This paper is not to be cited without prior reference to the author(s)

International Council for the Exploration of the Sea | Incorporation of External Factors in Marine Resource Surveys CM 2000/K:19

Accounting for spatial-scale in research surveys: analyses of 2-year old cod from English, German and International groundfish surveys in the North Sea

[1]C.M. O'Brien, [2]S. Adlerstein, and [3]S. Ehrich

[1]CEFAS Lowestoft Laboratory, Pakefield Road, Lowestoft, Suffolk NR33 0HT, United Kingdom [fax: +44 (0) 1502 513865, e-mail: c.m.obrien@cefas.co.uk]

[2]Universität Hamburg, Institut für Hydrobiologie und Fischereiwissenschaft, Olbersweg 24, D-22767 Hamburg, Germany [fax: +49 40 42838 6618, e-mail: sadlerstein@uni-hamburg.de]

[3]Bundesforschungsanstalt für Fischerei, Institut für Seefischerei, Palmaille 9, D-22767 Hamburg, Germany [fax: +49 40 38905 263, e-mail: ehrich.ish@bfa-fisch.de]

ABSTRACT

Routine research surveys are a major source of fisheries independent information for estimating stock abundance. Bottom trawl surveys are routinely performed at a coarse spatial scale and their abundance estimates consist of global means derived with standard formulae that give imprecise estimates. Model-based approaches have been proposed to increase precision but when estimates are derived from coarse surveys they do not account for spatial variation occurring at finer scales. In this paper we present the results of a study to investigate spatial indices and the empirical distribution of catch data collected at different spatial scales. The value of using high-resolution spatial catch data and environmental information to improve model–based estimates of distribution are discussed. The scale of the three surveys from which data are considered ranges from a coarse spatial resolution with a grid covering the whole North Sea (International bottom trawl survey, IBTS and English groundfish survey, EGFS) to a fine-scale in several delimited areas (German small-scale bottom trawl survey, GSBTS). North Sea cod (*Gadus morhua* L.) is presented as a timely case study and catch rates of 2-year old cod in the central North Sea are analysed from the third quarter of 1995. Results indicate the suitability of using a negative binomial distribution to describe the catch distribution at the different spatial scales considered. The benefits of using a general assumption about mixing processes are discussed.

Keywords: aggregation, conditionality, groundfish survey, heterogeneity, mixture distribution, scale

INTRODUCTION

Routine research surveys are a major source of fisheries independent information for estimating stock abundance. Bottom trawl surveys are routinely performed at a coarse spatial scale that provides regular spatial coverage. Abundance estimates consist of global means derived with standard formulae but these estimates are usually rather imprecise. Model-based approaches have been proposed as an option to increase precision. When estimates are derived from coarse spatial scale surveys they do not account for spatial variation occurring at finer scales.

In this paper we present the results of a study to investigate the value of using high-resolution spatial catch data. Intuitively, information that improves understanding of the variability of fish distribution should be helpful in specifying models to describe fish abundance. Three surveys will be considered which are conducted by research vessels but at different spatial scales. The scale of the surveys ranges from a coarse spatial resolution with a grid covering the whole North Sea (International Bottom Trawl Survey, IBTS and English GroundFish survey, EGFS) to a fine-scale in several delimited areas (German Small-scale Bottom Trawl Survey, GSBTS).

Statistical methods will be used to estimate spatial indices and the empirical distribution of catch data, and to investigate the spatial distribution of stock abundance. The benefits of combining the small- and large-scale surveys are discussed in the context of the analyses presented. North Sea cod (*Gadus morhua* L.) is presented as a timely case study. Catch rates of 2-year old cod in the central North Sea are analysed from the 3rd quarter of 1995.

MATERIAL AND METHODS

In the North Sea, the International Bottom Trawl Survey (IBTS) is conducted annually and provides estimates of fish abundance used to calibrate stock assessments (ICES, 1996). The design-based indices simply consist of stratified means of catch per unit effort, the strata corresponding to standard species-specific areas. Other routine surveys with higher spatial resolution have been conducted in the region. Amongst these is the German Small-scale Bottom Trawl Survey (GSBTS) conducted since 1987, together with others to study the fine spatial distribution of fish abundance and the variability of catch rates within areas of the IBTS grid (Gröger and Ehrich, 1992). This paper will use these data with unique characteristics collected by North Sea surveys conducted at different spatial resolutions. An indication of the survey coverage is shown by the third quarter surveys in 1995 (Figure 1).

Research vessel surveys

GSBTS: A series starting in 1987 of catch data at a fine-scale spatial resolution for North Sea stocks conducted annually mainly during the summer months. Data available are from hauls collected in 8 areas of 10nm by 10nm (about 25 hauls per area by year). Originally, the boxes were selected to sample in areas of high cod catches but this situation has changed within the last 10 years. Since 1999, two additional survey boxes in the northern North Sea have been sampled. Catch-at-length data and age information are available for cod, haddock, whiting, saithe, herring, mackerel, and Norway pout. The exact location of each haul is available.

