Characterization and classification of acoustically detected fish spatial distributions

Julian M. Burgos and John K. Horne

Burgos, J. M., and Horne, J. K. 2008. Characterization and classification of acoustically detected fish spatial distributions. – ICES Journal of Marine Science, 65: 1235–1247.

High-resolution, two-dimensional measurements of aquatic-organism density are collected routinely during echo integration trawl surveys. School-detection algorithms are commonly used to describe and analyse spatial distributions of pelagic and semi-pelagic organisms observed in echograms. This approach is appropriate for species that form well-defined schools, but is limited when used for species that form demersal layers or diffuse pelagic shoals. As an alternative to metrics obtained from school-detection algorithms, we used landscape indices to quantify and characterize spatial heterogeneity in density distributions of walleye pollock (*Theragra chalcogramma*). Survey transects were divided into segments of equal length and echo integrated at a resolution of 20 m (horizontal) and 1 m (vertical). A series of 20 landscape metrics was calculated in each segment to measure occupancy, patchiness, size distribution of patches, distances among patches, acoustic density, and vertical location and dispersion. Factor analysis indicated that the metric set could be reduced to four factors: spatial occupancy, aggregation, packing density, and vertical distribution. Cluster analysis was used to develop a 12-category classification typology for distribution patterns. Visual inspection revealed that spatial patterns of segments assigned to each type were consistent, but that there was considerable overlap among types.

Keywords: aggregations, echotrace classification, landscape indices, spatial pattern, walleye pollock.

Received 26 July 2007; accepted 9 April 2008; advance access publication 27 May 2008.

J. M. Burgos and J. K. Horne: School of Fishery and Aquatic Science, University of Washington, Box 35520, Seattle, WA 98355, USA. Correspondence to J. M. Burgos: tel: +1 206 221 6864; fax: +1 206 685 7471; e-mail: jmburgos@u.washington.edu.

Introduction

The use of scientific echosounders allows the collection of high-resolution observations (e.g. 10 m in the horizontal direction, 0.2 m in the vertical direction) on density distributions of aquatic organisms in the water column. When acoustic data are displayed in echograms, aggregations (i.e. areas of the echogram with multiple, non-resolved organisms) are evident features. Aggregations display a high diversity of spatial patterns, including schools, shoals, pelagic and benthic layers, and diffuse "clouds" (Scalabrin *et al.*, 1994; Petitgas and Levenez, 1996; Reid *et al.*, 2000). A common methodology used to quantify these patterns and to extract information over a range of spatial scales (hundreds of metres to tens of kilometres) is echotrace classification (ETC; Reid *et al.*, 2000). ETC identifies aggregations in echograms and characterizes them using descriptive metrics. ICES (2000) reviews the methodology used in ETC and documents case studies. Usually, ETC uses school-detection algorithms based on image-processing techniques that detect and provide quantitative descriptors on aggregations, including size, total acoustic density, and location in the water column (see Burgos and Horne, 2007, and references therein). The description of aggregations obtained from school-detection algorithms has been used to study processes in marine systems, including shoal formation (Rose *et al.*, 1995), schooling behaviour (Brehmer *et al.*, 2007), predator–prey interactions (Nøttestad *et al.*, 2002), and vertical migrations (Gauthier and Rose, 2002).

School-detection algorithms may not be appropriate to characterize all spatial patterns observed in acoustic data (Reid *et al.*, 2000). These algorithms are designed to isolate discrete aggregations with defined boundaries, utilizing input parameters that define minimum aggregation size, minimum density, and minimum separation from other aggregations. Boundaries are then set around groups of adjacent pixels that satisfy criteria defined by the input parameters (ICES, 2000; Burgos and Horne, 2007). The operating assumption is that detected groups of pixels correspond to biologically meaningful structures (i.e. schools or shoals). This assumption is not valid for species that do not form discrete aggregations. For example, gadoids form diffuse pelagic shoals and continuous pelagic and benthic layers that may extend for long (>100 nautical mile) distances (e.g. Walline, 2007) and lack well-defined edges. Owing to the lack of clear boundaries, aggregation metrics obtained from school-detection algorithms are highly sensitive to the values of input parameters (Burgos and Horne, 2007). In these cases, although they can be used to obtain relative measures of spatial pattern (Wilson *et al.*, 2003), it is difficult to assign biological significance to the structures detected.

One possible approach to characterizing spatial patterns in echograms when aggregations cannot be delimited clearly is to describe the overall distribution of acoustic density within echogram segments of equal length, also known as elementary sampling distance units (ESDUs). Reid *et al.* (2000) suggest that the presence of these structures should be inspected visually and

classified into four echo types. Similarly, Petitgas and Levenez (1996) proposed a classification system of nine echo types. Visual inspection can be used to discriminate and categorize aggregations, but this approach is subjective and may be time-consuming.

Visual methods of classification can be combined with the use of descriptive metrics to quantify patterns within ESDUs. For example, to analyse the distribution of myctophids, Bertrand *et al.* (2002) combined visual inspection and classification of echograms with a set of metrics calculated for echogram sections of equal length. Spatial-aggregation categories included the sound-scattering layer, the nucleus in the scattering layer, large aggregated structures, stick-shaped structures, and small aggregates. Metrics calculated in each echogram section included total acoustic density and the percentage of cells with non-zero values. Although this approach allowed quantification of structures lacking defined edges, it still required visual examination of echograms. An extension of this approach would expand the number and types of aggregation metric, and use numerical classification methods to obtain a classification typology.

In recent years, landscape-ecology researchers have proposed a large number of metrics to measure spatial heterogeneity (e.g. McGarigal and Marks, 1995). Usually, information on the spatial composition of the landscape, a region of interest, is displayed in raster maps, where each pixel is assigned a category. This approach defines homogeneous regions (Gustafson, 1998) and calculates indices that quantify their spatial heterogeneity, including composition (i.e. the number and proportion of patch types), and spatial configuration (i.e. the size, shape, and location of patches and interspersion among patch types). Usually, these metrics have been used to describe spatial patterns of landscapes at relatively large ($\sim$100 km) scales (e.g. Schumaker, 1996), although in a limited number of studies, they have been used to describe small-scale ($\sim$1 m) spatial patterns in benthic communities (e.g. Teixidó *et al.*, 2002). The representation of fish distributions detected acoustically in an echogram is analogous to a raster map, and landscape metrics could be used to characterize their spatial patterns.

Here, we develop methods to quantify and classify spatial patterns of acoustic scatterers, using a combination of landscape indices and additional metrics appropriate for acoustic data. We divided each echogram into a series of segments or ESDUs, and calculated metrics to describe the spatial distribution of acoustic density in each segment. We utilized ordination and clustering techniques to analyse the relationship among metrics and to assign echogram segments to a reduced number of categories. Walleye pollock (*Theragra chalcogramma*) in the Bering Sea are used as the test species because they form a diverse array of spatial-distribution patterns, including patterns known to commercial fishers as "hay stacks", i.e. vertically elongated benthic schools, "cherries", i.e. spherical pelagic schools, or "carpets", i.e. benthic layers. Walleye pollock is an important species in the pelagic and demersal components of the North Pacific and Bering Sea ecosystems (Springer, 1992), and sustains a large commercial fishery with landings fluctuating between 1.18 and 1.55×10^6 t in the period 2000–2005 (National Marine Fisheries Service, 2007).

Methods

Data

The acoustic data used in this study were collected between 2000 and 2002 during five echo-integration trawl surveys in the eastern Bering Sea. Surveys were carried out as part of the regular survey programme of the NOAA Alaska Fishery Science Center (AFSC) to estimate the biomass of walleye pollock (Figure 1). Surveys were conducted between February and March (winter surveys) and between June and July (summer surveys), following standardized procedures (Karp and Walters, 1994; Honkalehto *et al.*, 2002). For our study, we used 52 transects from the summer surveys, with a combined total of 28 828 km, and 53 transects from the winter surveys, totalling 19 140 km. We used data collected during spawning (winter surveys) and feeding (summer surveys) periods to obtain a diversity of spatial patterns. Acoustic data were collected using a Simrad EK500 echosounder, operating a split-beam transducer at a frequency of 38 kHz, with a beam angle of 7° between half power points. The ping rate was 1 s^{-1}, and the pulse duration was set at 1.024 ms (Honkalehto *et al.*, 2002). The echosounder was calibrated using copper spheres following the procedures outlined by Foote *et al.* (1987). Staff from the AFSC analyzed the data using a minimum threshold of -70 dB re 1 m^2 m^{-2}. Acoustic returns were classified as walleye pollock or other species, based on the analysis of trawl samples and on the morphology of echogram patterns observed in current and previous surveys.

