a}f} \sum_{tgs} n_{tfgs}) \tag{4}$$

where $c_{\text{tot}}$ is total sampling costs, $c_{\text{fix}}$ is fixed costs for collecting one sample, $c_{\text{s}}$ is cost per sample for sorting a sample into species categories, $c_{\text{m}f}$ is the cost for measuring length and weight of one fish and $c_{\text{a}f}$ is the cost for ageing one fish in fishery $f$.

The calculation of the optimum allocation that minimizes (3) with respect to the constraint, (4), depends on the functional relationship between the objective function and $m_{qf}$, $l_{tfgs}$ and $n_{tfgs}$. Therefore, the calculation of the optimum allocation will be considered after the objective function has been calculated.

Preliminary estimates of the costs are based on the assumptions: $c_{\text{fix}}$ is 1000 DKK, while the cost of analysis of one biological sample, $c_{\text{bio}} = c_{\text{s}} + \sum (c_{\text{m}f} + c_{\text{a}f})$, is 1500 DKK. Furthermore, it is assumed that the costs of sorting into species categories, length-measuring and aging of the fish are of equal size, 500 DKK for each of the three steps. The cost per fish, $c_{\text{m}f}$ and $c_{\text{a}f}$, are obtained as average costs by:

$$c_{\text{m}f} = \frac{500}{\sum_{tgs} l_{tfgs}} \text{ DKK} ; \quad c_{\text{a}f} = \frac{500}{\sum_{tgs} n_{tfgs}} \text{ DKK.}$$

Note, that the total costs of aging the fish include costs of presupposed length-measurering of the fish, $c_{\text{m}f}$, and the additional costs, $c_{\text{a}f}$.

*Minimum level of precision of estimates combined with optimum allocation*
For fixed total sampling costs, which corresponds to the costs in 1993, the precision of estimated catch in numbers for the by-catches of whiting and haddock associated with the optimum sampling may be unsatisfactorily low. On the other hand, the precision of e.g. the sandeel estimates may very high. This situation can be changed by combining a selected minimum level of precision of estimated annual catch in numbers at age, $\hat{N}_{sa}$, for all species and selected age groups with a type of limited optimum allocation. In the present paper the precision of estimates has been defined by the coefficient of variation, c.v., such that a minimum level of precision is expressed by a maximum c.v. This modified optimum sampling procedure is described in the following:

*A*. Firstly, the optimum sampling allocation has been calculated assuming that only the species sandeel, Norway pout, sprat and herring have been taken into account in the sampling allocation. This means that the priority weights, $v_{sa}$, have been set to zero for all age groups of whiting and haddock, while the remaining priority weights were set according to the above definition. Total cost has been set to an arbitrary value.

Now the number of samples by fishery has been adjusted, keeping the optimum number of fish per sample length-measured and aged unchanged. In that way the optimum allocation has partly been kept. This adjustment of the number of samples and the number of fish length-measured and aged has been done so that the minimum level of precision has been obtained for sandeel, Norway pout, sprat and herring and for all age groups considered. For the sprat fishery, including both sprat and herring, the adjustment has been chosen, which ensures that the minimum level of precision has been reached for both species for all relevant age groups.

For the biological samples which corresponds to this sampling scheme the optimum number of whiting and haddock to be measured and aged per sample can be calculated by setting the priority weights to zero for the four other species and the weights for whiting and haddock to those defined above. This is done for each of the Norway pout and sprat fisheries in which by-catches of haddock and whiting occur. The adjusted number of samples obtained above is now supplemented with the optimum number of whiting and haddock measured and aged. This constitutes the preliminary sampling scheme for these two species.

*B*. If the precision for whiting and haddock is still too low compared with the selected minimum level of precision, additional samples are required with the only purpose of increasing the precision of these two by-catch species. Since the two species are not taken in the sandeel fishery, the number of samples should be increased in the Norway pout and the sprat fisheries only. The number of samples in these two fisheries has now been adjusted keeping the optimum number of whiting and haddock measured and aged per sample constant. The adjustment has been made such that the precision of the whiting and haddock estimates – in conjunction with the preliminary sampling scheme obtained above – satisfies the required minimum level of precision for both species and for the selected age groups.

The costs of estimating the proportion of whiting only in the sprat fishery, $c_s$, was reduced by a factor tentatively set to 0.5. This was done because the process of sorting a sample into species categories in this case avoids the time consuming work of separating herring and sprat.

Finally, some minimum requirements to the sampling scheme were assumed irrespective of the optimum allocation. If the catch by month, fishery and fishing area in 1993 was larger than 1000 tonnes, at least one sample should be taken and a minimum of 200 fish should be aged in order to obtain useful estimates of the age distribution and mean weight at age. Due to the hierarchical order of the sampling system described above, the number of fish measured with respect to weight and length was set equal to the number of fish aged in the cases where the optimum allocation of the number of fish weight- and length-measured was less than the optimum number of fish aged.

In some cases the optimum number of especially whiting and haddock sampled may be larger than the number of fish available in the biological samples. In such cases additional samples of these two species must be collected at random from the landings.

The minimum level of precision of estimates of $\hat{N}_{sa}$, the maximum c.v., was preliminary set to 0.1.

*Calculation of the objective function*

The objective function is:

$$O = VAR(\sum_{sa} v_{sa} \hat{N}_{sa}) = VAR\left( \sum_{tfgsa} v_{sa} \frac{L_{tfg} \hat{E}_{q_t fs}}{\bar{w}_{tfgs}} \hat{p}_{tfgsa} \right) . \qquad (5)$$

To make it possible to calculate the objective function, (5), and thereby the optimum sampling allocation the following assumptions have been made:

1. The landings by fishery, $L_{tfg}$, are known without error;
2. for each fishery the vessels – from which the samples are taken – are randomly selected;
3. the samples from the cargo are taken at random;
4. the species composition of a sample is independent of a) the age distribution for all species in the sample and b) the weight of the individual fish;
5. the age distribution of one species is independent of the age distribution of another species;
6. the weight of one fish is independent of the weight of other fish.