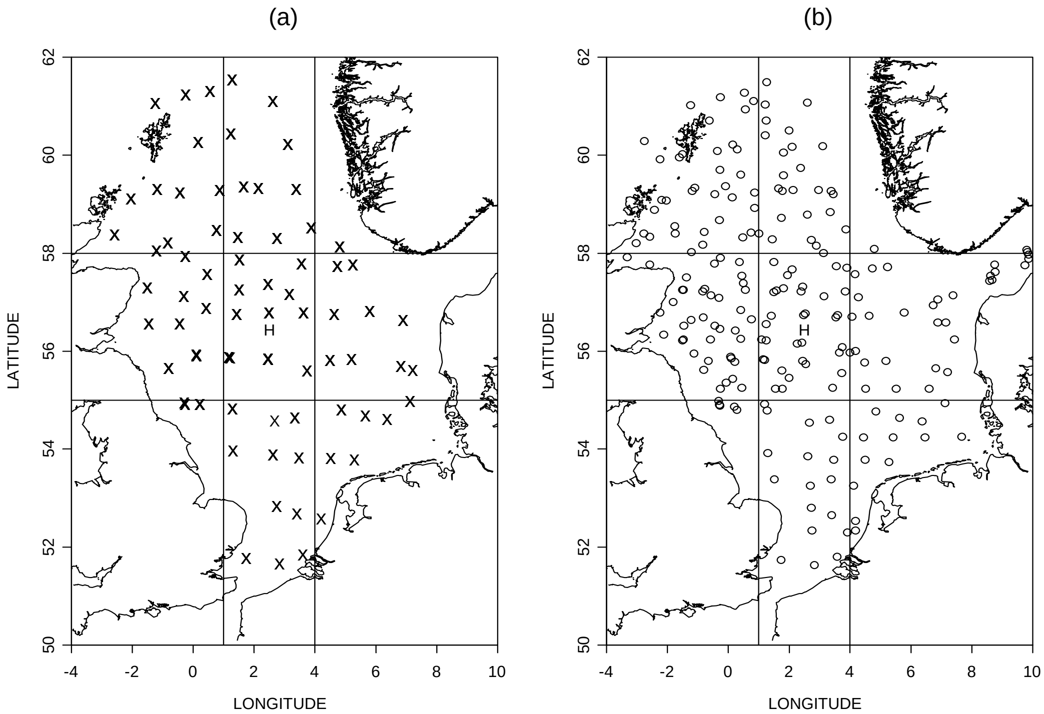

Figure 1. Location of all survey stations during the third quarter of 1995: (a) EGFS trawls (shown by the crosses), and (b) IBTS trawls (shown by the circles). The location of the GSBTS (Box H) indicated by the single letter H in each map.

EGFS: A series of trawling surveys for demersal fish in the North Sea started in the summer (3rd quarter) of 1977. The broad aim was to establish a time series of data that could be used to study the biology and ecology of demersal fish and certain problems of fisheries management (Harding *et al.*, 1986).

IBTS: Since 1991 international collaboration has made it possible to undertake surveys in each of the four quarters throughout the year with at least one haul on a 1° longitude *x* 0.5° latitude grid. The surveys are undertaken during January/February (corresponding to the 1st quarter), May/June (corresponding to the 2nd quarter), August/September (corresponding to the 3rd quarter) and October/November (corresponding to the 4th quarter) to carry out a groundfish survey of the North Sea using a standard GOV trawl gear. However, since 1997 co-ordinated surveys during the 2nd and 4th quarter of the year have not taken place. The exact location of each haul is available.

Sources of catch information for comparative study

The GSBTS and the EGFS/IBTS are undertaken within the North Sea at different levels of spatial coverage and at different months of the year. To enable an investigation of the variability of catch rates within these surveys it was decided to consider catch rates of 2-year old cod in the central North Sea. Survey data are available at a similar time of year for all three surveys during the 3rd quarter of 1995.

The GSBTS has a designated Box H in the central North Sea. Trawl stations of the EGFS/IBTS were selected to be within an area defined between longitude 1° and 4°, and between latitude 55° and 58° (Figure 2). Box H of the GSBTS provides catch data for 26 hauls, the area identified in close proximity to Box H covered by the EGFS provides 14 hauls, and the area identified in close proximity to Box H covered by the IBTS provides 36 hauls. These choices allow the variability of catch rates within the selected areas to be investigated over several day/week/month periods for different levels of aggregation (c.f. Ehrich *et al.*, 1998).

Since 1991 the EGFS data has been a subset of the IBTS data set so the 14 hauls identified from the EGFS in 1995 are included within the 36 hauls identified from the IBTS. The fine-scale data is from the research cutter *Solea* that has a towing speed of 3.5 knots when fishing, in comparison to the 4 knots used for fishing tows within the EGFS and IBTS. The IBTS uses a standard GOV trawl gear, whereas *Solea* is equipped with a smaller but similar otter trawl. No significant differences for cod have been detected between the two gears.

