

PRECISION IN SYSTEMATIC TRAWL SURVEYS AS ASSESSED FROM REPLICATE SAMPLING BY PARALLEL TRAWLING OFF NAMIBIA

T. STRØMME* and T. HILENDE†

Following independence of Namibia in 1990, abundance of that country's hake stocks was monitored by trawl surveys conducted by the Norwegian F.R.V. *Dr Fridtjof Nansen* using a systematic survey design. Precision from such designs is considered to be better than with a random design, but it cannot be quantified through standard analysis of variance. In 1998 and 1999, the trawl surveys were duplicated in full using a commercial trawler that operated in parallel with the *Dr Fridtjof Nansen*. Both vessels had the same fishing gear and rigging. In both years, biomass and distribution patterns estimated by the vessels were similar. The paired datasets collected from the surveys were used to analyse variability in the point samples. Absolute differences between pairs of catches were, in general, proportional to the catch, but they varied randomly following a normal distribution around the catch level. Local variability was analysed as random noise, modelled and later reapplied on single vessel data series to evaluate the effect of sample size and replicates on the survey mean. In a survey with 200 stations, local sample noise accounted for about 4% of the variability in the survey mean. Alternatively, running series of simulated surveys by bootstrapping on pairs of catch data gave similar results, when a small systematic vessel effect was adjusted for. The main statistical techniques applied were less susceptible to outlier catches than straightforward correlations or regressions and could therefore, perhaps with some advantage, also be used to estimate the vessel factor when intercalibrating trawl survey vessels.

Keywords: intercalibration, point precision, replicate sampling, trawl surveys

At Independence in 1990, Namibia declared an Exclusive Economic Zone (EEZ) of 200 miles and took exclusive control over one of the biggest hake *Merluccius* spp. resources in the world, which until then had been exploited by an international fleet and managed through the International Commission for the Southeast Atlantic Fisheries (ICSEAF). Prior to Independence Namibia's hake resource had been declining, the annual Total Allowable Catch (TAC) being consistently considerably higher than the catch taken (Sætersdal *et al.* 1999). At Independence, all rights of foreign fleets to fish hake off Namibia were cancelled, the stock being considered to be at a low level and to consist mainly of young fish. A strategy for rebuilding it rapidly was implemented; its main elements were monitoring its state through trawl surveys, limited access through licensing, and recording of the annual catch through landing statistics. From 1997, intensive biological sampling of the catch was implemented through an observer programme.

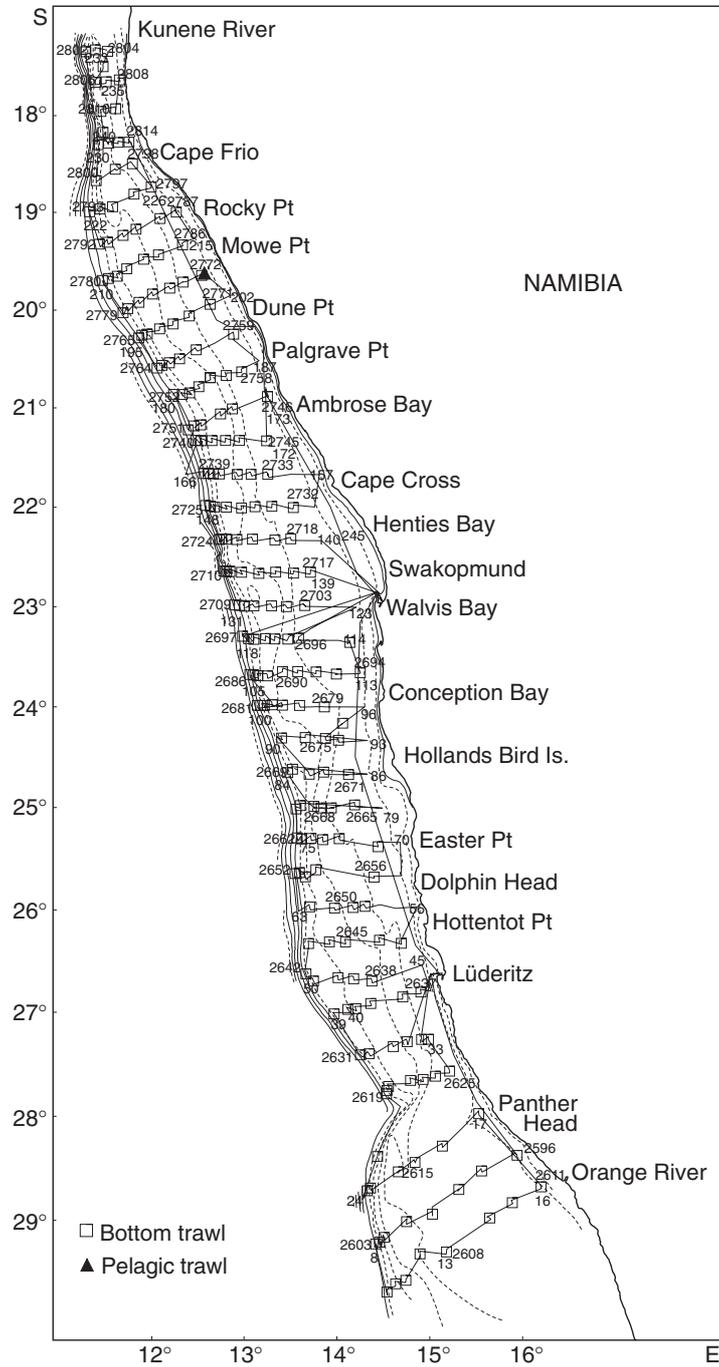
For the period 1990–1997, the annual hake TAC was set at 20% of adult biomass, based on biomass indices obtained from surveys that were treated as absolute estimates. ICSEAF statistics were not considered a reliable data source for stock assessment. From 1998, an interim management procedure was implemented, in which the trend in biomass index

from the survey and the mean catch per unit effort (cpue) from a reference fleet were combined to adjust the TAC up or down from the previous year (Geromont *et al.* 1999). For these reasons, the survey index has been and continues to be a vital element in the management of hake, and there is a need to obtain more information on the precision and bias of the estimates. This paper focuses on the precision in the point samples of the Namibian trawl surveys, between-point precision being analysed separately (Strømme in prep.). Bias in these surveys is the subject of several other studies (Huse *et al.* 1998, 2001, Strømme and Voges in prep.).

