Using mixed models to combine smooth size-selection and catch-comparison curves over hauls

R.J. Fryer, A.F. Zuur, and N. Graham

Abstract: Parametric size-selection curves are often combined over hauls to estimate a mean selection curve using a mixed model in which between-haul variation in selection is treated as a random effect. This paper shows how the mixed model can be extended to estimate a mean selection curve when smooth nonparametric size-selection curves are used. The method also estimates the between-haul variation in selection at each length and can model fixed effects in the form of the different levels of a categorical variable. Data obtained to estimate the size-selection of dab by a Nordmøre grid are used for illustration. The method can also be used to provide a length-based analysis of catch-comparison data, either to compare a test net with a standard net or to calibrate two research survey vessels. Haddock data from an intercalibration exercise are used for illustration.

Résumé : Les courbes paramétriques de sélection de la taille sont souvent combinées dans toutes les récoltes pour obtenir une courbe de sélection moyenne qui utilise un modèle mixte dans lequel la variation de la sélection entre les récoltes est considérée comme un effet aléatoire. Notre étude démontre comment le modèle mixte peut être étendu de façon à estimer une courbe de sélection moyenne en utilisant des courbes lisses de sélection non paramétriques. La méthode estime aussi la variation de la sélection entre les récoltes pour chacune des longueurs et elle peut modéliser des effets fixes sous forme de niveaux différents d'une variable catégorique. Un exemple utilise des données obtenues pour estimer la sélection de taille chez la limande par une grille de Nordmøre. La méthode peut aussi servir à analyser des données de comparaisons de récoltes d'après la longueur, soit pour comparer un filet expérimental avec un filet standard ou pour calibrer deux navires d'inventaire de recherche. Un exemple utilise des données sur l'aiglefin provenant d'un exercice d'intercalibration.

[Traduit par la Rédaction]

Introduction

It is widely accepted that the size selection of a fishing net can vary from haul to haul even though the net has not been changed (Fryer 1991). Such between-haul variation in selection might arise in response to changes in sea state, towing direction, fishing grounds, etc. Fryer (1991) showed that between-haul variation must be accounted for when modelling selection data to avoid making incorrect inferences about the selection of the net. Further, he described a model that incorporates between-haul variation when the selection of each haul can be described by a logistic curve. Essentially, the parameters of the logistic curve for each haul are assumed to vary stochastically about mean selection parameters that characterise the average selection of the net. The model is an example of a mixed model (e.g., Hocking 1996) in which hauls are treated as random effects. Fixed effects such as changes in mesh size can also be modelled. The mixed-model methodology is easily extended to other parametric selection curves (Millar and Fryer 1999) and has been

R.J. Fryer[1] and A.F. Zuur. Fisheries Research Services, Marine Laboratory, P.O. Box 101, 375 Victoria Road, Torry, Aberdeen, AB11 9DB, U.K.
N. Graham. Institute of Marine Research, Fish Capture Division, P.O. Box 1870, Nordnes, N-5817 Bergen, Norway.

[1]Corresponding author (e-mail: R.Fryer@marlab.ac.uk).

widely applied (e.g., Reeves et al. 1992; Madsen et al. 1999; Zuur, G., et al. 2001).

Bootstrapping methods (Efron and Tibshirani 1993) have been used to incorporate between-haul variation in studies using nonparametric selection curves. Nonparametric curves are typically used when parametric curves are not sufficiently flexible to describe the selection of each haul. For example, Millar (1993) modelled the selection of a scallop dredge using four parametric curves, but found the fit of all four curves to be inadequate. He then switched to isotonic regression curves. These curves are a series of flat lines with jumps between them, the only constraint being that selection must either stay constant or increase as fish size increases. A mean selection curve was estimated by pooling the data across hauls and fitting a single isotonic regression curve. Confidence intervals for summary measures such as the 50% retention length were then obtained by bootstrapping. Between-haul variation was incorporated by bootstrapping at the haul level, i.e., by taking a random sample of hauls (with replacement). A second stage of bootstrapping was required to incorporate within-haul variation, i.e., variation at the fish level. Munro and Somerton (2001) used a different type of nonparametric curve, the smoothing spline, to model the footrope selectivity of four species from various trials. Again, the data were pooled over hauls and a single smoothing spline was fitted. Pointwise 95% confidence bands around the fitted curves were then constructed by bootstrapping over hauls.

The smoothing spline used by Munro and Somerton (2001) is an example of a linear smoother (Hastie and Tibshirani 1990). Other common examples include running-mean and

kernel smoothers (Hastie and Tibshirani 1990) and the loess smoother of Cleveland (1979). Linear smoothers occupy the middle ground between parametric curves and isotonic regression curves. They are nonparametric in the sense that their shape is determined, to a large extent, by the data themselves, although the user controls the amount of smoothing. On the other hand, in common with parametric curves, they are amenable to formal statistical inference. For example, the fit of a smoother can be compared with that of a linear model (a special case of the smoother), and the precision of a smoother can be presented as, e.g., pointwise 95% confidence bands. In particular, smoothers provide exactly the same summary information that is used by the mixed-model methodology to combine parametric selection curves over hauls. Our first objective in this paper is to show how we can generalise the mixed-model methodology to exploit this information and thus combine smooth selection curves over hauls. The method estimates the mean selection curve of the net, with pointwise standard errors or confidence bands, and the between-haul variation in selection. Fixed effects, in the form of the levels of a categorical variable, can also be modelled. The methodology is developed in the following section and then illustrated using data obtained to estimate the selection of dab (*Limanda limanda*) by a Nordmøre selection grid (grate).

Our second objective is to show how the same techniques can be used to model catch-comparison data, collected either to compare a test net with a standard net (Briggs 1992; Armstrong et al. 1998) or to calibrate two research survey vessels (e.g., Ehrich 1991; Pelletier 1998). Most analyses of such data have compared catches that have been aggregated over all lengths or over a wide range of length classes (e.g., all lengths above the minimum landing size); see the review of Pelletier (1998). However, a length-based analysis could provide more insight into the relative performance of the two nets or vessels. For example, Warren (1997) modelled catch ratios from two research vessels as a parametric function of length. Modelling catch ratios as a function of length is a means of estimating the relative size selection of the two nets or vessels. When the data have been collected under a parallel or alternate haul design, smoothers provide a natural way of modelling the catch ratios for each paired haul and can be used when a suitable parametric curve can not be found. We show how the mixed-model methodology can be used to combine these smooth curves over paired hauls and illustrate the methodology using data for haddock (*Melanogrammus aeglefinus*) collected in an intercalibration exercise.

Size-selection curves

Parametric curves

We begin by reviewing the mixed-model methodology for combining parametric size-selection curves. We develop the theory in the context of a covered cod-end experiment (Wileman et al. 1996) in which H replicate hauls are made with the same test cod-end and we wish to estimate the mean selection of the cod-end. Generalising to other types of selectivity experiments, such as trouser-trawl or twin-trawl experiments, is then straightforward. For simplicity, we leave the theory for fixed effects — controlled changes such as a change in mesh

size — until later. Full methodological details can be found in Fryer (1991) and Millar and Fryer (1999).

Suppose that, in haul h of a covered cod-end experiment, we measure n_{1hl} fish of length l in the cod-end and n_{2hl} fish in the cover. Often catches are subsampled, so let q_{1hl} and q_{2hl} be the subsampling fractions of length l fish in the cod-end and cover, respectively. Assuming that fish escape or are retained by the cod-end independently of each other, then conditional on the total numbers of fish measured, the numbers of fish measured in the cod-end can be modelled as having binomial distributions:

(1) $n_{1hl}|n_{1hl} + n_{2hl} \sim \text{Binomial}(n_{1hl} + n_{2hl}, \phi_{hl})$

where

(2) $\phi_{hl} = \dfrac{r_h(l)q_{1hl}}{r_h(l)q_{1hl} + (1 - r_h(l))q_{2hl}}$

and $r_h(l)$ is the selection (retention) curve of the cod-end, i.e., the probability that a fish of length l is retained by the cod-end given that it entered the cod-end. Note that ϕ_{hl} is the probability that a fish is measured in the cod-end, given that it is measured in either the cod-end or the cover. When there is no subsampling, or equal subsampling in the cod-end and cover, ϕ_{hl} simplifies to $r_h(l)$.