Assumption 1 means that the landings statistics are correct, which will not be questioned in the present paper. The assumptions 2 and 3 depends on the sampling and are probably correct. The assumptions 4-6 have not specifically been investigated. One cannot exclude that some of the assumptions are wrong under certain circumstances. For instance the age distribution of a species in a specific area may be correlated to the age distribution of another species (herring and sprat caught in shallow areas). Even though as a starting point it seems reasonable to apply these assumptions and thereby avoiding complicated studies on the covariance structure of the variable in question.

The objective function, (5), has been calculated in Appendix II, in which it has been shown that the result can be approximated by:

$$O \simeq O_1 + O_{21} + O_{22} + O_3 = \sum_{qf} \frac{C_{qf}}{m_{qf}} + \sum_{fgs} \frac{D_{tfgs}}{l_{tfgs}} + \sum_{tfgs} \frac{E_{tfgs}}{l_{tfgs}} + \sum_{tfgs} \frac{F_{tfgs}}{n_{tfgs}} \tag{6}$$

where

$$O_1 = \sum_{qf} \frac{C_{qf}}{m_{qf}} \quad O_{21} = \sum_{fgs} \frac{D_{tfgs}}{l_{tfgs}} \quad O_{22} = \sum_{tfgs} \frac{E_{tfgs}}{l_{tfgs}} \quad O_3 = \sum_{tfgs} \frac{F_{tfgs}}{n_{tfgs}}$$

The four components of the objective function corresponds to four sources of errors:

$O_1$ denotes the error due to estimation of the species composition;
$O_{21}$ denotes the error due to estimation of the mean weight;
$O_{21}$ denotes the error due to the correlation between mean weight and age;
$O_3$ denotes the error due to estimation of the age distribution.

The coefficients $C_{qf}$, $D_{tfgs}$, $E_{tfgs}$ and $F_{tfgs}$ are given in Appendix II.

Only terms of order $m^{-1}$, $l^{-1}$ and $n^{-1}$ have been included in the objective function because of the insignificance of the second order terms for the relevant size of the three quantities.

### *Solution to optimum sampling allocation*

Minimization of the approximated objective function, (6), with respect to fixed total costs as expressed in (4) corresponds to optimum allocation in standard stratified random sampling with fixed costs, see for example Cochran (1977). The optimum allocation is:

$$\begin{aligned} m_{qf}^{optimum} &= G\sqrt{\frac{C_{qf}}{c_{\text{fix}} + c_{\text{s}}}} \\ l_{tfgs}^{optimum} &= G\sqrt{\frac{D_{tfgs} + E_{tfgs}}{c_{\text{mf}}}} \\ n_{tfgs}^{optimum} &= G\sqrt{\frac{F_{tfgs}}{c_{af}}} \end{aligned} \tag{7}$$

where

$$G = \frac{c_{\text{tot}}}{\sqrt{c_{\text{fix}} + c_{\text{s}}}\sum_{qf}\sqrt{C_{qf}} + \sum_{tfsg}\sqrt{(D_{tfgs}+E_{tfgs})c_{\text{m}f}} + \sum_{tfgs}\sqrt{F_{tfgs}\, c_{af}}}$$

and where the coefficients again are given in appendix II.

*The variance of the catch by weight and numbers at age*
These can be calculated similarly to the calculations of the objective function. The formulae are given in Appendix III. The variance of numbers at age can be split into the error components similarly to the objective function.

## Results

*Precision of estimates for the 1993 sampling scheme*
Coefficient of variation of landings in weight by species is given in Table 4. The table shows that for all species except for the by-catches of whiting and haddock the annual c.v. by weight is about 6% or lower and 10-15% for the by-catches.

Table 4. Coefficient of variation of catch in weight by quarter and species in the Danish North Sea industrial fishery in 1993. Per cent.

| Quarter | Sandeel | Norway pout | Herring | Sprat | Whiting | Haddock |
|---|---|---|---|---|---|---|
| 1 | 0.6 | 4.5 | 15.6 | 15.2 | 24.0 | 20.3 |
| 2 | 0.9 | – | – | – | – | – |
| 3 | 5.8 | 6.1 | 8.6 | 7.7 | 15.4 | 29.9 |
| 4 | 5.5 | 4.1 | 12.3 | 5.5 | 15.8 | 30.5 |
| Total | 1.1 | 2.8 | 6.5 | 4.6 | 10.2 | 15.6 |

Table 5, c.v. of catch in numbers at age, shows that the precision of estimated annual catch in numbers for the first four species is high compared with other sources of error as the c.v. is less than 8% (implying that the 95% confidence limits are less than ±16%). The high c.v. value of sandeel in the first quarter of age group 2 is not important as the catch is insignificant. The precision is much lower for the by-catch species whiting and haddock. For both species especially the 0- and 1-group are important as the Danish catches constitute significant parts of total international catches. For these age groups the annual c.v. lies between about 15 and 30%. The quarterly c.v.'s are of course higher, especially for whiting for which the 1-group in third quarter is about 77%.

It is also of interest to identify the errors caused by the three steps of the sampling procedure, namely the determination of the species composition, the mean weight and the age. Furthermore, the correlation between determination of mean weight and age is also considered. The annual proportions of these four sources of error are described by the relative value of $V_1$, $V_{21}$, $V_{22}$ and $V_3$ defined in Appendix III. The results are given in Table 6.

Table 6 shows that for all species except for sandeel and for all age groups the error due to the variation in the species composition constitutes by far the most important component of the total variance of estimated catch in numbers at age. For sandeel the situation is less clear. For the age groups 3 and older, the error due to the age distribution dominates. For the age groups 0, 1 and 2 there are no common

Table 5. Coefficient of variation of estimated catch in numbers by quarter, age and species in the Danish North Sea industrial fishery in 1993. Percent.