In all, 17 trawl surveys for hake were carried out by two Norwegian research vessels, both named *Dr Fridtjof Nansen*, during the respective periods 1990–1993 and 1994–1999. After the first pilot survey in February/March 1990, a systematic survey design with transect sampling was implemented. Transects ran almost perpendicular to the coast and were approximately 20 miles apart. Stations along transects were semi-randomly distributed over the shelf, with stations 10–15 miles apart. The slope (200–700 m) was sampled systematically by 100 m depth interval, with a random starting depth along each transect. Pelagic fish above the trawl were sampled concurrently by means of acoustics. A more detailed description of the sampling methods is given in Sætersdal *et al.* (1999),

* Institute of Marine Research, P.O. Box 1870 Nordnes, N-5817 Bergen, Norway. E-mail: tore.stroemme@imr.no

† Ministry of Fisheries and Marine Resources, National Marine Information and Research Centre, P.O. Box 912, Swakopmund, Namibia



and Figure 1 shows an example of a cruise track and station grid.

Variance of abundance indices in trawl surveys is usually assessed by calculating standard errors (*SEs*) and confidence limits of the mean (Saville 1977, Gunderson 1993). An underlying assumption for these techniques is that the samples are collected at random from the population, an assumption violated in systematic surveys. Generally, systematic surveys are considered to give more precise values of a mean than random surveys, except where there is spatial periodicity in the population and the spatial distribution has the same period as the sampling frequency (Cochran 1977, Hilborn and Walters 1992, Hayek and Buzas 1997). Such a phenomenon is seldom found in wild populations. However, the main weakness in systematic surveys is that the precision cannot be quantified owing to the lack of a reasonable estimate of variance in the population if there is only one systematic sample (Hayek and Buzas 1997).

If there is no spatial structure in an animal population, a systematic grid can be viewed as random with respect to the distribution of the animals; in such cases, standard error and confidence limits can be calculated using the formula for random sampling (Cochran 1977, Hayek and Buzas 1997). In that case, no gain in precision over random sampling is obtained, but the sampling can be more effective. Suggestions have been made to treat systematic surveys as "pseudo-random" surveys, assuming that the individuals in a population randomly occupy a location within the stratum (Saville 1977), then allowing analysis of variance using the *SE* and confidence limits of the mean. Following this rationale, the "confidence interval" for indices of hake abundance derived from the *Dr Fridtjof Nansen* surveys is typically 20–25% (Strømme *et al.* 1998, 1999). However, treating data collected from transect sampling through a gradient, as if they were randomly collected, would overestimate the variance of the mean because the variance in the data also incorporates change through the range in a continuous gradient.

One alternative conceptual approach when sampling a gradient is to consider the sampling through the gradient as stratified sampling, with one sample in each stratum (Cochran 1977), but then the data are insufficient to calculate the variance within each stratum. Cochran (1977) suggests that neighbouring stations be grouped into pairs, thus obtaining measures of local variance, and discusses methods of calculating variance in such cases. Another approach is to take two, or preferably more, repeated samples at each location (stratum), from which estimates of precision can be obtained. Such a technique is called replicate sampling and is much used in social science (Sudman 1976,

Levy and Lemeshow 1999), geology (Davis 1986), palaeontology, and in surveys of terrestrial animal populations (Hayek and Buzas 1997), but it has not yet been applied in fisheries research. Hilborn and Walters (1992) advocate repeated sampling along randomly selected transect or grid starting points in systematic trawl survey design to obtain an estimate of precision to be applied to the survey estimate, but they provide no examples of the method in practice. This follows the same rationale as for replicate sampling, except that Hayek and Buzas (1997) perhaps do not emphasize the relative positioning of the replicates as strongly as Hilborn and Walters (1992). The main reason why replicates are less frequently used in fishery surveys than in other fields of science is probably the high cost of such surveys, which precludes doubling or trebling sampling effort. However, in the case of a series of systematic surveys, where a structure with a continuous variation is sampled, occasional checks on the sampling error may be sufficient (Cochran 1977).

An opportunity to replicate survey results arose in the course of the Namibian hake surveys. In 1997, trawl sampling procedures undertaken by the *Dr Fridtjof Nansen* were successfully copied by Namibian commercial trawlers. This formed part of a programme to develop Namibian capacity for undertaking its own trawl surveys, and led to a decision to transfer responsibility for hake surveys to Namibian scientists on board commercial trawlers after an overlapping period of two years. In 1998 and 1999, the hake surveys were duplicated in full, using *Dr Fridtjof Nansen* and a commercial trawler working in parallel with identical gear and rigging. The performance of the sampling gear was similar and led to almost identical estimates of abundance. The results suggested that the datasets from the two vessels could be treated as an instance of replicate sampling, with two replicates at each station. Analysis of these replicate surveys was expected to shed light on topics such as precision in systematic trawl surveys, the representativeness of a point sample for the population at a location and local distribution patterns of hake. The findings are reported here.

MATERIAL AND METHODS

Standard trawl samples were of 30 minutes duration, but shorter if the substratum was inclement. The distance covered during a trawl was measured as the distance covered over the seabed, estimated with GPS. SCANMAR sensors mounted on the trawl were used to determine the starting point of the haul and to monitor the performance of the sampling gear

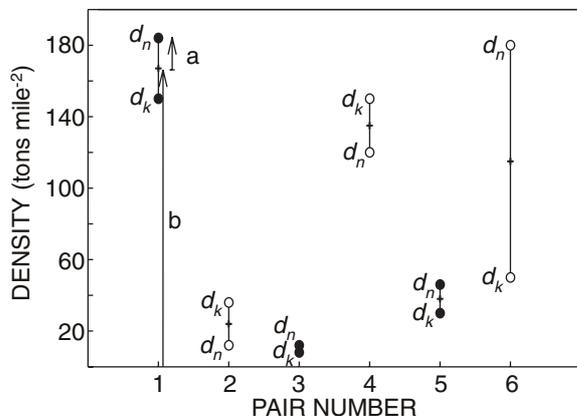


Fig. 2: Examples of density estimates from paired trawl data. The ratio estimator used in the paper is the ratio between a and b in the Figure (d_n – density *Dr Fridtjof Nansen*, d_k – density *Katima*; illustrative values only)

and bottom contact during the haul. A constraining rope was mounted on the warps 100 m in front of the doors in order to keep the door distance constant at all depths, following a procedure suggested by Engås (1994), based on findings by Godø and Engås (1989). The catch was sorted by species, and hake were classified into the three species *Merluccius capensis*, *M. paradoxus* and *M. polli*, and sorted by sex. For all groups, samples were taken to determine length frequency distributions.

All data were stored in the NANSIS database (Strømme 1992), together with a quality control index. Data from both the *Dr Fridtjof Nansen* and the commercial trawler were exported from NANSIS into ASCII tables and analysed by the S-plus statistical package. All hake of both sexes were summed into one catch record for hake and the catch records from the two vessels were paired. Only pairs in which both vessels successfully completed a haul were included in the analysis. Paired hauls where quantities of mud were brought aboard in one or both hauls were excluded, as well as stations in which one or both of the vessels had a zero catch. The last case happened in three instances at the extremities of hake distribution; the catch of the other vessel was very small ($<10 \text{ kg h}^{-1}$) on all three occasions. Pelagic fish observed above the headrope of the trawl were not included in the point estimates, contrary to the standard analysis of the *Dr Fridtjof Nansen* surveys (Sætersdal *et al.* 1999). Totals of 176 and 189 pairs from the 1998 and 1999 surveys respectively were used in the analysis.