Data were imported into Echoview 3.10 (SonarData, 2004), and displayed in echograms. A threshold of -55 dB re 1 m^2 m^{-2} was applied to the data, because pixels with lower mean volume-backscatter strength (S_v) were unlikely to contain returns from walleye pollock (Burgos and Horne, 2007). Echoes from 14 m below the surface and 0.5 m above the sounder-detected seabed were not included in the analysis. Regions of the echogram containing returns from species other than adult walleye pollock were removed using virtual echograms (Higginbottom *et al.*, 2000). The spatial resolution of the raw data is variable. In the horizontal direction, the resolution, i.e. one ping, is equivalent to 5–20 m, depending on vessel speed. In the vertical direction, the resolution is a function of bottom depth. To have a regular

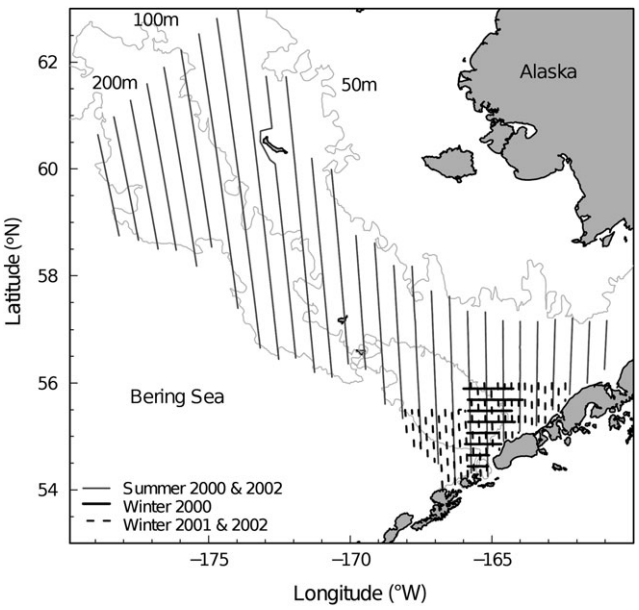

Figure 1. Survey transects of two summer (2000 and 2002) and three winter (2000, 2001, and 2002) walleye pollock surveys in the eastern Bering Sea.

and consistent spatial resolution, the acoustic data were echo-integrated at a horizontal resolution of 20 m (equivalent to 1–4 pings) and a vertical resolution of 1 m, the maximum vertical resolution available in Echoview for echo integration. A visual comparison between echograms using the raw acoustic data, as collected during the surveys, and echograms using the echo-integrated data revealed that the overall spatial pattern within each segment was not affected by the reduction in spatial resolution. The echo-integrated data from each transect were stored as a matrix, where each cell contained the mean S_v of each 20 m × 1 m echogram region. Because each cell in the matrix contained the integrated S_v value of 5–20 pixels from the original echogram, it was possible to have cells with S_v values below the -55 dB re 1 m^2 m^{-2} threshold. Cells with at least 1 pixel with a value of S_v above threshold were considered occupied cells. Matrix cells corresponding to pixels in the original echogram whose values were all below -55 dB re 1 m^2 m^{-2} were considered empty or background cells. Cells with mean S_v values higher than -39.67 dB re 1 m^2 m^{-2} were considered errors in seabed detection, or noise in the water column. This upper threshold corresponds to a walleye pollock of 40-cm fork length (i.e. the mean size of adult pollock captured on trawl samples during the surveys) and an average spatial separation of one body length. These cells were coded as empty cells or as cells below the seabed, depending on their position.

Characterization of spatial patterns in echograms

To quantify spatial patterns in density distributions of walleye pollock, we used a connected component algorithm to detect patches of occupied cells within echogram segments of equal length. A series of descriptive metrics was calculated in each segment to characterize spatial patterns.

Echogram segmentation

Echograms from each transect were divided into segments of 1.84 km (~1 nautical mile), starting at the beginning of and continuing until the end of each transect. Most transects were intermittently interrupted for fishing or other vessel manoeuvres, resulting in spatially, but not temporally, continuous data. All transect interruptions were identified using cruise-log notes. Whenever there was an interruption, the current segment was ended and a new segment started after the location of the interruption. This avoided possible artefacts from echogram segments containing data that were not continuous in time. Echogram segments <900-m long were removed from further analysis.

Patch detection and characterization

We utilized a connected component algorithm (Han and Wagner, 1990) to identify patches in each echogram segment. A patch was any single or group of adjacent cells, following the eight-neighbour rule, corresponding to pixels with a volume-backscatter coefficient equal to or above a threshold value of -55 dB re 1 m^2 m^{-2}, i.e. occupied cells. The concept of a patch is analogous to "schools" or "aggregations" in ETC, but is less restrictive because no minimum-size criteria were applied, and there was no assumed correspondence to biological aggregations, i.e. schools or shoals. Patches were considered only within the boundaries of each segment. In cases where patches extended beyond the limits of a segment, only the portion within the segment was considered. The area- and line-backscattering strength (S_L, MacLennan *et al.*, 2002) were calculated for each patch.

Calculation of metrics

Spatial patterns in each echogram segment were quantified using a set of descriptive metrics. Metrics were selected by visually inspecting echograms of walleye pollock and identifying aspects of spatial distributions that varied among them. These included occupancy, patchiness, size distribution of patches, distances among patches, acoustic density, mean vertical location, and deviation in vertical location. The number of metrics that could be used to quantify these characteristics is large, because each characteristic can be described in several ways. As an example, spatial occupancy can be described using the proportion of pixels above threshold, the number of patches, or the total patch area. We initially calculated 47 metrics in each echogram segment. It is likely that some of these metrics were redundant and were highly correlated. To identify redundant metrics, we calculated Spearman's "measure-of-association" and, following Riitters *et al.* (1995), grouped those that were highly correlated ($\rho >$ 0.9), and selected one from each group. We avoided metrics with highly skewed values and preferred those that were simple to compute and to interpret. The final number of metrics was 20. A complete list of metrics is presented in the Appendix, together with calculation details.

Exploratory factor analysis

To analyse relationships among metrics, we used exploratory factor analysis (EFA; Fabrigar *et al.*, 1999). EFA models the covariance among observed variables through a reduced number of unobserved variables or factors. The variables observed are assumed to be a linear function of one or more factors. The interpretation of each factor is based on variables that contribute high loadings (usually with absolute values >0.5) to each factor.

The first step in EFA is to select the number of factors to include in the common factor model. To select the appropriate number, we used parallel analysis (Montanelli and Humphreys, 1976). This technique calculates eigenvalues of multiple sets of random data: in our case 1000 sets. These eigenvalues were compared with those calculated from the reduced correlation matrix of the observed data, i.e. a correlation matrix with the square of the multiple correlation coefficient between each variable and all other variables placed in the diagonal. The number of factors included in the model corresponds to the number of eigenvalues in the observed data that are larger than the median eigenvalues obtained from the random set. The common-factor model was fitted following the iterative principal factor method, using a convergence criterion of 10^{-4} (Rencher, 2002).

EFA models with more than two factors do not have a unique solution. A common approach to select a solution is to search a simple structure, in which each factor is defined by a subset of observed variables with high loadings, and each observed variable has high loadings on a subset of factors (Fabrigar *et al.*, 1999). This is achieved by rotating the factors in multidimensional space. There are many rotation methods available. We used the promax rotation (Hendrickson and White, 1964), a non-orthogonal rotation that allows for correlation among factors. The model was fitted using the principal-factor method, which does not assume multivariate normality in the observed variables. Factor scores for each echogram were predicted from the common-factor model using the regression method, minimizing the mean square error (Rencher, 2002).

If factor scores summarize the spatial patterns observed in echograms, then echograms with similar factor scores should have similar patterns. To verify this assumption, we divided the range of each factor score into ten intervals and randomly selected up to 20 echograms from each combination of intervals. Echograms were inspected visually for consistency in their spatial patterns. Aspects considered included overall echogram arrangement, e.g. proportion of occupied cells, patchiness, and vertical distribution, and the presence of similar structures, e.g. schools, layers, or shoals.

