

This is a pre-copy-editing, author-produced PDF of an article accepted for publication in ICES Journal of Marine Science following peer review. The definitive publisher-authenticated version JUL 2007; 64 (5) 1028-1032 is available online at: <http://dx.doi.org/10.1093/icesjms/fsm077>.

Detecting sampling outliers and sampling heterogeneity when catch-at-length is estimated using the ratio estimator

Joël Vigneau¹ and Stéphanie Mahévas²

¹ IFREMER, Avenue du Général de Gaulle, 14520 Port-en-Bessin, France

² IFREMER, rue de l'île d'Yeu, 44311 Nantes, France

Correspondence to J. Vigneau: tel: +33 231 515600; fax: +33 231 515601; e-mail:
joel.vigneau@ifremer.fr

Abstract:

Measuring fish on board fishing vessels or at fish markets to collect data for stock assessment purposes is one of the most straightforward actions carried out by fisheries scientists worldwide. However, such samples are not straightforward to handle and analyse because of their vector-type structure. A generic tool that allows investigation in any multinomial-like sampling scheme is provided, as long as the scheme is built on a ratio estimator, which is the case for most length sampling in the fisheries sector. The use of this tool is discussed using data obtained from two different sampling designs, one consisting of commercial market samples by category and the other on fishing activity or métier. The identification of outliers, misallocated samples, or potential bias as well as the analysis of heterogeneity within and between strata are discussed. The objective of such exploratory analyses is to help sampling coordinators design the best sampling scheme and improve the quality of input data for stock assessment models. The statistics described here are easy to implement and their use is recommended as a necessary stage before any use of sampling data at an international level.

Keywords: length structure, market sampling, sampling design, stratification

Detecting sampling outliers and sampling heterogeneity when catch-at-length is estimated using the ratio estimator

65

5

Joël Vigneau and Stéphanie Mahévas

10 Vigneau, J., and Mahévas, S. 2007. Detecting sampling outliers and sampling heterogeneity when catch-at-length is estimated using the ratio estimator. – ICES Journal of Marine Science, 64.

70

Measuring fish on board fishing vessels or at fish markets to collect data for stock assessment purposes is one of the most straightforward actions carried out by fisheries scientists worldwide. However, such samples are not straightforward to handle and analyse because of their vector-type structure. A generic tool that allows investigation in any multinomial-like sampling scheme is provided, as long as the scheme is built on a ratio estimator, which is the case for most length sampling in the fisheries sector. The use of this tool is discussed using data obtained from two different sampling designs, one consisting of commercial market samples by category and the other on fishing activity or métier. The identification of outliers, misallocated samples, or potential bias as well as the analysis of heterogeneity within and between strata are discussed. The objective of such exploratory analyses is to help sampling coordinators design the best sampling scheme and improve the quality of input data for stock assessment models. The statistics described here are easy to implement and their use is recommended as a necessary stage before any use of sampling data at an international level.

75

80

Keywords: length structure, market sampling, sampling design, stratification.

Received 6 October 2006; accepted 30 April 2007.

85

25 J. Vigneau: IFFREMER, Avenue du Général de Gaulle, 14520 Port-en-Bessin, France. S. Mahévas: IFFREMER, rue de l'île d'Yeu, 44311 Nantes, France. Correspondence to J. Vigneau: tel: +33 231 515600; fax: +33 231 515601; e-mail: joel.vigneau@ifremer.fr.

30 Introduction

Monitoring fisheries is essential to providing diagnostics on the state and dynamics of marine resources exploited by fishing activities. Measuring fish length at fish markets or on board fishing vessels mobilizes a lot of manpower all around the world. Together with age-reading estimates, these operations constitute the basic data source for estimating population dynamics in fisheries science (Pastoors *et al.*, 2001). Inherent to any sampling procedure, estimating the fish length distribution in a sample may contain bias and uncertainties. The dissemination of errors or uncertainties in input data for assessment models has been studied by Kimura (1989), Pelletier (1991), Restrepo *et al.* (1992), and more recently by Patterson *et al.* (2001) and Reeves (2003). Those studies showed that the quality of biological advice is highly dependent on the quality of the underlying data. Assessment of the accuracy in the sampling design and optimization of the sampling intensity are therefore prerequisite to achieving any targeted level of precision and to certifying the representativity of the estimation of fish landings-at-length.