In area sampling, the sampling unit is often referred

to as a sample quadrat, irrespective of its shape. For the current study, the sampling quadrat is a rectangle of about 22 by 2 800 m, the area swept by the trawl during a normal 30-minute haul.

Figure 2 shows examples of paired catch data for analysis. Several attempts were made to express similarity of paired stations numerically by means of an index. Exploratory analysis showed that catch rates were highly skewed following a negative binomial distribution (Elliott 1971), with variance/mean ratios in the range 1 490–6 170 for the two vessels in the two years. Also, when grouping catch rates by catch levels, the variance increased with the catch level (heteroscedasticity), so violating assumptions for standard ANOVA and *t*-tests (Sokal and Rohlf 1981). A similar pattern showed up in the pairs: high catch rates seemed to be associated with big differences. It was concluded that the index should preferably be independent of the catch rate, so removing the effects of skewness and heteroscedasticity. To allow summary statistics on the index, the latter should preferably be symmetrical, in the sense that pairs with a larger catch of vessel A should be balanced with pairs with a similarly larger catch of vessel B. A simple ratio between catch rates or density estimates would not perform symmetrically around the 1:1 ratio. A logarithmic transformation of the ratio would be symmetrical and probably remove most of the problem with heteroscedasticity, but means and variances of such a ratio perform differently from similar statistics on non-exponential numbers. Finally, an estimator was chosen that expressed the similarity of a pair of stations by taking the ratio between the deviation from their means and the mean. This is demonstrated as a/b in Figure 2. The ratio is referred to here as a “similarity index” (SI), and it can be calculated directly from the catch data or the density estimates by the formula

$$SI = \frac{d1 - d2}{d1 + d2}$$

where $d1$ and $d2$ are the catch rates or density estimates from the two vessels operating in tandem. After statistical treatment the ratio $d1/d2$ can, if needed, be recalculated from the SI , as follows:

$$\frac{d1}{d2} = \frac{1 + SI}{1 - SI}$$

The index has a continuous performance between -1 and +1 and is symmetrical around 0. Where there is no difference between the results of the vessels the index would be zero, and a negative index would signify that the second vessel had the higher catch rate. An index of +1 or -1 would signify that the catch from respectively the second and the first vessel was zero.

Table I: Results from ANOVA analysis on pairs

Comparison	Degrees of freedom		Mean square		Variance ratio (<i>F</i>)		Cumulative probability	
	1998	1999	1998	1999	1998	1999	1998	1999
Between pairs of replicates	350	376	1 430 980	3 993 429	13.9	16.7	1.0	1.0
Within pairs of replicates	175	188	102 995	239 469				
<i>Dr Fridtjof Nansen</i>								
Between pairs of stations along transects	328	336	1 077 734	2 538 266	1.16	1.44	0.86	0.996
Within pairs of stations along transects	164	168	966 323	1 765 137				
<i>Commercial vessels</i>								
Between pairs of stations along transects	328	336	722 239	1 461 397	1.17	1.47	0.87	0.997
Within pairs of stations along transects	164	168	619 590	989 439				

For convenience, mean *cpue* was used as an index of abundance. This is not strictly correct, because the sampling density was not uniform in the survey area; the slope had a higher sampling density than the shelf. However, the main objective of this paper is to look at the relative precision of biomass estimates, and the error introduced by use of the mean *cpue* in this manner was considered minor and negligible.

An estimate of biomass from replicate sampling in a systematic survey is straightforwardly the mean of two or more replicates:

$$\hat{\mu}_{\text{sys}} = \sum_{i=1}^m \frac{\mu_i}{m}$$

assuming a random start position for each replicate. The sampling variability of the mean can be estimated by

$$\hat{\sigma}_{\hat{\mu}}^2 = \frac{\sum (\hat{\mu}_i - \hat{\mu}_{\text{sys}})^2}{m(m-1)} \times \frac{N-n}{N-1},$$

where *m* is the number of replicates, *N* the population size and *n* the sample size. Thus, for systematic sampling with a random start for the replicates, variance and *SE* can be estimated from the observed data (Hayek and Buzas 1997).

The series of similarity indices for the two years were subsequently normalized by subtracting the index with its mean for each year, so removing the expected systematic vessel effect between pairs of vessels. Thereafter, the two series were linked in a pool of similarity indices. From this pool, the performance of the mean of the index was analysed by a series of bootstraps in which sample size was incremented from *n* = 5 to 200 in steps of 5. This was done in order to investigate the accumulated effect of sample size on the estimated mean *SI*. A model for the similarity index was con-

structed, and this was applied to the original data repeatedly to simulate replicate sampling. The performance of these replicates was compared with the original data.

The robustness of the mean *cpue* from the surveys was tested by combining the data from the two vessels. One of the two samples from each location was selected at random, a sort of bootstrapping from the replicates that is sometimes referred to as “pseudo-replication” (Sudman 1976). The resultant mean *cpue* from a survey was repeatedly calculated in 1 000 runs, for each of the two years independently. This was an attempt to introduce randomly the noise associated with point samples to investigate its aggregated effect on the estimate of the mean.

RESULTS

The results from a simple ANOVA on the catch rates are shown in Table I. The mean square ratios *between pairs/within pairs* were 13.9 and 16.7 for the replicate datasets in 1998 and 1999 respectively, indicating, not unexpectedly, that differences in catch within pairs on average tend to be much less than catch differences between stations. For comparison with the replicate pairs, single vessel datasets were organized in along-transect neighbouring pairs to test if local variance between such pairs would also be significantly lower than the overall variance in the data. The variance ratios were 1.16 and 1.44 for the data from the *Dr Fridtjof Nansen* in 1998 and 1999 respectively (Table I). The ratios for the commercial trawler data were similar, 1.17 and 1.47 for the same two years. The ratio is significant at a 95% level in 1999 but not in 1998. The rather weak signal on similarity between pairs along transects did not support further analysis using replicate techniques on these data combi-