Now suppose that $r_h(l)$ is an m-parameter selection curve with parameters $\mathbf{v}_h = (v_{h1},...,v_{hm})^{\mathrm{T}}$, where the T superscript denotes the transpose of a vector. For example, if $r_h(l)$ is the logistic selection curve

(3) $r_h(l) = \dfrac{\exp(a_h + b_h l)}{1 + \exp(a_h + b_h l)}$

we have $m = 2$ and $\mathbf{v}_h = (a_h, b_h)^{\mathrm{T}}$. Between-haul variation is incorporated by assuming that the selection parameters $\mathbf{v}_h$ vary randomly from haul to haul about a mean selection curve according to some probability distribution. Specifically, the $\mathbf{v}_h$ are assumed to be (multivariate) normally distributed with mean $\boldsymbol{\theta} = (\theta_1,...,\theta_m)^{\mathrm{T}}$, an m-vector of parameters that characterise the mean selection of the cod-end, and variance $\mathbf{D}$, an $m \times m$ matrix that measures the between-haul variation in selection, i.e.,

(4) $\mathbf{v}_h \sim \mathrm{N}(\boldsymbol{\theta}, \mathbf{D})$

Variation in selection between hauls is thus treated as a random effect.

In the special case of the logistic selection curve, the mean selection parameters $\boldsymbol{\theta}$ and the between-haul variance matrix $\mathbf{D}$ can be estimated in a single iterative procedure as a generalised linear mixed model. (Strictly, this is a random effects model, because there are no fixed effects at this stage.) More generally $\boldsymbol{\theta}$ and $\mathbf{D}$ can be estimated in two stages. First, the selection parameters $\mathbf{v}_h$ are estimated separately for each haul by maximum likelihood. Let $\hat{\mathbf{v}}_h$ be the maximum likelihood estimator of $\mathbf{v}_h$ and $\mathbf{R}_h$ be the variance of $\hat{\mathbf{v}}_h$ (conditional on $\mathbf{v}_h$). Thus, $\mathbf{R}_h$ measures the within-haul binomial sampling variability from haul h. Assuming that there are reasonable numbers of fish in haul h, $\hat{\mathbf{v}}_h$ is approximately unbiased and normally distributed:

(5) $\hat{\mathbf{v}}_h|\mathbf{v}_h \sim \mathrm{N}(\mathbf{v}_h, \mathbf{R}_h)$

450

Can. J. Fish. Aquat. Sci. Vol. 60, 2003

Combining eqs. 4 and 5 gives

$$(6) \qquad \hat{\mathbf{v}}_h \sim \mathrm{N}(\boldsymbol{\theta}, \mathbf{D} + \mathbf{R}_h)$$

from which $\boldsymbol{\theta}$, its standard error, and $\mathbf{D}$ are estimated by residual maximum likelihood.

Although this methodology was developed for a covered cod-end experiment, there is nothing that links it specifically to this type of experiment. It can thus be generalised to any selectivity experiment in which selection during haul (or deployment) h can be characterised by an m-parameter curve with parameters $\mathbf{v}_h = (v_{h1},\ldots,v_{hm})^{\mathrm{T}}$ and estimates $\hat{\mathbf{v}}_h$ with variance matrix $\mathbf{R}_h$ can be obtained. For example, Zuur, G., et al. (2001) describe an experiment to estimate the size selection of a cod-end and a square-mesh panel using covers over both the cod-end and the panel. Selection on each haul was modelled using five parameters that were then combined over hauls.

Smooth curves

We now return to the covered cod-end experiment and show how the theory can be extended to incorporate smooth functions of length. Taking the logistic transformation of eq. 2 gives

$$(7) \qquad \mathrm{logit}(\phi_{hl}) = \mathrm{logit}(r_h(l)) + \mathrm{log}(q_{1hl}/q_{2hl})$$

When $r_h(l)$ is the logistic curve (eq. 3), this becomes

$$(8) \qquad \mathrm{logit}(\phi_{hl}) = a_h + b_h l + \mathrm{log}(q_{1hl}/q_{2hl})$$

and eqs. 1 and 8 form a generalised linear model with binomial errors, a logistic link, and offset $\mathrm{log}(q_{1hl}/q_{2hl})$ (McCullagh and Nelder 1989). More generally, we might assume that $r_h(l)$ is an arbitrary smooth function of length $s_h(l)$. Equation 7 now becomes

$$(9) \qquad \mathrm{logit}(\phi_{hl}) = s_h(l) + \mathrm{log}(q_{1hl}/q_{2hl})$$

and eqs. 1 and 9 form a generalised additive model, again with binomial errors, a logistic link, and offset $\mathrm{log}(q_{1hl}/q_{2hl})$ (Hastie and Tibshirani 1990). Methods for estimating $s_h(l)$ and the precision of these estimates are described by Hastie and Tibshirani (1990) and are implemented in several statistical packages such as S-PLUS (Insightful Corp., Seattle, Wash.) and GenStat (GenStat VSN International Ltd., Herts, U.K.). We discuss computational and smoothing issues later. For now it is sufficient to recognise that selection during each haul can be modelled as a smooth function of length, thus capturing a wide variety of selection patterns that cannot be described by a single parametric curve. The next stage is to combine these smooth functions over hauls to estimate mean selection in a way that incorporates between-haul variation. The random effects model (eq. 6) suggests a natural way of doing this.

Suppose that the sampled length classes run from 1 to L and let $\mathbf{s}_h$ be the L vector of selection probabilities (on the logistic scale) $\mathbf{s}_h = (s_h(1),\ldots,s_h(L))^{\mathrm{T}}$. Now let $\hat{\mathbf{s}}_h$ be the estimate of $\mathbf{s}_h$ obtained by fitting a smoother to the data for haul h and let $\mathbf{R}_h = \mathrm{var}(\hat{\mathbf{s}}_h)$. Assuming an adequate amount of smoothing, we have approximately

$$(10) \qquad \hat{\mathbf{s}}_h | \mathbf{s}_h \sim \mathrm{N}(\mathbf{s}_h, \mathbf{R}_h)$$

Between-haul variation can now be incorporated by assuming that

$$(11) \qquad \mathbf{s}_h \sim \mathrm{N}(\boldsymbol{\theta}, \mathbf{D})$$

where $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_L)^{\mathrm{T}}$ is an L vector of parameters representing mean selection at each length and $\mathbf{D}$ is an $L \times L$ matrix measuring between-haul variation. Combining eqs. 10 and 11 then gives

$$(12) \qquad \hat{\mathbf{s}}_h \sim \mathrm{N}(\boldsymbol{\theta}, \mathbf{D} + \mathbf{R}_h)$$

the same random effects model as before.

In principle we could estimate $\boldsymbol{\theta}$ and $\mathbf{D}$ by residual maximum likelihood using the algorithms for combining parametric selection curves (e.g., Fryer 1991). However, we have found numerical difficulties with this approach because the matrices $\mathbf{D}$ and $\mathbf{R}_h$ can become quite large and the elements of $\hat{\mathbf{s}}_h$ can be highly correlated (particularly the elements corresponding to adjacent length classes) so that $\mathbf{R}_h$ can be "close to singular". Our solution, which works well in practice, is to consider each length class separately in a univariate analysis. Thus, for length class l, we have

$$(13) \qquad \hat{s}_h(l) \sim \mathrm{N}(\theta_l, D_{ll} + R_{hll})$$

where D_{ll} and R_{hll} are the lth diagonal elements of $\mathbf{D}$ and $\mathbf{R}_h$, respectively. From this model, we estimate θ_l, its standard error, and D_{ll} by residual maximum likelihood using the algorithms for combining parametric selection curves (see Appendix A). This is then repeated for each length class in turn. Although the estimation procedure is conducted one length class at a time, it still leads to an unbiased estimator of the mean selection curve and gives appropriate pointwise standard errors (assuming that all of the underlying assumptions are met). However, there will clearly be some loss of efficiency compared with an estimation procedure that uses the information from all length classes simultaneously.

Again, the random effects model (eq. 12) is not specific to the covered cod-end experiment, and the theory can therefore be generalised to any experiment in which selection during haul h can be characterised by a smooth curve $s_h(l)$ and $\hat{\mathbf{s}}_h$ with variance matrix $\mathbf{R}_h$ can be estimated.

Fixed effects

Fixed effects are usually incorporated by allowing the mean selection parameters to vary between hauls according to some design matrix, just as fixed effects are incorporated in linear regression (e.g., Draper and Smith 1981). Fryer (1991) describes the approach for selectivity data when parametric selection curves can be used. However, unless the design matrix has a particularly simple structure, the mean selection parameters must be estimated simultaneously, rather than one length class at a time. For smooth selection curves, we therefore restrict attention to a particular class of fixed effects where the design matrix does allow straightforward estimation.