Development of a classification typology

We were interested in classifying patterns observed in echograms into a relatively small number of categories and in developing a classification typology. We used cluster analysis to detect groups of echogram segments within the multidimensional space defined by the factor axes. Factor scores were first standardized, subtracting the mean and dividing by the standard deviation, to ensure that all factors were equally weighted during the cluster analysis. We used model-based clustering (Fraley and Raftery, 2002), where observations are assumed to come from a mixture of multivariate normal densities. Each cluster is centred at its mean, where the density of points is highest. Depending on the cluster covariance structure, the shape, volume, and orientation of clusters can be the same or vary among clusters. Model selection includes selecting the number of clusters and the covariance structure and is based on the Bayesian Information Criteria (BIC; Raftery, 1995). A maximum BIC score provides evidence of a model and an associated number of clusters (Fraley and Raftery, 2002). Model fitting uses the expectation-maximization (EM) method (Dempster *et al.*, 1977). The algorithm was initialized assigning each observation to a cluster using hierarchical clustering. Next, a series of iterations was performed, each consisting of two steps. The first step estimates the conditional probability for each observation belonging to each cluster, based on the current parameter estimates. In the second step, model parameters are updated using a maximum-likelihood criterion. Steps are repeated until convergence, defined as the change in likelihood of the model parameters estimated in successive iterations. Cluster membership is assigned to the cluster with greatest probability. A measure of classification uncertainty can be obtained by subtracting this probability from 1. Cluster analysis was carried out using the library "mclust" (Fraley and Raftery, 2006) in the R statistical environment (R Development Core Team, 2006). To describe spatial patterns assigned to each cluster, a random sample of 60 echogram segments was selected from the samples with a classification uncertainty of 0.2 or lower, a value arbitrarily chosen to define low classification uncertainty. Echograms were examined visually for consistency in their spatial patterns, including overall echogram arrangement and the presence of similar structures.

Results

Data analysis and metric calculations

We obtained 10 802 echogram segments, after dividing transects from the five surveys and removing segments that did not contain acoustic returns from walleye pollock, or were <900 m in length ($n = 2348$). A few ($n = 358$) segments were also removed because some of the metrics could not be calculated. As an example, nearest-neighbour-distance (NND) could not be computed for segments containing a single patch. Table 1 provides summary statistics of the 20 metrics calculated in each echogram segment. To illustrate the potential range of metric values for spatial patterns common in walleye pollock, we calculated representative metric values for three 18.4-km long echograms, corresponding to ten consecutive echogram sections of 1 nautical mile. These sections contained typical aggregation patterns

Table 1. Summary statistics for descriptive metrics calculated for 10 802 echogram segments, each 1 nautical-mile-long, obtained from five echo-integration trawl surveys in the eastern Bering Sea.

Metric	Mean	Standard deviation	Kurtosis	Skewness
Patch density (PD)	2.59E−04	2.61E−04	5.65	1.54
Percentage of landscape (PLAND)	0.023	0.036	31.58	4.19
Landscape division index (LDI)	0.87	0.15	8.64	−2.26
Patch area—10th percentile (A10)	51.19	29.81	3.41	0.90
Patch area—50th percentile (A50)	96.49	71.00	3.59	1.03
Patch area—90th percentile (A90)	139.20	97.53	3.41	0.96
Relative patch area—10th percentile (RA10)	0.25	0.11	7.23	2.02
Relative patch area—50th percentile (RA50)	0.68	0.05	3.07	0.43
Relative patch area—90th percentile (RA90)	0.96	0.02	2.22	−0.34
Biomass depth—10th percentile (BD10)	0.77	0.19	3.75	−1.04
Biomass depth—50th percentile (BD50)	0.88	0.14	9.70	−2.25
Biomass depth—90th percentile (BD90)	0.94	0.09	35.42	−4.80
Median distance to nearest neighbour (medNND)	62.81	139.33	49.05	6.01
Volume-backscattering strength (S_v mean)	−72.14	8.79	3.61	−0.78
Acoustic density—10th percentile (S_v 10)	−64.77	3.18	15.74	2.19
Acoustic density—50th percentile (S_v 50)	−57.28	4.27	8.98	1.86
Acoustic density—90th percentile (S_v 90)	−48.39	6.05	4.07	1.07
Inertia in the horizontal direction (inx)	213 877.90	109 700.80	2.96	0.06
Inertia in the vertical direction (iny)	21 355.65	63 322.51	106.40	8.96
Index of aggregation (lagg)	3.89E−04	4.75E−04	6.91	1.96

observed among walleye pollock: schools (i.e. small aggregations with well-defined boundaries), a large pelagic shoal (i.e. large aggregation with diffuse boundaries), and a benthic layer (i.e. a large aggregation with limited vertical extension located near the bottom; Figure 2). Metric values reflect differences among these type patterns. The pelagic shoal is characterized by high spatial occupancy [measured by percentage of landscape (PLAND)], and low values of landscape division index (LDI), indicating high aggregation levels. Well-defined schools are characterized by high values of LDI, indicating high fragmentation, high S_v90, indicating the presence of pixels with high acoustic density, and low PLAND. Values for the benthic-layer segments are intermediate. Differences in the vertical distribution are well characterized by the percentiles of biomass depth (BD10, BD50, and BD90). Values were high for the benthic layer, indicating that most of the biomass was near the bottom, low for the pelagic layer, and intermediate for schools.

Factor analysis

Parallel analysis, based on comparing the eigenvalues of the reduced correlation matrix for observed data with the median eigenvalue of multiple sets of random data, indicated that the optimal number of factors for the EFA was four (Figure 3). The fact that differences between consecutive eigenvalues were larger among the first four eigenvalues than among all other eigenvalues supports this conclusion (Fabrigar *et al.*, 1999).

Results of the EFA calculations are summarized in Table 2. The four factors explained 63.6% of the variance of the dataseries. The promax rotation successfully produced factors characterized by variables where high loadings of any variable were restricted to one factor. Each factor can be interpreted based on variables that have loadings >0.5. In our case, the four factors measure spatial occupancy, aggregation level, packing density, and vertical distribution. Most metrics had correlations >0.5 with at least one of the factors. Two exceptions included inertia in the horizontal (inx) and vertical (iny) directions that, owing to their high variability, did not have high loadings on any factor. The results of the EFA were validated by randomly splitting the dataset in two and repeating the analysis in each half (Fabrigar *et al.*, 1999). Factor loadings obtained in both halves matched those obtained with the full dataset. The scores of variables with high loadings differed by a maximum of 1.78% relative to the scores obtained with the full dataset. Sample echograms with scores ranging from low to high for each factor are shown in Figure 4. Echograms with similar factor scores had similar patterns, suggesting that factor scores adequately summarize the range of aggregation types for walleye pollock.

