Technical University of Denmark

DTU

Estimating the relationship between abundance and distribution

Rindorf, Anna; Lewy, Peter

DTU Library
Technical Information Center of Denmark

Estimating the relationship between abundance and distribution

Anna Rindorf and Peter Lewy

Abstract: Numerous studies investigate the relationship between abundance and distribution using indices reflecting one of the three aspects of distribution: proportion of area occupied, aggregation, and geographical range. Using simulations and analytical derivations, we examine whether these indices provide unbiased estimates of the relationship when estimated from count data. The indices investigated include the proportion of empty samples, the proportion of structurally empty samples, Lloyds index of patchiness, measures derived from Lorenz curves (such as D_{95} and the Gini index), and measures based on Euclidean distance to the centre of gravity of the spatial distribution. Only the proportion of structurally empty areas, Lloyds index, and indices of the distance to the centre of gravity of the spatial distribution are unbiased at all levels of abundance. The remaining indices generate relationships between abundance and distribution even in cases where no underlying relationships exists, although the problem decreases for measures derived from Lorenz curves when samples contain more than four individuals on average. To illustrate the problem, the indices are applied to juvenile North Sea cod, *Gadus morhua*.

Résumé : De nombreuses études examinent la relation entre l'abondance et la distribution à l'aide d'indices qui représentent l'un des trois aspects de la distribution, soit la proportion du territoire occupé, l'agrégation et la répartition géographique. Au moyen de simulations et de dérivations analytiques, nous vérifions si ces indices fournissent des estimations justes de la relation lorsqu'ils sont basés sur des données de dénombrements. Les indices examinés incluent la proportion d'échantillons vides, la proportion d'échantillons structurellement vides, le coefficient de répartition en taches de Lloyd, des mesures dérivées des courbes de Lorenz, telles que D_{95} et le coefficient de Gini, ainsi que des mesures basées sur la distance euclidienne jusqu'au centre de gravité de la distribution spatiale. Seuls la proportion de sites structurellement vides, le coefficient de Lloyd et les coefficients de distance au centre de gravité de la distribution spatiale restent non biaisés à tous les niveaux d'abondance. Les autres indices génèrent des relations entre l'abondance et la distribution même lorsqu'il n'existe pas de relation sous-jacente; cependant, dans le cas de mesures dérivées du coefficient de Lorenz, le problème diminue lorsque les échantillons contiennent plus de quatre individus en moyenne. Nous utilisons les indices avec des données sur de jeunes morues, *Gadus morhua*, de la mer du Nord pour illustrer le problème.

[Traduit par la Rédaction]

Introduction

The relationship between abundance and distribution is a key issue in fisheries management and ecology, as a population that contracts into a small area at low abundance becomes highly vulnerable to both exploitation and climatic changes. The contraction renders catch rates high in spite of stock declines and has facilitated the overexploitation of several fish stocks (Hamre 1978; Beverton 1990; Hutchings 1996). Furthermore, the presence or absence of contraction affects the success of different management measures. A contracting stock can be safeguarded through protecting the area where the stock is concentrated, whereas stocks that become increasingly dispersed as density declines will not obtain the same beneficial effect of area-based management. In contrast, measures such as effort limitation may be efficient for non-contracting stocks but have little effect on contracting stocks, as sustained high catch rates will cause fishing mortality per unit effort to increase as abundance declines. The relationship is of even greater importance to stocks faced with climatically induced decreases in stock size; the decrease in abundance increases the risk of local extinction and the decrease in the number of local aggregations increases the risk of stock extinction ("double jeopardy"; Lawton 1996). Hence, detecting the presence or absence of a concentration of the stock as abundance declines is of great importance to successful fisheries management in a changing environment.

Numerous studies have attempted to determine whether a relationship between abundance and distribution exists in natural populations. Most of these studies are based on measuring one or several of three different aspects of distribution: the proportion of the distribution area occupied, aggregation, and geographical range. In the case where the individuals can be detected and counted without error, these aspects can be estimated directly and related to abundance. However, in the majority of marine cases, complete census is impossible. Instead, distribution is inferred from samples collected by nets, trawls, and traps or from counts of individuals in a specified area of the bottom. In these cases, the number of individuals detected varies as a result of both random variation associated with sampling and as a result of differences in local abundance.

A. Rindorf and P. Lewy. Technical University of Denmark, National Institute of Aquatic Resources, Charlottenlund Castle, 2920 Charlottenlund, Denmark.

Corresponding author: Anna Rindorf (e-mail: ar@aqua.dtu.dk).

Spatial statistical analyses can be used to separate sampling variation from changes in local abundance (Rivoirard et al. 2000; Stelzenmüller et al. 2005; Lewy and Kristensen 2009). However, to demonstrate a relationship between abundance and distribution, simpler distribution indices are frequently used. It is essential for the interpretation of the results based on such indices that random variation due to sampling does not generate a spurious link between the index and abundance. Such a link will cause the index to change when abundance changes even though the parameters of the underlying distribution of the individuals remain the same.

This study presents a statistical investigation of the properties of a number of different indices frequently used to study the relationship between distribution and abundance directly from samples consisting of counts of individuals. The indices are examined to determine whether they are statistically independent of the mean number of individuals per sample when the sampled number shows either a Poisson or negative binomial distribution. When individuals are randomly distributed in space, the Poisson distribution is the appropriate distribution to describe the variation in the number of individuals per sample. However, when individuals are patchily distributed, the number of individuals per sample will follow a negative binomial distribution where the aggregation of individuals is reflected in the size parameter k (Lloyd 1967; Taylor et al. 1988). The indices and the aspect of distribution they reflect are listed in Table 1. Among the numerous published indices of spatial distribution available, we include only those that are frequently found in the literature and have been used to infer abundance–distribution relationships directly from samples (without prior smoothing or transformation). The properties of these indices are investigated both analytically and by simulation (Table 1). Further, the ability of the indices to detect changes in distribution is investigated in situations where the requirements of the indices with respect to statistical properties and spatial independence of samples are violated.

To illustrate the practical implications of using biased indices, we estimate the relationship between abundance and distribution of juvenile North Sea cod (*Gadus morhua*).

Distribution indices investigated

Area occupied

The proportion of an area that is occupied by a given species is often determined from samples thst each contain a number of individuals. The occupied proportion is assumed to be equal to the proportion of samples that contain counts of at least one individual (occurrence) or alternatively as the proportion of samples containing more than a fixed number of individuals. However, both of these indices are intrinsically linked to the mean number of individuals per sample when individuals are randomly distributed in space (Murawski and Finn 1988; Hartley 1998; Swain and Sinclair 1994) (Table 2) or in some proportion of space (Sileshi et al. 2009). Comparing occurrence with mean number of individuals in non-empty samples only (Macpherson 1989; Eycott et al. 2006; Webb et al. 2007) does not eliminate the link (Wright 1991). Occasionally, empty samples occur through a combination of random effects and sampling in truly empty areas (structurally empty samples). In these cases, the relative importance of the two types of empty samples can be estimated using zero inflated Poisson and negative binomial distributions (Sileshi et al. 2009).

Assume that the data consist of individual counts in a number of samples, N, of which a proportion p_e are taken in areas uninhabited because of, for example, adverse habitat conditions or contraction of the population. If the individuals within the inhabited area are distributed according to a negative binomial distribution with a parameter m reflecting the mean of number of individuals in a sample, the probability density function describing the probability of obtaining n individuals in a sample is

$$(1) \quad P_{\text{Negbin}}(n) = \begin{cases} p_e + (1-p_e)\left(\dfrac{k}{k+m}\right)^k & \text{if} \quad n = 0 \\ (1-p_e)\dfrac{\Gamma(n+k)}{\Gamma(n)\Gamma(k)}\left(\dfrac{k}{k+m}\right)^k\left(\dfrac{m}{k+m}\right)^n & \text{if} \quad n > 0 \end{cases}$$

where k is the size parameter of the negative binomial distribution. The parameters p_e, k, and m can be estimated using the R package zeroinfl (R Development Core Team 2010) or by maximizing the likelihood function.

Aggregation

Lloyds index of patchiness

Lloyds index of patchiness is a function of the empirical mean number of individuals per sample $\overline{n}$ and the sampling variance s^2 (Lloyd 1967):

$$(2) \quad I_{\text{LP}} = 1 + \frac{s^2}{\overline{n}^2} - \frac{1}{\overline{n}}$$

Unlike the proportion of non-empty samples, Lloyds index

is independent of the mean number of individuals per sample when applied to Poisson or negative binomially distributed data (Lloyd 1967). Lloyds index of patchiness can be regarded as an empirical estimate of the true index $1 + \frac{\sigma^2}{m^2} - \frac{1}{m}$, where m and σ^2 are the true mean and variance, respectively, of the distribution. The index estimated from the empirical mean and variance converges to the true index when the number of samples increases towards infinity. The true index takes the value 1 when the number of individuals in a sample is Poisson distributed and takes the value $1 + 1/k$ when data are negative binomially distributed, where k is the size parameter. In these cases, Lloyds index is approximately independent of the mean (Lloyd 1967) and can hence be used to compare patchiness of distributions with different means. However, this property is not retained when the distribution

Table 1. Overview of the measures investigated.

Type	Measure	Method used
Occupied area	Proportion of empty samples	Analytical derivation
	Proportion of samples containing more than a fixed number of individuals	Analytical derivation
	Proportion of structurally empty samples	Simulation[b] and subsequent estimation
Aggregation	Lloyds index of patchiness[a]	Simulation[b] and subsequent estimation of maximum likelihood and moment estimates
	Measures based on the area below the Lorenz curve (Gini and A)	Simulation[b] and subsequent estimation
	Measures based on points on the Lorenz curve (D_y)	Simulation[b] and subsequent estimation
Geographical spread	Average distance to centre of gravity	Analytical supplemented by simulation[b] to determine accuracy
	Area of the contour ellipse	Analytical supplemented by simulation[b] to determine accuracy

[a]Lloyds index of patchiness converges to $1 + 1/k$ when data are negative binomially distributed, where k is the size parameter in the distribution.

[b]Simulations varied in number of samples simulated (20 to 500), the mean number of individuals per sample (0.1–10), and the distribution across samples to ensure that the greatest possible resolution was obtained in the region with significant bias. Simulations did not cover parameter combinations with constant bias (high densities) or for which bias could not be estimated (very low densities).

Table 2. The relationship between mean m and two indicators of aggregation (proportion of empty samples and Lloyds index of patchiness) for a range of statistical distributions.

Distribution	Variance	Proportion of zeros	Lloyds index
Binomial	$m\left(1 - \frac{m}{n}\right)$	$\left(1 - \frac{m}{n}\right)^n$	$1 - \frac{1}{n}$
Poisson	m	e^{-m}	1
Negative binomial	$m + \frac{m^2}{k}$	$\left(\frac{k}{k+m}\right)^k$	$1 + \frac{1}{k}$
Geometric	$m(m-1)$	$\frac{1}{m}$	$2\left(1 - \frac{1}{m}\right)$
Normal[a]	σ^2	0	$1 + \frac{\sigma^2}{m^2} - \frac{1}{m}$
Log-normal[b]	$m^2\left(e^{\sigma^2} - 1\right)$	0	$e^{\sigma^2} - \frac{1}{m}$
Gamma[c]	ms	0	$1 + \frac{1}{m}(s - 1)$
Exponential	m^2	0	$2 - \frac{1}{m}$

[a]σ^2 denotes the variance.