| | | Age | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Quarter | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| Sandeel | 1 | – | 5.6 | 35.6 | – | – | – | – |
| | 2 | 12.5 | 2.5 | 2.0 | 4.5 | 7.8 | 3.8 | 5.8 |
| | 3 | 3.4 | 3.1 | 3.3 | 3.4 | 4.5 | 4.9 | – |
| | 4 | 5.4 | 5.4 | – | – | – | – | – |
| | Total | 2.8 | 2.3 | 2.0 | 4.3 | 7.5 | 3.7 | 5.8 |
| Norway pout | 1 | – | 5.0 | 4.8 | 3.2 | – | – | – |
| | 2 | – | – | – | – | – | – | – |
| | 3 | 26.7 | 5.4 | 4.5 | – | – | – | – |
| | 4 | 4.8 | 4.1 | 4.1 | – | – | – | – |
| | Total | 4.7 | 3.3 | 2.6 | 3.2 | – | – | – |
| Herring | 1 | – | 12.8 | 15.1 | – | – | – | – |
| | 2 | – | – | – | – | – | – | – |
| | 3 | 6.8 | 8.6 | – | – | – | – | – |
| | 4 | 9.0 | 15.4 | – | – | – | – | – |
| | Total | 5.7 | 12.0 | 15.1 | – | – | – | – |
| Sprat | 1 | – | 15.6 | 15.3 | 16.3 | 15.2 | – | – |
| | 2 | – | – | – | – | – | – | – |
| | 3 | – | 7.9 | 8.2 | 7.7 | 7.7 | – | – |
| | 4 | 6.9 | 5.5 | 5.7 | 8.4 | – | – | – |
| | Total | 6.9 | 4.4 | 6.6 | 7.5 | 7.5 | – | – |
| Whiting | 1 | – | 50.5 | 24.7 | 21.7 | 109.5 | – | – |
| | 2 | – | – | – | – | – | – | – |
| | 3 | 17.7 | 76.7 | 37.4 | 15.0 | 15.8 | – | – |
| | 4 | 24.6 | 22.2 | 17.8 | 15.3 | 17.9 | 17.9 | 17.9 |
| | Total | 15.7 | 32.6 | 13.7 | 10.1 | 13.7 | 17.9 | 17.9 |
| Haddock | 1 | 20.0 | 30.7 | 23.0 | 20.0 | – | – | – |
| | 2 | – | – | – | – | – | – | – |
| | 3 | 36.7 | 33.5 | 61.4 | – | – | – | – |
| | 4 | 31.3 | 31.3 | 30.5 | – | – | – | – |
| | Total | 28.2 | 20.9 | 18.4 | 20.0 | – | – | – |

patterns. The reason why the determination of the species composition generates small errors for sandeel is that the species constitutes a relatively constant fraction of about 95% of the catches.

*Optimum stratified sampling*

The resulting precision of the estimated number of fish by species and age for the optimum sampling allocation is shown in Table 7. As seen the optimum allocation increases the precision for sandeel and haddock whereas the other species only changed slightly. For sandeel the precision increased for the age groups 1-4 for which the precision was high already (Table 6).

Table 6. The annual percentage of the errors generated by determination of species composition, mean weight, correlation between mean weight and age distribution, and age determination, $V_1$, $V_{21}$, $V_{22}$ and $V_3$.

| | | Age | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Error | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| Sandeel | Species comp. | 90.2 | 3.4 | 10.5 | 1.7 | 0.5 | 2.3 | 1.1 |
| | Weight | 6.4 | 50.8 | 36.4 | 5.8 | 0.5 | 1.8 | 0.2 |
| | Correlation | 2.6 | 6.6 | 4.2 | − 3.1 | − 0.5 | − 3.6 | − 0.1 |
| | Age | 0.8 | 39.3 | 48.8 | 95.6 | 99.6 | 99.5 | 98.8 |
| | Total | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Norway pout | Species comp. | 54.2 | 77.4 | 80.9 | 100.0 | – | – | – |
| | Weight | 18.0 | 15.3 | 2.8 | 0 | – | – | – |
| | Correlation | 18.9 | 3.0 | − 5.5 | 0 | – | – | – |
| | Age | 8.9 | 4.2 | 21.8 | 0 | – | – | – |
| | Total | 100 | 100 | 100 | 100 | – | – | – |
| Herring | Species comp. | 99.0 | 98.9 | 98.5 | – | – | – | – |
| | Weight | 0.9 | 0.7 | 0 | – | – | – | – |
| | Correlation | 0 | 0.2 | − 0.3 | – | – | – | – |
| | Age | 0 | 0.2 | 1.7 | – | – | – | – |
| | Total | 100 | 100 | 100 | – | – | – | – |
| Sprat | Species comp. | 62.3 | 96.4 | 95.5 | 89.4 | 100.0 | – | – |
| | Weight | 0 | 0.8 | 0.5 | 0.1 | 0 | – | – |
| | Correlation | 0.2 | 0.4 | − 0.3 | − 0.1 | 0 | – | – |
| | Age | 37.5 | 2.5 | 4.3 | 10.5 | 0 | – | – |
| | Total | 100 | 100 | 100 | 100 | 100 | – | – |
| Whiting | Species comp. | 97.6 | 99.0 | 81.4 | 89.4 | 98.4 | 100.0 | 100.0 |
| | Weight | 2.1 | 0.4 | 18.6 | 1.4 | 0.1 | 0 | 0 |
| | Correlation | 0.2 | − 0.3 | − 42.9 | − 2.8 | − 0.5 | 0 | 0 |
| | Age | 0.1 | 0.9 | 42.9 | 12.0 | 2.0 | 0 | 0 |
| | Total | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Haddock | Species comp. | 68.5 | 51.2 | 68.2 | 100.0 | – | – | – |
| | Weight | 22.8 | 27.2 | 63.6 | 0 | – | – | – |
| | Correlation | 7.7 | 14.9 | − 180.7 | 0 | – | – | – |
| | Age | 1.0 | 6.7 | 148.8 | 0 | – | – | – |
| | Total | 100 | 100 | 100 | 100 | – | – | – |

Table 7. Coefficient of variation using optimum allocation relative to actual coefficient of variation in 1993 of estimated annual catch in numbers by age and species.

| | Age | | | | | | |
|---|---|---|---|---|---|---|---|
| Species | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| Sandeel | 0.93 | 0.61 | 0.50 | 0.58 | 0.61 | 1.03 | 1.09 |
| Norway pout | 1.00 | 0.91 | 1.08 | 0.94 | – | – | – |
| Herring | 1.05 | 0.87 | 0.87 | – | – | – | – |
| Sprat | 1.29 | 1.11 | 0.92 | 0.93 | 1.03 | – | – |
| Whiting | 1.03 | 0.89 | 0.92 | 0.96 | 1.01 | 1.08 | 1.08 |
| Haddock | 0.92 | 0.72 | 0.85 | 0.93 | – | – | – |

These results can be explained by the change of the optimum sampling allocation relative to the actual allocation in 1993, which is described in the Tables 8-10.