The question of sampling intensity (i.e. the quantity of samples to collect) has been widely studied to achieve a certain level of precision (Lai, 1987; Quinn and Deriso, 1999). A procedure or a robust tool for investigating the accuracy of a sampling design is currently not available to ensure a validated stock-standardized approach at regional and European stock scales (ICES, 2006). Analysis of a sampling design aimed at estimating fish length distribution requires a statistical procedure appropriate to vector type estimates. To investigate the quality of sampling design in a vector-type estimator, an Euclidean distance (Lele and Cole,

1996) or a Mahalanobis distance (Dryden and Mardia, 1998) can be calculated that would take into account the multicollinearity between all values. Generalization of the Hotelling T^2 (Hotelling, 1931) to p variables can also be used as a measure of distance, because it is the counterpart of the Student's t -test used in control charts. Even if this last statistic depends on the normality of multinomial samples, Srivastava and Mudholkar (2001) made it more robust by trimming the samples. Bootstrapping is a technique used for assessing the accuracy of almost any statistical estimate and testing the correlation coefficient between observed and theoretical proportions such as proposed by Morales *et al.* (2004).

All these statistics are relevant for comparing the significance of the difference between two (sets of) samples. However, they are designed to compare sample means when they are assumed to be of equal distribution. The distance between samples can then be calculated, but they would be difficult to interpret because they cannot be linked to the variance formula of the landings-at-length. To scrutinize the samples, an index known as Δ is presented here; it is derived directly from the analytical expression of the landings-at-length variance when a ratio estimator is used. Reference to the ratio estimator is justified by the fact that the vast majority of raising procedures used to derive landings-at-length estimates from market, harbour or at-sea sampling is based on the ratio between the number of fish measured and the related weight of the sample. This has been shown for the ICES region by Pastoors *et al.* (2001). The purpose of the Δ index is to summarize information about the discrepancy between one vector-type sample and the overall vector derived from all samples. Such transformation of a vector-type

90

95

100

105

110

115

120

estimate to a scalar-type estimate will permit sampling exploratory analyses according to the common rules of sampling theory (Cochran, 1977; Thomson, 1992). The objectives of this tool are twofold: (i) to understand and/or to quantify the contribution of individual samples to the overall variance and (ii) to estimate the similarities between samples. For the latter objective, the discrepancy between samples is not squared to preserve the relative order between each sample. For this reason, the index presented here cannot be considered as a distance.

The use of the Δ index will be considered for samples derived from both commercial category sampling and métier-based sampling, which correspond to the known technical stratification utilized in Europe (ICES, 2004). The difference between the two strategies relates to the sampling unit, in this case a box of fish, and the total landings of that species during a fishing trip, which takes place using specific gear in a known area and targeting a given (assemblage of) species, respectively (EC, 2006). In general, sampling by commercial categories requires stability in specifications over the harbours sampled, whereas sampling by métier implies an assumption of similarity in exploitation pattern. Two case studies obtained from French fish market samplings were used to demonstrate the use of the Δ index: commercial category sampling of sole (*Solea vulgaris*) from the eastern Channel, ICES Division VIIId, and métier sampling of hake (*Merluccius merluccius*) from the eastern Atlantic, ICES Subareas VII and VIII.

Methods

Variance of landings-at-length

The Δ index is directly derived from formulation of the variance in landings-at-length. Let j be the length class index ($j = 1, \dots, J$), so the total landings by number \hat{D} is expressed as the sum of the total landings at length j as

$$\hat{D} = \sum_j \hat{D}_j \quad (1)$$

and its variance is given by

$$\text{Var}(\hat{D}) = \sum_j \text{Var}(\hat{D}_j) + \sum_{j \neq j'} \text{Cov}(\hat{D}_j, \hat{D}_{j'}). \quad (2)$$