[b]σ^2 denotes the log variance.

[c]s denotes the scale parameter.

differs from that of a binomial, Poisson, or negative binomial distribution (Table 2). For individuals distributed according to the continuous lognormal and gamma distributions, Lloyds index depends on the mean, except in the special case where the gamma distribution has the scale parameter 1 (Table 2).

Lorenz curves

A Lorenz curve is the cumulative distribution of the samples ordered by ascending size. Indices based on Lorenz curves include the minimum area containing a fixed percentage of the population, D_y, and the Gini index (Poulin 1993; Swain and Sinclair 1994; Woillez et al. 2007, among others). The indices describe aggregation as the difference between the observed distribution and a distribution where every sample contains the same number of individuals.

The measures D_{50}, D_{75}, D_{90}, and D_{95}, together denoted D_y, are defined as the minimum area containing $y\%$ of the individuals and is determined by a point on the Lorenz curve (Fig. 1). D_y can be estimated by

$$(3) \quad D_y = \frac{l_y + \frac{y/100 - z(l_y)}{z(l_y + 1) - z(l_y)}}{N}$$

where $z(l_y) = \sum_{i=1}^{l_y} n_{[i]} / \sum_{i=1}^{n} n_i$, l_y fulfil that $z(l_y) \le 1 - y/100 \le z(l_y + 1)$, N is the number of observations, and $(n_{[1]},$

$n_{[2]}, \ldots, n_{[N]}$ are the observations in descending order. Hence, $z(l_y)$ is the proportion of all individuals recorded in the smallest l_y samples. It is assumed here that the observations represent equally sized areas. D_y takes values between zero and $y/100$. If D_y is zero or close to zero, the individuals are concentrated in a few samples. Conversely, if $D_y = y/100$, all samples take equal values. Hence, increasing aggregation results in a decrease in D_y. The indices D_{50}, D_{75}, D_{90}, and D_{95} were introduced by Swain and Sinclair (1994) and have subsequently been reintroduced by Atkinson et al. (1997) and Anderson and Gregory (2000).

Another commonly used measure based on Lorenz curves is the Gini index, G, defined as twice the area between the Lorenz curve and the diagonal (Fig. 1). The index is derived from socioeconomics and ranges from 0 to 1. It takes the value 0 when all samples are equal and 1 when all individuals are recorded in a single sample. It is estimated as

$$(4) \quad G = \frac{\sum_{i=1}^{N-1} i(N - i)\left[n_{(i+1)} - n_{(i)}\right]}{(N - 1)\sum_{i=1}^{n} n_i}$$

where N is the number of observations, and $(n_{(1)}, n_{(2)}, \ldots, n_{(N)})$ are the observations in ascending order. As above, the

observations are assumed to represent equally sized areas. Another measure that is occasionally used is the area A below the Lorenz curve (Laurel et al. 2004; Woillez et al. 2007). This area can be expressed directly as $(1 - G)/2$, and therefore the bias in this index can be estimated directly from that of the Gini index. The Gini index and the area below the Lorenz curve have the advantage compared with indices based on single points on the Lorenz curve (such as D_{95}), as they integrate aggregation over the entire Lorenz curve and therefore alleviate the need to examine a range of indices (e.g., D_{50}, D_{90}, and D_{95}).

Geographical spread

Indices of the geographical spread are defined here as indices that depend on the geographical distribution (e.g., latitude and longitude) and include the minimum and maximum latitude at which the species is observed (Gaston 1991), the average distance to the centre of gravity of its distribution (Murawski and Finn 1988; Marshall and Frank 1994), and indices based on contour ellipses (Atkinson et al. 1997; Brodie et al. 1998, Woillez et al. 2007). Minimum and maximum observed latitude are intrinsically linked to mean abundance and can therefore not be used to estimate the relationship between the two (Brown 1984).

Murawski and Finn (1988) suggested measuring the geographical spread of a population as the average distance d from an individual sampled to the centre of gravity of all samples:

$$d = \frac{\sum_{i=0}^{N} n_i \, \text{dist}(i, C)}{\sum_{i=0}^{N} n_i}$$

where $\text{dist}(i, C)$ is the geographical distance between the location of sample i and the centre of gravity C, n_i is the number of individuals in sample i, and N is the number of samples taken. Their hypothesis was that a concentration of the population would lead to a decrease in the average distance to centre.

A similar measure is the average squared distance to centre of gravity of latitude and longitude of individuals (also termed inertia; Woillez et al. 2007):

$$s^2 = \frac{\sum_{i=0}^{N} n_i [\text{dist}(i, C)]^2}{\sum_{i=0}^{N} n_i}$$

which together with the covariance between weighted latitude and longitude can be used to construct contour ellipses (Atkinson et al. 1997). If individuals are normally distributed in space, the area of the contour ellipse can be used directly as an estimate of the area containing a certain percentage of the population. The area can be calculated from the average squared distance to the centre and the covariance between the distance in latitudinal and longitudinal directions (as shown in Appendix A). Clearly, d, s^2, and the area of the contour ellipse are heavily dependent on the dis-

Fig. 1. Illustration of the Gini index (G) and D_y. The Gini index is twice the grey area or, equivalently, $1 - 2A$, where A is the area below the Lorenz curve, $L(x)$. As indicated on the figure, D_y is determined by the first coordinate of the point $(x, 1 - y/100)$ on the Lorenz curve ($D_y = 1 - L^{-1}(1 \text{ year}/100)$). Values on the x axis can be the cumulative proportion of samples as shown here, the cumulative number of individuals used in socioeconomics, or the cumulative area used by Swain and Sinclair (1994).
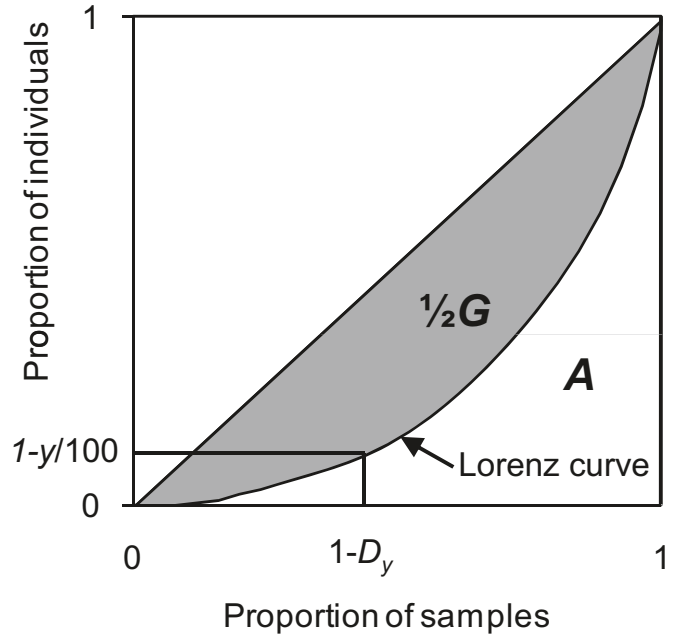
tribution of sampling sites. If sampling at a given time is only conducted in, say, the southern half of the usual sampling area, the average distance to centre and the area of the contour ellipse will both decrease even when no change in distribution occurs.

Materials and methods

Area occupied

The proportion of empty samples as a function of mean sample sizes when no areas were structurally empty was estimated directly from the relationships in Table 2. To investigate the case where a significant proportion of the samples are taken in structurally empty areas, accuracy and bias of the estimate of p_e was investigated by simulating binomial distributed samples with a probability of 0.25 of being structurally empty (p_e). For the nonstructurally empty samples, number of individuals in each sample was simulated as negatively binomial distributed observations with a given size and mean. For each number of samples (250 and 500), mean (1, 2, …, 10), and k (0.5, 1 and 2), 1000 data sets were simulated, and the mean and the 2.5% and 97.5% quartiles of $\hat{p}_e$ were estimated using maximum likelihood estimation in R (R Development Core Team 2010). The number of samples was initially set at lower values to allow direct comparison with the other measures, but this resulted in estimates of 2.5% and 97.5% quartiles $\hat{p}_e$, which included the full range (0–1), and hence the number of samples had to be increased.

Aggregation

Lloyds index of patchiness

Lloyds index can either be estimated from moment estimates (eq. 2) or from maximum likelihood estimates of k (Lloyds index $= 1 + 1/k$). We used simulations to estimate the bias of the two different estimates at different number of samples (20, 40, ..., 100), mean (0.1, 0.3, ..., 0.9, 1.0, 1.5, ..., 3.0), and k (0.1, 0.2, 0.5, and 1). First, we simulated negative binomial distributed samples at all combinations of number of samples, mean, and k. We then used these data to estimate Lloyds index from moments and maximum likelihood estimates of k. 1000 data sets were generated for each combination and the median, 2.5%, and 97.5% quantiles of the estimates estimated for each method. A number of the simulations resulted in a total number of individuals < 2 in the data set, and the parameters were therefore not estimable. These data sets were discarded. This occurred in a maximum of 30% of all cases.

Lorenz curves

We used simulations to estimate the indices for data from different distributions and parameter combinations. We restricted the analyses to statistical distributions of counts only and used a constant density in all areas and then sampled data assuming a Poisson or negative binomial distribution. A total of 10 000 samples were simulated at each average density. Average density varied between 0.1 and 10, and k varied between 0.1 and 1.

Geographical spread

To investigate the bias and precision of the indices of geographic range, we considered the estimate in one dimension with no loss of generality. The centre of gravity, C, the average distance, d, and the area of the contour ellipse (equivalent in the one-dimensional case to the square root of the average squared distance to centre) estimated from samples can be shown analytically to be unbiased or approximately unbiased, and the geographical spread of the underlying distribution does not affect their coefficient of variation, CV (Appendix A). The effect of varying mean density, standard deviation of the distribution of the underlying spatial distribution, and number of sampling locations on the uncertainty of d and s^2 estimated from samples were investigated using simulations. In these analyses, density was assumed to follow a normal distribution in the interval $I = (-10; 10)$ with a centre of gravity equal to zero. Average density was changed by multiplying densities throughout the interval by a constant. Density $\lambda(x)$ at each location x in the interval was calculated at n equidistant locations, and observations $Y(x_1) \dots Y(x_n)$ were generated assuming $Y(x)$ to be Poisson distributed with mean $\lambda(x)$; d and s^2 were then calculated using the generated observations. This was repeated 1000 times and the standard deviation of the indices estimated. Three values of geographical spread (standard deviation in the underlying normal distribution $= 1, 2,$ and 3) and sampling intensity (number of samples $= 20, 50,$ and 100) were investigated at mean densities ranging from 1 to 10.