Table 8 shows that the optimum sampling allocation of biological samples is close to the actual allocation. This seems surprising as one would expect that the optimum stratified sampling should increase the number of biological samples for estimation of the species composition as this step generates the largest error. However, the optimum of course depends on the price relations for the three sampling steps. Apparently, the results show that when the number of biological samples is increased then it is profitable to measure and age most of the fish sampled because of the relative low expenses. Alternative runs, where for instance $c_{fix}$ and $c_s$ (see equation (4)) are reduced to 200 and 150 DKK, respectively, results in an overall increase of the number of biological samples by about 60%.

Table 9, giving the optimum allocation of fish sampled, shows that the number of fish measured and aged should be increased for the by-catch species whiting and

Table 8. Optimum sampling allocation of the number of biological samples relative to the actual allocation in 1993.

| | Fishery | | | |
|---|---|---|---|---|
| Quarter | Sandeel | Norway pout | Sprat | Total |
| 1 | 0.1 | 1.2 | 1.3 | 1.0 |
| 2 | 1.1 | – | – | 1.1 |
| 3 | 1.7 | 0.9 | 1.0 | 1.1 |
| 4 | 0.4 | 0.9 | 0.6 | 0.7 |
| Total | 1.1 | 0.9 | 0.9 | 1.0 |

Table 9. Optimum sampling allocation of the number of fish weight measured and aged by quarter relative to the actual allocation in 1993.

| | Sandeel | | Norway pout | | Herring | | Sprat | |
|---|---|---|---|---|---|---|---|---|
| Quarter | Weighed | Aged | Weighed | Aged | Weighed | Aged | Weighed | Aged |
| 1 | 1.9 | 0.1 | 0.8 | 0.6 | 0.6 | 0.5 | 1.1 | 0.5 |
| 2 | 3.0 | 1.2 | – | – | – | – | – | – |
| 3 | 0.7 | 0.2 | 0.8 | 0.8 | 1.1 | 0.1 | 1.5 | 0.8 |
| 4 | 0.5 | 0.2 | 0.9 | 0.9 | 0.7 | 0.4 | 0.7 | 0.5 |
| Total | 1.7 | 0.6 | 0.8 | 0.7 | 0.9 | 0.3 | 1.0 | 0.5 |

| | Whiting | | | | Haddock | |
|---|---|---|---|---|---|---|
| | Norway pout fishery | | Sprat fishery | | Norway pout fishery | |
| Quarter | Weighed | Aged | Weighed | Aged | Weighed | Aged |
| 1 | 7.7 | 7.7 | 10.2 | 6.8 | 2.2 | 2.2 |
| 2 | – | – | – | – | – | – |
| 3 | 14.3 | 23.1 | 4.3 | 3.4 | 4.2 | 4.2 |
| 4 | 8.2 | 8.2 | 1.2 | 1.2 | 6.0 | 6.0 |
| Total | 9.1 | 9.9 | 3.3 | 2.8 | 3.3 | 3.3 |

Table 10. Optimum sampling allocation (%) of the annual number of fish weight measured and aged by fishing area.

| Fishing area | Sandeel* | | Norway pout ** | | Herring ** | |
|---|---|---|---|---|---|---|
| | Weighed | Aged | Weighed | Aged | Weighed | Aged |
| 1 | 40 | 37 | – | – | 35 | 63 |
| 2 | 6 | 11 | – | – | 8 | 12 |
| 3 | 2 | 9 | – | – | 57 | 25 |
| 4 | 24 | 14 | 57 | 50 | – | – |
| 5 | 15 | 14 | 43 | 50 | – | – |
| 6 | 5 | 5 | – | – | – | – |
| 7 | 7 | 11 | – | – | – | – |
| Total | 100 | 100 | 100 | 100 | 100 | 100 |

* The indices for fishing area refer to Figure 1A.
** The indices for fishing area refer to Figure 1B.

haddock while it should be reduced for the other species. Sandeel in the second quarter is, however, an exception as the number of fish measured should be tripled. Apparently, the very large increase in the number of whiting and haddock weighed and aged causes surprisingly small improvements in the precision of the estimates. This is due to the fact that the main part of the errors is generated by determination of the species composition (Table 6).

Table 10 shows the optimum allocation distributed by the fishing area. As the prices for measuring and aging the fish are the same for all fishing areas the proportions in the table are proportional to the product of the number of fish measured or aged and the variation within the fishing area. For sandeel, fishing area 1 and 4 are the most important. For herring fishing area 3 is most important concerning the weight variation, while fishing area 1 is most important for the aged fish.

*Minimum level of precision of estimates combined with optimum allocation*

The above procedure was carried out for obtaining a minimum level of precision for selected age groups in conjunction with optimum allocation. Tentatively, the minimum level of precision was set to a maximum coefficient of variation of 0.1. The age groups were selected such that the annual Danish North Sea industrial catches by age made up at least 10% of total international landings. The age groups selected, the precision of estimates and the gain in the precision are given in the Tables 11, 12 and 13 respectively. Table 13 shows that the gain in precision defined as the optimum coefficient of variation relative to 1993 is about 0.3-0.4 for whiting and haddock. For sandeel and Norway pout, the precision has been reduced, while only small changes have taken place for herring and sprat.

Some of the main changes in this combined sampling allocation are: in general total sampling costs should be tripled. Regarding the total number of biological samples these should be quadrupled; the sandeel samples should be drastically reduced while the number of samples in the other two fisheries should be multiplied by 7. The number of whiting and haddock measured and aged should be multiplied by more than 20.