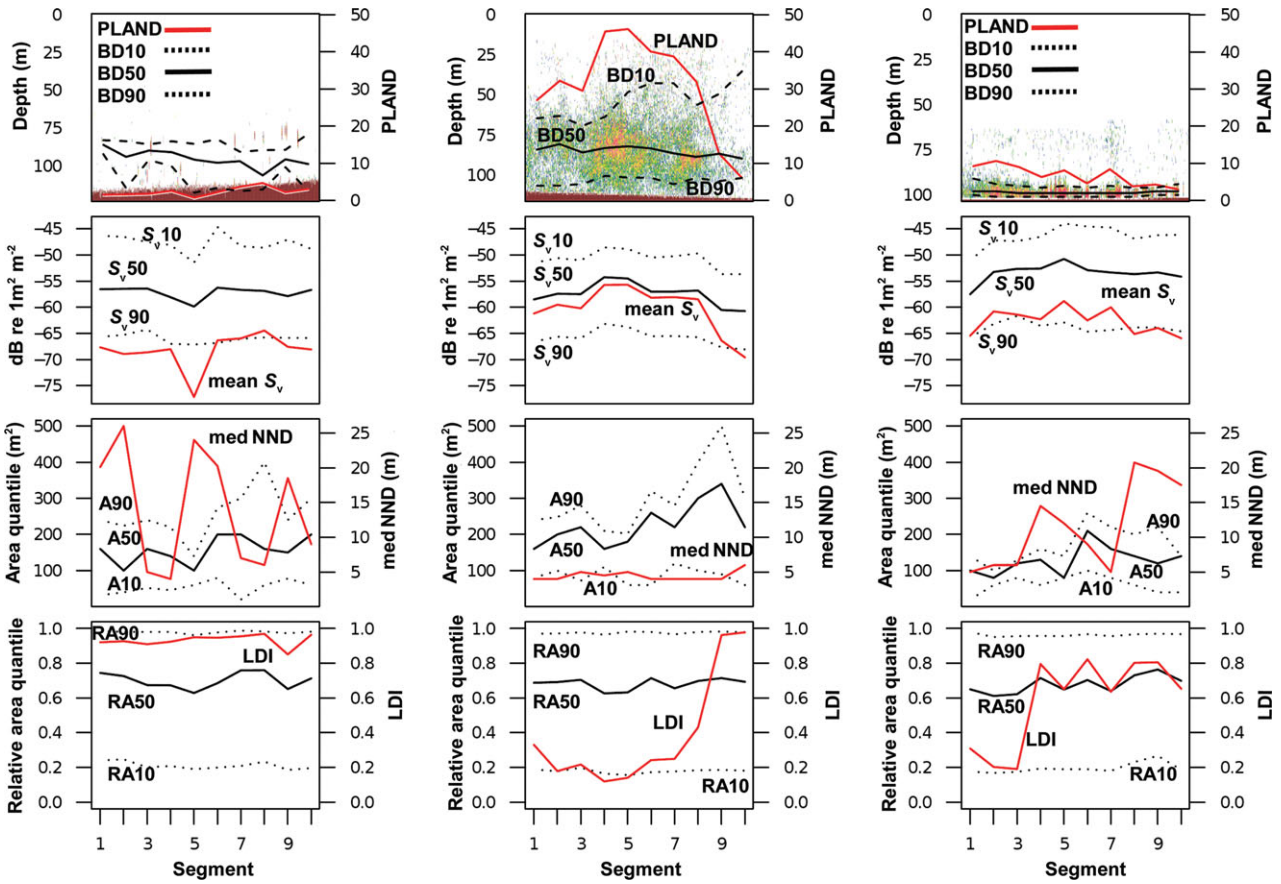

Figure 2. Spatial-density distributions of walleye pollock: well-defined schools, a large pelagic shoal, and a benthic layer, and corresponding values of descriptor metrics. Each echogram represents 10 nautical miles of survey track, divided into segments of one nautical mile. Metrics include PLAND, 10th, 50th, and 90th percentiles of biomass depth (BD10, BD50, and BD90), volume-backscattering strength (S_v mean), 10th, 50th, and 90th percentiles of acoustic density (S_v10, S_v50, and S_v90), landscape-division index (LDI), and 10th, 50th, and 90th percentiles of relative patch area (RA10, RA50, and RA90).

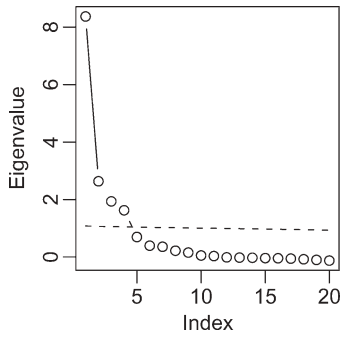

Figure 3. Plot of eigenvalues, sorted in decreasing value, of the reduced correlation matrix of the 20 descriptive metrics calculated in 10 802 echogram segments. The dotted line is the mean eigenvalues of 1000 sets of random data. The number of eigenvalues in the observed data greater than the eigenvalues of the random sets indicates the appropriate number of factors to be used in an EFA.

Factor 1: spatial occupancy/patch size

The first factor explained 24.7% of the variance and was interpreted as a combination of spatial occupancy and patch size. It represents a gradient from empty to fully occupied echograms. Measures of patch density (PD), percentage of landscape (PLAND), 10th, 50th, and 90th percentiles of patch size (A10, A50, and A90), and mean acoustic density (S_v mean) had loadings >0.5, indicating that highly occupied echograms tend to have more biomass, larger patches, and a higher number of patches. The high correlation with the 90th percentile of relative patch area (RA90) indicates that in highly occupied echograms, a larger proportion of the total patch area corresponded to larger patches. The negative loading assigned to the index of aggregation (Iagg) is a result of the large number of patches observed with high occupancy levels.

Factor 2: aggregation level

Factor 2 was interpreted as a measure of aggregation, explaining 15.8% of the variance. Metrics with high positive loadings on this factor included two measures of relative patch size (RA10 and RA50), and the median distance to the nearest neighbour (medNND). The LDI had a high negative loading. Low scores of factor 2 represented echograms that contained highly fragmented acoustic backscatter. Fragments were small aggregations or individual pixels. In these echograms, a larger proportion of the occupied area corresponded to small patches, and distances among patches tended to be small. High scores in factor 2 corresponded to echograms with high aggregation levels.

Factor 3: packing density

Metrics with high loadings on the third factor included S_v10, S_v50, and S_v90, three metrics that characterize the distribution of acoustic density within the 20×1-m cells. Echogram segments with higher scores in factor 3 tended to have cells with high values of acoustic density. Factor 3 explained 12.1% of the total variance. This factor was interpreted as an index of packing density (i.e. the number of fish per unit volume of water).

Factor 4: vertical distribution

The fourth factor explained 11.0% of the variance and was interpreted as an index of vertical distribution. Quantiles of biomass

Table 2. Results of the EFA.

Parameter	Factor 1	Factor 2	Factor 3	Factor 4
Proportion of variance	0.247	0.158	0.121	0.11
Cumulative variance	0.247	0.405	0.526	0.636
Metric	**Loadings**			
Patch density (PD)	0.685*	−0.321	−0.22	−0.2
Percentage of landscape (PLAND)	0.718*			−0.123
Landscape division index (LDI)	0.115	−0.707*	−0.17	
Patch area—10th percentile (A10)	0.662*	−0.37	−0.116	−0.12
Patch area—50th percentile (A50)	0.923*	−0.165	−0.116	
Patch area—90th percentile (A90)	0.962*	−0.125		
Relative patch area—10th percentile (RA10)	−0.419	0.9*		
Relative patch area—50th percentile (RA50)	0.148	0.752		
Relative patch area—90th percentile (RA90)	0.757	0.168		
Biomass depth—10th percentile (BD10)	−0.17	0.133	0.197	0.789*
Biomass depth—50th percentile (BD50)				0.945*
Biomass depth—90th percentile (BD90)		−0.131		0.714*
Median distance to nearest neighbour (medNND)	−0.302	0.593*		
Volume-backscattering strength (S_v mean)	0.633*	−0.421	0.439	0.15
Acoustic density—10th percentile (S_v 10)	−0.248	0.31	0.616*	−0.125
Acoustic density—50th percentile (S_v 50)		0.126	0.943*	
Acoustic density—90th percentile (S_v 90)			0.858*	0.177
Inertia in the horizontal direction (inx)	0.276	−0.313	−0.112	
Inertia in the vertical direction (iny)	−0.141	0.153	0.22	
Index of aggregation (Iagg)	−0.515*	0.409		0.156

The four-factor model had 116 degrees of freedom and explained ~64% of the variance in the metrics dataset. Metrics with absolute loadings >0.5 (marked with an asterisk) were considered important in the interpretation of each factor. Factor scores with absolute values <0.1 are not displayed.