Failing to meet assumptions

The distribution of individuals in space is rarely random

Fig. 2. Bimodal distribution in space. Solid line, high density distribution with $D_{95} = 0.74$; short dashed line, low density distribution with $D_{95} = 0.74$; long dashed line, low density distribution with $D_{95} = 0.39$; dash-dotted line, low density distribution with $D_{95} = 0.54$.
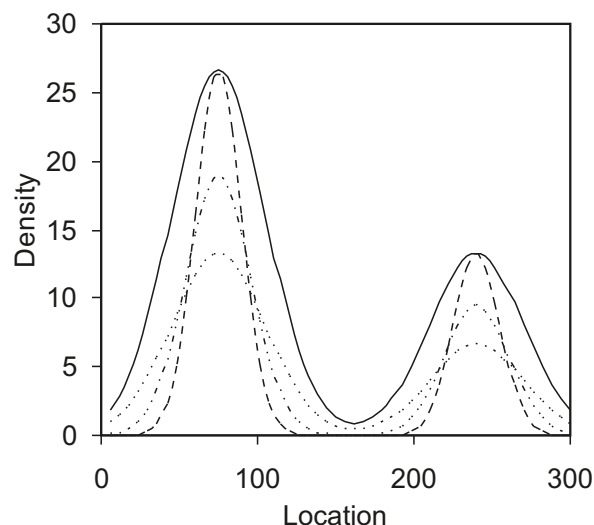
with a common mean in all areas. To investigate the effect of a more ecologically likely spatial distribution, we investigated the ability of Lloyds index, the area of the contour ellipse, D_{95}, and the Gini index to detect changes in aggregation of individuals with a bimodal spatial distribution (Fig. 2). We estimated the indices based on 500 simulated surveys sampling at 300 equidistant points along the distribution. Samples were drawn from a negative binomial distribution with a spatially varying mean described by the bimodal distribution and a common size parameter $k = 0.4$. Two different effects of a decrease in abundance were investigated: proportional decrease in all areas (proportional density model; Hilborn and Walters 1992; Shepherd and Litvak 2004) and disproportional decrease in low abundance areas (constant density and basin model; Hilborn and Walters 1992; Shepherd and Litvak 2004). Furthermore, the effect on the indicators of increasing the level of aggregation of the underlying spatial distribution was investigated. The level of aggregation of the underlying spatial distribution (Fig. 2) was quantified by estimating D_{95} of the underlying (true) distribution. D_{95} is an unbiased index of aggregation in this case, as the distributions are continuous normal distributions unaffected by sampling error (Swain and Sinclair 1994) rather than the number of individuals observed in samples. To be a reliable index, there must be a tight relationship between the index and D_{95}, and index values obtained at other average densities should fall inside the quantiles (corresponding to no falsely detected changes in distribution).

Data example: North Sea cod

The data used to illustrate the problems resulting from using biased indices consisted of survey catch rates of 1-year-old cod derived from catches in the International Council for the Exploration of the Sea (ICES) International Bottom Trawl Survey (ICES International Bottom Trawl Survey Database, February 1983 to 2010, http://www.ices.dk). Haul duration generally varies between 0.5 and 1 h, but to assure that catch rates were comparable, only hauls

lasting between 25 and 35 min were used in this study. The survey follows a stratified-random design with two hauls taken within each ICES statistical rectangle (0.5° latitude × 1° longitude) on average. Catch of 1-year-old cod was estimated by applying an age–length key to the observed catch at length. If the total number of fish caught of all ages is negatively binomial distributed, multiplying this number by a fraction results in a number that is not strictly discretely distributed. In our study, we focused on 1-year-olds, as their length distribution is generally well separated from older fish, and hence the assumption of observing an integer is reasonable. There were, however, cases where a caught fish was of a length with a significant probability of being age 2. In these cases, we rounded the number caught to the nearest integer. We believe that the error introduced by this is small compared with the variation we would introduce by examining the total (integer) number of cod caught of all ages.

From catches of 1-year-old cod, we estimated mean catch rate, Lloyds index, the proportion of empty samples, D_{95}, the Gini index, average distance to centre, and the area of the contour ellipse ($P = 0.40$) for every year. The proportion of structurally empty samples was not estimated because of the low accuracy at low values of k. For the two geographic indices, catch rates were averaged within statistical rectangles prior to estimation to avoid bias due to changes in sampling intensity. Finally, a model of the form $\ln(\text{measure}) = a + b \times \ln(\text{mean catch})$ was fitted, and the statistical test of $b = 0$ was performed.

Results

Area occupied

The intrinsic link between mean sample size and the proportion of nonzero samples diminishes rapidly with increasing mean if individuals are Poisson distributed. At a mean sample size of 5, the proportion of nonzero samples is greater than 0.99 (Table 2, Fig. 3). However, if data are negative binomial distributed, the problem caused by "false" empty samples alone is aggravated (Fig. 3). Hence, conclusions regarding distribution can only be based directly on the occurrence of nonzero samples when the mean number of individuals per sample is large (>4 for Poisson distributed data, >10 for negative binomial distributed data), and the probability of obtaining zero samples by random effects alone is small.

The maximum likelihood method provides unbiased estimates of p_e (bias on average 2% with no trend), which are reasonably precise for mean sample sizes above 5 (Fig. 4). The precision of the method relies on both mean and size parameter of the negative binomial distribution. Hence, distributions with low mean or a size parameter below 1 are unlikely to provide precise estimates of p_e unless the number of samples is very large.

Aggregation

Lloyds index of patchiness

The moment estimate of Lloyds index of patchiness was consistently biased at high index values (>2), even in the 200 samples case (Fig. 5). In contrast, the bias in the maxi-

Fig. 3. Proportion of non-empty samples as a function of mean number of individuals in a sample for Poisson (solid line) and negative binomially distributed observations (broken lines).
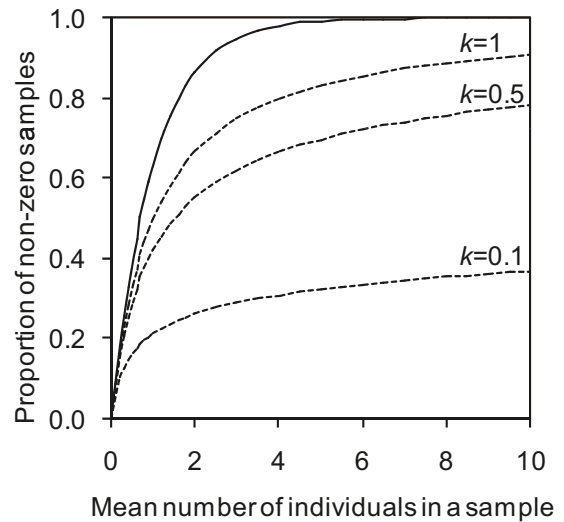
mum likelihood estimate was small when more than 40 samples were taken and the mean exceeded 0.5 (Fig. 5). When the number of samples exceeded 100, the maximum likelihood estimate was approximately unbiased for all means. The variation in both estimates was large when Lloyds index was large or mean was small, but consistently larger for moment estimates than for maximum likelihood estimates (Appendix A). When setting the acceptable upper limit for the CV of the estimate at 20%, none of the two methods are sufficiently accurate to allow the estimation of Lloyds index from less than 60 samples at any value of k (Appendix A). Hence, to estimate Lloyds index, it is a requirement that the number of samples taken is large, and this is increasingly important as either the mean or k decreases. Further, it is preferable to use maximum likelihood estimates of k rather than moments to estimate the index.

Lorenz curves

D_{95} and the Gini index are both highly dependent on the mean at low mean numbers of individuals in a sample (Fig. 6). If the distribution of individuals is known without error, D_{95} or the Gini index can obviously be calculated without bias using eqs. 3 and 4, respectively. However, this is rarely the case, and when observations are subject to sampling variation, indices based on Lorenz curves are systematically biased when the mean number of individuals per sample is below 4 (Fig. 6).

Geographical spread

Though both indices are unbiased, the CV of both d and s^2 increases as average density is decreased (Fig. 7). The effect is greatest at average densities below 5 but levels off at average densities between 5 and 10. As expected, increasing the number of samples taken decreases the CV of the indices, and with a sampling intensity of 100 samples, the CV of d and s^2 remains below 0.08 and 0.15, respectively, even at an average density of 1. The spread of the underlying distribution does not affect the CV of d and s^2 (Fig. 7). However, their precision depends heavily on sampling intensity and

388

Can. J. Fish. Aquat. Sci. Vol. 69, 2012

Fig. 4. Estimated uninhabited proportion of samples from simulated data where the true proportion, p_e, is 0.25. Panels show the number of simulated samples (250: panels a, b, and c; or 500: panel d) and size parameter k of the negative binomial distribution (0.5, panel a; 1, panel b; and 2, panels c and d). Solid line denotes the mean and broken lines are 2.5% and 97.5% quantiles.
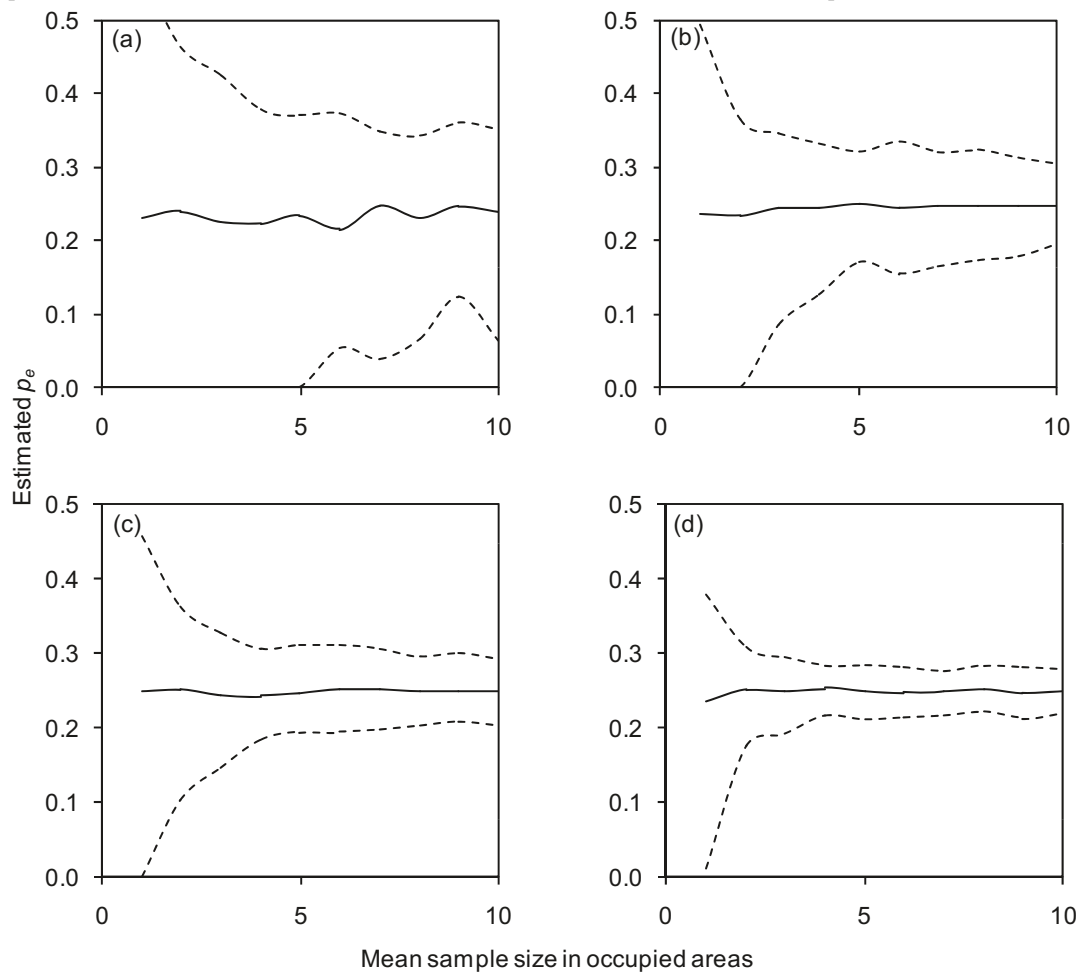
mean density, and the CV increases as average density or sampling intensity is decreased (Fig. 7).