Table 11. Selected age groups for which coefficient of variation was required less than 0.1. + indicates selected.

| | Age 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Sandeel | + | + | + | + | + | + |
| Norway pout | + | + | + | | | |
| Herring | + | + | | | | |
| Sprat | | + | + | + | | |
| Whiting | + | + | | | | |
| Haddock | + | + | | | | |

Table 12. Coefficient of variation (%) of estimated annual catch in numbers by age and species for optimum sampling combined with a maximum coefficient of variation of 0.1 for selected age groups.

| Species | Age 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Sandeel | 10 | 3 | 2 | 3 | 5 | 4 | 7 |
| Norway pout | 10 | 7 | 7 | 8 | – | – | – |
| Herring | 5 | 10 | 13 | – | – | – | – |
| Sprat | 8 | 4 | 6 | 7 | 7 | – | – |
| Whiting | 5 | 10 | 5 | 4 | 5 | 6 | 6 |
| Haddock | 10 | 6 | 6 | 8 | – | – | – |

Table 13. Coefficient of variation of estimated annual catch in numbers by age and species for optimum sampling combined with a maximum coefficient of variation of 0.1 for selected age groups relative to 1993.

| Species | Age 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Sandeel | 4.0 | 1.3 | 2.7 | 1.7 | 07. | 1.1 | 1.2 |
| Norway pout | 2.1 | 2.1 | 2.7 | 2.5 | – | – | – |
| Herring | 0.9 | 0.8 | 0.9 | – | – | – | – |
| Sprat | 1.2 | 0.9 | 0.9 | 0.9 | 0.9 | – | – |
| Whiting | 0.3 | 0.3 | 0.4 | 0.4 | 0.4 | 0.3 | 0.3 |
| Haddock | 0.4 | 0.3 | 0.3 | 0.4 | – | – | – |

## Discussion

Optimum allocation in the fishery sampling programme constitutes a framework for the implemented programme, which may deviate from the planned. Apart from practical circumstances (illness, accidents, bad weather conditions etc.) the actual programme also has to cope with changes in the fishery compared to previous years which formed the basis for the optimum programme. Such changes do not need to involve bias in the estimates, but may only affect the optimum allocation of resources. However, in the fishery sampling programme and the subsequent estimation procedure at least two factors may result in bias.

The first factor involves classification of total catch into three target fisheries which were carried out for each vessel trip. The classification criteria used were obtained by analysis of the biological samples with known species composition as information on gear and mesh size used, fishing area and time was combined. Especially in transition periods between the target fisheries, when more than one fishery take place in the same area, the classification might be wrong. Furthermore, the criteria may change from year to year because the fisheries change, which may require annual modifications. In principle one could remove this bias by asking the fishermen to record the type of fishery in their logbooks. If this is unattainable more reliable information on, for instance, mesh size used may be needed together with updated information on the fisheries. Visual inspections of landings are especially important during periods of unreliable classification of fisheries. Finally, one could investigate the validity of the criteria by analysing instances of misclassification using data for other years.

The second factor which may result in biasness is errors associated with age determination of the fish. This kind of bias has not been included in the present analysis. It only accounts for the stochastic variation in the age distribution of fish in a random sample, based on the assumption that the age is correctly determined and therefore reflects the true age distribution of the fish in the sea.

The sampling programme depends on the objectives or priorities of the managers and the politicians. In the above, minimization of the variance of the weighted sum of catch in numbers combined with an upper limit of the error has been considered, where the priority weights were optional. In addition to the difficulties in selecting the priority weights, many different objectives can be considered as well. For instance Gulland (1955) considered the objective to achieve a constant (low) coefficient of variation for significant age groups. Sparre *et. al.* (1977) used the sum of variances over age of estimated fishing mortality at age as objective function while Kimura (1977) and Lai (1987) used the sum over age of variances of the proportion of fish at age as objective function. Minimum variance of estimated projected landings or biomass are two other examples which may be useful because they directly involve the application of estimates. Objectives of this type would give priority to catches by species and age where the national proportions of total international catches constitute a significant part and where fishing mortality is at the same level or larger than the natural mortality. The drawback is more complicated non-linear objective functions.

Using the mean weight or the value by age of the fish as the priority weights implies that larger fish get a larger priority. Other priority weights could be selected by putting higher priority to the small fish for instance in order to obtain good recruitment estimates. Another criterion of priority may be to obtain good estimates of total international landings. If Danish catch in numbers at age for selected species constitute significant parts of the total international catches, these species and age groups should be given high priorities.

It should be stressed that the sampling costs used are preliminary guesstimates only and the optimum results should be treated as an example. In practice more reliable cost data need to be calculated.

In the present sampling system the age has been determined by reading the otholiths of all fish for which the length does not uniquely separate the age groups. A system of length samples combined with age-length keys has not been investigated. It is an open question whether such a procedure would be superior to the present method. However, some studies indicate that the system used may be preferred. Kutkuhn (1963) showed that simple random sampling including otolith sampling should only be preferred when the cost of ageing is at least five times the cost of length-measurements. This is in accordance with the results of Gudmundsdóttir *et al.* (1988), who found indications of 'that not much is gained from adding length sampling to an otolith sampling scheme' for the Icelandic cod and capelin. Sparre *et al.* 1977 came to similar conclusions for cod in the North Sea.

Estimates of catch in numbers at age, $\hat{N}_{tfgsa}$, could be obtained in another way than done by (2). Instead of first estimating the weight proportion by species (formula (1)) and then combining this with the age distribution one could estimate directly from the biological samples

$$\hat{N}_{tfgsa} = \frac{L_{tfg}}{w_{tfg}} n_{tfgsa}$$

where $L_{tfg}$ and $w_{tfg}$ are the weight of the landings and samples in month $t$ in fishery $f$ in fishing area $g$, respectively, and $n_{tfgsa}$ the number of fish by species and age group in the same samples. The properties of this estimate have not been studied.

## Acknowledgements

I am grateful to Jørgen Dalskov from the Danish Institute of Fisheries Research for helpful discussions about the Danish industrial sampling system.

## References

*Cochran, W.G.*, 1977: Sampling techniques. – John Wiley & Sons, New York. 3rd ed.

*Deriso, R.B., Quinn II, T.J. & P.R. Neal*, 1985: Catch-age analysis with auxiliary information. – Can. J. Fish. aq. Sci. *42*: 815-824.

*Doubleday, W.G.* 1976: A least square approach to analysing catch at age data. – ICNAF Res. Bull. *12*: 69-81.

*Doubleday, W.G. & D. Rivard* (eds), 1983: Sampling commercial catches of marine fish and invertebrates. – Can. Spec. Publ. Fish. aq. Sci. *66*: 290 pp.

*Fournier, D. & C.P. Archibald*, 1982: A general theory for analysing catch at age data. – Can. J. Fish. aq. Sci. *39*: 1195-1207.