Specifically, we consider the situation in which there are G groups of hauls and we wish to estimate and compare the mean selection curve for each group. Often the G groups will correspond to G different cod-ends. Let $s_{gh}(l)$ be the selection curve (on the logistic scale) for haul h in group g, $h = 1,\ldots,H_g$, $g = 1,\ldots,G$. Further, let $\mathbf{s}_{gh} = (s_{gh}(1),\ldots,s_{gh}(L))^{\mathrm{T}}$ be

the vector of selection probabilities for length classes 1 to L, let $\hat{\mathbf{s}}_{gh}$ be the smooth estimate of $\mathbf{s}_{gh}$, and let $\mathbf{R}_{gh} = \mathrm{var}(\hat{\mathbf{s}}_{gh})$. Assuming that the between-haul variance $\mathbf{D}$ is constant over groups, we have

$$\hat{\mathbf{s}}_{gh} \sim \mathrm{N}(\boldsymbol{\theta}_g, \mathbf{D} + \mathbf{R}_{gh})$$

where $\boldsymbol{\theta}_g = (\theta_{g1},\dots,\theta_{gL})^T$ is the mean selection curve for group g. To estimate the mean curves, we again consider each length class separately in a univariate analysis. Thus, for length class l, we have

(14) $\hat{\mathbf{s}}_{gh}(l) \sim \mathrm{N}(\theta_{gl}, D_{ll} + R_{ghll})$

from which we estimate θ_{gl}, $g = 1,\dots,G$, their standard errors, and D_{ll} by residual maximum likelihood (Appendix A). This process is then repeated for each length class in turn.

Large differences between the mean curves of each group are easily recognised by plotting the estimated mean curves with their pointwise 95% confidence bands. However, more subtle differences can only be established by formal hypothesis tests. Often we need to compare the hypotheses

H$_0$: $\boldsymbol{\theta}_g = \boldsymbol{\theta}$, $g = 1,\dots,G$

H$_1$: $\boldsymbol{\theta}_g$ unconstrained

i.e., there are no differences between the mean curves against some unspecified difference. The natural approach here would be to use a likelihood ratio test. However, this is not possible because, by estimating the elements of one length class at a time, we cannot construct the full likelihood of the data. It is also unclear how many degrees of freedom are associated with $\hat{\boldsymbol{\theta}}_g$. Instead, a bootstrap hypothesis test can be used (Efron and Tibshirani 1993). For this, we need to calculate a suitable test statistic that measures the difference between the estimated mean curves and then assess the significance of the test statistic by comparing it with a bootstrap reference distribution.

To motivate the test statistic, we again look to the univariate analysis for length class l and suppose that we are interested in comparing the hypotheses

H$_0$: $\theta_{gl} = \theta_l$, $g = 1,\dots,G$

H$_1$: θ_{gl} unconstrained

In the univariate case, there is no difficulty in using a likelihood ratio test. Here, the test statistic T_l is minus twice the difference in log-likelihood obtained by fitting model 14 under the two hypotheses, and H$_0$ is rejected if T_l is large relative to a χ^2 distribution on $G - 1$ degrees of freedom. By combining the test statistics T_l over all length classes, we can measure the difference between the two mean curves as a whole. Two candidate test statistics are

$$T_{\max} = \max_l T_l$$
$$T_{\mathrm{ave}} = (1/L)\sum_l T_l$$

The choice between them will depend on the departures from H$_0$ that might be expected a priori. $T_{\max}$ will tend to have greater power when the mean curves differ in only part of the length range, whereas T_{ave} will have greater power when there are differences at all lengths. In practice, both test statistics might be used.

The bootstrap reference distribution for a test statistic T (either $T_{\max}$ or T_{ave}) is constructed by generating $b = 1,\dots,B$ data sets that satisfy the null hypothesis and, for each, calculating the corresponding test statistic, denoted T^b. The values T^b, $b = 1,\dots,B$, are bootstrap realisations of the null distribution of T and are thought of as "typical" values of T if the null hypothesis is true. If the observed statistic is similar in magnitude to the bootstrap realisations, then the observed data are consistent with the null hypothesis. Conversely, if the observed statistic is "large" relative to the bootstrap realisations, then there is evidence against the null hypothesis. The achieved significance level (ASL) is the bootstrap equivalent of a "p value" and is simply the proportion of bootstrap realisations that exceed the observed statistic.

The details of the bootstrap procedure are as follows. First, we generate a data set that satisfies the null hypothesis by calculating

$$\hat{\mathbf{s}}^*_{gh} = \hat{\mathbf{s}}_{gh} - \hat{\boldsymbol{\theta}}_g + \hat{\boldsymbol{\theta}}$$

where $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_g$ are the estimates of $\boldsymbol{\theta}_g$ under the null and alternative hypotheses, respectively. The curves $\hat{\mathbf{s}}^*_{gh}$ have similar shape and variability to the original curves and satisfy the null hypothesis because they are centred on the common curve $\hat{\boldsymbol{\theta}}$. Each bootstrap realisation is obtained by taking H_g of the centred curves at random and with replacement from each group g and calculating the corresponding value of T.

Note that residual maximum likelihood is generally used to estimate and visualise the mean curves $\boldsymbol{\theta}_g$, because residual maximum likelihood gives unbiased estimates of the between-haul variance $\mathbf{D}$ and hence of the standard errors of $\hat{\boldsymbol{\theta}}_g$, whereas ordinary maximum likelihood does not. However, the residual likelihood can only be used to assess the significance of random effects, not fixed effects; therefore, ordinary maximum likelihood should be used in all stages of the bootstrap hypothesis test.

Computational and smoothing issues

We have used a loess smoother (Cleveland 1979) to implement the methods described above. For normally distributed data, loess smoothers are constructed by sequentially fitting a weighted linear regression to the data in a neighbourhood around each length class. As the width of the neighbourhood increases, so does the smoothness of the fitted line. Hastie and Tibshirani (1990) give a clear description of the fitting process and discuss some of the properties of loess smoothers. For binomial data, as here, the loess smoother is again constructed using a series of weighted linear regressions, but an iterative fitting procedure is now required (Hastie and Tibshirani 1990).

We chose loess smoothers because they deal naturally with two practical problems that arise when, as is usually the case, a different number of length classes is sampled on each haul. First, it is necessary to ensure that the amount of smoothing is compatible with the number of sampled length classes. Within loess, this can be done by fixing the number of nearest neighbours (i.e., the number of length classes in the neighbourhood that contribute to each weighted linear regression) to be the same for all hauls. As a result, the degrees of freedom of the smoother increases with the number

452

Can. J. Fish. Aquat. Sci. Vol. 60, 2003

of sampled length classes, that is, the shape of the smoothers can be more flexible when more length classes are available.

Second, the mean curve **θ** estimated from model 13 (or the mean curves **θ**$_g$ estimated from model 14) can be rather jagged when different length classes are sampled in each haul. To overcome this, we restrict the estimation to those length classes that are sampled in at least half of the hauls. In addition, we interpolate or extrapolate the fitted curves $\hat{s}_h(l)$ to ensure that the selection probability is estimated at each length class (within the selected length range) for each haul. Within loess, interpolation is achieved by fitting a locally weighted straight line at the internal length class. Extrapolation to small (large) lengths is achieved by projecting the locally weighted straight line fitted at the smallest (largest) sampled length class, respectively.

The choice of amount of smoothing (i.e., the number of nearest neighbours) is, to a certain extent, a matter of common sense and trial and error. In practice, we try several sizes of neighbourhood and choose the largest neighbourhood (i.e., the greatest amount of smoothing) that gives an adequate fit to the data based on residual plots, while still leaving some degrees of freedom for the smooth. More formally, one might use Akaike's information criterion (AIC) to compare the fits obtained using different neighbourhoods. (S-PLUS code implementing the methods is available from the first author.)