Failing to meet assumptions

In spite of the fact that the data were not derived from a negative binomial distribution, the maximum likelihood estimate of Lloyds index reflected both the direction and level of change in distribution accurately (Fig. 8). There was a tight relationship between Lloyds index and D_{95} of the underlying distribution, indicating that these indices reflect the same aspect of the distribution (Fig. 8). Though both D_{95} and the Gini index produced values with a tight relationship with the underlying distribution, their actual values were far from that of the underlying distribution (Figs. 8b and 8c). Furthermore, varying only density produced indices outside the quantiles though no change in distribution occurred (Fig. 8). The proportional density model ($D_{95} = 0.74$ at different means) produced values of Lloyds index that did not differ significantly, confirming the absence of bias in this measure. The difference between the distributions caused no change in ellipse area, and this indicator did not reflect any of the changes in the bimodal distribution. Hence, only Lloyds index fulfilled both of the requirements to the index, as it reflected distribu-

tional change and was robust to changes in abundance in the absence of changes in distribution.

An overview of the indices investigated and their ability to provide unbiased estimates of the relationship between abundance and distribution and their sensitivity to violations of the assumptions is provided (Table 3).

Example: North Sea cod

The conclusions reached differed widely depending on whether biased or unbiased indices were used (Fig. 9). If the proportion of empty samples or D_{95} was used as the indicator of the distribution, one would draw the (erroneous) conclusion that the area occupied is decreasing as abundance decreases (Figs. 9d and 9f; $P < 0.01$ in both cases). In contrast, there is no significant relationship between abundance and aggregation or between abundance and geographic extent when the unbiased indices are used (Figs. 9a, 9b, and 9c; $P > 0.05$ in all cases). The Gini index showed no significant relationship with the mean (Fig. 9e; $P = 0.3238$) and would thus not have produced a false positive in this case. Though the relationship between the mean of the two was not significant, variation in both Lloyds index and average distance to centre appeared to increase as mean catch decreased.

Fig. 5. Median bias of Lloyds index of patchiness estimated by moment estimates (*a* and *b*) and maximum likelihood *k* (*c* and *d*). Effect of varying the true Lloyds index on bias when the number of samples taken is 40 (*a* and *c*): Lloyds index = 2 (solid), 3 (dash-dotted), 6 (dashed), and 11 (dotted); and effect of varying the number of samples taken when the true Lloyds index is 6 (*b* and *d*): number of samples = 20 (dotted), 40 (thin dashed), 60 (dash-dotted), 80 (thin solid), 100 (thick dashed), and 200 (thick solid).
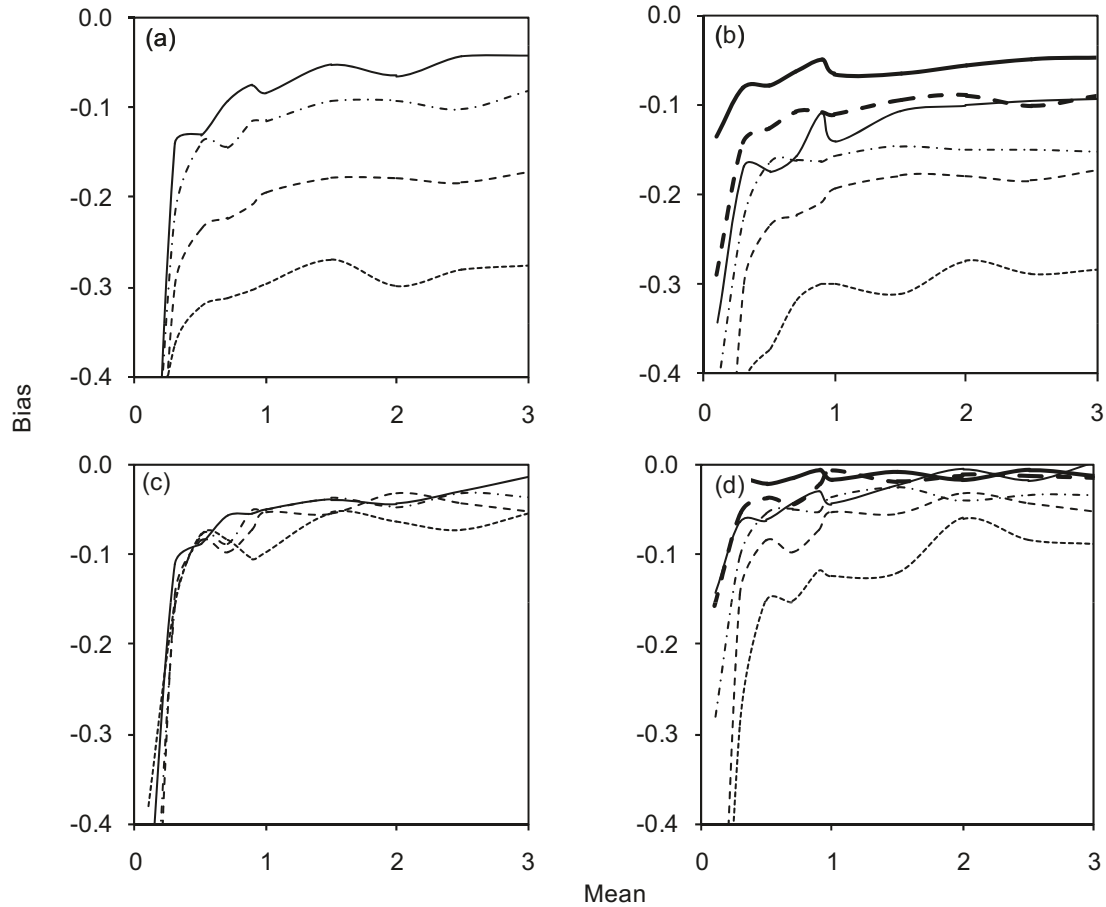
Fig. 6. Gini index (*a*) and D_{95} (*b*) as a function of mean for different error distributions: binomial (sample size = 20: solid line, marked B), Poisson (broken line, marked P), and negative binomial (remaining broken lines) with different values of *k*.
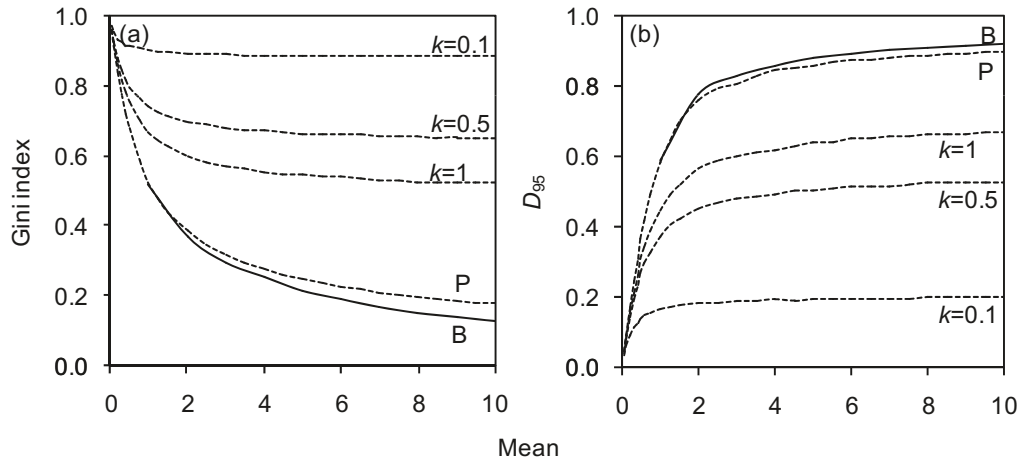

Discussion

We have shown here that a number of frequently used indices of distribution are intrinsically linked to the mean sample size when the mean is small. Using such indices may lead to the conclusion that a change in distribution has occurred when in fact there has only been a change in bias. Apart from the effect that this has on the understanding of the ecology of individual species, it may lead to inefficient conservation of natural populations if conservation management is erroneously based on assumptions of aggregation at low stock size. The problem can be avoided by using the proportion of structurally empty areas, Lloyds index, and the area of the contour ellipse as indices of area occupied, aggregation,

Published by NRC Research Press

390

Can. J. Fish. Aquat. Sci. Vol. 69, 2012

Fig. 7. Investigating the effect of sampling on the bias of the measured mean squared distance to centre s^2. Data include median bias ((estimated index – true index)/true index; middle line) and 2.5% quantile (bottom line) and 97.5% quantile (top line) of bias. (*a*) Effect of spread of the underlying distribution: standard deviation = 1 (solid line), 2 (dashed line), and 3 (dotted). Number of samples = 50 in all cases. (*b*) Effect of number of sampling sites along the transect: 20 (solid), 50 (hatched) and 100 (dotted). Standard deviation of the underlying distribution = 2 in all cases.
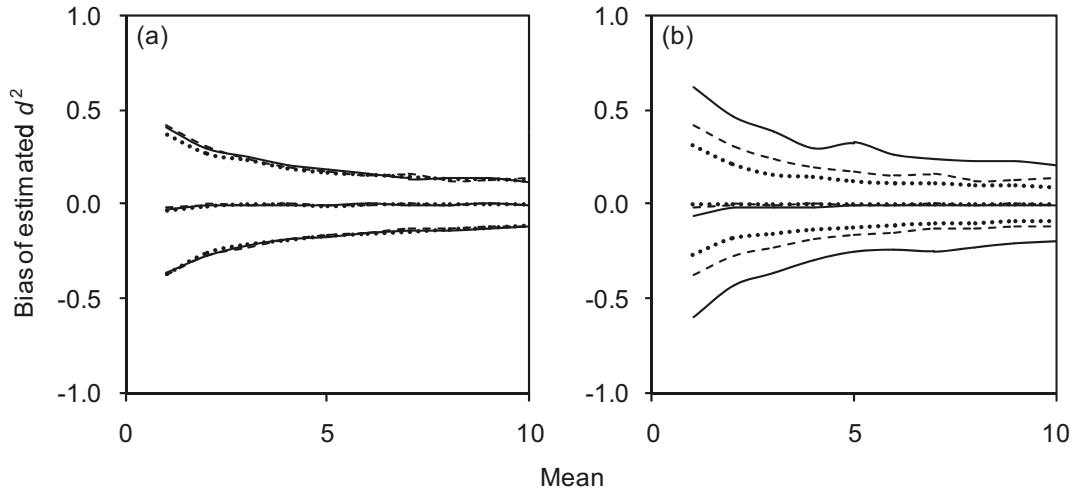
Fig. 8. Relationship between D_{95} of the underlying distribution and Lloyds index (*a*), D_{95} (*b*), Gini index (*c*), and the area of the contour ellipse (*d*) estimated from simulated samples of the population. Diamonds, estimated mean index; dashed lines lines, 2.5% and 97.5% quantiles of the estimated index; solid line, power function describing the relationship between the two; dash-dotted line, values of the underlying (true) distribution (D_{95} and Gini index only). All solid symbols depict indices estimated at an average density of 4.9, while open symbols are derived from average densities varying between 1.2 and 9.9 at a constant D_{95} of 0.74.
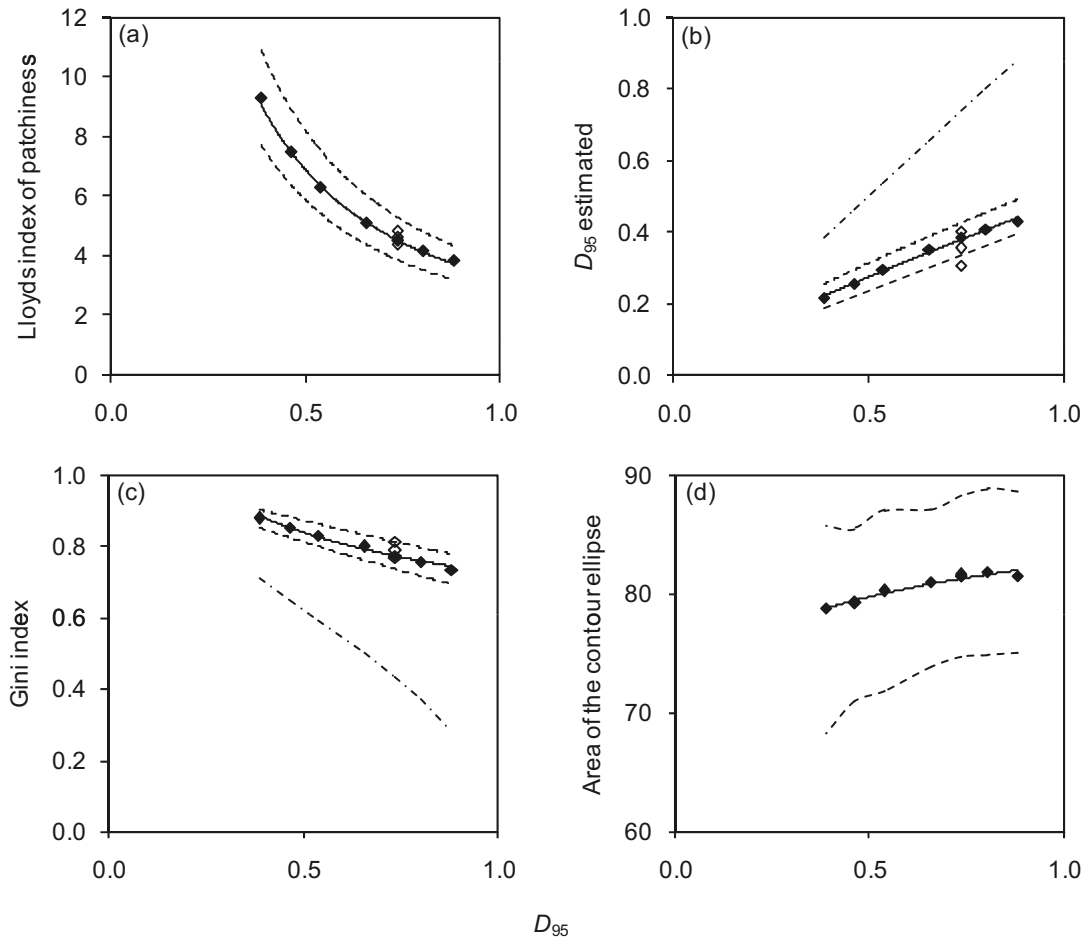
Table 3. Overview of the bias of all measures and requirements of unbiased measures.