*Gavaris, S. & C.A. Gavaris*, 1983: Estimation of catch at age and its variance for groundfish stocks in the Newfoundland Region. – *In* W.G. Doubleday & D. Rivard (eds): Sampling commercial catches of marine fish and invertebrates, pp. 178-182. Can. Spec. Publ. Fish. aq. Sci. *66*: 290 pp.

*Gudmundsdóttir, Á., B.Æ. Steinarsson & G. Stefánsson*, 1988: A simulation procedure to evaluate the efficiency of some otolith and length sampling schemes. – ICES C. M. 1988/D: 14.

*Gulland, J.A.*, 1955: Estimation of growth and mortality in commercial fish populations. – Fish. Invest., London, 2, *18*(9): 49 pp.

*ICES*, 1994: Report of the workshop on sampling strategies for age and maturity. – ICES C.M. 1994/D: 1.

*Kimura, D.K.*, 1977: Statistical assessment of age-length key. – J. Fish. Res. Bd Can. *34*: 317-324.

*Kutkuhn, J.H.*, 1963: Estimating absolute age composition of California salmon landings. – Fish. Bull. *120*: 47 pp.

*Lai, H.L.*, 1987: Optimum allocation for estimating age composition using age-length key. – Fish. Bull. *85*(2).

*Lassen, H., B. Mørck-Larsen & A.F. Elkarachily*, 1973: A method of obtaining estimates of catch composition in a mixed fishery. – ICES C.M. 1973/H: 18.

*Lewy, P.*, 1981: Optimum sampling concerning catch composition in the Danish North Sea industrial fishery. – ICES C.M. 1981/D: 9.

*Lewy, P.*, 1988: Integrated stochastic virtual population analysis: Estimates and their precision of fishing mortalities and stock sizes for the North Sea whiting stock. – J. Cons. int. Explor. Mer *44*: 217-228.

*Lewy, P.*, 1995: A generalized Dirichlet distribution accounting for singularities of the variable. – Biometrics. In press.

*Lewy, P. & H. Lassen*, 1995: Should total landings be used to correct estimated catch in numbers or mean weight at age? In prep.

*Lewy, P. & M. Vinther*, 1994: Identification of Danish North Sea trawl fisheries. – ICES J. mar. Sci. *51*: 263-272.

*Murawski, S.A., A.M. Lange, M.P. Sissenwine & R.K. Mayo*, 1983: Definitions of analysis of multispecies ottertrawl fisheries off the northeast coast of the United States. – J. Cons. int. Explor. Mer *41*: 13-27.

*Pope, J.G.*, 1972: An investigation of the accuracy of virtual population analysis using cohort analysis. – ICNAF Res. Bull. *9*: 65-74.

*Pope, J.G. & D. Gray*, 1983: An investigation of the relationship between the precision of assessment data and the precision of total allowable catch. – *In* W.G. Doubleday & D. Rivard (eds): Sampling commercial catches of marine fish and invertebrates, pp. 151-157. Can. Spec. Publ. Fish. aq. Sci. *66*.

*Rivard, D.*, 1983: Effects of systematic, analytical, and sampling errors on catch estimates: A sensitivity analysis. – *In* W.G. Doubleday & D. Rivard (eds): Sampling commercial catches of marine fish and invertebrates, pp. 114-129. Can. Spec. Publ. Fish. aq. Sci. *66*.

*Pelletier, D. & P. Gros*, 1991: Assessing the impact of sampling error on model-based management advice: Comparison of equilibrium yield per recruit variance estimators. – Can. J. Fish. aq. Sci. *48*: 2129-2139.

*Rogers, J.B. & E.K. Pikitch*, 1992: Numerical definition of groundfish assemblages caught off the coasts of Oregon and Washington using commercial fishing strategies. – Can. J. Fish. aq. Sci. *43*: 325-342.

*Sparholt, H.*, 1990: A stochastic integrated VPA for herring in the Baltic Sea using acoustic estimates as auxiliary information for estimating natural mortality. – J. Cons. int. Explor. Mer *46*: 325-332.

*Sparre, P., H. Knudsen, & S. Munck-Petersen*, 1977: Optimization of sampling programs for estimation of age distribution and fishing mortality. – ICES C.M. 1977/F: 43.

*Sparre, P. & S.C. Venema*, 1992: Introduction to tropical fish stock assessment. Part 1. Manual. – FAO Fisheries Technical Paper No. 306.1 Rome, FAO. 1992. 376 pp.

## Appendix I. Notation

*Indices*

$t$ indicates month;
$q$ indicates quarter;
$q_t$ indicates the quarter which corresponds to month $t$;
$f$ indicates fishery;
$g$ indicates fishing area;
$s$ indicates species;
$u$ indicates species;
$a$ indicates age.

*Parameters and observations*

$m_{qf}$ indicates the number of biological samples by quarter and fishery;
$n_{tfgs}$ indicates the number of fish aged in the samples;
$l_{tfgs}$ indicates the number of fish weighed and measured by length;

$n_{tfgsa}$ indicates the number of fish at age $a$ in the samples;
$L_{tfg}$ indicates landing by month, fishery and fishing area;
$\bar{w}_{tfg}$ indicates mean weight of the fish in the sample;
$\bar{w}_{tfgsa}$ indicates mean weight at age of the fish in the sample;
$H_{qfs}$ indicates the proportion of species $s$ of one sample;
$v_{sa}$ indicates the priority weight;
$\hat{E}_{qfs}$ indicates estimated proportion of species $s$ in fishery $f$ in quarter $q$ based on the $m_{qf}$ samples;
$\hat{L}_{tfgs}$ indicates estimated landing;
$\hat{p}_{tfgsa}$ indicates the proportion $n_{tfgsa}/n_{tfgs}$;
$\hat{N}_{tfgsa}$ indicates estimated number of fish;
$\hat{N}_{sa}$ indicates the sum: $\sum_{tfg} \hat{N}_{tfgsa}$;
$O$ indicates the objective function;
$c_{\text{tot}}$ indicates total sampling costs;
$c_{\text{fix}}$ indicates fixed costs for collecting one sample;
$c_{\text{s}}$ indicates cost per sample for dividing a sample into species categories;
$c_{\text{m}f}$ indicates cost per fish in fishery $f$ for measuring length and weight of one fish;
$c_{\text{a}f}$ indicates cost per fish in fishery $f$ for aging one fish.