Here, the covariance term is considered as negligible and the focus is placed on $\text{Var}(\hat{D}_j)$. Let d and w be the number of fish in the sample and the sampled weight, respectively, and W the total weight of landings. The term k refers to the stratum index ($k = 1, \dots, K$), and v is the sample index ($v = 1, \dots, n_k$). Respectful of field sampling and assuming that the number of fish measured is dependent on the sample weight, the estimator of landings-at-length j may be decomposed as follows:

$$\hat{D}_j = \sum_{k=1}^K \frac{W_k}{\sum_{v=1}^{n_k} w_{kv}} \left(\sum_{v=1}^{n_k} d_{j_{kv}} \right) = \sum_{k=1}^K W_k \frac{\sum_{v=1}^{n_k} d_{j_{kv}}}{\sum_{v=1}^{n_k} w_{kv}}. \quad (3)$$

Table 1. Possible variants of the Δ index.

	One length class	Several or all length classes
One stratum	$\Delta_{jkv} = d_{j_{kv}} - \frac{\sum_v d_{j_{kv}} w_{kv}}{\sum_v w_{kv}}$	$\Delta_{kv} = \sum_j \left(d_{j_{kv}} - \frac{\sum_v d_{j_{kv}} w_{kv}}{\sum_v w_{kv}} \right) + \sum_{j'} \Delta_{j'kv}$
Several or all strata	$\Delta_{jv} = d_{j_{kv}} - \frac{\sum_k \sum_v d_{j_{kv}} w_{kv}}{\sum_k \sum_v w_{kv}}$	$\Delta_v = \sum_j \left(d_{j_{kv}} - \frac{\sum_k \sum_v d_{j_{kv}} w_{kv}}{\sum_k \sum_v w_{kv}} \right) + \sum_{j'} \Delta_{j'v}$

Formulation (3) expresses the raising procedure suggested that calculates the ratio between the number of fish measured and the related sampled weight, which is then re-scaled to total weight landed to provide estimates of landings-at-length. The estimate of variance is given by

$$\text{Var}(\hat{D}_j) = \sum_{k=1}^K W_k^2 \text{Var} \left(\frac{\sum_{v=1}^{n_k} d_{j_{kv}}}{\sum_{v=1}^{n_k} w_{kv}} \right). \quad (4)$$

Analytical development of the variance estimate is based on the formula of Cochran (1977) relating to the approximation of the variance of a ratio:

$$\text{Var} \left(\frac{\sum_{v=1}^{n_k} d_{j_{kv}}}{\sum_{v=1}^{n_k} w_{kv}} \right) = \frac{1 - \left(\frac{\sum_{v=1}^{n_k} w_{kv}}{W_k} \right)}{\left(\frac{\sum_{v=1}^{n_k} w_{kv}}{W_k} \right)^2 / n_k} \times \frac{\sum_{v=1}^{n_k} \left(d_{j_{kv}} - \frac{\sum_{v=1}^{n_k} d_{j_{kv}} w_{kv}}{\sum_{v=1}^{n_k} w_{kv}} \right)^2}{n_k - 1}. \quad (5)$$

Use of the Δ index for exploratory analysis

The Δ index, which is directly derived from the principal component of Equation (5), compares the number-at-length in the sample ($d_{j_{kv}}$) with the number-at-length of all samples re-scaled to the sampled weight. Four variants can be adapted to explore the samples of a stratum regarding one length class in one stratum (Δ_{jkv}), several or all length classes in one stratum (Δ_{kv}), the samples over all strata regarding one length class (Δ_{jv}), or the samples over all strata considering several or all length classes (Δ_v). See Table 1 for a full expression of these four variants.

There are two ways of using the Δ index, squared or not squared. When not squared, the relative order between each element is preserved and can be used to show sample similarities. In that case, it is noticeable that values of Δ are always centered on 0, whatever the variant used. When squared, the Δ index quantifies the exact contribution of a single sample to the overall variance. It then allows identification of the most influential sample(s) or possible outlier(s). A sample is regarded as an outlier if, after careful verification, it is considered as an accidental occurrence not representative of the population being sampled. However, such possible outliers can also be detected with the non-squared Δ index by checking the validity of extreme values. The non-squared Δ index is therefore more informative and is the form upon which the results are discussed in this document.