Illustration: grid selection

We illustrate the methodology using data collected to investigate the selection of a Nordmøre grid placed at the entrance to the lower tier of a separator trawl (Fig. 1). The motivation for this gear design comes from the North Sea trawl fishery for *Nephrops norvegicus* where the relatively small minimum mesh size (currently 80 mm) can result in the by-catch of large quantities of fish. Many of the fish are below the minimum landing size and must be discarded, but a significant quantity can be landed and is of economic importance to the fishermen. It would clearly be desirable to separate the *Nephrops* and fish catches into two cod-ends that had different size-selection properties. The separator trawl (Main and Sangster 1985) attempts to do this by exploiting behavioural differences between species. The trawl is divided horizontally and attached to two separate cod-ends with mesh sizes that can be manipulated to suit the species entering them. The separator trawl is partially successful in the North Sea *Nephrops* fishery, with *Nephrops* tending to follow the lower portion of the trawl and haddock and whiting (*Merlangius merlangus*) tending to follow the upper portion. However, cod (*Gadus morhua*) and flatfish behave similarly to *Nephrops* and are typically retained in the lower cod-end. The Nordmøre grid (Isaksen et al. 1992) is an alternative method of separating species and has generally been used in shrimp and prawn fisheries to reduce levels of unwanted by-catch. It consists of a series of longitudinal bars that act to prevent the passage of "larger" fish while admitting "smaller" shrimps or prawns. The bar spacing determines the selection properties of the grid in the same way that the mesh size determines the selection properties of a cod-end. However, in the North Sea *Nephrops* fishery, a grid used in isolation would lose the landed component of the by-

catch as well as the discarded component. Thus, the combination of grid and separator trawl was designed to partition the *Nephrops* and fish by-catch into the lower and upper cod-ends, respectively. As well as providing "clean" catches, this arrangement would allow the mesh sizes of the two cod-ends to be adjusted to reflect the different target species.

Trials with this gear arrangement were conducted on commercial *Nephrops* grounds in the North Sea in August 1999. Various grid configurations were investigated. Here, we consider a subset of the data collected using two grids with bar spacings of 25 and 30 mm, respectively. A gap of 150 mm was placed in the lower part of each grid (Fig. 1) to prevent debris from blocking the grid. The upper and lower cod-ends of the separator trawl were both made of 40-mm mesh to retain all fish and *Nephrops* that entered them. Reasonable numbers of *Nephrops*, haddock, and dab were caught during the trials, and here we present the results for dab. Full details of the trials can be found in EU Report FAIR CT-98-4164 (available from the authors upon request).

To formulate an appropriate selection model, we follow Millar and Fryer (1999) and assume that dab of length l enter the trawl (during haul h) according to a Poisson process with rate λ_{hl}. Let $a_h(l)$ be the probability that a dab of length l is retained in the upper cod-end given that it entered the trawl. Then the numbers of dab of length l measured in the upper and lower cod-ends, n_{1hl} and n_{2hl}, respectively, have Poisson distributions

$$n_{1hl} \sim \text{Poisson}(\lambda_{hl} a_h(l) q_{1hl})$$

$$n_{2hl} \sim \text{Poisson}(\lambda_{hl} (1 - a_h(l)) q_{2hl})$$

where q_{1hl} and q_{2hl} are the subsampling fractions of length l dab in the upper and lower cod-ends, respectively. Conditioning on the total numbers of dab measured eliminates the nuisance parameters λ_{hl}, giving

$$n_{1hl} | n_{1hl} + n_{2hl} \sim \text{Binomial}(n_{1hl} + n_{2hl}, \phi_{hl})$$

where

$$\text{logit}(\phi_{hl}) = \text{logit}(a_h(l)) + \log(q_{1hl}/q_{2hl})$$

Note that $a_h(l)$ describes the available selection rather than the contact selection of the grid (cf. Millar and Fryer 1999; Zuur, G., et al. 2001), because some dab could pass into the upper cod-end without contacting the grid by always swimming in the upper part of the trawl.

The observed proportions of dab retained in the upper cod-end generally increased with length (Fig. 2); although often "large" dab were found in the lower cod-end, presumably because they had passed through the lower gap in the grid. We fitted a loess smoother to the data for each haul using a window that gave positive weight to the nine nearest neighbours around each length class. This resulted in between 2.0 and 3.2 degrees of freedom for each smooth, depending on the number of sampled length classes (Table 1). For comparison, we also fitted a linear logistic function of length. The smooth curves provide a reasonable fit to the data (Table 1; Fig. 2) and are preferable to the linear logistic curves that tend to overpredict selection at small and large lengths (see hauls 5, 6, 10, and 12 in Fig. 2). More formally, the smooth model is a better fit than the linear logistic model based on AIC (Table 1). There is some evidence of over-

Fig. 1. Illustration of the test gear, in which the Nordmøre grid was placed at the entrance to the lower tier of the separator trawl. The grid was 1.1 m high × 0.8 m wide and was made from 30-mm-thick nylon sheet. There was a gap of 150 mm in the lower part of the grid to prevent blockage by debris. All of the netting was made from 40-mm mesh. The section from the grid to the cod-ends was ~12 m long and the separator panel was ~2 m wide. The upper and lower cod-ends were ~4 and 8 m long, respectively.
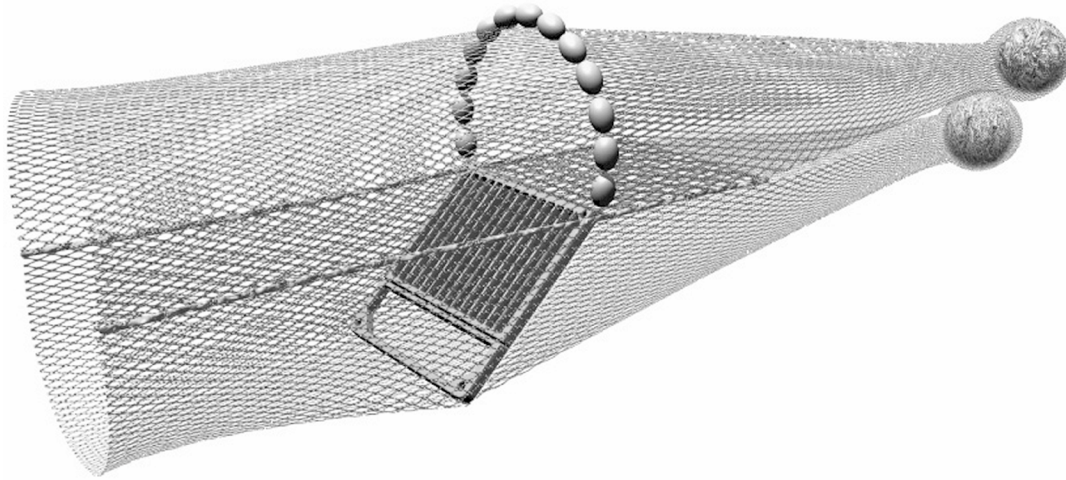

Fig. 2. For each haul, the points (solid circles) indicate the numbers at length retained in the upper cod-end, expressed as a proportion of the total numbers at length that entered the gear. The numbers at length are raised, i.e., the measured numbers at length are divided by the corresponding subsampling fraction. The fit of the smoother (solid line) with 95% pointwise confidence bands (shaded area) and the fit of the linear logistic model (broken line) are also shown. Hauls 2–5 used 25-mm bar spacing and hauls 6–12 used 30-mm bar spacing. The hauls are numbered from 2 to 12 for consistency with the full data set (EU Report FAIR CT-98-4164, available from the authors upon request).
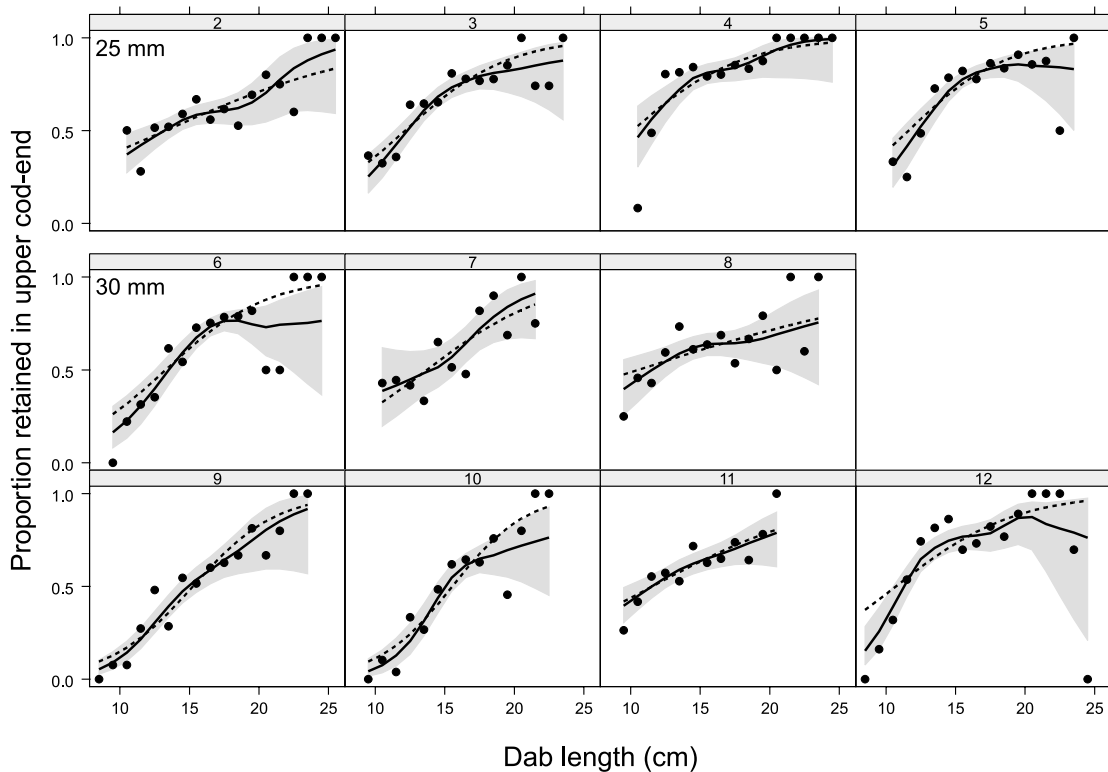
dispersion in some hauls (Table 1), so we adjusted the variance matrices **R** accordingly (see Millar and Fryer 1999) whenever the residual deviance exceeded the residual degrees of freedom.