## Appendix II. Calculation of the objective function

The Objective function is:

$$O = VAR\left(\sum_{sa} v_{sa}\hat{N}_{sa}\right) = VAR\left(\sum_{tfgsa} v_{sa}\frac{L_{tfg}\hat{E}_{q_t fs}}{\bar{w}_{tfgs}}\hat{p}_{tfgsa}\right) = VAR\left(\sum_{qfs}\hat{E}_{qfs}B_{qfs}\right)$$
$$= \sum_{qf} VAR\left(\sum_{s}\hat{E}_{qfs}B_{qfs}\right) \tag{8}$$

where

$$B_{qfs} = \sum_{t\epsilon q}\sum_{g} A_{tfgs}, \quad A_{tfgs} = \sum_{a} v_{sa}\,\hat{p}_{tfgsa}\frac{L_{tfg}}{\bar{w}_{tfgs}} \quad \text{and} \quad \hat{E}_{q_t fs} = \hat{E}_{qfs} \text{ for } t\epsilon \text{ quarter } q\,.$$

The last sign of equation in (8) follows from the fact that samples taken at different times or from different fisheries are independent, which imply that the covariances are zero.

The sampling assumptions are given in the section on methods. The first three assumptions ensure that the estimates of $\hat{L}_{tfgs}$ and $\hat{N}_{sa}$ are unbiased estimates. The fourth assumption means that $COV(\hat{E}_{q_t fs}, B_{qfs}) = 0$ for all $q$, $f$ and $s$. The last two assumptions imply that $COV(A_{tfgs}, A_{tfgu}) = 0$ and as a consequence $COV(B_{qfs}, B_{qfu}) = 0$ for all $t$, $q$, $g$ and $s \neq u$.

The formulae

$$VAR(XY) \sim (EX)^2 VAR(Y) + (EY)^2 VAR(X) + VAR(X)VAR(Y) - COV(X,Y)(COV(X,Y) - 2EX\,EY)$$

and

$$VAR\left(\frac{1}{X}\right) \sim \frac{VAR(X)}{(EX)^4}$$

have been used to obtain approximations of the objective functions. The latter formula has been used in a case where P($X$ = 0) = 0. Terms where the number of samples and the number fish measured or aged are of the order less than $n^{-2}$ have been ignored because in practice they do not affect the optimum allocation solution.

For convenience the indices $q$ and $f$ have been left out in the following. For the term $A_{tfgs}$ the index $g$ has also been omitted. The last term of (8) can be written as

$$VAR(\sum_s \hat{E}_s B_s) = \sum_s VAR(\hat{E}_s B_s) + \sum_{s \neq u} COV(\hat{E}_s \hat{B}_s, E_u B_u) \simeq \sum_s (\hat{E}_s^2 VAR(B_s) + B_s^2 VAR(\hat{E}_s)) + \sum_{s \neq u} B_s B_u COV(\hat{E}_s, \hat{E}_u) \quad (9)$$

As described by Lewy (1995), the variance $VAR(\hat{E}_s)$ and the covariance $COV(\hat{E}_s, \hat{E}_u)$ can be approximated by

$$\frac{VAR(H_s)}{m} \text{ and } \frac{COV(H_s, H_u)}{m}, \text{ respectively,}$$

where $H_s$ and $H_u$ denote theproportion of one sample of species $s$ and $u$.We therefore get:

$$VAR(\sum_s \hat{E}_s B_s) \simeq \frac{\sum_s B_s^2 VAR(H_s) + \sum_{s \neq u} B_s B_u COV(H_s, H_u)}{m} + \sum_{t \epsilon q} \sum_s \hat{E}_s^2 \sum_g VAR(A_s). \quad (10)$$

We still need to calculate the last term in (10):

$$VAR(A_s) = VAR\left(\sum_a v_{sa} \hat{p}_{sa} \frac{L}{\bar{w}_s}\right) \sim (\sum_a v_{sa} \hat{p}_{sa})^2 L^2 VAR\left(\frac{1}{\bar{w}_s}\right) + \left(\frac{L}{\bar{w}_s}\right)^2 VAR(\sum_a v_{sa} \hat{p}_{sa})$$
$$\frac{-2L^2}{\bar{w}_s} COV\left(\sum_a v_{sa} \hat{p}_{sa}, \frac{1}{\bar{w}_s}\right) \sum_a v_{sa} \hat{p}_{sa} \,. \quad (11)$$

In (11) $VAR\left(\frac{1}{\bar{w}_s}\right)$, $VAR(\sum_a v_{sa} \hat{p}_{sa})$ and $COV\left(\sum_a v_{sa} \hat{p}_{sa}, \frac{1}{\bar{w}_s}\right)$ need to be calculated.

The variance of the reciprocal of the mean weight is approximated by

$$VAR\left(\frac{1}{\bar{w}_s}\right) \sim \frac{VAR(\bar{w}_s)}{\bar{w}_s^4} \,. \quad (12)$$

If the weight of each individual fish had been known, the variance of the average weight could simply be estimated from the sum of squared deviations. Since this is not the case, but the weight by length group is known for several samples, the variance is instead estimated as follows:

Assume that the month, fishery, fishing area and the species are given, and let these indices be left out in the following. For a length group $l$ assume further that the mean weight, $w_{jl}$, and the corresponding number of fish, $n_{jl}$, are known for the samples $j = 1,2,\ldots,k_l$. Finally, assume that:

$$E(\bar{w}_{jl}) = \mu_l \text{ and } VAR(\bar{w}_{jl}) = \frac{\sigma_l^2}{n_{jl}} \text{ for } j = 1,2,\ldots,k_l$$

where $\sigma_l^2$ denotes the variance of the weight of the individual fish of length $l$.

The Maximum likelihood estimates of $\mu_l$ and $\sigma_l$ are respectively

$$\hat{\mu}_l = \bar{w}_l = \sum_{j=1}^{k_l} \frac{n_{jl}}{n_l} \bar{w}_{jl} \text{ and } \hat{\sigma}_l^2 = \frac{1}{k_l - 1} \sum_{j=1}^{k_l-1} n_{jl} (\bar{w}_{jl} - \bar{w}_l)^2 \quad \text{where } n_l = \sum_j n_{jl}.$$

Based on these ML estimates it is possible – for a given species – to find the variance of

$$\bar{w} = \sum_l \frac{n_l}{n} \bar{w}_l$$

in formula (12), conditioned by total number of fish, $n = \sum_l n_l$.