Type	Measure	Biased?	Requirements
Occupied area	Proportion of empty samples	Approximately unbiased when mean > 5	
	Proportion of structurally empty samples	No	Large number of samples, mean > 3 and $k > 1$
	Proportion of samples containing more than a fixed number of individuals	Yes	
Aggregation	Lloyds index of patchiness (moment estimate)	Yes	
	Lloyds index of patchiness (maximum likelihood estimate)	Approximately unbiased when number of samples > 100	Poisson or negative binomial distributed data
	Measures based on points or area below the Lorenz curve (D_y, Gini, and A)	Approximately unbiased when mean > 4	
Geographical spread	Average distance to centre of gravity	No but CV large when number of samples < 100	Constant area sampled and unimodal spatial distribution
	Area of the contour ellipse	No but CV large when number of samples < 100	Constant area sampled and unimodal spatial distribution

Fig. 9. Relationship between abundance and distribution of juvenile Atlantic cod in the North Sea using unbiased (*a*, *b*, and *c*) and biased (*d*, *e*, and *f*) measures of distribution.
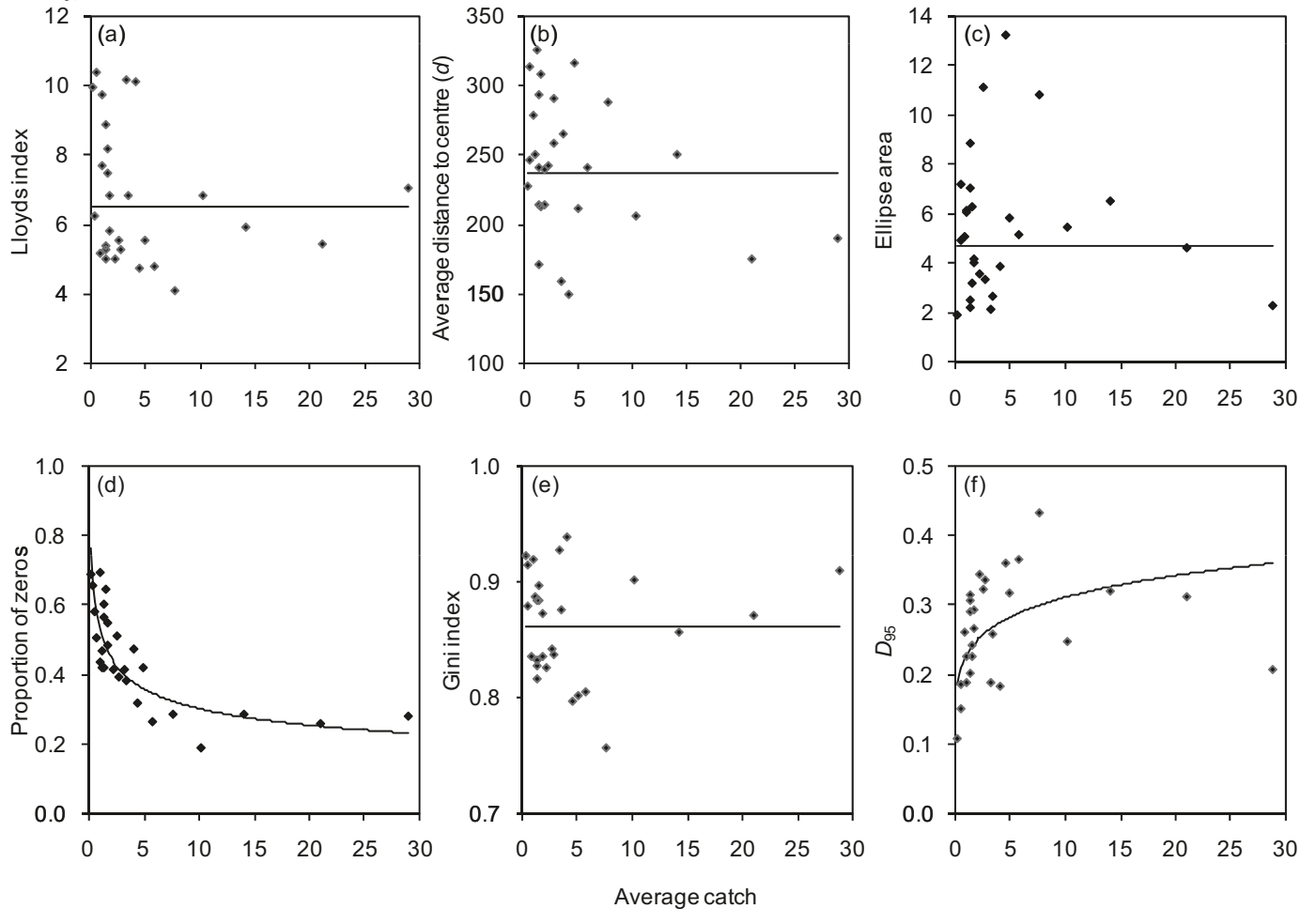
and geographical spread, respectively. However, although these indices are unbiased, they all showed consistent increases in variance at low mean sample size. Furthermore, the proportion of structurally empty areas was effectively impossible to estimate at values of k generally observed in analyses of marine fish (Murawski and Finn 1988; this study) unless the number of samples was very large and the area of

the ellipse only revealed changes in unimodal distributions. Lloyds index was remarkably robust to changes in the underlying distribution and would appear to be the best choice for low-mean data. It furthermore also reflected modest changes in occupied area.

Lloyds index is robust to deviations from a negative binomial distribution. However, two conditions must be fulfilled

392

Can. J. Fish. Aquat. Sci. Vol. 69, 2012

for the estimator to remain unbiased. Firstly, estimating the common form of Lloyds index on non-empty samples only (Hewitt 1981; Stabeno et al. 1996) provides biased results (Bradbury et al. 2003), and eq. 1 or Lloyds index for positive samples (k_+; Lloyd 1967) should be used instead if there are structurally empty samples. Secondly, the observations used must be counts rather than concentrations. Concentrations are often used, in particular when the sampled volume is unequal between samples (McGurk 1986; Stoffels et al. 2003; Bez 2000). However, the derived concentrations are not negative binomial distributed counts, and Lloyds index is therefore no longer independent of the mean (Appendix A). Furthermore, the problems with small counts are retained when numbers are divided by, for example, volume of water filtered. This means that the bias in D_{95} and the Gini index is retained (Appendix A).

To justify the use of the proportion of non-empty samples as a measure of occupied area, two statements are frequently made: The first claim is that it is impossible to distinguish true zeros from false (Hewitt 1981; Hanski et al. 1993; Gaston et al. 1998). It is clear from this and other studies (Lloyd 1967; MacKenzie et al. 2003; Sileshi et al. 2009) that it is indeed possible to estimate the proportion of true zeros in many cases. The second is that for an area to be occupied, there must be at least one individual present in the samples. However, if a species is wide ranging, samples with fewer than one individual will occur by chance when one or more individuals spend only part of their time in the area or when the probability of detecting an individual is less than 1. It is clear that if there are only a few specimens left of the species, these few specimens are unlikely to inhabit the entire area. However, the question is not whether the few remaining specimens occupy all areas but rather whether we would be able to detect a change in their distribution. In the case where we observe one or two specimens, their distribution will provide us with virtually no information of the distribution of the population. Hence, our conclusion in this case must be that the species is rare and that the current distribution cannot be evaluated because of the low number of observations. The result of the general acceptance of the two statements is that the intrinsic correlation between the abundance and proportion of non-empty samples is often ignored (Planque and Fromentin 1996; Blackburn et al. 1998; Freckleton et al. 2006, among others), though some authors have tried to avoid the problem by comparing occurrence with mean number of individuals in occupied samples only (Macpherson 1989; Eycott et al. 2006; Webb et al. 2007). The latter procedure does not eliminate the problem (Wright 1991) though it does decrease the correlation between abundance and occupancy considerably (Wilson 2008, 2011), in some cases leading to negative relationships (Blackburn et al. 2006; Wilson 2011).

Biased distribution indices have been used in studies in a wide number of ecological groups, including parasites (Poulin 1993), plants (Eycott et al. 2006), invertebrates (Foggo et al. 2003), fish (Blanchard et al. 2005), birds (Zuckerberg et al. 2009), and mammals (Holt and Gaston 2003), and it is likely that they have seriously affected the conclusions in a number of cases. For example, Winters and Wheeler (1985) and Saville and Bailey (1980) used the area fished by commercial fishermen as a measure of distribution, a measure

that is likely to be equivalent to the area in which a catch rate higher than some minimum level can be obtained. Crecco and Overholtz (1990) used the proportion of the area where survey catch rate was above a fixed level. However, as the proportion of samples that contain more than a fixed catch is linked directly to mean density, the relationship between abundance and distribution cannot be derived from these studies. Turning to indices based on Lorenz curves, studies of species showing low mean catch rates or greatly skewed data have demonstrated significant relationships between abundance and both D_{95} (Fisher and Frank 2004; Blanchard et al. 2005) and the Gini index (Laurel at al. 2004) of the type shown herein.