The mean selection curves (Figs. 3a, 3b) show that, for both bar spacings, the proportion of dab retained in the up-per cod-end increases with length, from less than 30% of 10-cm dab to more than 75% of 20-cm dab. The mean curve for the 25-mm bar spacing lies above that for the 30-mm bar spacing (Fig. 3c), which is plausible because the smaller bar spacing should force more dab of a given length into the up-per cod-end. However, pointwise confidence bands (Fig. 3c)

© 2003 NRC Canada

454

Can. J. Fish. Aquat. Sci. Vol. 60, 2003

Table 1. Summary of the linear logistic and smooth model fits to the grid selection data.

Bar spacing	Haul	Length classes	df of smooth	Linear logistic		Smooth	
				Residual df	Deviance	Residual df	Deviance
25 mm	2	16	3.0	14	20.1	12.0	15.7
	3	15	2.7	13	14.5	11.3	10.3
	4	15	2.5	13	28.1	11.5	20.8
	5	14	2.5	12	23.7	10.5	13.2
30 mm	6	16	2.9	14	17.7	12.1	9.3
	7	12	2.1	10	19.4	8.9	17.2
	8	15	2.6	13	22.2	11.4	18.4
	9	16	2.7	14	28.3	12.3	20.2
	10	14	2.4	12	19.1	10.6	11.1
	11	12	2.0	10	7.8	9.0	6.9
	12	17	3.2	15	38.5	12.8	18.0
Total		162	28.6	140	239.4	122.4	161.1
AIC					283.4		240.3

Note: The number of sampled length classes, the degrees of freedom (df) associated with the smoother (one degree of freedom was always associated with the linear logistic fit), and the residual degrees of freedom and residual deviance for the two model fits are given for each haul. The total Akaike's information criterion (AIC) for each model, defined as the total residual deviance plus twice the total number of model parameters, is also given. The smooth model has the lower AIC, indicating that it gives the better fit.

suggest that the difference is not statistically significant, a result confirmed formally by bootstrap hypothesis tests ($T_{\max} = 3.53$, ASL = 0.40; $T_{\mathrm{ave}} = 1.69$, ASL = 0.32). Unfortunately, the tests are likely to have low power because only 11 hauls were possible in total. For example, a simulation showed that a constant difference between the mean curves of 0.5 on the logistic scale (cf. Fig. 3c) would only be detected about 30% of the time given 11 hauls and the levels of between-haul variation observed in the data (Fig. 3d).

The smoothers provide a satisfactory analysis of the data, but would an improved choice of parametric curve have done a better job? An obvious contender here is a modified logistic curve (which can not be expressed conveniently on the logistic scale)

$$a_h(l) = (1 - \gamma_h) \exp(a_h + b_h l)/(1 + \exp(a_h + b_h l))$$

At small lengths, this curve behaves like a linear logistic curve. However, at large lengths $a_h(l) \approx (1 - \gamma_h)$, where γ_h can be loosely interpreted as the proportion of large dab that pass through the lower gap. The modified logistic gives a comparable fit to the smoother for many of the hauls in Fig. 2. However, there are some hauls in the full data set (EU Report FAIR CT-98-4164, available from the authors upon request) for which the modified logistic curve cannot be fitted and others (e.g., hauls 2 and 7 in Fig. 2) for which the fit of the modified logistic coincides with that of the linear logistic and a variance matrix $\mathbf{R}_h$ cannot be formed (because $\hat{\gamma}_h = 0$ and is on its lower bound). Both of these pose problems for a parametric analysis that are easily avoided by the analysis using smoothers.

Catch-comparison curves

Catch-comparison trials are usually conducted to compare a test net with a standard net or to calibrate two research survey gears or vessels. Typically, the trials consist of a series of paired hauls in which the two nets (gears or vessels) are fished either in parallel or one after the other. In theory, the same population will be fished throughout a paired haul, allowing the catches of the two nets to be meaningfully compared. Although Warren (1997) and Warren et al. (1997) have modelled catch ratios as a parametric function of length, it is often difficult to find a parametric curve that will adequately capture the quite complicated relationships that are observed. A natural way forward is then to use the flexibility of smoothers to model the catch ratios as a nonparametric function of length. A smooth curve can be fitted to the data for each paired haul and then combined to estimate a mean curve that measures the relative catch rates of the two nets.

Suppose that we are interested in comparing a test net with a standard net. During paired haul h, assume that fish of length l become available to the two nets according to a Poisson process with rate λ_{hl}. Let $a_{1h}(l)$ and $a_{2h}(l)$ be the probabilities that a fish of length l is retained in the test and standard nets, respectively, given that it was available to the two nets. The numbers of fish of length l measured in the test and standard nets, n_{1hl} and n_{2hl}, respectively, then have Poisson distributions

$$n_{1hl} \sim \mathrm{Poisson}(\lambda_{hl} a_{1h}(l) q_{1hl})$$

$$n_{2hl} \sim \mathrm{Poisson}(\lambda_{hl} a_{2h}(l) q_{2hl})$$

where q_{1hl} and q_{2hl} are the subsampling fractions of length l fish in the test and standard nets, respectively. Conditioning on the total numbers of fish measured gives

$$(15) \quad n_{1hl} | n_{1hl} + n_{2hl} \sim \mathrm{Binomial}(n_{1hl} + n_{2hl}, \phi_{hl})$$

where

$$(16) \quad \mathrm{logit}(\phi_{hl}) = \log(a_{1h}(l)/a_{2h}(l)) + \log(q_{1hl}/q_{2hl})$$