It can be shown that this variance is

$$VAR(\bar{w}|n) = \frac{1}{n} \left[ \sum_l \frac{n_l}{n} (\bar{w}_l^2 + \hat{\sigma}_l^2) - \bar{w}^2 \right]. \tag{13}$$

The second term in (8) can be calculated using that $n_{sa}$ for given $n_s$ is binomially distributed:

$$VAR(\sum_a v_{sa} \hat{p}_{sa}) = \frac{\sum_a v_{sa}^2 \hat{p}_{sa} - (\sum_a v_{sa} \hat{p}_{sa})^2}{n_s}. \tag{14}$$

The third term of (10), the covariance, is also conditioned in this case by both the number of fish aged and measured, i.e. $n$ and $l$, where $l$ is less than or equal to $n$ because the aged fish is a subsample of the measured fish. As the first step this covariance can be approximated by

$$COV\left(\sum_a v_{sa} \frac{n_{sa}}{n_s}, \frac{1}{\bar{w}_s} \middle| n_s, l_s\right) \sim -\frac{1}{n_s \bar{w}_s^2} \sum_a v_{sa} \, COV(n_{sa}, \bar{w}_s | n_s, l_s) \tag{15}$$

The last covariance term is:

$$COV\left(n_{sa}, \bar{w}_s | n_s, l_s\right) = \frac{1}{l_s} \sum_b \bar{w}_{sb} \, COV(n_{sa}, l_{sb} | n_s, l_s) = \frac{1}{l_s} \Big[ -\sum_{b \neq a} \bar{w}_{sb} n_s \hat{p}_{sa} \hat{p}_{sb}$$

$$+ \bar{w}_{sa} n_s \hat{p}_{sa} (1 - \hat{p}_{sa}) \Big] = -\frac{\hat{p}_{sa} n_s}{l_s} (\bar{w}_s - \bar{w}_{sa}) \tag{16}$$

Insertion of the formulae (9) - (16) into (8) results in the following approximation to the objective function:

$$O \simeq O_1 + O_{21} + O_{22} + O_3$$

where

$$O_1 = \sum_{qf} \frac{C_{qf}}{m_{qf}} \qquad C_{qf} = \sum_s [VAR(H_{qfs})B_{qfs}^2] + \sum_{s \neq u} [COV(H_{qfs}, H_{qfu})B_{qfs}B_{qfu}]$$

$$O_{21} = \sum_{tfgs} \frac{D_{tfgs}}{l_{tfgs}} \qquad D_{tfgs} = \left( \hat{E}_{q_t fs}, A_{tfgs} \frac{\hat{\sigma}_{tfgs}}{\bar{w}_{tfgs}} \right)^2$$

$$O_{22} = \sum_{tfgs} \frac{E_{tfgs}}{l_{tfgs}} \qquad E_{tfgs} = 2\, \hat{E}_{q_t fs}^2 A_{tfgs} \frac{L_{tfg}}{\bar{w}_{tfgs}} \sum_a v_{sa} \hat{p}_{tfgsa} \left( 1 - \frac{\bar{w}_{tfgsa}}{\bar{w}_{tfgs}} \right)$$

$$O_3 = \sum_{tfgs} \frac{F_{tfgs}}{l_{tfgs}} \qquad F_{tfgs} = \hat{E}_{q_t fs}^2 \left( \frac{L_{tfg}}{\bar{w}_{tfgs}} \right)^2 [\sum_a v_{sa}^2 \hat{p}_{tfgsa} - (\sum_a v_{sa} \hat{p}_{tfgsa})^2]$$

$O_1$ denotes the variation due to the errors related to the species composition;
$O_{21}$ denotes the variation due to errors in the mean weight;
$O_{21}$ denotes the variation due to correlation between mean weight and age;
$O_3$ denotes the variation due to errors in the age distribution;
$H_{qfs}$ denotes the variable for the proportion of species $s$ in one sample in quarter $q$ in fishery $f$.

The variance/covariance of $H_{qfs}$ is calculated using the D-D distribution and the formulae given by Lewy (1995).

## Appendix III. The variance of the catch by weight and numbers at age

The variance of the landings of species $s$ in month $t$ can easily be calculated:

$$VAR(\hat{L}_{ts}) = VAR(\sum L_{tfg} \hat{E}_{q_t fs}) = \frac{\sum_f L^2_{tf} VAR(H_{q_t fs})}{m_{qtfs}} \quad . \qquad (17)$$

Here the calculation of the variance, $VAR(H_{q_t fs})$, is based on the D-D distribution.

When calculating the corresponding variance by quarter, $VAR(\hat{L}_{qs})$, the formula (17) can be used if $L_{tf}$ is replaced by $L_{q_t}$. The variance by year is simply the sum of the quarterly variances.

The variance of the catch in numbers by month, species and age, $VAR(\hat{N}_{tsa})$, can be calculated similarly to the objective function. The result is:

$$VAR(\hat{N}_{tsa}) = V_1 + V_{21} + V_{22} + V_3$$

where

$$V_1 = \sum_{fg} \frac{R_{tfgsa}{}^2 \, VAR(H_{q_t fs})}{m_{qf}}$$

$$V_{21} = \sum_{fg} \frac{(\hat{E}_{q_t fs} R_{tfgsa})^2 \left( \frac{\sigma_{tfgs}}{\bar{w}_{tfgs}} \right)^2}{l_{tfgs}}$$

$$V_{22} = \sum_{fg} \frac{2(\hat{E}_{q_t fs} R_{tfgsa})^2 (1 - \bar{w}_{tfgsa} / \bar{w}_{tfgs})}{l_{tfgs}}$$

$$V_3 = \sum_{fg} \frac{(\hat{E}_{q_t fs} R_{tfgsa})^2 (1/\hat{p}_{tfgsa} - 1)}{n_{tfgs}}$$

and

$$R_{tfgs} = \hat{p}_{tfgsa} \frac{L_{tfg}}{\bar{w}_{tfgs}}$$

$V_1$ denotes the variation due to the errors related to the species composition;
$V_{21}$ denotes the variation due to errors in the mean weight;
$V_{22}$ denotes the variation due to correlation between mean weight and age;
$V_3$ denotes the variation due to errors in the age distribution.