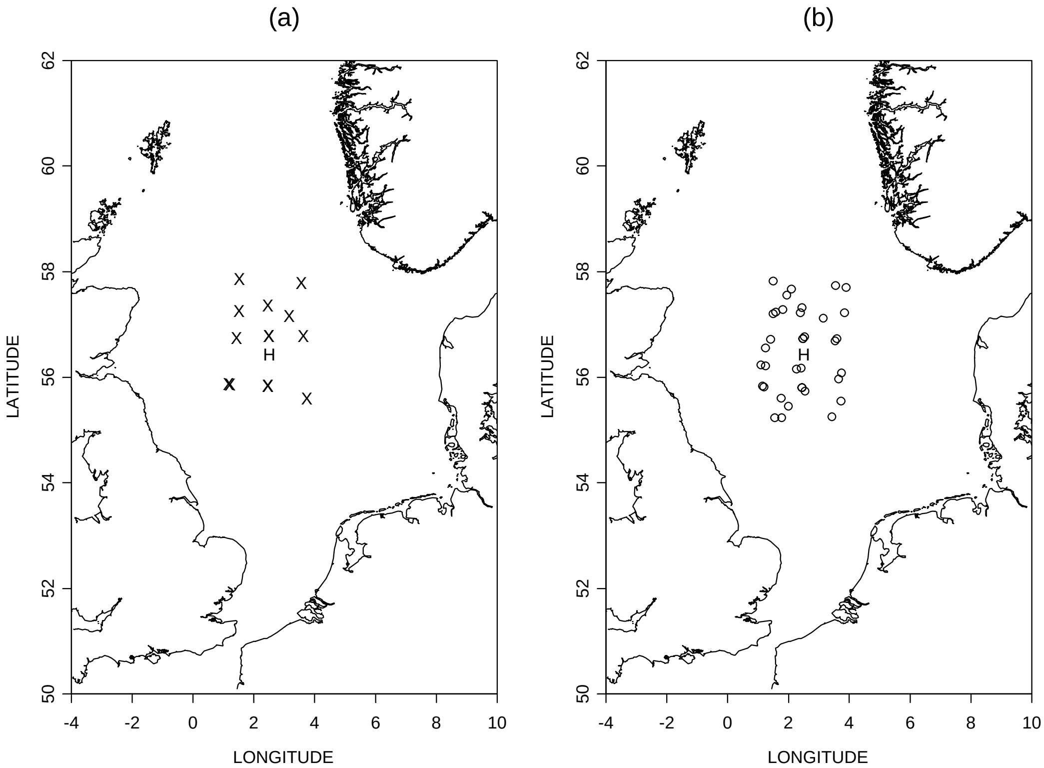

Figure 2. Location of the survey stations during the third quarter of 1995 in close proximity to the sampled Box H: (a) EGFS trawls (shown by the crosses), and (b) IBTS trawls (shown by the circles). The location of the GSBTS (Box H) indicated by the single H in each map.

Empirical distribution of catch

The empirical distribution of catch per tow can vary and depends on a number of different factors. In many cases, scientists assume the Poisson or related distributions as processes which may have generated the catch data (Taylor, 1953) but data transformations may seem appropriate; e.g. the non-linear logarithmic transformation in the case of an assumed log-normal distribution. Other scientists, in contrast, remove zero catches and after suitable transformation, treat both categories of data (zero and non-zero catches) separately (Pennington and Grosslein, 1978). Often, however, transformations are not without problems; conclusions representative of the original (untransformed) catch data might be difficult and at best *approximate*. Clumping, aggregation or contagion characterizes most populations in the wild and counts of individuals $x_1, x_2, ..., x_n$ on small sub-areas of known size are among the oldest techniques in ecology employed to aid in the understanding of such processes. The approach to the statistical analysis of spatial patterns based on quadrat counts is simple and straight-forward. The quadrat of interest is defined to be the sampled station - each station corresponding to a standard tow of fixed duration.

Spatial indices and over-dispersion

Spatial aggregation in the distribution of fish *between* the stations sampled might be conjectured on the basis of calculated indices. Many indices have been proposed and the commonly used ecological and statistical indices are adopted in this paper ($\bar{x}$, s^2, I, ICS, ICF, IP, U) through their application to catch data of 2-year old cod from the central North Sea. The various indices are first defined and discussed in general terms.

Fisher *et al.*(1922) and many others since have suggested the index of dispersion

$$I = s^2 / \bar{x} \qquad (1)$$

where $s^2 = \sum (x_i - \bar{x})^2 / (n - 1)$ and $\bar{x} = \sum x_i / n$ as usual. The summations are over the values of i from 1 to n. This ratio sample variance/sample mean was termed the *relative variance* by Clapham(1936). The index I affords a test of the hypothesis that the x_i are an independent random sample from a Poisson distribution with unspecified mean. If individuals (fish) are uniformly spaced, the variance will be much less than the mean, and the index of dispersion will be close to zero. If individuals (fish) are aggregated, the variance will be greater than the mean, and the index of dispersion will be much larger than 1. The statistic $(n - 1)$ I is often referred to a chi-square (χ^2-) distribution with $(n - 1)$ degrees of freedom (Hoel, 1943; Kathirgamatamby, 1953).

David and Moore(1954) introduced their index of clumping (or *contagiousness*)

$$ICS = I - 1 \qquad [2]$$

where I is given by [1]. For a Poisson distribution, ICS has mean zero and an accepted interpretation of a positive value for ICS is as the number of other individuals associated with a randomly chosen individual. If the individuals (fish) are clustered, the index will be large, whereas if the individuals are regularly spaced the index will be negative. The sampling distribution of ICS is unknown. The notation ICS and ICF = $\bar{x}$ / ICS was introduced by Douglas(1975). ICS is his *index of cluster* *size*, discussed earlier under the index of David and Moore(1954). The *index of cluster frequency* ICF should measure the mean number of clusters per quadrat.

Lloyd(1967) defined an *index of mean crowding*

$$IC = \bar{x} + ICS \qquad [3]$$

to represent the average number of other individuals that are contained in the quadrat that contains a randomly chosen individual and an *index of patchiness*

$$IP = 1 + (1 / ICF) \qquad [4]$$

as the ratio of IC / $\bar{x}$. Values of IP < 1 indicate a regular distribution, values of IP equal to 1 a random distribution and values of IP > 1 an aggregated distribution.

Once over-dispersion or aggregation has been identified, there is a need to account for it, possibly by use of either a probabilistic distribution or model. Representing the over-dispersion in catch data by a specific model leads naturally to the theory of mixtures of Poisson distributions. The development of the theory of such mixtures has its roots in the early seminal works of Greenwood and Yule(1920) and Lundberg(1940) whose authors introduced the *idea of conditionality* on a realisation of a variable parameter. This idea of conditionality is of fundamental importance and leads to the development of models based on the identity:

$$P_n(t) = \int_0^\infty \{ e^{-\alpha t} (\alpha t)^n / n! \} \, u(\alpha) \, d\alpha \qquad [5]$$

n such processes the parameter α appears as a random variable with a density u(.), and the probability of exactly n events during time t is denoted $P_n(t)$. Typically, such processes are considered for a standard period of time and αt replaced by a single parameter. Suppose, for example, that the discrete variable Y representing collected catch data might be Poisson with mean Z. Furthermore, this mean is itself a random variable which might be described by the gamma distribution with mean μ and index $k\mu$; i.e. $E\{Z\} = \mu$ and $var\{Z\} = \mu / k$. Then this mixture process leads to the negative binomial distribution (Plackett, 1981) with mean, $E\{Y\} = \mu$, and variance, $var\{Y\} = \mu (1 + \mu) / k$.