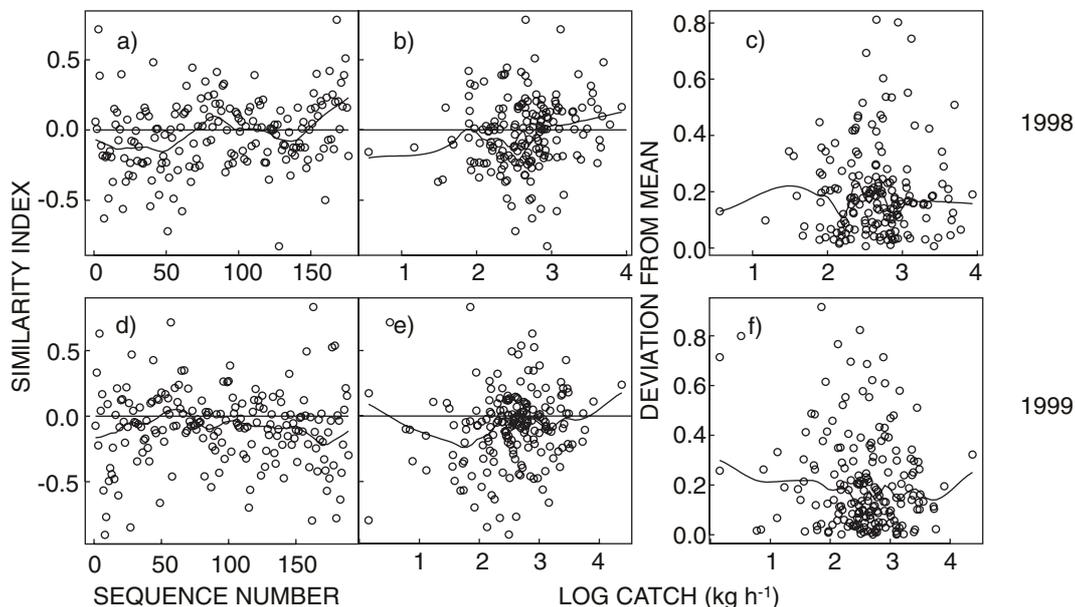


Fig 3: Properties of the similarity index SI derived for 1998 and 1999 – (a, d) performance of the SI during the course of the surveys; (b, e) SI plotted as a function of the log of catch rate; (c, f) absolute deviation of the SI from its mean plotted against the log of catch rate

nations.

Figure 3 shows some performance features of the similarity index for the two surveys. A chronological plot of the index would show if there were systematic differences with time, for instance through training gained in using the research vessel gear by the crew on the commercial trawler or by changing the commercial trawler, which occurred midway through the 1998 survey. Data from 1998 seem to group roughly in two sets, one for the set including the first trawler (until sequence 72) and one for the second vessel thereafter (Fig. 3a). The mean for the first set is -0.105 , with the trend-line consistently below the zero-line. This could point to a systematic effect in which the commercial trawler tended to catch slightly more than the *Dr Fridtjof Nansen*. The second set in 1998 had a mean of 0.028 with a trend-line oscillating around zero, indicating that the vessel did not systematically perform differently from the *Dr Fridtjof Nansen*. For 1999, data are randomly distributed around a line slightly below zero; the mean from that dataset is -0.08 (Fig. 3d). This could again point to a slight bias, i.e. that the commercial vessel was systematically catching slightly more than the *Dr Fridtjof Nansen*. That the difference is not attributable to random factors is again confirmed by the trend-

line, which was generally below the zero-line for the duration of the survey. The SE of the estimated mean of SI was 0.031 , 0.022 and 0.021 for 1998 (up to sequence 72), 1998 (after sequence 72) and 1999 respectively. The respective 95% confidence intervals are ± 0.06 , ± 0.04 and ± 0.04 .

In Figures 3b and 3e the similarity index is plotted against catch, the latter expressed on a logarithmic scale because of its wide range. The trend-line shows that the index increased slightly along with the catches in 1998, but this seems mainly to be a random effect of the paucity of catches at the extremities of the catch distribution and is therefore not real. In 1999, there was no apparent trend. A consistent increasing or decreasing trend could indicate that the absolute difference in catch rates between pairs does not generally increase in proportion to the catch level, as was assumed when the index was formulated. However, it seems the similarity index performed as expected, i.e. independent of the catch, so successfully removing heteroscedasticity in the original catch records.

In Figures 3c and 3f, the absolute deviation of the similarity index from its mean is plotted against catch. The trend-line in this respect would show the average level of the index, irrespective of which vessel had the highest catch rate. On average, the absolute value of

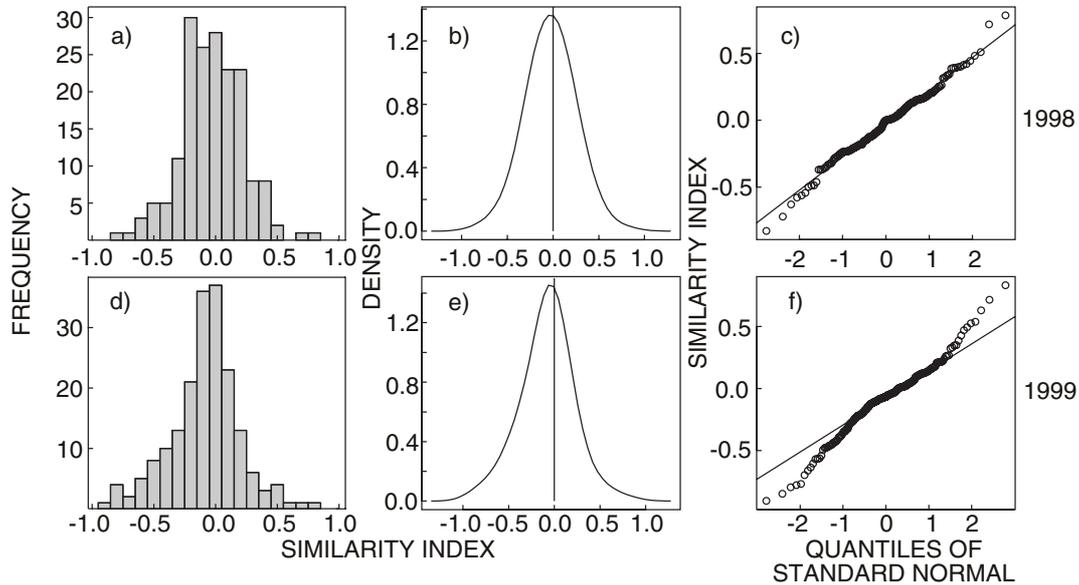


Fig. 4: Properties of the similarity index SI in 1998 and 1999, showing (a, d) frequency distribution of the SI , (b, e) density function of the SI , and (c, f) quantiles of standard normal against quantiles of SI to test for normality

the index was some 0.2 in both years, i.e. that the catch in each vessel was about 20% away from (higher or lower) the mean for the two vessels. The exact arithmetic means of the absolute deviation values were 0.20 and 0.21 for 1998 and 1999 respectively.

Figures 4a and 4d are frequency distributions of the similarity index for the two surveys. Such plots often give a false impression of discontinuity if the class divisions are not grouped symmetrically around population modes (Venables and Ripley 1994), whereas density plots do not suffer such a disadvantage. Figures 4b and 4d show the density plots for the similarity index for the same data. The modes for the distributions were -0.05 and -0.06 in 1998 and 1999 respectively. Figures 4c and 4f show the probability distribution of the similarity index in a qq-plot (Venables and Ripley 1994) for the two years of analysis. For 1998, the index performed close to a normal distribution, whereas in 1999 the curve had a stronger central tendency.