Case studies

Two case studies are described here to represent commercial category and métier sampling: eastern Channel sole (*Solea vulgaris*) and Northeast Atlantic hake (*Merluccius merluccius*). Sole landings

245 in the eastern Channel are mostly shared between trawlers and gill-
 250 netters. The objective for sampling is based on commercial categories distributed among the principal harbours and stratified by quarter. Sorting into five commercial categories, as based on EU standards, is assumed to be stable between auctions and over the whole year.

Hake are fished with gillnets, trawls, and lines, representing fishing activities classified in internationally specified Fishery Units (FU) (Artexe *et al.*, 2002). French landings of hake are sampled within five FUs: FU05 (inshore fish trawler in ICES Subarea VII), FU09 (*Nephrops* trawlers in ICES Subarea VIII), FU10 (trawlers in ICES Subarea VIII), FU12 (longliners in ICES Subarea VIII), and FU13 (gillnetters in ICES Subarea VIII). The sampling scheme is distributed among six harbours, from south of the Bay of Biscay to south of Brittany, stratified by quarter.

Results

Outliers and misallocated samples

265 The general assumption that samples represent the underlying population may be distorted by one or more samples taking most of the variance in a single stratum. Removing or reallocating such samples has to be done objectively following careful verification. In the example shown in Figure 1a, the investigation of odd values of the Δ index revealed that sample 89 151 was an outlier because of an error in sample weight, and that sample 85 284 was misallocated and should have been a sample of market category 50. No error was identified for samples 91 346 and 92 785, so they were retained in the analysis. The Δ index values obtained for sampling for sole in VIId are more homogeneous after removal of the outliers and misallocation correction (Figure 1b).

275 M tier sampling is an alternative to market category sampling. The main difference between the two methods is that m tier-based

sampling requires rising to total landings for a fishing trip to reconstitute the overall distribution of landed fish length, something that is not carried out for market category sampling. The difference is shown when comparing the magnitude of the Δ values for hake (Figure 2) with those for sole (Figure 1b). Moreover, the uncertainty attributable to raising the sampling to the total landings of one vessel is unknown because of the absence of replicates during the sampling of each individual commercial category. The heterogeneity also tends to be enhanced when one mixes trips with low and high landing volumes. We did not find any outliers or misallocated samples in this second sampling design.

Heterogeneity between strata

When sorting Δ values by commercial category (Figure 1a), ordination between the biggest fish (category 10) and the smallest fish (category 50) is clear. Shading discriminates the northeastern from the southwestern part of Division VIId. No evidence of difference in Δ values was found between the two geographical areas. Moreover, sorting the data into five commercial categories may not be optimal because of the lack of obvious difference between consecutive categories. Splitting in two or three categories may have improved the precision for the same sampling effort, because more samples are allocated to each stratum. This reveals the potential danger of defining too many strata in a sampling design.

In hake sampling by m tier (Figure 2), the most heterogeneous stratum observed is FU13, perhaps because of the longer trips undertaken when operating this m tier as opposed to the shorter trips for other m tiers. The Δ index is linked linearly to the number of fish measured, which corresponds to the number of fish raised to the whole trip sampled for m tier sampling. The optimal sampling allocation combines the heterogeneity and the size of a stratum. In this case, the sampling effort allocated to