The example of cod in the North Sea illustrates the effects of using biased indices to estimate abundance–distribution relationships. Had we used only the proportion of empty samples and D_{95}, we would have concluded that cod concentrated in a few areas as density declined. However, this conclusion is caused by the use of biased indices rather than by changes in distribution as is evident when we use unbiased indices to investigate the abundance–distribution relationship. Even if we had used unbiased indices, we would have been tempted to conclude that though cod did not aggregate, they exhibited a greater variability in aggregation behaviour at low density. However, it is clear from the simulation studies that this type of increase in variance is to be expected from the decreased accuracy of the estimator even in the absence of a change in aggregation. The lack of change in aggregation of juvenile cod as density declines is in accordance with a study of this species in Placentia Bay, which found no significant difference in k (Robichaud and Rose 2006). Similarly, Murawski and Finn (1988) found no significant correlation between average distance to centre and abundance of age 0 and 1+ on Georges Bank. Older cod off the east coast of Canada showed pronounced aggregation as density declined, and the proportions of high catch rates remained virtually constant as abundance declined (Hutchings 1996). Similar aggregative relationships have been found for southern Gulf of St. Lawrence cod age 3 and older (Swain and Sinclair 1994). Determining whether these differences in aggregative behaviour are caused by ontogenetic shifts in behaviour or by differences between stocks will require more detailed analyses of unbiased abundance–distribution relationships.

In conclusion, we recommend that the frequency distribution and mean of the samples is examined in detail before deciding on an unbiased measure of distribution. The proportion of empty samples should not be used to estimate abundance – occupied area relationships in any case. Instead, the proportion of structurally empty samples can be used to estimate empty areas except in cases where the data are highly skewed. In these cases, it is not possible to estimate the proportion of structurally empty samples unless the number of samples is very large (>500), and even in this case, mean sample sizes < 1 are unlikely to provide reliable results. If the mean number of individuals in the samples is >4, D_{95} or the Gini index should provide approximately unbiased results of the relationship between abundance and aggregation. Both indices are furthermore unbiased if applied to smoothed data derived from a spatial model (Swain and Sinclair 1994; Lewy and Kristensen 2009). The average distance to centre, the average squared distance to centre, and the area

of the ellipse should only be used as indices of geographical spread when the spatial distribution is strictly unimodal, as these indices are insensitive to changes in other distributions. For indices of aggregation based on count data, Lloyds index estimated using maximum likelihood methods is the preferable choice, possibly supplemented by the area of the ellipse to indicate geographical spread when the distribution is unimodal. Both indices are unbiased down to very low mean samples sizes and suffer only from increased variation as mean sample size is reduced.

Acknowledgements

References

Anderson, J.T., and Gregory, R.S. 2000. Factors regulating the survival of northern cod (NAFO 2J3KL) during their first 3 years of life. ICES J. Mar. Sci. **57**(2): 349–359. doi:10.1006/jmsc.1999.0530.

Atkinson, D.B., Rose, G.A., Murphy, E.F., and Bishop, C.A. 1997. Distribution changes and abundance of northern cod (*Gadus morhua*), 1981–1993. Can. J. Fish. Aquat. Sci. **54**(S1): 132–138. doi:10.1139/f96-158.

Beverton, R.J.H. 1990. Small marine pelagic fish and the threat of fishing; are they endangered? J. Fish Biol. **37**(Suppl. sA): 5–16. doi:10.1111/j.1095-8649.1990.tb05015.x.

Bez, N. 2000. On the use of Lloyds index of patchiness. Fish. Oceanogr. **9**(4): 372–376. doi:10.1046/j.1365-2419.2000.00148.x.

Blackburn, T.M., Gaston, K.J., Greenwood, J.J.D., and Gregory, R.D. 1998. The anatomy of the interspecific abundance-range size relationship for the British avifauna. II. Temporal dynamics. Ecol. Lett. **1**(1): 47–55. doi:10.1046/j.1461-0248.1998.00005.x.

Blackburn, T.M., Cassey, P., and Gaston, K.J. 2006. Variations on a theme: sources of heterogeneity in the form of the interspecific relationship between abundance and distribution. J. Anim. Ecol. **75**(6): 1426–1439. doi:10.1111/j.1365-2656.2006.01167.x. PMID:17032375.

Blanchard, J.L., Mills, C., Jennings, S., Fox, C.J., Rackham, B.D., Eastwood, P.D., and O'Brien, C.M. 2005. Distribution–abundance relationships for North Sea Atlantic cod (*Gadus morhua*), observation versus theory. Can. J. Fish. Aquat. Sci. **62**(9): 2001–2009. doi:10.1139/f05-109.

Bradbury, I.R., Snelgrove, P.V.R., and Pepin, P. 2003. Passive and active behavioural contributions to patchiness and spatial pattern during the early life history of marine fishes. Mar. Ecol. Prog. Ser. **257**: 233–245. doi:10.3354/meps257233.

Brodie, W.B., Walsh, S.J., and Atkinson, D.B. 1998. The effect of stock abundance on range contraction of yellowtail flounder (*Pleuronectes ferruginea*) on the Grand Bank of Newfoundland in the Northwest Atlantic from 1975 to 1995. J. Sea Res. **39**(1–2): 139–152. doi:10.1016/S1385-1101(97)00056-7.

Brown, J.H. 1984. On the relationship between abundance and distribution of species. Am. Nat. **124**(2): 255–279. doi:10.1086/284267.

Crecco, V., and Overholtz, W.J. 1990. Causes of density-dependent catchability for Georges Bank haddock *Melanogrammus aeglefinus*. Can. J. Fish. Aquat. Sci. **47**(2): 385–394. doi:10.1139/f90-040.

Eycott, A.E., Watkinson, A.R., and Dolman, P.M. 2006. Ecological patterns of plant diversity in a plantation forest managed by clearfelling. J. Appl. Ecol. **43**(6): 1160–1171. doi:10.1111/j.1365-2664.2006.01235.x.

Fisher, J.A.D., and Frank, K.T. 2004. Abundance–distribution relationships and conservation of exploited marine fishes. Mar. Ecol. Prog. Ser. **279**: 201–213. doi:10.3354/meps279201.

Foggo, A., Frost, M.T., and Attrill, M.J. 2003. Abundance–occupancy patterns in British estuarine macroinvertebrates. Mar. Ecol. Prog. Ser. **265**: 297–302. doi:10.3354/meps265297.

Freckleton, R.P., Noble, D., and Webb, T.J. 2006. Distributions of habitat suitability and the abundance–occupancy relationship. Am. Nat. **167**(2): 260–275. doi:10.1086/498655. PMID:16670985.

Gaston, K.J. 1991. How large is a species' geographic range? Oikos, **61**(3): 434–438. doi:10.2307/3545251.

Gaston, K.J., Blackburn, T.M., and Lawton, J.H. 1998. Aggregation and interspecific abundance–occupancy relationships. J. Anim. Ecol. **67**(6): 995–999. doi:10.1046/j.1365-2656.1998.6760995.x.

Hamre, J. 1978. The effect of recent changes in the North Sea mackerel fishery on stock and yield. Rapp. P.-V. Reun. Cons. Int. Explor. Mer, **172**: 197–210.

Hanski, I., Kouki, J., and Halkka, A. 1993. Three explanations of the positive relationships between distribution and abundance of species. *In* Species diversity in ecological communities, historical and geographical perspectives. *Edited by* R. Ricklefs and D. Schluter. University of Chicago Press, Chicago, Ill. pp. 108–116.

Hartley, S. 1998. A positive relationship between local abundance and regional occupancy is almost inevitable (but not all positive relationships are the same). J. Anim. Ecol. **67**(6): 992–994. doi:10.1046/j.1365-2656.1998.6760992.x.

Hewitt, R. 1981. The value of pattern in the distribution of young fish. Rapp. P.-V. Reun. Cons. Int. Explor. Mer, **178**: 229–236.

Hilborn, R., and Walters, C.J. 1992. Quantitative fisheries stock assessment. Chapman and Hall, New York.

Holt, A.R., and Gaston, K.J. 2003. Interspecific abundance–occupancy relationships of British mammals and birds: is it possible to explain the residual variation? Glob. Ecol. Biogeogr. **12**(1): 37–46. doi:10.1046/j.1466-822X.2003.00315.x.

Hutchings, J.A. 1996. Spatial and temporal variation in the density of northern cod and a review of hypotheses for the stock's collapse. Can. J. Fish. Aquat. Sci. **53**(5): 943–962. doi:10.1139/f96-097.

Laurel, B.J., Gregory, R.S., Brown, J.A., Hancock, J.K., and Schneider, D.C. 2004. Behavioural consequences of density-dependent habitat use in juvenile cod *Gadus morhua* and *G. ogac*: the role of movement and aggregation. Mar. Ecol. Prog. Ser. **272**: 257–270. doi:10.3354/meps272257.

Lawton, J.H. 1996. Population abundances, geographic ranges and conservation: 1994 Witherby Lecture. Bird Study, **43**(1): 3–19. doi:10.1080/00063659609460991.

Lewy, P., and Kristensen, K. 2009. Modelling the distribution of fish accounting for spatial correlation and overdispersion. Can. J. Fish. Aquat. Sci. **66**(10): 1809–1820. doi:10.1139/F09-114.

Lloyd, M. 1967. Mean crowding. J. Anim. Ecol. **36**(1): 1–30. doi:10.2307/3012.

MacKenzie, D.I., Nichols, J.D., Hines, J.E., Knutson, M.G., and Franklin, A.B. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. Ecology, **84**(8): 2200–2207. doi:10.1890/02-3090.

Macpherson, E. 1989. Influence of geographical distribution, body size and diet on population density of benthic fishes off Namibia (South West Africa). Mar. Ecol. Prog. Ser. **50**: 295–299. doi:10.3354/meps050295.

Marshall, C.T., and Frank, K.T. 1994. Geographic responses of groundfish to variation in abundance: methods of detection and their interpretation. Can. J. Fish. Aquat. Sci. **51**(4): 808–816. doi:10.1139/f94-079.

394

Can. J. Fish. Aquat. Sci. Vol. 69, 2012

McGurk, M.D. 1986. Natural mortality of marine pelagic fish eggs and larvae: role of spatial patchiness. Mar. Ecol. Prog. Ser. **34**: 227–242. doi:10.3354/meps034227.

Murawski, S.A., and Finn, J.T. 1988. Biological bases for mixed-species fisheries: species co-distribution in relation to environmental and biotic variables. Can. J. Fish. Aquat. Sci. **45**(10): 1720–1735. doi:10.1139/f88-204.

Planque, B., and Fromentin, J.-M. 1996. Calanus and the environment in the eastern North Atlantic. I. Spatial and temporal patterns on *C. finmarchicus* and *C. helgolandicus*. Mar. Ecol. Prog. Ser. **134**: 101–109. doi:10.3354/meps134101.

Poulin, R. 1993. The disparity between observed and uniform distributions: a new look at parasite aggregation. Int. J. Parasitol. **23**(7): 937–944. doi:10.1016/0020-7519(93)90060-C. PMID: 8106186.