Figure 5a shows the normalized series of similarity indices for both years combined, sorted in increasing sequence. The 365 indices could represent the stochastic noise in the point samples from the two surveys. Overlaid on the series as a line is a series of 365 normally distributed data points, with a mean of 0.0 and a standard deviation (SD) of 0.26, which matches the SI series

almost exactly. This normal distribution becomes a model of the similarity index and is used below.

From the pool of noise data (normalized SIs), random samples with replacement were taken. As stated in the Methods section above, sample size n was increased stepwise from 5 to 200 with a step of 5. In all, 1 000 sets of sample size n were taken and the mean calculated for each. These 1 000 means form a distribution of sample means, from which the mean of the means and the 0.025 and the 0.975 quantiles were extracted. The results are shown in Figure 5b. The empirical upper and lower quantiles correspond to the 95% confidence interval of the mean from the 1 000 samples with sample size n . Figure 5b shows how the confidence interval narrows as sample size increases. For a sample of 200, the 95% interval is ± 0.037 of the mean SI . Analytically, confidence limits can be calculated as the SE times the critical value for a Student's t -distribution ($n = 200$):

$$SE P(|t| \geq 0.95) = \frac{0.26}{\sqrt{200}} 1.96 = 0.036 \quad ,$$

which is fairly close to the empirical figure.

Figures 5c and 5d show the results of inferring artificial noise in the original *Dr Fridtjof Nansen* data. Normalized SIs at all stations in the surveys of 1998

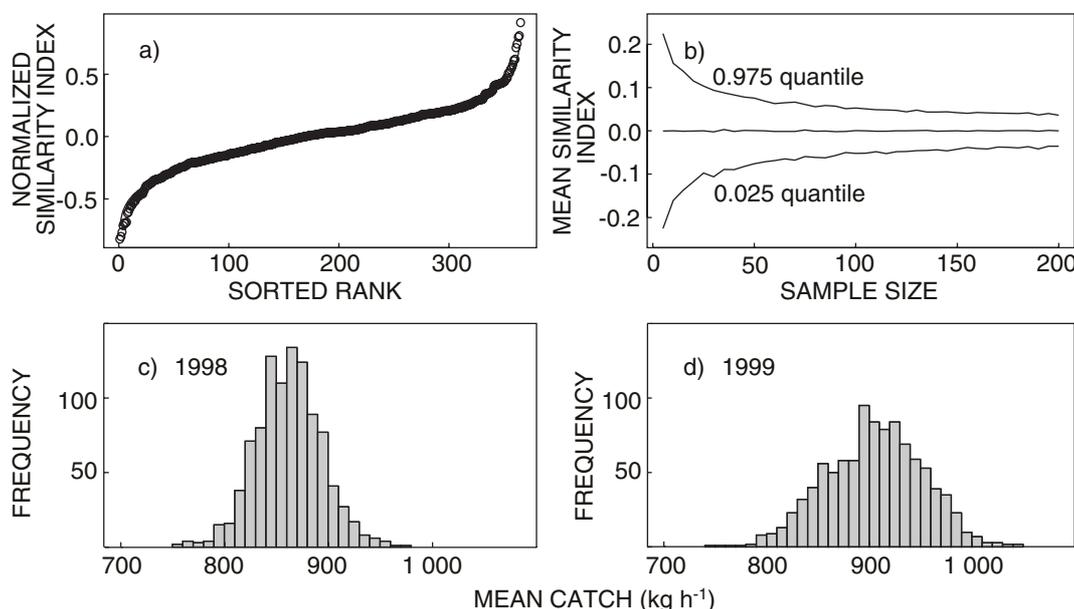


Fig. 5: (a) Battery of normalized similarity indices from both years pooled and sorted with (overlaid) a line showing the results by sampling from a normal distributed model of the SI ; (b) performance of the mean obtained by sampling 1 000 units from the battery with sample size n , which increases from 5 to 200; and distribution of the mean catch from the *Dr Fridtjof Nansen* after applying random noise from the model on the original catch records in 1 000 runs for (c) 1998 and (d) 1999

and 1999 were simulated with random noise generated from the model of SI , and a mean was calculated. The noise was randomly sampled from a normal distribution, with a mean equal to the estimated density at the point and the SD 0.26 times the same density. The sample model then becomes

$$y_i = \hat{\mu}_i + \varepsilon_i, \varepsilon_i \approx N\left(0, (0.26\hat{\mu}_i)^2\right).$$

Sampling was done in 1 000 runs. Figures 5c and 5d show the distributions of the survey means from these runs. The mean of the means is, not unexpectedly, identical to the mean in the original data. The 95% interquartiles are ± 7.4 and $\pm 10.1\%$ of the means for the runs from 1998 and 1999 respectively.

Figure 6 shows the main results from the bootstrap analysis applied to pairs of catch data, where the candidate in each pair was chosen at random (such pseudo-replication is described in the Methods section). Figures 6a and 6b show the distribution of the mean abundance index for 1 000 runs on the data for 1998 and 1999 respectively. In 1998, 950 (95%) of the runs lay within the range 790–880, with a mean of 835. This is equivalent to a confidence limit of $\pm 5.4\%$ of the mean. In 1999, the 95% range was 850–980, which gave a confidence limit of $\pm 7.1\%$. The distri-

bution of the mean in Figure 6b has a bimodal feature, attributable largely to an extraordinarily large catch at one station, where the *Dr Fridtjof Nansen* caught 27 tons and the commercial trawler 17 tons. The higher mode is formed by the 50% random presence in the series of the *Dr Fridtjof Nansen* catch at that station. If this station is “neutralized” by setting the research catch equal to the catch of the commercial trawler (17 tons), the bimodal feature disappears (Fig. 6c). The calculated 95% interquartile distance reduces from ± 7.1 to $\pm 3.7\%$ of the mean. Figures 6d–f show the qq-plots for the three distributions. For 1998, the bootstrapped abundance index performed close to a normal distribution (Fig. 6d), but for 1999, it deviated strongly from a normal distribution (Fig. 6e). The main cause was again clearly the extraordinarily large catch at one station because, when that result was neutralized, the bootstrap performed very close to a normal distribution (Fig. 6f).