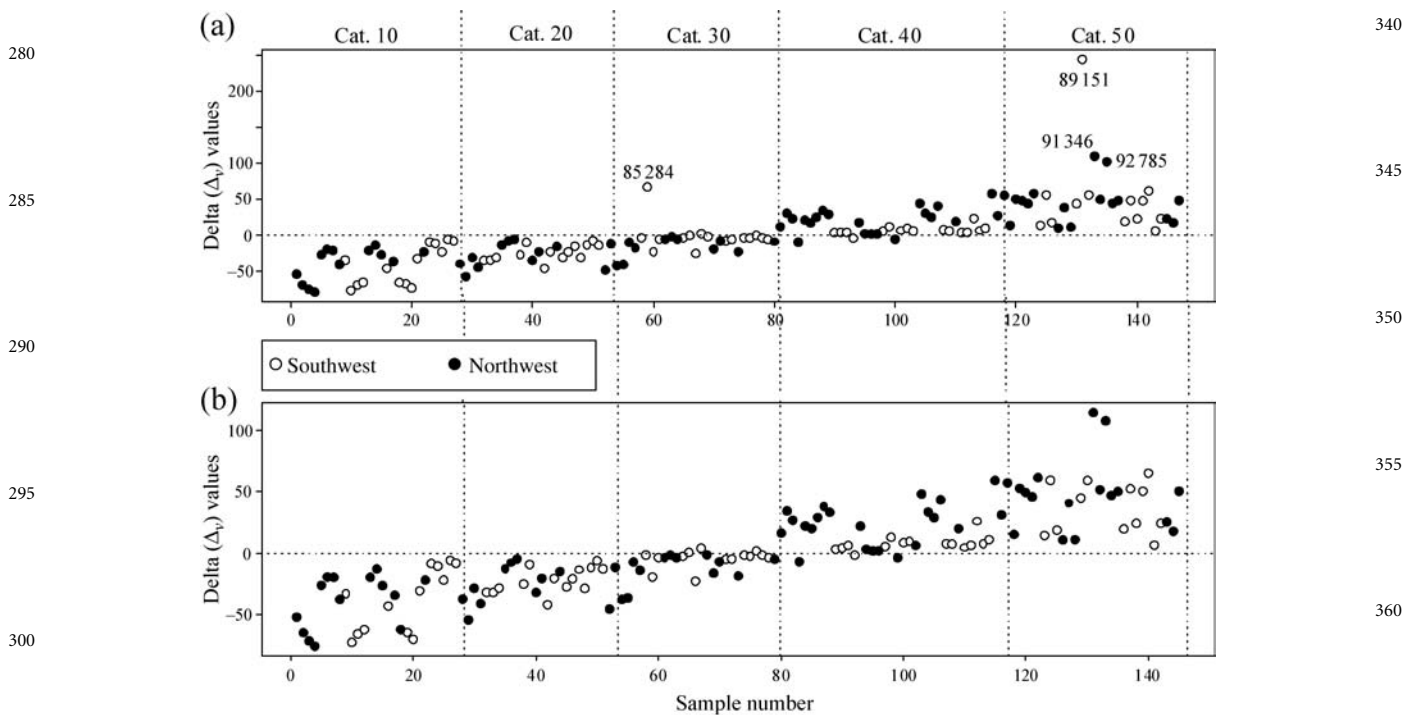


Figure 1. Sole in VIId in 2003. Samples ordered by market category: (a) all samples and (b) after correction of the dataset. Shading of the points represents a split of Division VIId into two parts, southwest and northeast. Note that the scale on the y-axes differs.