For a negative binomial distribution, ICF gives the method of moments estimate of *k* (Anscombe, 1950), a parameter inversely related to the clumping of the population. Myers(1978) showed that indices based on *k* were strongly correlated with population density and consequently, the use of the index of cluster frequency, ICF, is generally recommended only once the suitability of a negative binomial distribution has been assessed through the calculation of a suitable goodness-of-fit test statistic such as the U-statistic.

U-statistic goodness-of-fit test

This test uses the observed and expected variances of the negative binomial distribution (Evans, 1953) and is defined by

U = sample variance - expected variance under a negative binomial distribution

$$= s^2 - \{ \bar{x} + \bar{x}^2 / k \} \qquad [6]$$

where the symbol s^2 denotes the sample variance and the symbol $\bar{x}$ the sample mean, both defined in the usual way. The standard error of U is defined as follows:

$$\text{s.e.}(U) = \sqrt{\frac{1}{n} \left\{ 2\bar{x}(\bar{x}+g_1)g_2 \frac{[g_2 g_2 \ln g_2 - g_1(1+2g_1)]}{[g_2 \ln g_2 - g_1]} + g_3 \right\} } \qquad [7]$$

where

$$g_1 = \bar{x} / k \qquad [8]$$

$$g_2 = 1 + g_1 \qquad [9]$$

$$g_3 = g_2\, g_1^4\, [g_2^{(1+k)} - (\bar{x} + g_2)] / [g_2 \ln g_2 - g_1]^2 \qquad [10]$$

and ln denotes the natural logarithm to the base e. The expected value of U is zero so one approach to test if the observed value of U is significantly different from zero is to compare the calculated value with 2 standard errors of U. If a calculated value of U exceeds 2 s.e.(U), then reject the null hypothesis that the negative binomial distribution is a suitable model for the observed data at the 5% level of significance. Once the suitability of the negative binomial distribution has been ascertained, the distribution and its estimated parameters may be used to simulate catch distributions.

Spatial abundance distribution and relationship with covariates

Generalized linear models, GLMs (McCullagh and Nelder, 1989), and generalized additive models, GAMs (Hastie and Tibshirani, 1990), will be used to test and quantify the relationships between the abundance distribution (catch rates) and environmental variables. Modelling will be restricted to the GSBTS (Box H) since the coarse scale surveys have limited information on environmental variables. Analysis of deviance will be used to test the significance of candidate covariates, determine the way these covariates should be incorporated in the model, the appropriate link function and error structure. GAMs will be used to explore any non-linearity in the relationship of catch rates with continuous variables by introducing them as non-parametric *smooths* in the model.

RESULTS

The analyses presented and discussed in this Section refer only to the catch data collected during the 3rd quarter of 1995 in the central North Sea on the GSBTS (Box H), and the EGFS and the IBTS both in the vicinity of Box H.

Catch statistics

The GSBTS stations were taken during the three days 25-27 July 1995, the EGFS stations were taken during the period 15 August to 4 September 1995 and the IBTS stations were taken during the period 6 August to 4 September 1995. Histograms of the cod catches are distinctive (Figure 3) but reassuringly those for the EGFS and IBTS are visually similar.

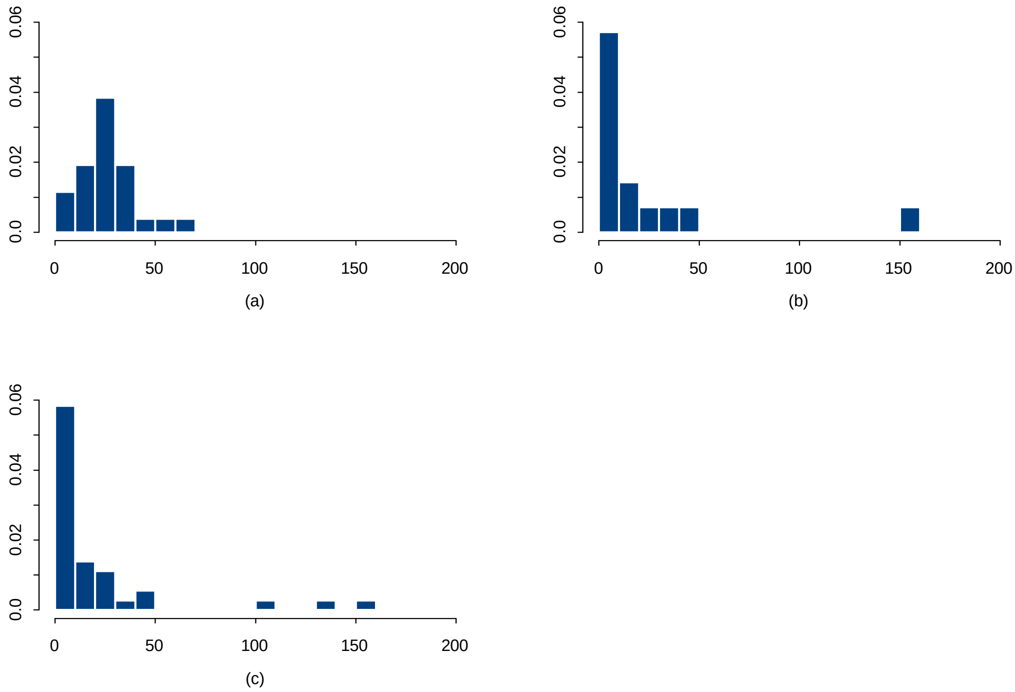

Figure 3. Histogram of the catch numbers of 2-year old cod in the central North Sea during the 3rd quarter of 1995: (a) GSBTS (Box H); (b) EGFS (stations near Box H); and (c) IBTS (stations near Box H).