R Development Core Team. 2010. R: a language and environment for statistical computing [online]. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available from http://www.R-project.org/.

Rivoirard, J., Simmonds, J., Footem, K.G., Fernandes, P., and Bez, N. 2000. Geostatistics for estimating fish abundance. Blackwell Science, Oxford, UK.

Robichaud, D., and Rose, G.A. 2006. Density-dependent distribution of demersal juvenile Atlantic cod (*Gadus morhua*) in Placentia Bay, Newfoundland. ICES J. Mar. Sci. **63**(4): 766–774. doi:10. 1016/j.icesjms.2005.12.002.

Saville, A., and Bailey, R.S. 1980. The assessment and management of the herring stocks in the North Sea and to the west of Scotland. Rapp. P.-V. Reun. Cons. Int. Explor. Mer, **177**: 112–142.

Shepherd, T.D., and Litvak, M.K. 2004. Density-dependent habitat selection and the ideal free distribution in marine fish spatial dynamics: considerations and cautions. Fish Fish. **5**: 141–152.

Sileshi, G., Hailu, G., and Nyadzi, G.I. 2009. Traditional occupancy–abundance models are inadequate for zero-inflated ecological count data. Ecol. Model. **220**(15): 1764–1775. doi:10.1016/j.ecolmodel.2009.03.024.

Stabeno, P.J., Schumacher, J.D., Bailey, K.M., Brodeur, R.D., and Cokelet, E.D. 1996. Observed patches of walleye pollock eggs and larvae in Shelikof Strait, Alaska: their characteristics, formation and persistence. Fish. Oceanogr. **5**(Suppl. s1): 81–91. doi:10.1111/j.1365-2419.1996.tb00084.x.

Stelzenmüller, V., Ehrich, S., and Zauke, G.-P. 2005. Effects of survey scale and water depth on the assessment of spatial distribution patterns of selected fish in the northern North Sea showing different levels of aggregation. Mar. Biol. Res. **1**(6): 375–387. doi:10.1080/17451000500361009.

Stoffels, R.J., Closs, G.P., and Burns, C.W. 2003. Multiple scales and the relationship between density and spatial aggregation in littoral zone communities. Oikos, **103**(1): 81–92. doi:10.1034/j.1600-0706.2003.12453.x.

Swain, D.P., and Sinclair, A.F. 1994. Fish distribution and catchability: what is the appropriate measure of distribution? Can. J. Fish. Aquat. Sci. **51**(5): 1046–1054. doi:10.1139/f94-104.

Taylor, L.R., Perry, J.N., Woiwod, I.P., and Taylor, R.A.J. 1988. Specificity of the spatial power-law exponent in ecology and agriculture. Nature, **332**(6166): 721–722. doi:10.1038/332721a0.

Webb, T.J., Noble, D., and Freckleton, R.P. 2007. Abundance–occupancy dynamics in a human dominated environment: linking interspecific and intraspecific trends in British farmland and woodland birds. J. Anim. Ecol. **76**(1): 123–134. doi:10.1111/j.1365-2656.2006.01192.x. PMID:17184360.

Wilson, P.D. 2008. The pervasive influence of sampling and methodological artefacts on a macroecological patter: the abundance–occupancy relationship. Glob. Ecol. Biogeogr. **17**(4): 457–464. doi:10.1111/j.1466-8238.2008.00385.x.

Wilson, P.D. 2011. The consequences of using different measures of mean abundance to characterize the abundance–occupancy relationship. Glob. Ecol. Biogeogr. **20**(1): 193–202. doi:10.1111/j.1466-8238.2010.00569.x.

Winters, G.H., and Wheeler, J.P. 1985. Interaction between stock area, stock abundance, and catchability coefficient. Can. J. Fish. Aquat. Sci. **42**(5): 989–998. doi:10.1139/f85-124.

Woillez, M., Poulard, J.-C., Rivoirard, J., Petitgas, P., and Bez, N. 2007. Indices for capturing spatial patterns and their evolution in time, with application to European hake (*Merluccius merluccius*) in the Bay of Biscay. ICES J. Mar. Sci. **64**(3): 537–550. doi:10.1093/icesjms/fsm025.

Wright, D.H. 1991. Correlations between incidence and abundance are expected by chance. J. Biogeogr. **18**(4): 463–466. doi:10.2307/2845487.

Zuckerberg, B., Porter, W.F., and Corwin, K. 2009. The consistency and stability of abundance–occupancy relationships in large-scale population dynamics. J. Anim. Ecol. **78**(1): 172–181. doi:10.1111/j.1365-2656.2008.01463.x. PMID:18717742.

Appendix A

Bias in Lloyds index estimated by moments and maximum likelihood

Figures A1 and A2 show bias and accuracy of estimates of Lloyds index of patchiness based on moments (eq. 2; Fig. A1) or maximum likelihood estimates of k (Fig. A2).

Estimating the area of the contour ellipse and bias of C and d due to sampling error

The area of the contour ellipse A_p containing p percent of the observations can be estimated as

$$A_p = c_p \sqrt{s_x^2 s_y^2 (1 - \rho^2)}$$

where $c_p = \chi^2(2)_p \pi$, and s_x and s_y are the standard deviations in the x and y directions and where ρ is the correlation coefficient.

To compare distribution with abundance, only an index of geographical distribution is needed, and it is of no consequence what probability level is used (the χ^2 value), since this is simply a scaling factor. In the example given below, we have used $c_p = 1$ without lack of generality. The area of the contour ellipse is independent of the mean number of individuals in a sample and of mean in the x and y directions because this is the case for both standard deviations and correlation.

To investigate whether the measures C and d and the area of the contour ellipse are unbiased, we consider the one-dimensional case for which $\lambda(x)$ denotes the intensity of a species distributed on the x axis in a given interval, I. In this case, an unbiased estimate of s is equivalent to an unbiased estimate of the contour ellipse. The true centre of gravity (C_{true}), average distance to centre (d_{true}), and variance (s_{true}^2) are

$$C_{\text{true}} = \frac{\int_I x\lambda(x)dx}{\int_I \lambda(x)dx}$$

$$d_{\text{true}} = \frac{\int_I |x - C|\lambda(x)dx}{\int_I \lambda(x)dx}$$

Fig. A1. Bias and accuracy of estimates of Lloyds index of patchiness based on moments (eq. 2). Data for median bias ((estimated index – true index)/true index; solid line), 2.5% quantile and 97.5% quantile of bias (dashed lines) of 1000 simulations for each combination of number of samples, mean, and k are shown. Horizontal dashed lines indicate 0.4, corresponding roughly to a coefficient of variation (CV) of the index of 20%.
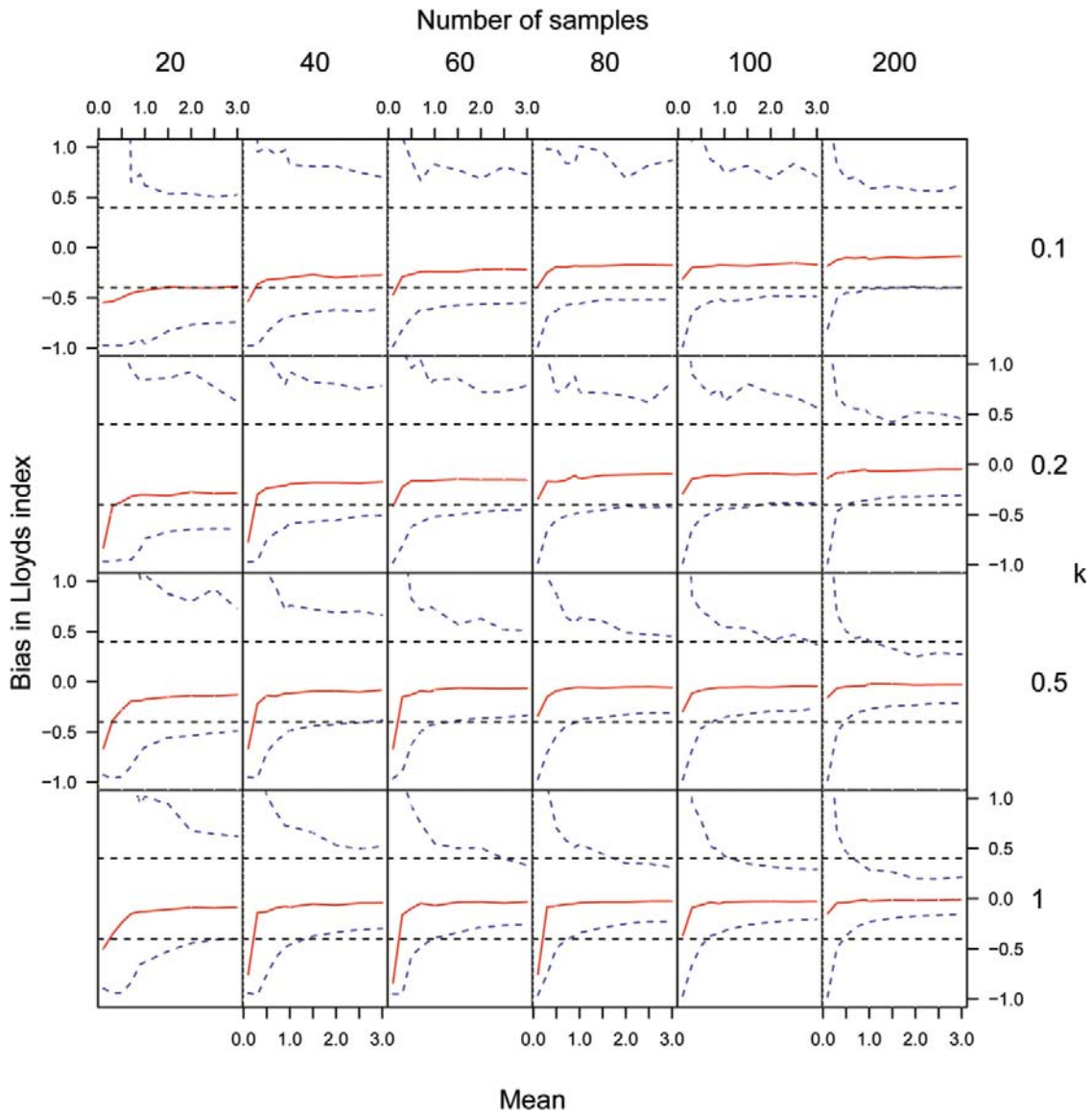
$$s^2_{\text{true}} = \frac{\int_I (x - C)^2 \lambda(x) dx}{\int_I \lambda(x) dx - 1}$$