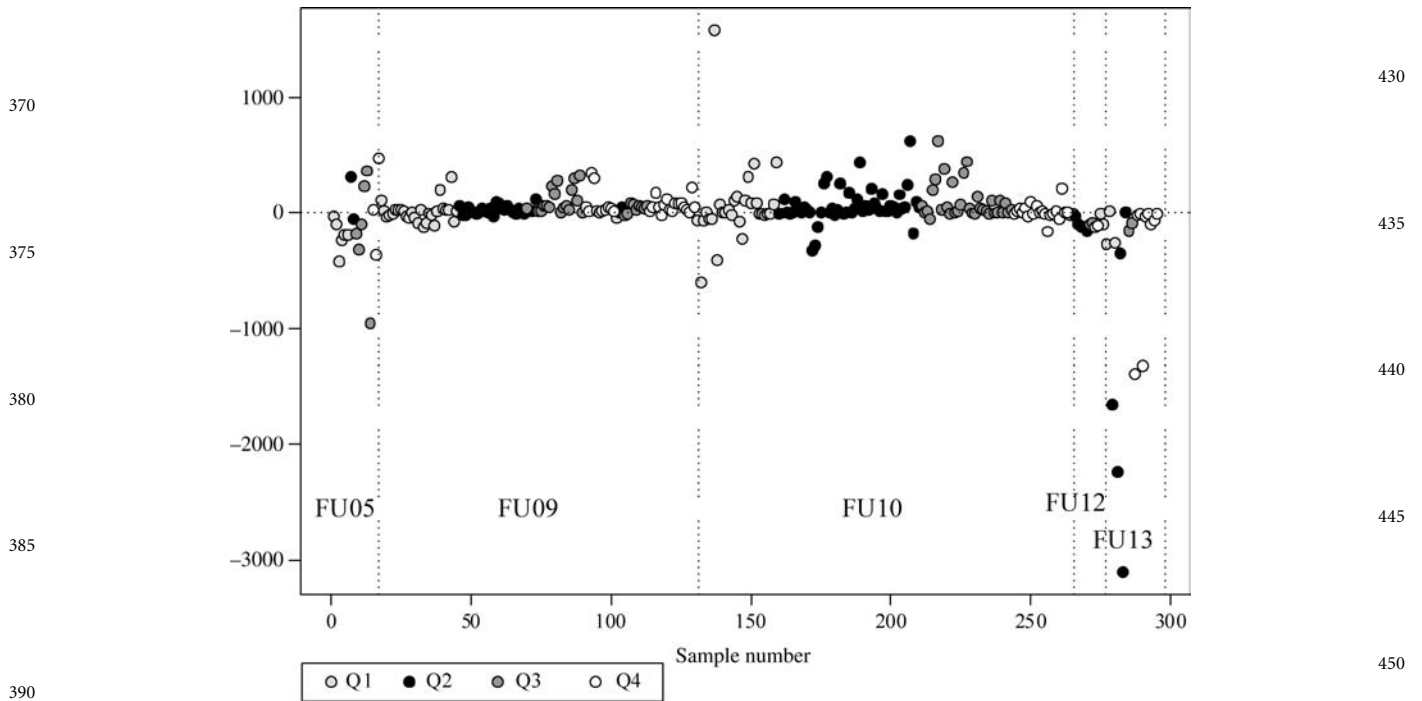


Figure 2. Hake in the Atlantic in 2003, with samples ordered by FU.

FU13 is likely to be insufficient if the total landings of the stratum are significant. Adjusting the focus may highlight differences undetected in the general picture. This can be observed for hake landed in FU09 and FU10 with seemingly similar results (Figure 2). However, if just quarter 4 is specifically taken into account, the hake landed in FU09 are much smaller than those in FU10.

Heterogeneity within strata

Searching for homogeneous patterns within one stratum or within a specific group of samples is possible with the variants Δ_{jkv} and Δ_{kv} of the index. For example, the difference in fishing small or less small fish by a given gear can be examined by focusing on the category containing the smallest fish (Figure 3a). The results are clearly counter-intuitive, beam trawlers showing a different mean length distribution than either gillnetters or otter trawlers, which have similar length distributions. However, the small number of samples here precludes firm conclusions being drawn from Figure 3a. Differences between geographical regions (Figure 3b) are probably attributable to beam trawlers being sampled mainly in the northeast of Division VIIId.

Discussion

Market sampling for fish length is an extensive and time- and manpower-consuming process. Outliers and misallocated samples incorrectly increase the variance and the bias of final estimates. The international age-structured aggregated data used in assessment models result from a multi-stage process beginning with market sampling for fish length. In the process, the variance attached to the total volume of landings and to the age-length keys dominates the variance attached to the length structure (Gavaris and Gavaris, 1983). This is not the case for bias, because a bias at the beginning of the process will strongly distort the final estimate. The issue becomes serious when considering technical measures based on gear selectivity or when aiming to provide data for length-based

stock assessment. The Δ index is a practical and flexible tool for exploring length distributions derived from sampling catches at sea or at ports. It constitutes an initial step towards development of a more complete tool for multi-stage analysis while quantifying final estimates of uncertainty. Moreover, merging the data at an international level should probably better be done at a disaggregated level, such as by métier or fishery unit, but this can only be achieved after careful data investigation.

Investigating the sources of variability is essential when developing a plan of sampling. Stratifying to reduce variance leads to clustering similarities in a population, i.e. in exploitation patterns among a multitude of métiers or well-discriminated commercial categories. The Δ index is flexible enough to explore whether métiers differ from one another based on the smallest and/or largest length classes, or whether two commercial categories are similar enough to be combined. Regrouping strata with similar length structure is therefore essential to minimizing the risk of overstratification.