The catches range from a minimum value of 7 fish per station to a maximum of 69 fish per station within Box H on the GSBTS and from a minimum of zero fish to a maximum of 153 fish on both the EGFS and the IBTS. The empirical catch distributions are similar for the EGFS and the IBTS but both are different from that of the GSBTS.

Spatial indices and over-dispersion

The catch data are clearly over-dispersed for the three surveys considered since the variance of each is greater than the corresponding mean (Table 1) and the indices of dispersion are all highly significant (95% points of the χ^2-distribution with 25, 13 and 35 degrees of freedom are 37.6, 22.4 and 49.8 respectively). Whilst the estimated mean catch rates are similar for the three surveys, the estimated variances are different. Reassuringly, once again, the estimated variances of the catch rates for the EGFS (stations near Box H) and the IBTS (stations near Box H) are of a similar magnitude.

Survey	*Mean* $\bar{x}$	*Variance* s^2	$(n - 1)$ I
GSBTS: Box H	26.83	210.77	196.43
EGFS: near Box H	22.16	1592.49	934.04
IBTS: near Box H	21.02	1286.14	2141.59

Table 1. Summary statistics of the catch numbers of 2-year old cod in the central North Sea during the 3rd quarter of 1995 for the three surveys: (a) GSBTS (Box H); (b) EGFS (stations near Box H); and (c) IBTS (stations near Box H). The statistic $(n - 1)$ I provides the homogeneity test statistic (probability, $p = 0$, for all values).

Calculated indices of clumping and aggregation are positive and indicative of the presence of extra-population heterogeneity within the catch rates on each of the three surveys (Table 2). Once again, the estimated indices for the EGFS (stations near Box H) and the IBTS (stations near Box H) are of a similar magnitude.

Survey	*Index of cluster size* ICS	*Index of cluster frequency* ICF	*Index of patchiness* IP
GSBTS: Box H	6.86	3.91	1.26
EGFS: near Box H	70.85	0.31	4.19
IBTS: near Box H	60.19	0.34	3.86

Table 2. Indices of clumping for the catch numbers of 2-year old cod in the central North Sea during the 3rd quarter of 1995 for the three surveys: (a) GSBTS (Box H); (b) EGFS (stations near Box H); and (c) IBTS (stations near Box H).

Applying the *U*-statistic goodness-of-fit test to each survey independently does not lead to the rejection of the null hypothesis that the negative binomial distribution fits the station count data for the 2-year old cod caught in the central North Sea (Table 3). The distributions are aggregated but a negative binomial distribution is a good fit to each survey's catch data.

Survey	*Sample variance* s^2	*Expected variance under a negative binomial distribution*	*U*
GSBTS: Box H	210.77	206.74	4.03
EGFS: near Box H	1592.50	1467.06	125.44
IBTS: near Box H	1286.14	1446.24	-160.10

Table 3. The *U*-statistic for the catch numbers of 2-year old cod in the central North Sea during the 3rd quarter of 1995 for the three surveys: (a) GSBTS (Box H); (b) EGFS (stations near Box H); and (c) IBTS (stations near Box H).

Maximum likelihood estimation of the two parameters of a negative binomial distribution for each of the three surveys independently yields estimates of 4 and 0.13, respectively, for the GSBTS (Box H) and estimates of 1 and 0.04, respectively, for both the EGFS (stations near Box H) and the IBTS (stations near Box H). The moment estimator of k; namely, ICF (Table 2), agrees with the maximum likelihood estimator since the parameter k of a negative binomial distribution is integer in value and hence must be rounded.

Simulated catch distributions

To illustrate the suitability of the negative binomial distribution and to indicate the inherent level of variability, simulated catch rates were generated using the appropriate negative binomial distribution for each of the three surveys (Figures 4-6). Each simulated realization is constrained to generate the same number of catch rates as stations encountered on each of the three surveys considered. The similarity between Figures 5 and 6 is indicative of the fact that the appearance of the simulated catch distributions is little affected by sample size.

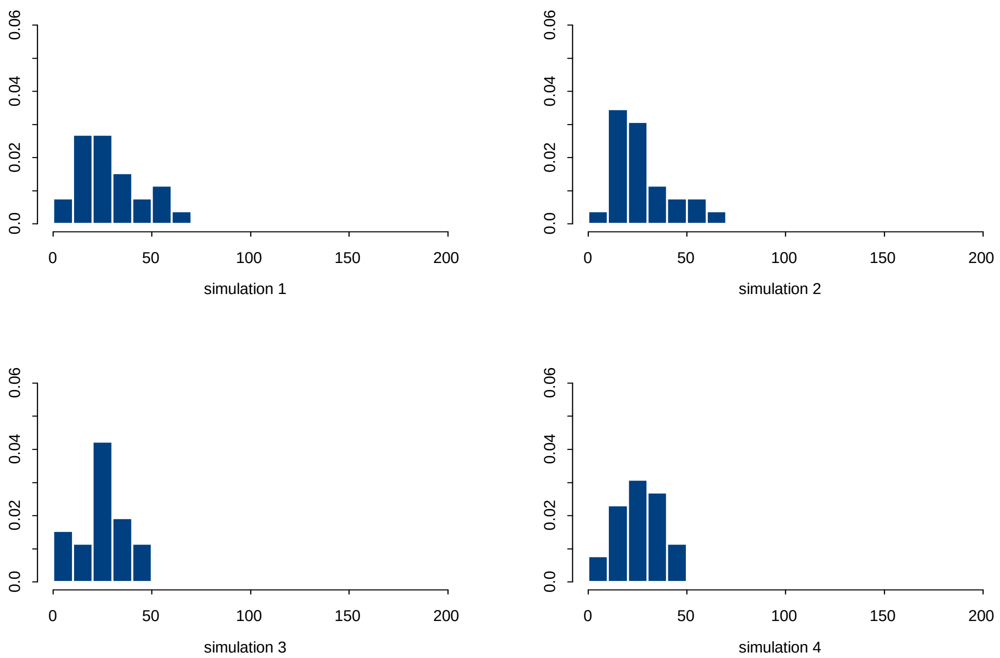

Figure 4. Four simulated realizations of a negative binomial distribution with parameters 4 and 0.13. Each simulation is constrained to generate the same number of values (namely, 26) as stations encountered by the GSBTS (Box H).