Fig. 3. Results from fitting the mixed model to the dab grid selection data. (*a*) The mean selection curve (solid line) with pointwise 95% confidence bands (shaded area) for the grid with 25-mm bar spacing; the individual haul selection curves (broken lines) are also shown. The mean curve is only estimated at lengths that were sampled in at least six hauls (one-half of the hauls in the trials) and therefore does not extend as far as some of the individual haul curves. The curves are plotted on the proportion scale for ease of interpretation (i.e., they have been transformed from the logistic scale on which they were estimated). (*b*) The corresponding plot for the grid with 30-mm bar spacing. (*c*) The mean selection curve for 30-mm bar spacing minus the mean selection curve for 25-mm bar spacing (solid line) with pointwise 95% confidence bands (shaded area). The negative estimates mean that the 25-mm selection curve lies above the 30-mm selection curve. The differences are plotted on the logistic scale. (*d*) The estimates of between-haul variance at each length (solid line) with approximate 95% confidence bands (shaded area) constructed by profile likelihood methods. The point estimates are zero at large lengths, but this primarily reflects the small number of dab caught at these lengths because the confidence bands give due warning that nonnegligible between-haul variation is also consistent with the data.
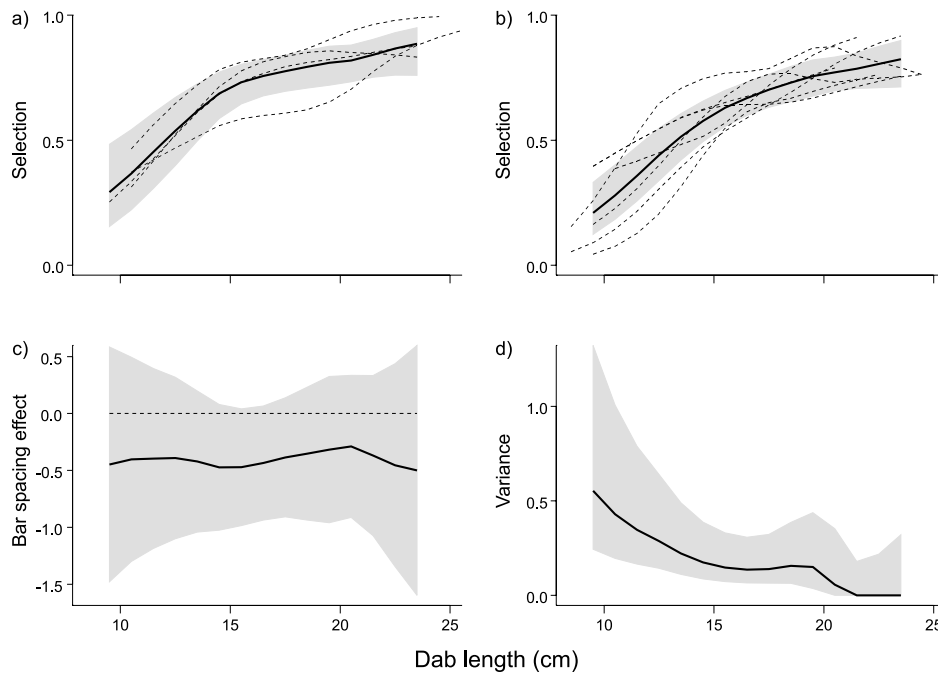
The term $\log(a_{1h}(l)/a_{2h}(l))$ is the log relative catch rate of the test gear relative to the standard gear. Assuming the log relative catch rate is a smooth function of length $s_h(l)$, eq. 16 becomes

(17) $\text{logit}(\phi_{hl}) = s_h(l) + \log(q_{1hl}/q_{2hl})$

and eqs. 15 and 17 form the same generalised additive model as was obtained for size-selection curves (eqs. 1 and 9). Smooth curves can be therefore be fitted to the data for each paired haul and then combined over hauls to estimate a mean curve **θ** using the techniques described earlier.

Note that the formulation of the catch-comparison model uses the term "available" in its broadest sense to indicate that fish are on the fishing grounds and subject to exploitation during the paired haul. Of course, it is likely that the two nets will trawl through different local aggregations of fish during a paired haul, particularly with an alternate haul design. This will affect the relative catch rate of the two nets for that paired haul. For example, if the test net trawls through a large patch of fish that is missed by the standard gear, then the relative catch rate of the test net will be inflated accordingly. However, such effects are assumed to occur at random and to average out over a series of paired hauls. They are thus implicitly accommodated, along with all other random effects, within the matrix **D** that measures the variation in relative catch rate between paired hauls. Other approaches

are possible. For example, the model described by Pelletier (1998) has a component that explicitly describes how different numbers of fish become available to the two nets within a paired haul. However, her approach does not lend itself so readily to a length-based analysis.

To illustrate the methodology, we use data from an intercalibration exercise to compare the research survey vessels *Scotia II* and *Scotia III*. Since the mid-1970s, *Scotia II* has taken part in annual groundfish surveys like the International Bottom Trawl Surveys (ICES 1999). This has involved fishing for 1 h with a Grande Ouverture Verticale (GOV) trawl (ICES 1999) at a series of designated stations each year. The time series of data obtained from these surveys plays an important role in the assessment of commercial groundfish stocks. *Scotia II* was replaced by *Scotia III* in 1998, with the intention that *Scotia III* would continue to participate in the groundfish surveys, again using the GOV trawl, but towing for only 30 min in line with the current practice of other European fisheries institutes. To maintain the continuity of the survey times, a comparative fishing trial was conducted in April 1998. This consisted of a series of parallel hauls with *Scotia III* fishing for 30 min and *Scotia II* fishing for 1 h. All things being equal, a 30-min tow with *Scotia III* would be expected to catch half as many fish as a 1-h tow with *Scotia II*. The hypothesis of interest is therefore whether the relative catch rate of *Scotia III* to *Scotia II* is equal to 0.5 for all

456

Can. J. Fish. Aquat. Sci. Vol. 60, 2003

length classes. Here, we consider the results for haddock. Full details of the trials can be found in Zuur, A.F., et al. (2001).

The relationship between relative catch rate and length varies widely between paired hauls (Fig. 4) and would be difficult to capture with a simple parametric curve. We fitted a loess smoother to the data for each haul, using a window that gave positive weight to the 13 nearest neighbours around each length class. This resulted in between 2.6 and 4.2 degrees of freedom for each smooth. The smooth curves generally provided a reasonable fit to the data (Fig. 4). However, there was some evidence of overdispersion in some fits, so the variance matrices **R** were adjusted accordingly (see Millar and Fryer 1999) whenever the residual deviance exceeded the residual degrees of freedom.

The mean curve (Fig. 5) shows that a relative catch rate of 0.5 is consistent with the data for lengths up to 35 cm. However, above 35 cm, there is a suggestion that *Scotia III* catches more haddock than might be expected relative to *Scotia II*. To formally test this requires a slight modification to the bootstrap hypothesis test used before. For each length class l, we now compare the hypotheses

$$H_0: \theta_l = \log(0.5)$$

$$H_1: \theta_l \text{ unconstrained}$$

by fitting the model

$$\hat{s}_h(l) \sim N(\theta_l, D_{ll} + R_{hll})$$

under the two hypotheses and computing a test statistic T_l given by minus twice the difference in the log-likelihoods. The test statistics T_l are then combined over all length classes to give $T_{\max}$ or T_{ave} as before. To bootstrap, we generate a data set that satisfies the null hypothesis by calculating

$$\hat{s}_h^* = \hat{s}_h - \hat{\theta} + \log(0.5)$$

where $\hat{\theta}$ is the estimate of θ under the alternative hypothesis. Applying the bootstrap hypothesis test reveals only marginal evidence against the hypothesis of a constant relative catch rate of 0.5 for all lengths ($T_{\max} = 6.26$, ASL = 0.07; $T_{\text{ave}} = 1.29$, ASL = 0.28).

Because of logistic constraints, only 24 paired hauls were possible during the trials, quite a small number for an intercalibration exercise. It is perhaps fortunate that the null hypothesis of a constant relative catch rate of 0.5 was sustained, or at least not strongly refuted, because the width of the confidence bands in Fig. 5 reveal that the mean curve is not estimated sufficiently precisely to provide acceptable conversion coefficients. To maintain the continuity of the survey time series, the pragmatic approach is therefore simply to halve *Scotia II* catches to provide *Scotia III* equivalents. This conversion should be reasonable for lengths up to 35 cm, and due warning that conversions above 35 cm might be suspect is illustrated in Fig. 5. The choice of sample size is discussed further by Pelletier (1998), who tabulates the number of paired hauls used by other intercalibration studies (range 40–285, median 72). Pelletier herself used 30 paired hauls, but these were replicated in four areas. Finally, Warren (1997), who modelled catch ratios as a parametric function of length, states that 285 paired hauls gives "confidence intervals that

are, perhaps, acceptably small", although the catch rates that he was modelling were much lower than those of *Scotia II* and *Scotia III*.

Discussion

We have used the techniques described here to analyse several size-selection and catch-comparison trials. We have found the techniques easy to apply, simple to adapt to address specific management questions, and capable of generating results that are accessible to both gear technicians and managers. Most importantly, they offer a potential method of analysis when there are difficulties with a parametric approach.