The unit of the Δ index represents a number of individuals, and careful analysis of results is recommended because the value is based on the sum of positive and negative values. A low Δ value for one sample means that (i) very few individuals have been measured; (ii) all lengths are close to the mean values; or (iii) the positive and negative values counterbalance each other. The first two cases are expected and can be analysed as such. However, to avoid confusion, one half of the length range should be selected to cancel the counterbalancing effect. Care needs to be taken when weighing a sample, the second variable of this ratio estimator.

The collection of sampling information in fisheries will likely not be reduced, if fishery stakeholders and managers are willing to consider the interactions between fishing and marine resources and ecosystems. However, the representivity and the precision of the sampling information will need to be better assessed to certify the quality of the data underpinning fisheries advice, and

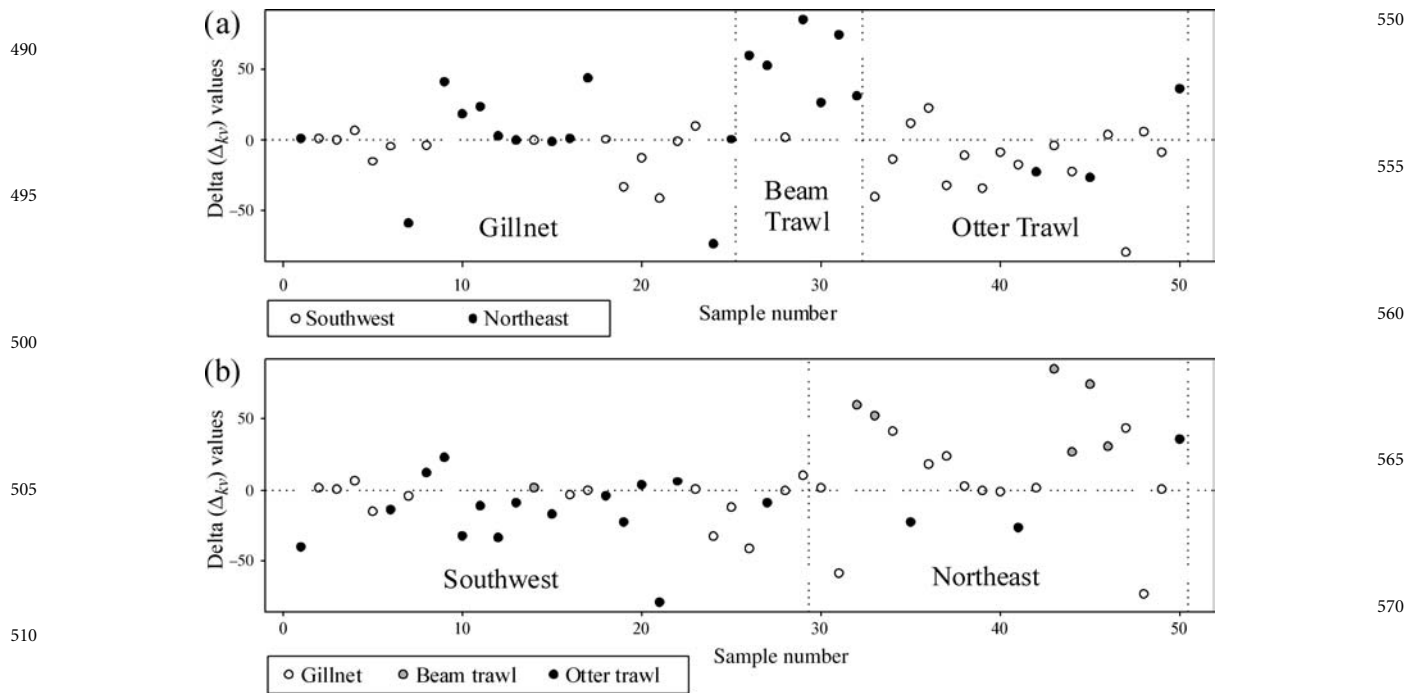


Figure 3. Sole in VIld, 2002–2004. Commercial category 50. Samples ordered by (a) gear and (b) area.

this process begins with scanning the raw data with a tool such as the Δ index.