The estimated mean catch rates obtained from the two replicates (data for *Dr Fridtjof Nansen* and the commercial trawler) were 836 and 911 kg h^{-1} for 1998 and 1999 respectively. The SE s from the replicates (see Methods section) were ± 27 and $\pm 4.5 \text{ kg h}^{-1}$, equivalent to 2.9 and 0.5% of the mean respectively. Using the SE of the replicate means as a measure of

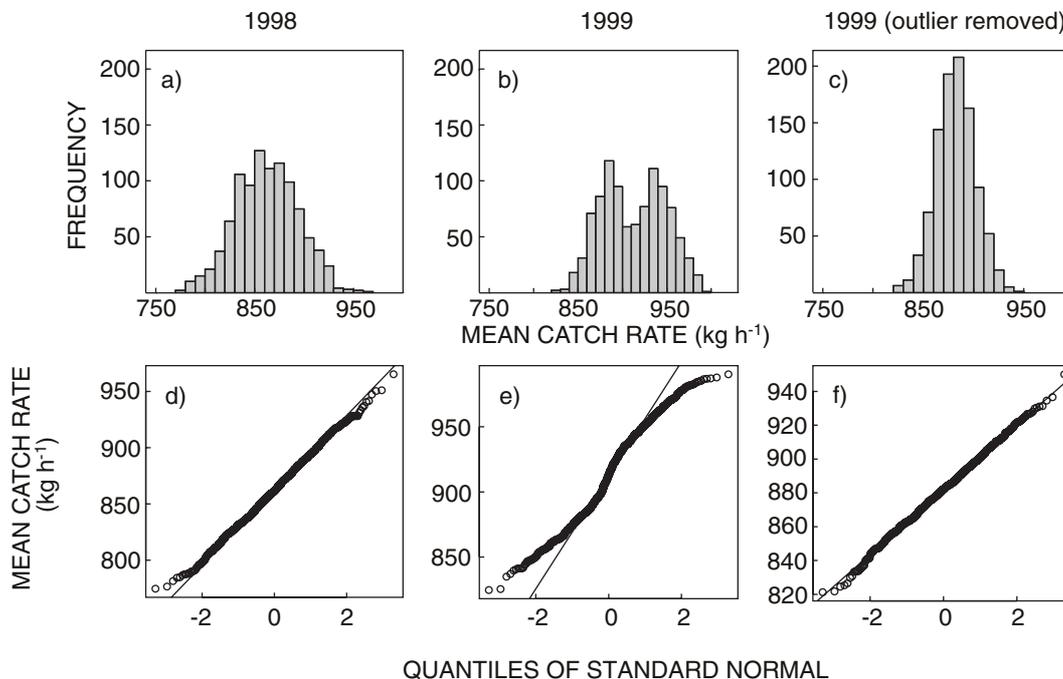


Fig. 6: Results from pseudo-replication by sampling all stations in a survey, but with random representation from the first or the second vessel in 1998, in 1999, and in 1999 with one outlier removed. Distributions of means from 1 000 runs are shown in (a)–(c) and quantiles of the mean compared against the quantiles of standard normal in (d)–(f) to test for normality (qq-test)

precision suffers from the limitation of having only two replicates each year. The low figures for both years indicate a small, but not exact, measure of precision in the point estimates that cannot be transferred to the estimate of the population mean, as discussed below. For comparison, stratified means by 100 m bottom depth were also calculated. The mean catch was 863 and 906 kg h⁻¹ for 1998 and 1999 respectively, with corresponding 95% confidence intervals of 22.9 and 37.1% when a sampling rate proportional to the area was assumed. Corrected for true strata areas and sampling densities, the confidence limits for the stratified mean became 28.9 and 67.4%. A summary of the main statistics of the study, apart from the ANOVA, is given in Table II.

DISCUSSION

The study shows that the chosen similarity index is generally independent of catch rate. The *SI* is also normally distributed around a mean value. The mean *SI* would indicate the presence of a systematic vessel

factor. For two identical vessels, the mean *SI* would be expected to be 0. The *SI* can also be used to estimate vessel calibration constants when intercalibrating survey vessels in trawl surveys. For such studies, regression or covariance techniques are often applied, both of which are highly sensitive to outliers. In contrast, the *SI* can be applied as a transforming factor to “normalize” catch rates by the use of ratios; it also removes density-dependent variance.

The variance of the *SI* would be a measure of the central tendency of the normal curve and would signify the average noise in a series of point observations. To measure the *SI*, replicate sampling is needed at each point, but the empirical study allowed inference of some general properties of point samples that could be used to model the error when, as usual, only single observations are drawn from a point. A point sample can be considered to consist of a signal from the underlying population plus some overlaid noise, the error term, which is distributed randomly around the signal (Biemer and Stokes 1991). The noise level seems proportional to the signal level, but the aggregated effect of the noise on the estimated mean would decrease with increasing sample size, because positive

Table II: A summary of the main statistics in the study undertaken to estimate precision in systematic trawl surveys

Parameter	1998	1999
Number of pairs (<i>n</i>)	176	189
<i>Analysis of catch</i>		
Mean <i>cpue</i> "Dr Fridtjof Nansen" (kg h ⁻¹)	863	906
Mean <i>cpue</i> commercial trawler (kg h ⁻¹)	809	915
Mean <i>cpue</i> from replicates (kg h ⁻¹)	836	911
<i>SE</i> from replicates (kg h ⁻¹)	27	4.5
Mean from density bootstrapping, 1 000 runs (kg h ⁻¹)	836	913
95% interquartiles range from bootstrapping relative to the mean (%)	±5.4	±7.1
95% confidence intervals (%; Strømme <i>et al.</i> 1997, 1998)	±28.2	±23.7
95% confidence interval, stratified by bottom depth (%)	22.9	37.1
<i>Analysis of similarity index</i>		
Mean similarity index (<i>SI</i>)	-0.026	-0.082
Absolute dispersion from mean <i>SI</i>	0.20	0.21
Standard error of <i>SI</i>	0.019	0.021
95% confidence interval for mean <i>SI</i>	±0.037	±0.041
<i>Analysis of noise (both years)</i>		
Sample size (<i>n</i>)	365	
Mean normalized noise level (<i>SI</i>)	0.0	
Standard deviation (<i>SI</i>)	0.26	
Mean from 1 000 bootstrap samples with sample size 200 with replacement (<i>SI</i>)	0.0	
95% interquartile distance from mean from sample size 200 (<i>SI</i>)	±0.037	
<i>Results from noise simulations (1 000 runs)</i>		
Mean <i>cpue</i> Dr Fridtjof Nansen (kg h ⁻¹)	864	905
95% interquartile range (kg h ⁻¹)	800–927	811–994
95% interquartile range relative to mean <i>cpue</i> (%)	±7.4	±10.1

and negative noise signals balance the mean more evenly. The noise pattern is probably related to the local distribution pattern of fish. Random or even patterns of distribution give low noise, whereas patchiness would increase the noise level. Noise is assumed to be species-specific and also to relate to the size of the sample quadrat, the area swept by the trawl in one haul. A long haul would even out local patchiness and tend to yield a lower noise level, whereas short hauls would reflect local patchiness and yield greater noise in the point sample. The average noise level would be expected to change if the local distribution pattern of the fish changed. In the present study on hake, the average absolute *SI* level was some 0.20 in both years (Figs 3c, f). On average, therefore, a catch deviated from the expected density by 20%. From the stable performance, it can be concluded that hake have a stable distribution pattern as regards patchiness/

non-patchiness throughout the survey area during the survey season, generally January/February.