The differences in the parameter estimates obtained for the GSBTS (Box H) and the EGFS/IBTS in the vicinity of Box H necessitates a comment but will be discussed later.

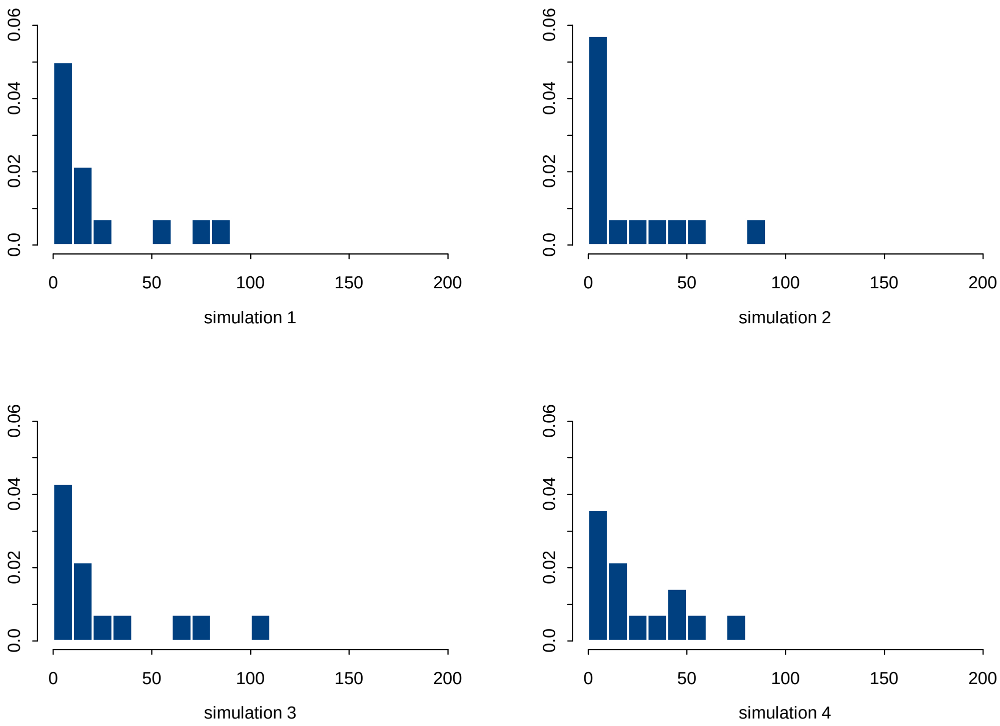

Figure 5. Four simulated realizations of a negative binomial distribution with parameters 1 and 0.04. Each simulation is constrained to generate the same number of values (namely, 14) as stations encountered by the EGFS in the vicinity of Box H.

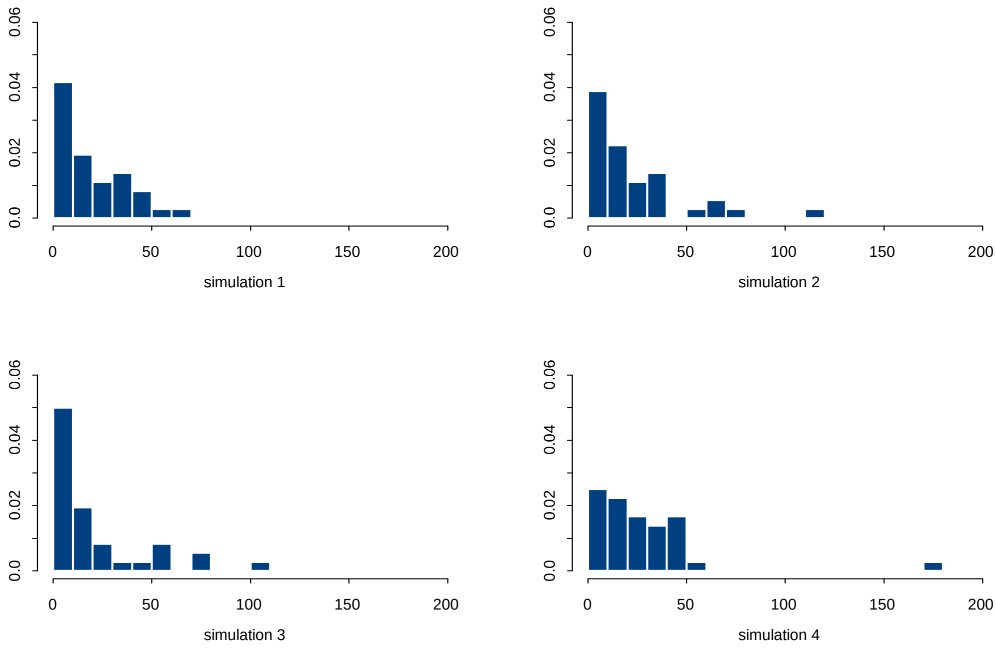

Figure 6. Four simulated realizations of a negative binomial distribution with parameters 1 and 0.04. Each simulation is constrained to generate the same number of values (namely, 36) as stations encountered by the IBTS in the vicinity of Box H.