Now assume that we have observations at N locations and denote the density at location x_i

$$\lambda_i = \lambda(x_i)$$

The centre C, average distance to centre d, and variance s^2 based on the locations x_1–x_N are correspondingly estimated by

$$C = \frac{\sum_i x_i \lambda_i}{\lambda_\bullet}$$

$$d = \frac{\sum_i |x_i - C| \lambda_i}{\lambda_\bullet}$$

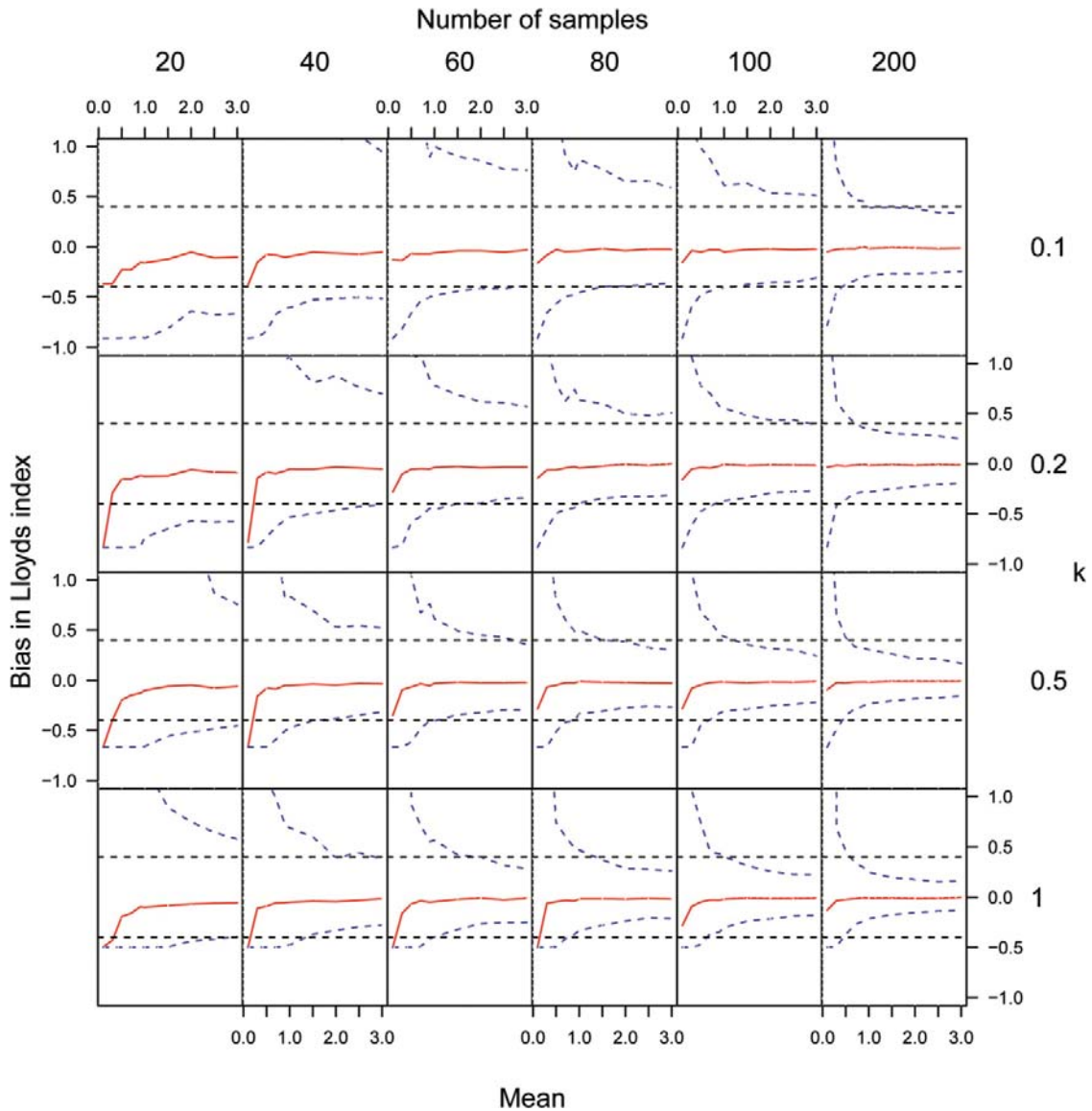$$s^2 = \frac{\sum_i (x_i - C)^2 \lambda_i}{\lambda_\bullet - 1}$$

where $\lambda_\bullet = \sum_i \lambda_i$. We now assume that the observed densities η_i include sampling noise.

The estimated centre, average distance to centre, and average squared distance to centre based on the observations $\eta_1 \ldots \eta_n$ are then

396

Can. J. Fish. Aquat. Sci. Vol. 69, 2012

Fig. A2. Bias and accuracy of estimates of Lloyds index of patchiness based on maximum likelihood estimates of *k*. Data for median bias ((estimated index – true index)/true index; solid line), 2.5% quantile and 97.5% quantile of bias (dashed lines) of 1000 simulations for each combination of number of samples, mean, and *k* are shown. Horizontal dashed lines indicate 0.4, corresponding roughly to a coefficient of variation (CV) of the index of 20%.

$$\widehat{C} = \frac{\sum_i x_i \eta_i}{\eta_\bullet}$$

$$\widehat{d} = \frac{\sum_i |x_i - \widehat{C}| \eta_i}{\eta_\bullet}$$

$$\widehat{s}^2 = \frac{\sum_i (x_i - \widehat{C})^2 \eta_i}{\eta_\bullet - 1}$$

where $\eta_\bullet = \sum_i \eta_i$. We assume that the observations for given densities are independent Poisson distributed variables, i.e., that

$$\eta_i | \lambda_i \sim \text{Poisson}(\lambda_i)$$

and hence $E(\eta_i) = \lambda_i$ for $i = 1, \dots, N$.
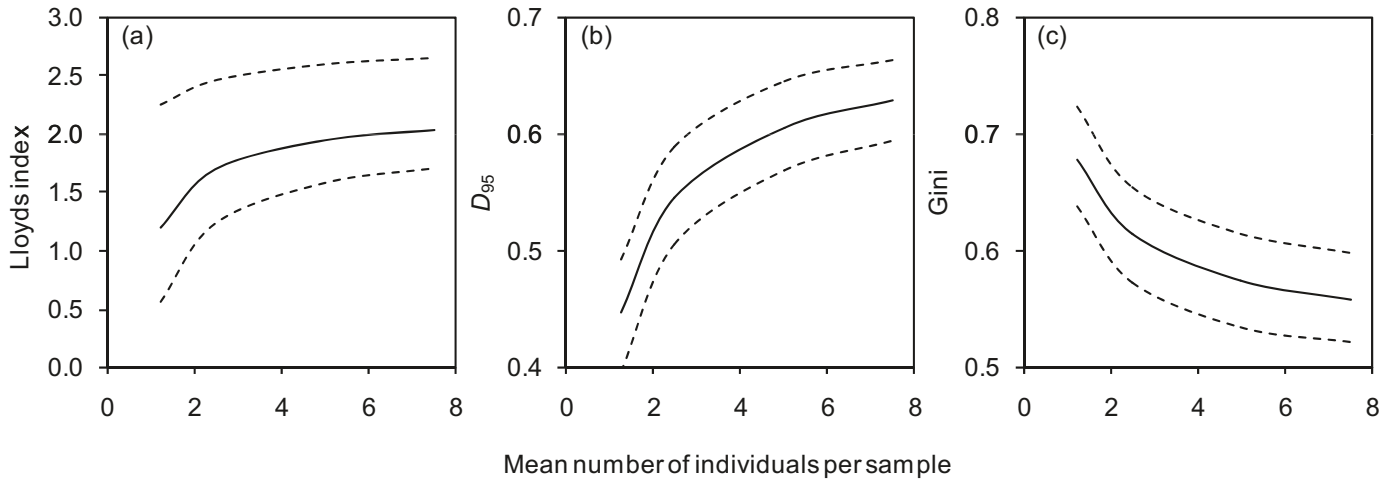
For the Poisson case, it can be shown that $\widehat{C}$, $\widehat{d}$, and $\widehat{s}^2$ are unbiased estimates of C, d, and s, respectively, i.e., $E(\widehat{C}) = C$, $E(\widehat{d}) = d$, and $E(\widehat{s}^2) = s^2$.

Using the fact that the Poisson distributed variables $\eta_1 \dots \eta_n$ for given sum $\sum_i \eta_i = \eta_\bullet$ follows the multinomial distribution with the number of trials equal to $\eta_\bullet$, probabilities $\frac{\lambda_i}{\lambda_\bullet}$, and mean $E(\eta_i | \eta_\bullet) = \eta_\bullet \frac{\lambda_i}{\lambda_\bullet}$, we find that

Fig. A3. Lloyds index (*a*), D_{95} (*b*), and the Gini index (*c*) estimated from concentrations in simulated samples with data derived from a negative binomial distribution with constant size parameter.

$$E(\widehat{C}) = E_{\eta_\bullet}\left[E(\widehat{C}|\eta_\bullet)\right] = E_{\eta_\bullet}\left[\frac{1}{\eta_\bullet}E\left(\sum_i x_i\eta_i|\eta_\bullet\right)\right] = E_{\eta_\bullet}\left(\frac{1}{\eta_\bullet}\sum_i x_i\eta_\bullet\frac{\lambda_i}{\lambda_\bullet}\right) = \frac{\sum_i x_i\lambda_i}{\lambda_\bullet} = C$$

$$E(\widehat{d}) = E_{\eta_\bullet}\left[E(\widehat{d}|\eta_\bullet)\right] = E_{\eta_\bullet}\left[\frac{1}{\eta_\bullet}E\left(\sum_i|x_i - \widehat{C}|\eta_i|\eta_\bullet\right)\right] = E_{\eta_\bullet}\left(\frac{1}{\eta_\bullet}\sum_i|x_i - \widehat{C}|\eta_\bullet\frac{\lambda_i}{\lambda_\bullet}\right) = \frac{\sum_i|x_i - \widehat{C}|\lambda_i}{\lambda_\bullet} = d$$

and

$$E(\widehat{s}^2) = E_{\eta_\bullet}\left[E(\widehat{s}^2|\eta_\bullet)\right] = E_{\eta_\bullet}\left\{\frac{1}{\eta_\bullet - 1}E\left[\sum_i(x_i - \widehat{C})^2\eta_i|\eta_\bullet\right]\right\} = E_{\eta_\bullet}\left[\frac{1}{\eta_\bullet - 1}\sum_i(x_i - \widehat{C})^2\eta_\bullet\frac{\lambda_i}{\lambda_\bullet}\right]$$

$$= \frac{\sum_i(x_i - \widehat{C})^2\lambda_i}{\lambda_\bullet - 1} = s^2$$

where E_η denotes expectation with respect to the distribution function of $\eta_\bullet$.

This proves the statement for the Poisson case. For any other distribution for which $E(\eta_i) = \lambda_i$, the three quantities are at least approximately unbiased, as both the numerators and the denominators of the quantities are unbiased.

As the number of observations tends towards infinity, $E(\widehat{C})$, $E(\widehat{d})$, and $E(\widehat{s}^2)$ tend towards the true values C_{true}, d_{true}, and s^2_{true}, respectively.

Bias when estimating Lloyds index, D_{95}, and the Gini index from concentrations

To demonstrate that the bias in D_{95} and Gini is retained even when transforming count data to concentrations by dividing with an uneven volume and that this also biases Lloyds index, a simulation study was performed to resemble an experiment with 25 samples of 1 L, 50 samples of 2 L, 50 samples of 3 L, and 25 samples of 4 L. The number of individuals in 1 L follows the negative binomial distribution with a fixed mean and size = 0.4. Hence, the number of copepods in 2 L was simulated as the sum of two simulations from a negative binomial distribution with mean *m* and size 0.4 and so forth. Concentrations were then estimated by dividing the simulated number of individuals by the volume of the sample (2 L in this case). After simulating concentration in all 150 samples, Lloyds index (moment estimate), D_{95}, and the Gini index were estimated. The simulation and estimation was performed 500 times, and the mean, 2.5, and 97.5 quantiles of each indicator were estimated (Fig. A3). It is clear that all three indicators show a change in aggregation, though the patchiness of the underlying distribution remains unchanged.