The mixed-model approach is not the only way of using nonparametric curves in a selectivity (or catch-comparison) analysis. As described earlier, Millar (1993) and Munro and Somerton (2001) pooled data over hauls, fitted a nonparametric selection curve, and assessed its precision by bootstrapping. Which approach is better? In our opinion, the two methods are complementary rather than competitive and the choice between them will be case specific and, to a large extent, a matter of personal preference. Both are simple to apply once the software is in place. The pooling approach makes fewer distributional assumptions than the mixed model; however, any concerns about the normality assumptions in the mixed model can be alleviated by fitting the mean curve as usual, but then estimating its precision by bootstrapping rather than by using the standard errors derived from the maximised residual log-likelihood. The mixed model will tend to be computationally faster than the pooling approach because no bootstrapping is involved (unless hypotheses are to be tested), but computational speed is rarely an issue these days. Perhaps the most important difference between the two approaches is that the mixed model provides a natural framework for hypothesis testing that is not available with the pooling approach. The mixed model does this by explicitly defining a population of selection curves that are distributed about a mean selection curve. Hypotheses can then be readily expressed in terms of the mean selection curve. On the other hand, the pooling approach works with a population of hauls that are characterised not by their selection, but by the joint distribution of the numbers at length retained in the different compartments of the fishing gear. The notion of a mean selection curve is only implicit in this population, and it is therefore much harder to formulate and test hypotheses.

Smooth size-selection curves can easily be fitted to data from covered cod-end experiments and are likely to be applicable to other selection trials using small mesh covers or cod-ends to retain all fish that enter the gear (as in the grid trials described earlier). Smooth size-selection curves are less likely to be applicable to data collected from a trouser-trawl or twin-trawl experiment (Wileman et al. 1996). Here, we have

$$n_{1hl} | n_{1hl} + n_{2hl} \sim \text{Binomial}(n_{1hl} + n_{2hl}, \phi_{hl})$$

$$(18) \quad \text{logit}(\phi_{hl}) = \log(r_h(l)) + \text{logit}(p_h) + \log(q_{1hl}/q_{2hl})$$

Fig. 4. For each haul, the points (solid circles) indicate the numbers at length retained by *Scotia III*, expressed as a proportion of the total numbers at length retained by both *Scotia II* and *Scotia III*. The numbers at length are raised, i.e., the measured numbers at length are divided by the corresponding subsampling fraction. The fit of the smoother (solid line) with 95% confidence bands (shaded area) is also shown. A constant relative catch rate of 0.5 would mean that one-third of the fish at length would be retained in *Scotia III*, as indicated by the broken horizontal line.
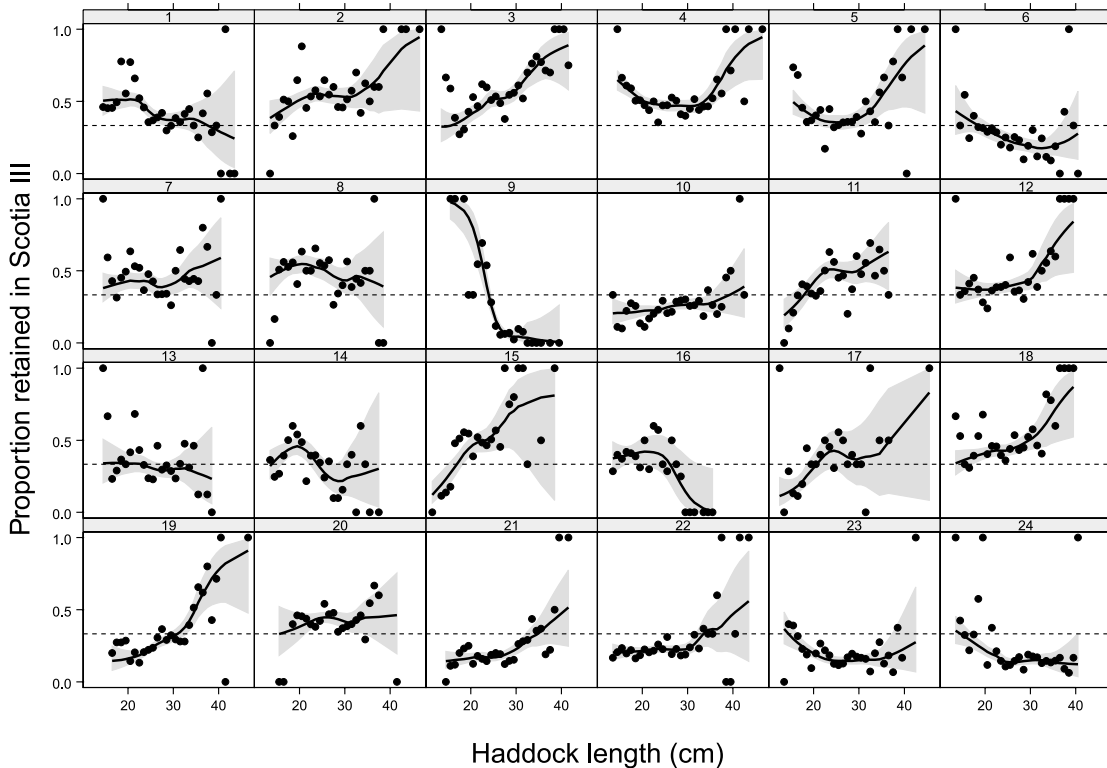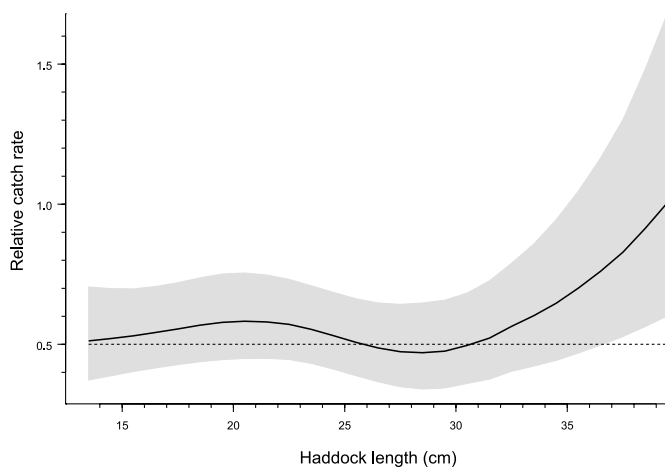

Fig. 5. The mean relative catch rate of *Scotia III* to *Scotia II* (solid line) with pointwise 95% confidence bands (shaded area). The mean relative catch rate was estimated on the log scale but has been back-transformed for presentation. The broken horizontal line indicates a constant relative catch rate of 0.5.

where n_{1hl} and n_{2hl} are the numbers of fish of length l measured in the test and small mesh cod-end, respectively; q_{1hl} and q_{2hl} are the corresponding subsampling fractions; $r_h(l)$ is the contact-selection curve of the test cod-end; and p_h is the split parameter, i.e., the probability that a fish enters the test cod-end given that it entered the gear (Millar and Walsh

1992). When the split parameter is known, $\log(r_h(l))$ can be modelled as a smooth function of length. Unfortunately, the split parameter is usually unknown and must be estimated from the data, in which case the split parameter and the smooth function of length become confounded. However, the techniques described in this paper would still be useful for trouser-trawl experiments if they were regarded as catch-comparison rather than size-selection experiments. In this case, eq. 18 can be rewritten as

$$\text{logit}(\phi_{hl}) = \log(p_h r_h(l)/(1 - p_h)) + \log(q_{1hl}/q_{2hl})$$

The term $\log(p_h r_h(l)/(1 - p_h))$ is the log relative catch rate of the test cod-end relative to the small mesh cod-end and can be modelled as a smooth function of length $s_h(l)$ as before. Clearly this approach would provide less information than a parametric analysis that estimates p_h and $r_h(l)$ separately. However, it might have advantages if few large fish are caught during the trials, in which case p_h can be confounded even with parametric selection curves. It might also have advantages when the selection processes are too complicated to be modelled by a single parametric curve, for example, when a panel is inserted in the cod-end and there is selection by both the panel and the cod-end.

Further work is required to develop efficient and stable numerical routines that will estimate all the elements of the mean curve simultaneously rather than one length class at a time. This would enable the full between-haul variance matrix to be estimated (rather than just the diagonal elements)

458

Can. J. Fish. Aquat. Sci. Vol. 60, 2003

and the full log-likelihood to be calculated and would allow the modelling of a wide range of fixed effects including continuous covariates. One way of reducing the numerical difficulties involved might be to impose some structure on the between-haul variance matrix, as is typically done in repeated-measures analysis (e.g., Crowder and Hand 1990). More fundamentally, the formulation of the mixed model could be improved by making the specification of the mean curve explicit. At present, the mean curve is constructed as the average of a population of smooth curves and is therefore implicitly a smooth curve in its own right, although all of its attributes (e.g., smoothness, degrees of freedom) are derived from the properties of the smooth curves that surround it. However, it would be more acceptable to define the mean curve explicitly as a smooth curve with a given degree of flexibility and then to specify how the individual haul curves are distributed about it. The development of such models is an area of active research within the statistical community (e.g., Verbyla et al. 1999) and it will only be a matter of time before they become generally available.