Spatial abundance distribution and relationship with covariates

GSBTS: The incorporation of temporal information into the modelling of the spatial abundance distribution within Box H was investigated for the three-day period sampled during the 3rd quarter of 1995. A GAM model was fitted with the covariates of time of sample and windspeed during sampling. Windspeed entered the model linearly but time required a non-parametric *smooth* and resulted in a residual deviance of 127.79 (20.18 degrees of freedom); in contrast to the null model with a deviance of 180.83 (25 degrees of freedom). Interestingly, once the covariates had been entered into the model correctly then the assumption of a Poisson distribution for the distribution of catch rates seemed reasonable.

DISCUSSION

Catch rates of 2-year old cod in the central North Sea have been analysed from third quarter research vessel surveys undertaken in 1995. It has been shown that the type of the frequency distribution is *inter alia* dependent upon the extent of the area in which the hauls took place. The idea of conditionality has been applied and a flexible mixture distribution – the negative binomial - proposed for the modelling of catch data. The distribution has a number of interesting properties and represents one type of a general process of aggregation.

A common form for the catch distribution of the small- and large-scale surveys has been adopted but estimates of the parameters of that distribution indicate that the mixing processes are different in the two cases. However, calculated indices of patchiness, IP, for the EGFS (stations near Box H) and the IBTS (stations near Box H) are of a similar magnitude, even though there are more than double the number of hauls for the IBTS in the vicinity of Box H. With hindsight, this result is only to be expected since the two surveys are both conducted at the same coarse spatial resolution.

It is important to identify the spatial scale over which sampling is contemplated and for which subsequent model-based inferences are to be made. Differences in the parameter estimates obtained for the mixture distribution fitted to the GSBTS (Box H) and the EGFS/IBTS in the vicinity of Box H were found. Figure 7 shows the spatial extent of the coverage by stations within each of the three surveys and the delimited area defined between longitude 1° and 4°, and between latitude 55° and 58° (Figure 2). Clearly, the GSBTS (Box H) may be viewed as a local survey consisting of repeatedly trawling within the same region; whilst the EGFS and IBTS may be viewed as both trawling within different regions. In principle, each EGFS/IBTS station may be considered as a single observation taken from a heterogeneous process of the type sampled within the GSBTS (Box H). From such a viewpoint, it is hardly surprising that a common distribution of catch rate seems appropriate for different surveys. The difference in the estimated values of parameters might be due to the estimation being related to different underlying processes. For instance, the GSBTS (Box H) might be viewed as providing information on *within* IBTS station variability; whilst the EGFS/IBTS might be viewed as providing information on *between* station variability. Alternatively, a spatially structured population can be divided into a series of subsets which may represent population clusters or arbitrary sampling areas (c.f. Thomas and Kunin, 1999). Much research on spatially structured populations has produced both theoretical and empirical evidence for a range of possible types of populations (Hanski and Gilpin, 1997). However, identification of the underlying process(es) remains a problem. Further fine-scale surveys would assist in the understanding of the processes of clumping, aggregation and contagion that are apparent from the catch rates of 2-year old cod analysed. Such research work immediately impact upon the optimal design of surveys, the calculation of abundance indices and associated confidence intervals for species at age.

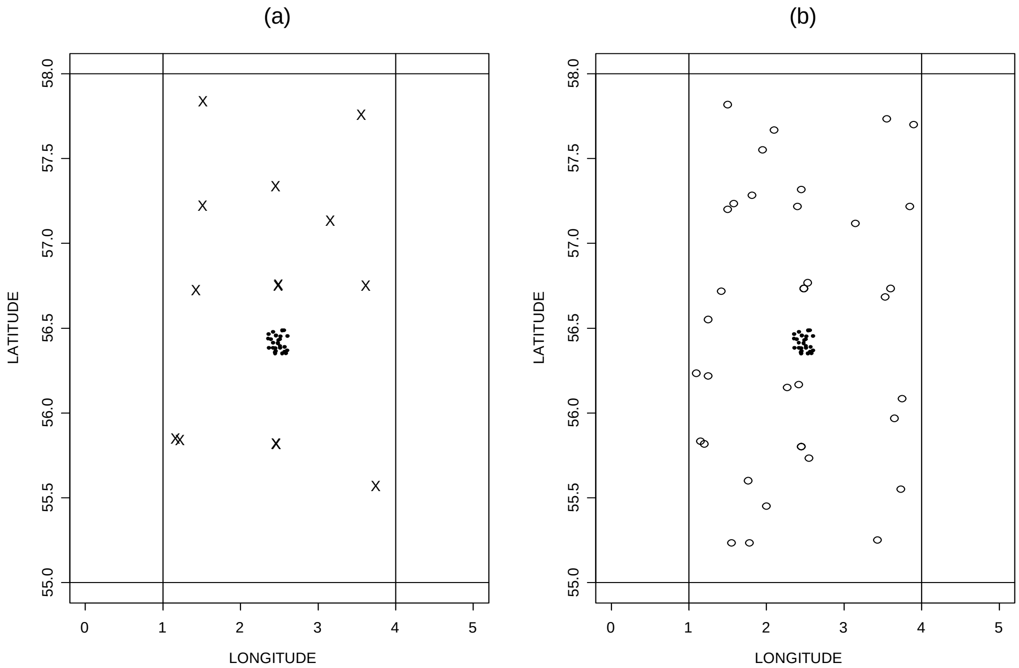

Figure 7. Location of the survey stations during the third quarter of 1995 in close proximity to the sampled Box H: (a) EGFS trawls (shown by the crosses), and (b) IBTS trawls (shown by the circles). The location of each of the survey stations on the GSBTS (Box H) is indicated by a dot in each map.