Random measurement errors would also contribute to the noise level. Inaccurate recordings of a haul's actual duration and distance covered, as well as variable bottom contact during the haul, would give errors in the sample obtained. When subsampling the catch on deck, imprecise subdivision of the catch would also generate random errors. During the surveys, effort was put into minimizing these factors, but it could never be expected that they would be zero. In a study by Hjellvik *et al.* (in prep.), patchiness in fish distribution was suggested as a source of measurement error.

Noise in the data collected can be simulated by a simple model in which each sample is considered to be noise randomly sampled from a normal distribution, with a mean equal to the density sampled at the point, and a *SD* equal to 0.26 times the density at the point. If it is assumed that the local distribution pattern of hake was the same in 1998 and 1999 as in earlier years, the model can be used to estimate the error derived from local noise in the surveys prior to 1998. It can also be applied to future surveys, making the same assumption.

As already mentioned, the size of the sample area would probably be important for the error term or the noise level. A small sample quadrat is more sensitive to local patchiness, whereas small-scale patchiness in a large quadrat is evened out. A typical trawl haul of 30 minutes covers a seabed rectangle of some 2 800 by 22 m, a length/width ratio of 140:1. Such a sample, almost like an integral sample in fisheries acoustics, probably evens out local patchiness, but it would be sensitive to variability on the mesoscale. Pennington and Vølstad (1994) recommend shorter hauls than the 30 minutes that seem to have become international standard. Shorter hauls would yield the same catch rates with less effort, allowing more samples to be collected during a survey and a resultant increased precision. The present study indicates that the precision of a single sample, i.e. to what extent it reflects the density at the location, is closely related to the size of the sample unit. Therefore, gain in overall standard error attributable to extra hauls must be balanced against loss in precision from small samples generating greater local variability.

The variation in the similarity index indicates that, on average, the catch of hake deviates by 20% from the expected mean catch in an area, i.e. that the average error term is about 20% of the expected catch rate. With a stochastic noise pattern, positive and negative errors tend to balance as more samples are taken. For increasing sample size, the accumulated error term should therefore approach zero, if it is stochastic. A bootstrap on the similarity index shows that 95% of

the bootstrapped means lay within ± 0.037 *SI* units from an estimated mean when the sample (survey) size is 200 stations. Transferred to a survey using one vessel, one would expect the 95% confidence limits of the mean *cpue*, derived from local error, to be $\pm 3.7\%$ under the same conditions.

The results from the bootstrap on pairs of catch data show that, in 95% of cases, the biomass index is within 5–7% of the mean *cpue* index (Table I). These figures include a vessel effect, which is probably the reason why they are 1–3% higher than the simulation studies on the *SI* index. The relatively good precision of the mean of the point samples is also reflected in the ANOVA (Table I). The high values of *F*, 13.9 and 16.7, confirm that the local sampling error contributes in only a small degree to the overall variance. Stratifying the data by 100 m bottom depth, confidence intervals of 22.9 and 37.1% of the mean were obtained for 1998 and 1999 respectively (Table I). In the cruise reports from the surveys, the fish aggregations were contoured according to density level and a “confidence limit” was calculated from these post-stratified “strata”. As post-stratifying on observed values reduces variance subjectively, the method underestimates the standard error compared to a pre-stratified sample based on prior criteria. Consequently, it is not a recommended procedure (Hayek and Buzas 1997). It has nevertheless been used in the survey reports as a rough indicator. The “confidence limits” from post-stratification are 28.2 and 23.7% for the two years (Strømme *et al.* 1998, 1999), a performance slightly better than the “random stratified” approach (Table I).

The above calculations are approximations because they are based on the assumptions that the data are collected by random and not, as in this case, with a systematic grid with transects across gradients of fish densities. Cape hake have a depth-dependent distribution (Gordoa *et al.* 1995, Sætersdal *et al.* 1999). As a fish cohort grows in body weight, but is reduced in number through fishing and by natural causes, it generally moves down the slope. Deep-water Cape hake *M. paradoxus* are found, as their common name signifies, deeper than shallow-water Cape hake, but the general trend is the same for both species.

If systematically sampled catch data are analysed as if they were randomly collected, the variance would also incorporate the differences attributable to the trend that would be interpreted as random variation. If the present point samples would pick up an underlying and strong trend signal, the mean square (*MS*) from neighbouring pairs along the transects should be considerably lower than the *MS* between all points. Neighbouring points could then be treated as replicates that could provide estimates on the overall mean, following the same methods used in this study for

true replicates, i.e. using *SI*s and bootstrapping. As seen in Table I, the *MS* from neighbouring pairs is only slightly lower than the *MS* between all pairs.

Three main sources of error that influence the precision of a biomass estimate of a surveyed population are suggested:

- (i) random measurement errors associated with the sampling procedure;
- (ii) local sampling errors associated with mesoscale patchiness;
- (iii) errors associated with extrapolating from point samples to the whole population.

The results in this paper have demonstrated that, in the case of hake off Namibia, the combined stochastic errors from sources (a) and (b), when aggregated over a number of stations, are small. In other words, the aggregated mean density index from the set of locations sampled is highly representative for those locations, provided the sample size is sufficient. A typical trawl survey in Namibia consists of some 220 trawl stations, of which about 180–190 have catches of hake. This number of stations should reduce local sample noise to <4%. However, the methods used in this study do not clarify the third source of error. The question whether selected locations properly model trends between observed points remains. Generally, systematic surveys are considered to track or map the underlying population better than random surveys when animals form aggregations or gradients and the survey track is laid across the gradient of the distribution (Cochran 1977, Hayek and Buzas 1997). These conditions are satisfied for hake off Namibia and the surveys applied to them. Quantifying the error associated with interpolating between point samples to the intervening non-sampled areas can be achieved through simulation by sampling from a virtual population with a defined structure. This is the subject of a parallel study (Strømme in prep.).

In a study on a series of sets of intercalibration hauls carried out at the start of bottom trawl surveys in the Barents Sea, Hjellvik *et al.* (in prep.) found that the measurement error at a sampling location accounted for about 2–5% of the total variance in log-transformed data, and they concluded that “the cod catch rate is a fairly precise measure of a fish density at a given site at a given time”. The present study confirms that this conclusion is also valid for hake off Namibia.