Acknowledgements

References

Armstrong, M.J., Briggs, R.P., and Rihan, D. 1998. A study of optimum positioning of square-mesh escape panels in Irish Sea *Nephrops* trawls. Fish. Res. **34**: 179–189.

Briggs, R.P. 1992. An assessment of nets with a square mesh panel as a whiting conservation tool in the Irish Sea *Nephrops* fishery. Fish. Res. **13**: 133–152.

Cleveland, W.S. 1979. Robust locally-weighted regression and smoothing scatterplots. J. Am. Stat. Assoc. **74**: 829–836.

Crowder, M.J., and Hand, D.J. 1990. Analysis of repeated measures. Chapman and Hall, London.

Draper, N.R., and Smith, H. 1981. Applied regression analysis. Second edition. John Wiley and Sons, New York.

Efron, B., and Tibshirani, R.J. 1993. An introduction to the bootstrap. Chapman and Hall, New York.

Ehrich, S. 1991. Comparative fishing experiments by research trawlers for cod and haddock in the North Sea. J. Cons. Int. Explor. Mer, **47**: 275–283.

Fryer, R.J. 1991. A model of between-haul variation in selectivity. ICES J. Mar. Sci. **48**: 281–290.

Hastie, T.J., and Tibshirani, R.J. 1990. Generalized additive models. Chapman and Hall, London.

Hocking, R.R. 1996. Methods and applications of linear models: regression and the analysis of variance. John Wiley and Sons, New York.

International Council for the Exploration of the Sea. 1999. Manual for the International Bottom Trawl Surveys, revision VI. ICES CM 1992/D:2 Addendum 2.

Isaksen, B., Valdermarsen, J.W., Larsen, R.B., and Karlsen, L. 1992. Reduction of fish by-catch in shrimp trawl using a rigid separator grid in the aft belly. Fish. Res. **13**: 335–352.

Madsen, N., Moth-Poulsen, T., Holst, R., and Wileman, D. 1999. Selectivity experiments with escape windows in the North Sea Nephrops (*Nephrops norvegicus*) trawl fishery. Fish. Res. **42**: 167–181.

Main, J., and Sangster, G.I. 1985. Trawling experiments with a two-level net to minimise the undersized gadoid bycatch in a *Nephrops* fishery. Fish. Res. **3**: 131–145.

McCullagh, P., and Nelder, J.A. 1989. Generalized linear models. 2nd edition. Chapman and Hall, London.

Millar, R.B. 1993. Incorporation of between-haul variation using bootstrapping and nonparametric estimation of selection curves. Fish. Bull. **91**: 564–572.

Millar, R.B., and Fryer, R.J. 1999. Estimating the size-selection of towed gears, traps, nets and hooks. Rev. Fish Biol. Fish. **9**: 89–116.

Millar, R.B., and Walsh, S.J. 1992. Analysis of trawl selectivity studies with an application to trouser trawls. Fish. Res. **13**: 205–220.

Munro, P.T., and Somerton, D.A. 2001. Maximum likelihood and non-parametric methods for estimating trawl footrope selectivity. ICES J. Mar. Sci. **58**: 220–229.

Pelletier, D. 1998. Intercalibration of research survey vessels in fisheries: a review and an application. Can. J. Fish. Aquat. Sci. **55**: 2672–2690.

Reeves, S.A., Armstrong, D.W., Fryer, R.J., and Coull, K.A. 1992. The effects of mesh size, cod-end extension length and cod-end diameter on the selectivity of Scottish trawls and seines. ICES J. Mar. Sci. **49**: 279–288.

Verbyla, A.P., Cullis, B.R., Kenward, M.G., and Welham, S.J. 1999. The analysis of designed experiments and longitudinal data by using smoothing splines (with discussion). Appl. Stat. **48**: 269–311.

Warren, W.G. 1997. Report on the comparative fishing trial between the *Gadus Atlantica* and *Teleost*. NAFO Sci. Counc. Stud. **29**: 81–92.

Warren, W., Brodie, W., Stansbury, D., Walsh, S., Morgan, J., and Orr, D. 1997. Analysis of the 1996 comparative fishing trial between the *Alfred Needler* with the Engel 145 trawl and the *Wilfred Templeman* with the Campelen 1800 trawl. NAFO SCR Doc. No. 97/68.

Wileman, D.A., Ferro, R.S.T., Fonteyne, R., and Millar, R.B. 1996. Manual of methods of measuring the selectivity of towed fishing gears. ICES Coop. Res. Rep. No. 215.

Zuur, A.F., Fryer, R.J., and Newton, A.W. 2001. The comparative fishing trial between *Scotia II* and *Scotia III*. Fisheries Research Services (FRS) Marine Laboratory Rep. No. 03/01.

Zuur, G., Fryer, R.J., Ferro, R.S.T., and Tokai, T. 2001. Modelling the size selectivities of a trawl cod-end and an associated square mesh panel. ICES J. Mar. Sci. **58**: 657–671.

Appendix A

Estimating mean selection curves

Appendix B of Fryer (1991) describes both maximum likelihood and residual maximum likelihood methods for estimating the mean selection parameters θ, their standard errors, and the between-haul variance matrix $\mathbf{D}$ from eq. 6 when θ is the mean of a two-parameter selection curve. The estimation of fixed effects is also described. To generalise the theory to an m-parameter selection curve, it is only necessary to replace the first term of the log-likelihood (in appendix B of Fryer (1991)) by $mH\log(2\pi)$ and the first term of the residual log-likelihood by $(mH - q)\log(2\pi)$.

The estimation procedures can be simplified in the special case of estimating θ_l, its standard error, and D_{ll} from eq. 13, or more generally θ_{gl}, $g = 1,\ldots,G$, their standard errors, and D_{ll} from eq. 14. Let $V_{gh} = D_{ll} + R_{ghll}$ and $W_{gh} = V_{gh}^{-1}$, where both V_{gh} and W_{gh} are scalars. Adapting the equations in Fryer (1991) gives the log-likelihood l^M to be

$$-2l^M(\hat{s}_{gh}(l) | \theta_{gl}, D_{ll}) =$$

$$\sum_g \left\{ H_g \log(2\pi) + \sum_h \log(V_{gh}) + \sum_h W_{gh}(\hat{s}_{gh}(l) - \theta_{gl})^2 \right\}$$

The maximum likelihood estimate of D_{ll} is the value $\hat{D}_{ll}^M$ that maximises $l^M(\hat{s}_{gh}(l) | \hat{\theta}_{gl}(D_{ll}), D_{ll})$, where

$$\hat{\theta}_{gl}(D_{ll}) = \sum_h W_{gh}\,\hat{s}_{gh}(l) \,/\, \sum_h W_{gh}$$

and can be found by a simple one-dimensional search. The maximum likelihood estimate of θ_{gl} is then $\hat{\theta}_{gl}^M = \hat{\theta}_{gl}(\hat{D}_{ll}^M)$.

The residual log-likelihood l^R is given by

$$-2l^R(\hat{s}_{gh}(l) | \theta_{gl}, D_{ll}) =$$

$$\sum_g \left\{ (H_g - 1) \log(2\pi) + \sum_h \log(V_{gh}) - \log(H_g) + \right.$$

$$\left. \log\left(\sum_h W_{gh} \right) + \sum_h W_{gh}(\hat{s}_{gh}(l) - \theta_{gl})^2 \right\}$$

The residual maximum likelihood estimate of D_{ll} is the value $\hat{D}_{ll}^R$ that maximises $l^R(\hat{s}_{gh}(l) | \hat{\theta}_{gl}(D_{ll}), D_{ll})$ and can again be found by a simple one-dimensional search. The residual maximum likelihood estimate of θ_{gl} is then $\hat{\theta}_{gl}^R = \hat{\theta}_{gl}(\hat{D}_{ll}^R)$.

The variances of $\hat{\theta}_{gl}^M$ and $\hat{\theta}_{gl}^R$ are estimated to be

$$\text{Var}(\hat{\theta}_{gl}) = 1 \,/\, \sum_h W_{gh}$$

where the scalars W_{gh} are constructed using $\hat{D}_{ll}^M$ and $\hat{D}_{ll}^R$, respectively.