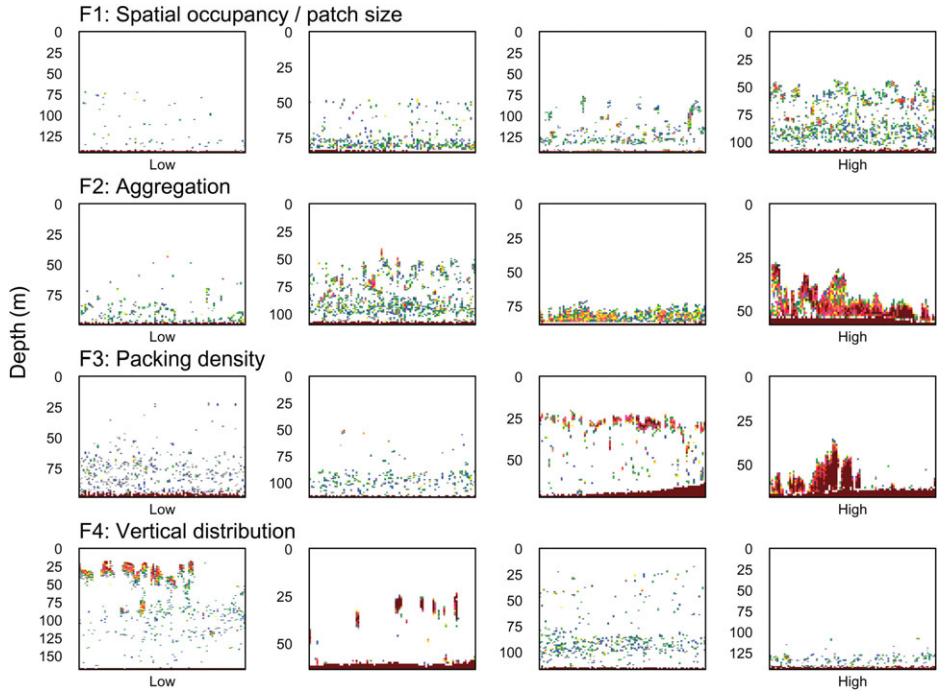

Figure 4. Sample echograms for each factor representing the range of factor scores. Echograms in each row correspond to a factor score approximately at each 25th percentile of the factor range.

depth (BD10, BD50, BD90), which describe the vertical distribution of acoustic backscatter, had high positive loadings. Because these metrics were expressed as a proportion of the average bottom depth in the segment, high scores in factor 4 indicated that biomass was located near the bottom, and low scores that biomass was distributed throughout the water column.

Relationships among factors

We described relationships among factors using only echogram segments with PLAND ≥ 0.006 ($n = 6575$). Visual inspection revealed that echograms with lower values of spatial occupancy had no discernible spatial patterns. Factor scores from those echograms do not provide information on spatial patterns.

To analyse relationships among factors, we used paired density plots (Figure 5), because scatterplots were too cluttered and trends were not easily discerned. The significance of pairwise relationships was tested using Spearman's rank-correlation test. Factor 1 had a significant positive correlation with factor 2 ($\rho = 0.69$, $p < 2.2\text{E}-16$), indicating that echogram segments with high spatial occupancy tended to have larger patches and high aggregation levels. Factor 1 also had a significant positive correlation with factor 3 ($\rho = 0.22$, $p < 2.2\text{E}-16$), indicating that echograms with high occupancy had, on average, cells with high S_v values. Despite the significant correlations, there was considerable variability in the relationship among factors. The highest scores in factor 3 were in echograms with intermediate scores in factor 1, indicating that there was a trade-off between spatial occupancy and packing density. Factor 1 was not significantly correlated with factor 4 (vertical distribution). Factor 2 had a significant positive correlation with factor 3 ($\rho = 0.34$, $p < 2.2\text{E}-16$), indicating that echograms with high aggregation tended to have cells with high values of packing density. Factor 2 also had a significant correlation with factor 4 ($\rho = 0.23$, $p < 2.2\text{E}-16$). This reflects the fact that walleye pollock is a semi-pelagic species and, usually, biomass is concentrated on or near the seabed (Karp and Walters, 1994). When aggregation levels were low, the biomass was distributed through a larger proportion of the water column. Finally, factors 3 and 4 were positively correlated ($\rho = 0.09$, $p = 6.75\text{E}-14$), indicating that there was a tendency in echograms with biomass concentrated near the bottom to have pixels with relatively high values of packing density.

Riitters *et al.* (1995) suggested that the metric with the highest loading in each factor can be used as a surrogate, by extension, of the complete set of metrics. In our case, these metrics for the four factors were the 90th percentile of patch area (A90), the 10th percentile of the relative patch area (RA10), the 50th percentile of pixel acoustic density (S_v50), and the 50th percentile of biomass distributions (BD50). The high (≥ 0.9) correlation among these metrics with their respective factors suggests that they are appropriate surrogates.

Development of classification typology

Cluster analysis was also performed on echogram segments with PLAND ≥ 0.006. The model with the greatest support corresponded to 12 ellipsoidal clusters with an unconstrained variance structure, allowing for differences in volume and shape (Figure 6). The BIC for this model was $-43\,911.52$, 41.6 units larger than the model with the next highest BIC, indicating that there was strong evidence of this model (Raftery, 1995). The fact that ~60% of the observations had a classification uncertainty of 0.2 or higher indicates that there was great overlap between clusters. There was no significant relationship between cluster size and classification uncertainty (linear regression, $r^2 = 0.04$, $p = 0.54$). For example, the largest cluster (Cluster 7, $n = 1105$) was also the best defined (with 97.0% of the observations having a classification uncertainty ≤ 0.2), whereas observations assigned to the second largest cluster

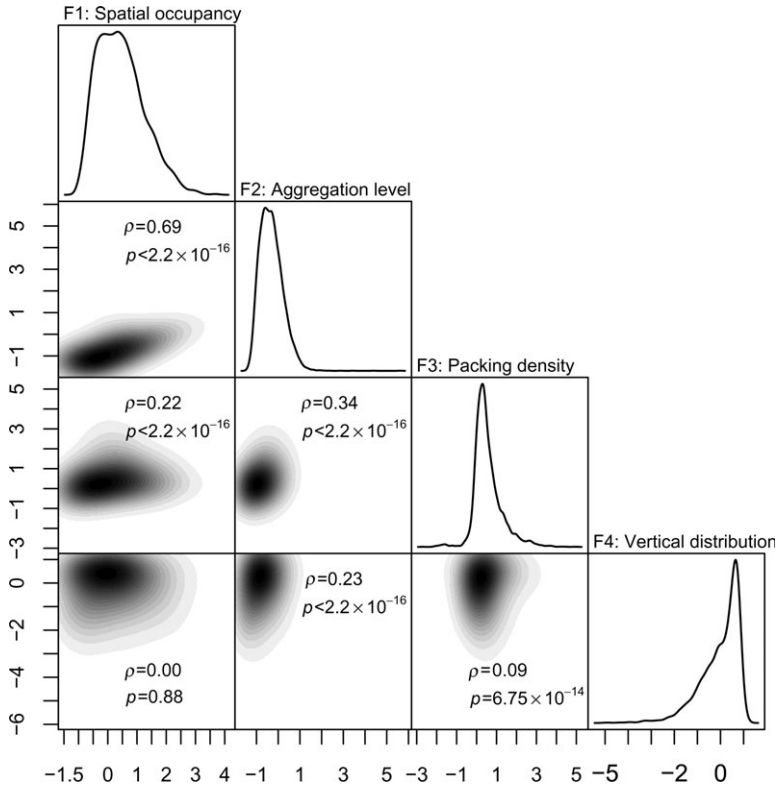

Figure 5. Paired density plots of factor scores calculated in echograms with PLAND > 0.006 (n = 6575). Association among factors was tested using Spearman's ρ statistic. Kernel-density estimates of each factor are included on the diagonal.

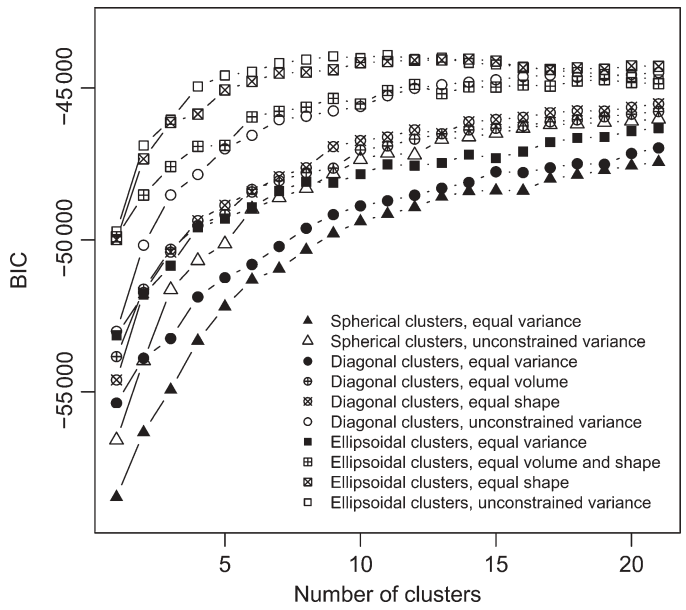

Figure 6. BIC scores for cluster models with varying covariance structure and number of clusters. The highest BIC value corresponds to an ellipsoidal model with unconstrained variance and 12 clusters.