The application of generalized linear models (McCullagh and Nelder, 1989) to the modelling of catch data follows naturally (O'Brien *et al.*, 1998). Additional comparisons of the catch distributions from the GSBTS, EGFS and IBTS and modelling the effect of environmental covariates in other years and areas would enable an appraisal of the consistency of the spatial distribution of cod to be investigated; possibly incorporating the ideas of simplicity and persistence (Houghton, 1987).

The need to decide whether catch data follow a log-normal, Poisson, gamma or some other distribution can be replaced by a general assumption about mixing processes and so, alleviate the need to have different estimation methods and procedures for each separate distribution assumption. To further substantiate this statement, if one were to relax the constraint that catch data follow discrete probability distributions then the negative binomial distributions identified in this paper for the catch data of the GSBTS, EGFS and IBTS might equally be replaced by:

- either a log-normal(3.13, 0.54) or a Gamma(3.69, 7.13) for the GSBTS (Box H),
- a log-normal(1.17, 2.89) for the EGFS (stations near Box H), and
- a log-normal(1.37, 2.63) for the IBTS (stations near Box H).

The introduction of such varied distributions is, however, not necessary!

Finally, the following general conclusions can be stated.

(i) The estimated mean catch rates of the three surveys investigated are similar but the estimated variances are different for the fine-scale and coarse scale surveys.
(ii) The type and appearance of the catch frequency distribution depend amongst others upon the extent of the area in which the hauls took place.
(iii) The negative binomial distribution is suitable to describe the catch distribution at the different spatial scales considered.
(iv) Fine-scale surveys like the GSBTS provide information on *within* IBTS station variability; whilst the coarse scale EGFS/IBTS provide information on *between* station variability.
(v) Further fine-scale surveys would be helpful to understand the processes of clumping, aggregation and contagion; improving the design of surveys and the calculation of abundance indices and associated confidence intervals.

ACKNOWLEDGEMENTS

This paper was prepared with funding support provided by the Ministry of Agriculture, Fisheries and Food (contract MF0316) to the first author (CMO'B); together with funding support provided to all three authors by the Commission of the European Communities Directorate General for Fisheries (DG XIV) under project contract 98/029 (Survey-based abundance indices that account for fine spatial scale information for North Sea stocks - FINE).

REFERENCES

Anscombe, F.J. (1950). Sampling theory of the negative binomial and logarithmic series distributions. *Biometrika*, **37**, 358-382.

Clapham, A.R. (1936). Over-dispersion in grassland communities and the use of statistical methods in plant ecology. *Journal of Ecology*, **24**, 232-251.

David, F.N. and Moore, P.G. (1954). Notes on contagious distributions in plant populations. *Annals of Botany*, **18**, 47-53.

Douglas, J.B. (1975). Clustering and aggregation. *Sankhya*, **37B**, 398-417.

Ehrich, S., Adlerstein, S., Götz, S., Mergardt, N. and Temming, A. (1998). Variation of meso scale fish distribution in the North Sea. *ICES C.M. 1998/J:25*.

Evans, D.A. (1953). Experimental evidence concerning contagious distributions in ecology. *Biometrika*, **40**, 186-211.

Fisher, R.A., Thornton, H.G. and Mackenzie, W.A. (1922). The accuracy of the plating method of estimating the density of bacterial populations, with particular reference to the use of Thornton's agar medium with soil samples. *Annals of Applied Biology*, **9**, 325-359.

Greenwood, M. and Yule, G.U. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal Statistical Society, Series A*, **83**, 255-279.

Gröger, J. and Ehrich, S. (1992). The importance of catch frequency distributions for the interpretation of catch data and the fit by the very adaptable and realistic beta distribution. *ICES C.M. 1992/D:18*.

Hanski, I.A. and Gilpin, M.E. (1997). *Metapopulation Biology: Ecology, Genetics, and Evolution*. San Diego:Academic Press.

Harding, D., Woolner, L. and Dann, J. (1986). The English groundfish surveys in the North Sea, 1977-85. *ICES C.M. 1986/G:13*.

Hastie, T.J. and Tibshirani, R.J. (1990). *Generalized Additive Models*. New York:Chapman & Hall.

Hoel, P.G. (1943). On indices of dispersion. *Annals of Mathematical Statistics*, **14**, 155-162.

Houghton, R.G. (1987). The consistency of the spatial distributions of young gadoids with time. *ICES C.M. 1987/G:15*.

ICES (1996). Manual for the International Bottom Trawl Survey. Revision V. Addendum to *ICES C.M. 1996/H:1*.

Kathirgamatamby, N. (1953). Note on the Poisson index of dispersion. *Biometrika*, **40**, 225-228.

Lloyd, M. (1967). Mean crowding. *Journal of Animal Ecology*, **36**, 1-30.

Lundberg, O. (1940). *On Random Processes and their Application to Sickness and Accident Statistics*. Uppsala:Almqvist and Wicksells.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*. London:Chapman & Hall.

Myers, J.H. (1978). Selecting a measure of dispersion. *Environmental Entomology*, **7**, 619-621.

O'Brien, C.M., Kell, L.T., Santiago, J. and V. Ortiz de Zarate, V.(1998). The use of generalized linear models for the modelling of catch-effort series. II. Application to north Atlantic albacore surface fishery. In ICCAT Collective Volume of Scientific Papers, *Vol. XLVIII(1), SCRS/97/49*, 170-183.

Pennington, M.R. and Grosslein, M.D. (1978). Accuracy of abundance indices based on stratified random trawl surveys. *ICES C.M. 1978/D:13*.

Plackett, R.L. (1981). *The Analysis of Categorical Data*. London:Griffin.

Taylor, C.C. (1953). Nature of variability in trawl catches. *U.S. Fish Wildlife Service, Fishery Bulletin*, **54**, 145-166.

Thomas, C.D. and Kunin, W.E. (1999). The spatial structure of populations. *Journal of Animal Ecology*, **68**, 647-657.