Acknowledgements

We thank the staff in IFREMER involved in the collection of sole and hake length samples, and Florence Nedelec, Ching Villanueva, and two anonymous reviewers for valued comments which improved the quality of the manuscript considerably. Free R-software was used for the work and we thank the R Development Core Team and all contributors to the R project (<http://www.R-project.org>).

References

Artexe, I., Belisario, A., Connolly, P., Gaudou, O., Jardim, E., Maxell, D., Millner, R. *et al.* 2002. Improving sampling of western and southern European Atlantic fisheries (SAMFISH). EU Study contract 99-099, Final Report. 184 pp.

Cochran, W. G. 1977. Sampling Techniques, 3rd edn. John Wiley, Chichester.

Dryden, I. L., and Mardia, K. V. 1998. Statistical Shape Analysis. John Wiley, Chichester. 347 pp.

EC. 2006. Report of the *ad hoc* meeting of independent experts on fleet–fishery based sampling. Commission Staff Working Paper. 101 pp.

Gavaris, S., and Gavaris, C. A. 1983. Estimation of catch at age and its variance for groundfish stocks in the Newfoundland Region. Canadian Special Publication of Fisheries and Aquatic Sciences, 66: 178–182.

Hotelling, H. 1931. The generalization of Student’s ratio. *Annals of Mathematical Statistics*, 2: 360–378.

ICES. 2004. Report of the Workshop on Sampling and Calculation Methodology for Fisheries Data. ICES Document CM 2004/ACFM: 12. 20 pp.

ICES. 2006. Report of the Planning Group on Commercial Catch, Discards and Biological Sampling. ICES Document CM 2006/ACFM: 18. 62 pp.

Kimura, D. K. 1989. Variability in estimating catch-in-numbers-at-age and its impact on cohort analysis. *Canadian Special Publication of Fisheries and Aquatic Sciences*, 108: 57–66.

Lai, H. L. 1987. Optimum allocation for estimating age composition using and age-length key. *Fishery Bulletin US*, 85: 179–185.

Lele, S., and Cole, T. M. 1996. A new test for shape differences when variance-covariance matrices are unequal. *Journal of Human Evolution*, 31: 193–212.

Morales, D., Pardo, L., and Santamaria, L. 2004. Bootstrap confidence regions in multinomial sampling. *Applied Mathematics and Computation*, 155: 295–315.

Pastoor, M. A., O’Brien, C. M., Flatman, S., Darby, C. D., Maxwell, D., Simmonds, E. J., Degel, H. *et al.* 2001. Evaluation of market sampling strategies for a number of commercially exploited stocks in the North Sea and development of procedures for consistent data storage and retrieval (EMAS). CFP Study Project 98/075: 298 pp., Appendices.

Patterson, K. R., Cook, R. M., Darby, C., Gavaris, S., Kell, L. T., Lewy, P., Mesnil, B. *et al.* 2001. Estimating uncertainty in fish stock assessment and forecasting. *Fish and Fisheries*, 2: 125–157.

Pelletier, D. 1991. Les sources d’incertitude en gestion des pêcheries. Evaluation et propagation dans les modèles. INAPG. 291 pp.

Quinn, T. J., and Deriso, R. B. 1999. Quantitative Fish Dynamics. Oxford University Press. 542 pp.

Reeves, S. A. 2003. A simulation study of the implications of age-reading errors for stock assessment and management advice. *ICES Journal of Marine Science*, 60: 314–328.

Restrepo, V. R., Hoenig, J. M., Powers, J. E., Baird, J. W., and Turner, S. C. 1992. A simple simulation approach to risk and cost analysis, with applications to swordfish and cod fisheries. *Fishery Bulletin US*, 90: 736–748.

Srivastava, D. K., and Mudholkar, G. S. 2001. Trimmed T2: a robust analog of Hotelling’s T2. *Journal of Statistical Planning and Inference*, 97: 343–358.

Thomson, S. K. 1992. Sampling. Wiley Interscience, New York. 343 pp.