(Cluster 5, n = 1010) had a relatively high classification uncertainty. Factor scores for each cluster are displayed using diamond plots (Figure 7), in which each axis represents one factor. Differences in the distribution of factor scores among clusters are characterized by the shape of the corresponding diamond plots. A description of spatial patterns observed in randomly selected echograms of each cluster is presented in Table 3. Patterns in some clusters were consistent (e.g. clusters 1, 5, and 10), whereas others included echograms with a high diversity of spatial patterns (e.g. clusters 4, 6, and 9). Clusters with low

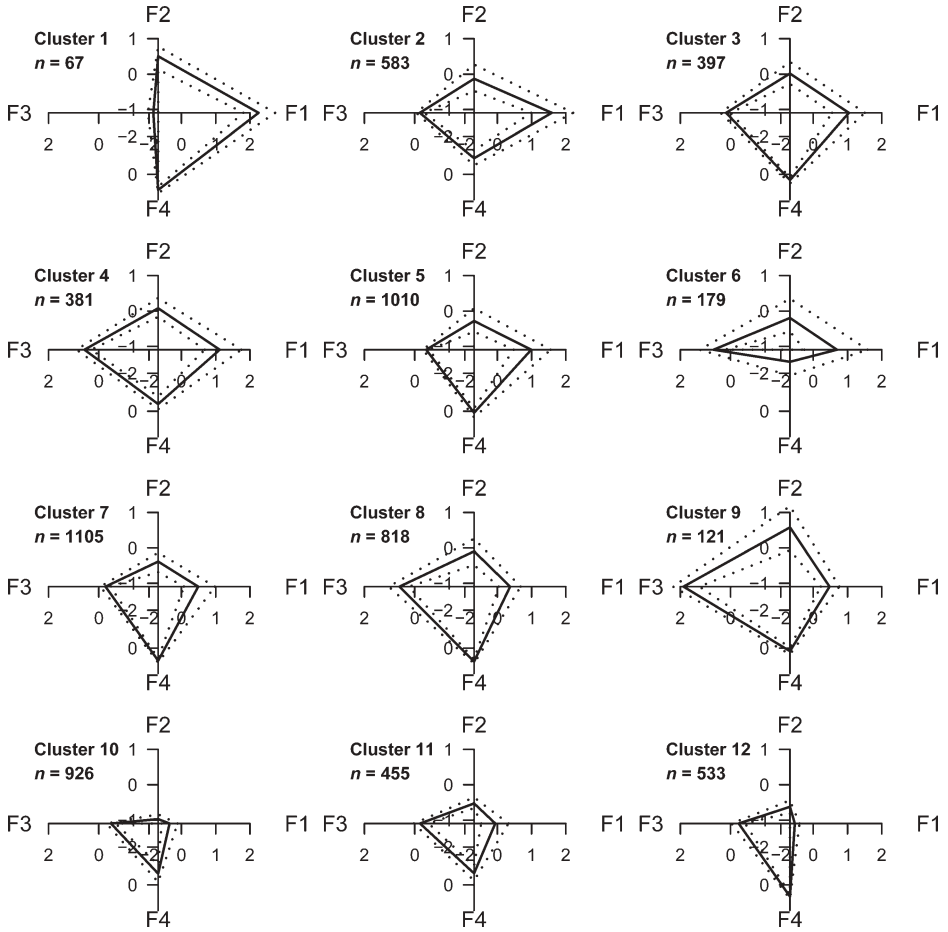

Figure 7. Diamond plots of factor scores from 12 clusters. Each axis in a plot represents one factor, starting from the horizontal positive axis and continuing anticlockwise. The continuous line indicates mean factor values, and dotted lines represent the 0.25 and 0.75 quantiles.

classification uncertainty had greater consistency in spatial patterns among echogram segments.

Discussion

The first objective of this study was to characterize spatial distributions of walleye pollock using acoustic data. Echogram segments with similar factor scores, i.e. within the same factor intervals, had similar spatial patterns, suggesting that factor scores provide an adequate quantification of these patterns. Because all factors had appropriate surrogate metrics, i.e. metrics with correlation values >0.9 with the factor, future studies could use these surrogate metrics, simplifying data analysis, and interpretation while retaining most of the information on spatial patterns.

Our second objective was to develop a classification typology of walleye pollock aggregations. Characteristic spatial patterns of walleye pollock include bottom layers, pelagic layers, and discrete schools. If these spatial patterns are distinct, then the cluster analysis should have separated types and grouped echogram segments into discrete categories. In our case, the optimal cluster model included 12 clusters, which is a larger number of categories than in other classification typologies (e.g. Petitgas and Levenez, 1996; Reid *et al.*, 2000; Bertrand *et al.*, 2002). The large proportion of observations with high classification uncertainty suggests that aggregations of walleye pollock cannot be classified into discrete, well-defined groups. Instead, the distributions form a continuum along each of the four factors. The 12 clusters represent regions of the four-dimensional space with high densities of observations. A typology is possible, even if clusters overlap (Legendre and Legendre, 1999), but overlap increases the classification uncertainty of individual samples. We propose that rather than using discrete categories, the analysis of spatial patterns in fish distributions should be conducted using continuous metrics.

Results are dependent on the set of metrics used to describe spatial patterns. The descriptive metrics that we used here were selected after visually inspecting echograms of walleye pollock and considering the echogram types proposed by Petitgas and Levenez (1996), Reid *et al.* (2000), and Bertrand *et al.* (2002). Gustafson (1998) proposed that pattern metrics should be selected according to the pattern or heterogeneity that is relevant to the process being studied. In our case, there is little information on the biological or physical processes that influence the spatial-distribution patterns of adult walleye pollock, in particular at spatial scales <10 km. Previous studies examined the influence of water temperature and thermocline depth (Swartzman *et al.*, 1994) and of prey abundance and water circulation (Hollowed *et al.*, 2007) on its distribution, but a comprehensive study has not been completed. In this case, a natural first step was to select a broad array of metrics that would quantify the spatial components that vary among echogram sections.

Table 3. Echogram-segment membership and uncertainty in clusters.

Cluster	*n*	% CI <0.2	Observed spatial patterns
1	67	65.0	Diffuse, low acoustic-density pelagic layers. Centre of mass is near the seabed, but scatters cover most of the water column
2	583	37.3	Diffuse, medium acoustic-density pelagic layers covering most of the water column
3	397	53.1	High acoustic-density benthic carpets and fragmented carpets of varying thickness, usually restricted to the bottom one-third of the water column. Benthic schools can also be found
4	381	9.6	High acoustic-density layers and schools separated from the seabed. High acoustic-density shoals and macrostructures on or off the seabed
5	1 010	18.8	Diffuse, low acoustic-density layers restricted to the bottom one-third of the water column
6	179	27.7	Patches of high acoustic density located in the upper two-thirds of the water column. Large diversity of patterns: pelagic schools, pelagic shoals, pelagic layers, and low-density scatters
7	1 105	97.0	Small scatters and patches located in the bottom one-third of the water column
8	818	18.4	Thin, high-density benthic layers. Small schools found occasionally.
9	121	26.0	Very dense aggregations located mostly in the bottom one-fifth of the water column. Patterns observed include thick benthic layers, shoals, benthic schools, and semi-pelagic layers
10	926	38.8	Small scatters in the lower half of the water column.
11	455	29.1	Small scatters in the lower two-thirds of the water column. Small pelagic schools and thin layers appear in few echograms
12	533	26.1	Almost empty echograms. Few small scatters in the bottom one-third of the water column

Number of echograms in each cluster (*n*), percentage of echograms low classification uncertainty (CI <0.2), and description of spatial patterns observed in echograms assigned to each cluster.