More use can be made of techniques associated with replicate sampling in fisheries research. The main reasons why such sampling has been little applied in fishery surveys are the associated high costs and intensive work required at sea. However, replicates can be obtained, as in this study, by repeated sampling at the

same location, or alternatively by dividing the original sample into subsamples of equal size. The latter method requires that the original structure in the sample be retained before subsampling. It is used frequently when dealing with palaeontological or geological samples (Davis 1986). During trawl surveys, the catch would have to be subsampled during the main sampling operation to retain its structure. This can be accomplished by the use of several codends that can be changed during trawling, so splitting the sample into subunits. Such a procedure would result in a smaller quadrat, but it would provide replicates. The multisampler (Engås *et al.* 1997) recently developed at the Institute of Marine Research, Norway, has made such a procedure possible.

ACKNOWLEDGEMENTS

We acknowledge with thanks the constructive reviews and valuable comments of Mr I. Hampton and an anonymous reviewer, the checking of the mathematical notations by Mr V. Hjellvik (Institute of Marine Research, Bergen, Norway), and the grammatical help of the guest editor, Dr A. I. L. Payne (CEFAS, Lowestoft, UK).

LITERATURE CITED

- BIEMER, P. and S. L. STOKES 1991 — Approaches to the modeling of measurement error. In *Measurement Errors in Surveys*. Biemer, P., Groves, R. M., Lyberg, L. E., Mathiowetz, N. A. and S. Sudman (Eds). New York; Wiley: 487–516.
- COCHRAN, W. G. 1977 — *Sampling Techniques*. New York; Wiley: 428 pp.
- DAVIS, J. C. 1986 — *Statistics and Data Analysis in Geology*. New York; Wiley: 646 pp.
- ELLIOTT, J. M. 1971 — Some methods for the statistical analysis of samples of benthic invertebrates. *Scient. Publ. Freshwat. Biol. Ass. UK* **25**: 144 pp.
- ENGÅS, A. 1994 — The effects of trawl performance and fish behaviour on the catching efficiency of demersal sampling trawls. In *Marine Fish Behaviour in Capture and Abundance Estimation*. Fernö, A. and S. Olsen (Eds). London; Fishing News Books: 45–68.
- ENGÅS, A., SKEIDE, R. and C. W. WEST 1997 — The 'multisampler': a system for remotely opening and closing multiple codends on a sampling trawl. *Fish. Res.* **29**: 295–298.
- GEROMONT, H. F., DE OLIVEIRA, J. A. A., JOHNSTON, S. J. and C. L. CUNNINGHAM 1999 — Development and application of management procedures for fisheries in southern Africa. *ICES J. mar. Sci.* **56**: 952–966.
- GODØ, O. R. and A. ENGÅS 1989 — Swept area variation with depth and its influence on abundance indices from trawl surveys. *J. Northw. Atl. Fish. Sci.* **9**(2): 133–139.
- GORDOA, A., MACPHERSON, E. and M-P. OLIVAR 1995 — Biology and fisheries of Namibian hakes (*M. paradoxus* and *M. capensis*). In *Hake: Biology, Fisheries and Markets*. Alheit, J. and T. J. Pitcher (Eds). London; Chapman & Hall: 49–88.
- GUNDERSON, D. R. 1993 — *Surveys of Fisheries Resources*. New York; Wiley: 248 pp.
- HAYEK, L-A. C. and M. A. BUZAS 1997 — *Surveying Natural Populations*. New York; Colombia University Press: 563 pp.
- HILBORN, R. C. and J. WALTERS 1992 — *Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty*. New York; Chapman & Hall: 570 pp.
- HJELLLVIK, V., GODØ, O. R. and D. TJØSTHEIM (in preparation) — The measurement error of marine survey catches: the bottom trawl case.
- HUSE, I., HAMUKUAYA, H., BOYER, D. C., MALAN, P. E. and T. STRØMME 1998 — The diurnal vertical dynamics of Cape hake and their potential prey. In *Benguela Dynamics: Impacts of Variation on Shelf-Sea Environments and their Living Resources*. Pillar, S. C., Moloney, C. L., Payne, A. I. L. and F. A. Shillington (Eds). *S. Afr. J. mar. Sci.* **19**: 365–376.
- HUSE, I., ILENDE, T. and T. STRØMME 2001 — Towards a catchability constant for trawl surveys of Namibian hake. In *A Decade of Namibian Fisheries Science*. Payne, A. I. L., Pillar, S. C. and R. J. M. Crawford (Eds). *S. Afr. J. mar. Sci.* **23**: 375–383.
- LEVY, P. S. and S. LEMESHOW 1999 — *Sampling of Populations: Methods and Applications*, 3rd ed. New York; Wiley: 525 pp.
- PENNINGTON, M. and J. H. VØLSTAD 1994 — Assessing the effect of intra-haul correlation and variable density on estimates of population characteristics from marine surveys. *Biometrics* **50**: 725–732.
- SAVILLE, A. (Ed.) 1977 — Survey methods of appraising fishery resources. *F.A.O. Fish. tech. Pap.* **171**: 76 pp.
- SOKAL, R. R. and F. J. ROHLF 1981 — *Biometry. The Principles and Practice of Statistics in Biological Research*, 2nd ed. New York; Freeman: 859 pp.
- STRØMME, T. 1992 — NAN-SIS: Software for fishery survey data logging and analysis. Users manual. *F.A.O. Computerized Information Series (Fisheries)* **4**: 103 pp.
- STRØMME, T. (in preparation) — Precision in biomass estimates from alternative trawl survey designs applied on a computer simulated demersal stock with a spatial structure.
- STRØMME, T., HAMUKUAYA, H., HUSE, I., VOGES, L. and L. MAARTENS 1998 — Surveys of the fish resources of Namibia: survey of the hake stocks. Cruise report 1/98. Reports on surveys with the R.V. *Dr Fridtjof Nansen*. Ministry of Fisheries and Marine Resources, Namibia: 89 pp. (mimeo).
- STRØMME, T., HUSE, I., and B. STAALESEN 1999 — Surveys of the fish resources of Namibia: survey of the hake stocks. Cruise report No. 1/99. Reports on surveys with the R.V. *Dr Fridtjof Nansen*. Ministry of Fisheries and Marine Resources, Namibia: 93 pp. (mimeo).
- STRØMME, T. and L. VOGES (in preparation) — Estimating the catchability coefficient in trawl survey time-series, by tuning with annual catch data.
- SUDMAN, S. 1976 — *Applied Sampling*. New York; Academic Press: 245 pp.
- SÆTERSDAL, G., BIANCHI, G., STRØMME, T. and S. C. VENEMA 1999 — The *Dr Fridtjof Nansen* Programme 1975–1993. Investigations of fishery resources in developing countries. History of the programme and review of results. *F.A.O. Fish. tech. Pap.* **391**: 434 pp.
- VENABLES, W. N. and B. D. RIPLEY 1994 — *Modern Applied Statistics with S-Plus*. New York; Springer: 642 pp.