An additional consideration for metric selection was their uniqueness (Haines-Young and Chopping, 1996). The term "uniqueness" refers to the unequivocal relationship between pattern and metric values, in other words knowledge of the metric value should allow the prediction of the pattern. We visually inspected echograms across the range of each metric, allowing the rest of the metrics to vary. In general, there was great variability in the patterns observed, suggesting that the uniqueness of each metric is relatively low. Patterns in echograms sampled across the range of each of the four factors were more consistent (Figure 4), indicating that the uniqueness of factor scores is greater than that of individual metrics, and supporting the use of factor scores to index distribution patterns.

The sensitivity of metrics should be considered when using indices to describe spatial patterns. Ideally, each metric should vary when the pattern changes and should be able to differentiate between patterns (Baskent and Jordan, 1995). It is difficult to evaluate the sensitivity of the metrics used here, because there is no independent quantification of the pattern being described. If we assume that variability in distribution patterns is being described, then the distribution of metric values provides information on its sensitivity. Metrics with a limited range of values, or whose distributions are highly skewed or kurtotic, have low sensitivity. From the variables used in this study, inertia in the vertical direction (iny) and median distance to the nearest neighbour (medNND) had high skewness and kurtosis, perhaps indicating low sensitivity.

An important step in our methodology was selecting the length of the echogram sections. It has been recognized that selecting the appropriate scale is a critical step in the analysis of spatial patterns (Horne and Schneider, 1995; Gustafson, 1998). Two components of scale are considered: the grain or resolution (i.e. the size of the sampling unit), and the extent or size of the study area (Turner *et al.*, 1989; Wiens, 1989). The length of echogram segments, together with the bottom depth, defines the area over which the descriptive metrics are applied. The echogram length also becomes the resolution of the data when analysing abundance and spatial patterns at larger scales. Ideally, the scale of observation should match the scales of the patterns or processes being described (Gustafson, 1998). Although scale-dependent descriptions of fish density have not shown a characteristic scale of high variability (Horne and Schneider, 1997), we wanted to match the length of the echogram segments to the average size of the walleye pollock distributions. Echogram segments should be long enough to include several aggregations, yet short enough to characterize spatial variability in aggregation patterns. Several studies have reported the sizes of walleye pollock aggregations. Kang *et al.* (2006) reported shoal lengths of the order of 5 km, and as large as 10 km off the coast of Japan, but provided no information on how representative these school lengths were. Wilson *et al.* (2003) reported aggregation lengths for adult walleye pollock in the Gulf of Alaska ranging between 200 and 400 m, with variogram ranges of 2.1–11.8 km. Although Mello and Rose (2005) related the variogram range to average patch size, results from Wilson *et al.* (2003) suggest that for the adults, there are at least two spatial structures, with ranges of approximately 100s of metres to 2–10 km. This latter scale matches observations by Horne and Walline (2005), who described scale-dependent variability of walleye pollock in the eastern Bering Sea, and inferred a mean patch size of 2–3 km, based on results of spectral analysis and variogram ranges. On the other hand, Walline (2007) reported larger (>100 nautical miles) aggregations in the same area. An appropriate choice for an echogram-segment length would match the lower limit of this spatial range, corresponding to the ESDU length of 1 nautical mile ($\sim$1.84 km) used in this study. Descriptive metrics calculated in echogram segments of this size would capture the small-scale heterogeneity described by Wilson *et al.* (2003), and would be appropriate to describe patterns over the range of the survey. As expected, echogram sections were large enough to include several small benthic or pelagic schools. At the same time, larger structures such as

pelagic shoals or benthic carpets were included in several contiguous echogram segments.

The reduction of spatial resolution in the data to 20×1 m may have introduced bias in the calculation of metric values. During data processing, we reduced the resolution of the acoustic data from one horizontal ping, between 5 and 20 m, depending on vessel speed, to 20 m, and from one vertical sampling bin, usually 0.2 m, depending on the sampling rate, to 1 m. Whenever a 20×1-m cell contained at least one pixel above threshold, the entire cell was considered an occupied cell. These steps potentially increase the size of patches and the proportion of the echogram containing occupied cells, which would affect metric values based on the number of cells and patch size (e.g. PLAND, A50). It is likely that there was some aliasing of patch edges and merging of nearby patches, reducing the number of patches and affecting the measurements of PD, PLAND, and fragmentation (Iagg and LDI), and reducing NND (Haines-Young and Chopping, 1996). Biases are assumed to be small because a 20×1-m resolution maintained the observed spatial patterns, and also constant because all echogram segments were analysed at the same resolution.

We have quantified and classified density distributions of walleye pollock using landscape and other metrics. Given the high diversity in the distribution types observed, we believe that the methods developed here could be used to analyse the spatial heterogeneity of any species, regardless of its spatial distribution. This approach is also applicable to any acoustically acquired, two-dimensional data. For species that form well-defined aggregations, it complements the results from school-detection algorithms (Reid *et al.*, 2000) by quantifying the spatial distribution around detected aggregations and providing a spatial context to aggregation-based descriptors (Brehmer *et al.*, 2007). For echograms where acoustic returns from multiple species are separated in space, analyses can be carried out for each species. Spatial relations among patches of different species could then be quantified. The next steps in our research are to characterize the spatial distribution of pattern descriptors at scales >2 km, to characterize seasonal and interannual differences in walleye pollock distributions, and to examine the correlation of these descriptors with physical variables.

The methods proposed here could also be applied to historical series of acoustic data, with the objective of describing spatial and temporal changes in distribution patterns. Pattern changes could be associated with fishing activity, short-term environmental factors (e.g. seasonal or annual fluctuations), and long-term processes such as climate change. This information could be used potentially to identify spatial or temporal trends as part of an ecosystem-based approach to fisheries management.

Acknowledgements

The data used here were provided by the Resource Assessment and Conservation Engineering Division of the Alaska Fishery Science Center (NOAA). Funding for the work was provided by NOAA through the Steller Sea Lion Research Initiative, Office of Naval Research (N0014-00-1-0180), and the Alaska Fishery Science Center (NA17RJ1232-AM01). We thank Ariel Gustavo Cabreira and an anonymous reviewer for useful comments and corrections.

References

Baskent, E. Z., and Jordan, G. A. 1995. Characterizing spatial structure of forest landscapes. Canadian Journal of Forest Research, 25: 1830–1849.

Bertrand, A., Bard, F. X., and Josse, E. 2002. Food habits related to the micronekton distribution in French Polynesia. Marine Biology, 140: 1023–1037.

Bez, N., and Rivoirard, J. 2001. Transitive geostatistics to characterize spatial aggregations with diffuse limits: an application on mackerel ichthyoplankton. Fisheries Research, 50: 41–58.

Brehmer, P., Gerlotto, F., Laurent, C., Cotel, P., Achury, A., and Samb, B. 2007. Schooling behaviour of small pelagic fish: phenotypic expression of independent stimuli. Marine Ecology Progress Series, 334: 263–272.

Burgos, J. M., and Horne, J. K. 2007. Sensitivity analysis and parameter selection for detecting aggregations in acoustic data. ICES Journal of Marine Science, 64: 160–168.

Dempster, A. P., Laird, N. M., and Rubin, D. B. 1977. Maximum likelihood for incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Society Series B, 39: 1–38.

Fabrigar, L. R., Wegner, D. T., MacCallum, R. C., and Strahan, E. J. 1999. Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods, 4: 272–299.

Fraley, C., and Raftery, A. E. 2002. Model-based clustering, discriminant analysis, and density estimation. Journal of the American Statistical Association, 97: 611–631.

Fraley, C., and Raftery, A. E. 2006. Technical Report no. 504. Department of Statistics, University of Washington, September 2006.

Foote, K. G., Knudsen, H. P., Vestnes, D. N., MacLennan, D. N., and Simmonds, E. J. 1987. Calibration of acoustic instruments for fish-density estimation: a practical guide. ICES Cooperative Research Report, 144. 69 pp.

Gauthier, S., and Rose, G. A. 2002. Acoustic observation of diel vertical migration and shoaling behaviour in Atlantic redfishes. Journal of Fish Biology, 61: 1135–1153.

Gustafson, E. J. 1998. Quantifying landscape spatial pattern: what is the state of the art? Ecosystems, 1: 143–156.

Haines-Young, R., and Chopping, M. 1996. Quantifying landscape structure: a review of landscape indices and their application to forested landscapes. Progress in Physical Geography, 20: 418–445.

Han, Y., and Wagner, R. A. 1990. An efficient and fast parallel-connected component algorithm. Journal of the Association for Computing Machinery, 37: 626–642.

Hendrickson, A. E., and White, P. O. 1964. Promax: a quick method for rotation to oblique simple structure. British Journal of Statistical Psychology, 17: 65–70.

Higginbottom, I. R., Pauly, T. J., and Heatley, D. C. 2000. Virtual echograms for visualization and post-processing of multiple-frequency, echosounder data. *In* Proceedings of the Fifth European Conference on Underwater Acoustics, ECUA 2000, 10–13 July 2000, pp. 1497–1502. Ed. by M. E. Zakharia, P. Chevret, and P. Dubail. Office for Official Publications of the European Communities, Luxembourg.

Hollowed, A. B., Wilson, C. D., Stabeno, P. J., and Salo, S. A. 2007. Effect of ocean conditions on the cross-shelf distribution of walleye pollock (*Theragra chalcogramma*) and capelin (*Mallotus villosus*). Fisheries Oceanography, 16: 142–154.

Honkalehto, T., Patton, W., de Blois, S., and Williamson, N. 2002. Echo integration, trawl- survey results for walleye pollock (*Theragra chalcogramma*) on the Bering Sea shelf and slope during summer 2000. NOAA Technical Memorandum, NMFS–AFSC–126.

Horne, J. K., and Schneider, D. C. 1995. Spatial variance in ecology. Oikos, 74: 18–26.

Horne, J. K., and Schneider, D. C. 1997. Spatial variance of mobile aquatic organisms: capelin and cod in Newfoundland coastal waters. Philosophical Transactions of the Royal Society of London, Series B, 352: 633–642.

Horne, J. K., and Walline, P. D. 2005. Spatial and temporal variance of walleye pollock (*Theragra chalcogramma*) in the eastern Bering Sea. Canadian Journal of Fisheries and Aquatic Sciences, 62: 2822–2831.

Hyndman, R. J., and Fan, Y. 1996. Sample quantiles in statistical packages. The American Statistician, 50: 361–365.

ICES. 2000. Report on echotrace classification. ICES Cooperative Research Report, 238. 107 pp.

Kang, M., Honda, S., and Oshima, T. 2006. Age characteristics of walleye pollock school echoes. ICES Journal of Marine Science, 63: 1465–1476.

Karp, W. A., and Walters, G. E. 1994. Survey assessment of semi-pelagic gadoids: the example of walleye pollock, *Theragra chalcogramma*, in the eastern Bering Sea. Marine Fisheries Review, 56: 8–22.

Legendre, P., and Legendre, L. 1999. Numerical Ecology, 2nd edn. Elsevier Science, New York.

MacLennan, D. N., Fernandes, P. G., and Dalen, J. 2002. A consistent approach to definitions and symbols in fisheries acoustics. ICES Journal of Marine Science, 59: 365–369.

McGarigal, K., and Marks, B. J. 1995. FRAGSTATS: spatial-pattern analysis program for quantifying landscape structure. General Technical Report PNW-GTR-351, USDA Forest Service. Pacific Northwest Research Station, Portland, OR.

Mello, L. G. S., and Rose, G. A. 2005. Using geostatistics to quantify seasonal distribution and aggregation patterns of fishes: an example of Atlantic cod (*Gadus morhua*). Canadian Journal of Fisheries and Aquatic Sciences, 62: 659–670.

Montanelli, R. G., and Humphreys, L. G. 1976. Latent roots of random-data correlation matrices with squared multiple correlations on the diagonal: a Monte Carlo study. Psychometrika, 41: 341–348.

National Marine Fisheries Service. 2007. Fisheries Statistics and Economics Division. http://www.st.nmfs.gov/st1/comm.ercial/index.html (accessed 26 February 2007).

Nøttestad, L., Ferno, A., Mackinson, S., Pitcher, T., and Misund, O. A. 2002. How whales influence herring-school dynamics in a cold-front area of the Norwegian Sea. ICES Journal of Marine Science, 59: 393–400.

Petitgas, P., and Levenez, J. J. 1996. Spatial organization of pelagic fish: echogram structure, spatial-temporal condition, and biomass in Senegalese waters. ICES Journal of Marine Science, 53: 147–153.

R Development Core Team. 2006. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org.

Raftery, A. E. 1995. Bayesian model selection in social research (with discussion by Andrew Gelman, Donald B. Rubin, and Robert M. Hauser). *In* Sociological Methodology, pp. 111–196. Ed. by P. V. Marsden. Blackwell Scientific Publications, Oxford, UK.

Reid, D., Scalabrin, C., Petitgas, P., Masse, J., Aukland, R., Carrera, P., and Georgakarakos, S. 2000. Standard protocols for the analysis of school-based data from echosounder surveys. Fisheries Research, 47: 125–136.

Rencher, A. C. 2002. Methods of Multivariate Analysis, 2nd edn. John Wiley, New York. 738 pp.

Riitters, K. H., O'Neill, R. V., Hunsaker, C. T., Wickham, J. D., Yankee, D. H., Timmins, S. P., Jones, K. B., *et al.* 1995. A factor analysis of landscape pattern and structure metrics. Landscape Ecology, 10: 23–29.

Rose, G. A., deYoung, B., and Colbourne, E. B. 1995. Cod (*Gadus morhua*) migration speeds and transport relative to currents on the north-east Newfoundland Shelf. ICES Journal of Marine Science, 52: 903–913.

Scalabrin, C., Diner, N., and Massé, J. 1994. Automatic shoal recognition and classification based on MOVIES-B software. *In* IEEE, Brest. Proceedings Oceans 94, Brest, 13–16 September 1994, pp. II 319–II 324.

Schumaker, N. H. 1996. Using landscape indices to predict habitat connectivity. Ecology, 77: 1210–1225.

SonarData. 2004. Echoview, version 3.10.129. http://www.sonardata.com/sonardata/WebHelp/Echoview.htm (accessed 12 July 2004).

Springer, A. M. 1992. A review: walleye pollock in the North Pacific – how much difference do they really make? Fisheries Oceanography, 1: 80–96.

Swartzman, G., Stuetzele, W., Kulman, K., and Powojowski, M. 1994. Relating the distribution of pollock schools in the Bering Sea to environmental factors. ICES Journal of Marine Science, 51: 481–492.

Teixidó, N., Garrabou, J., and Arntz, W. E. 2002. Spatial-pattern quantification of Antarctic benthic communities using landscape indices. Marine Ecology Progress Series, 242: 1–14.

Turner, M. G., O'Neill, R. V., Gardner, R. H., and Milne, B. T. 1989. Effects of changing spatial scale on the analysis of landscape pattern. Landscape Ecology, 3: 153–162.

Walline, P. 2007. Geostatistical simulations of eastern Bering Sea walleye-pollock spatial distributions, to estimate sampling precision. ICES Journal of Marine Science, 64: 559–569.

Wiens, J. A. 1989. Spatial scaling in ecology. Functional Ecology, 3: 385–397.

Wilson, C. D., Hollowed, A. B., Shima, M. S., Walline, P., and Stienessen, S. 2003. Interactions between commercial fishing and walleye pollock. Alaska Fishery Research Bulletin, 10: 61–77.

Appendix

Metrics used to characterize patterns in echogram segments (ESDUs)

The area of each echogram segment does not include the area below the seabed. The density of individuals in each pixel is assumed to be proportional to the volume-backscatter coefficient (S_v; MacLennan *et al.*, 2002). For percentile-based metrics, percentiles were calculated from the inverse of the empirical distribution function (Hyndman and Fan, 1996).

Patch density (PD): the number of patches divided by the area of the echogram segment.

Percentage of landscape (PLAND): the total patch area expressed as a percentage of the area of the echogram segment. PLAND can vary between 0, for empty echograms, and 1, for fully occupied echograms with no pixels below threshold.

Landscape division index (LDI): a measure of patch fragmentation. LDI is calculated as

$$\mathrm{LDI} = 1 - \sum_{i=1}^{n} \left(\frac{a_i}{A}\right)^2,$$

where n is the number of patches, a_i the area of patch i, and A the area of the echogram segment. LDI can vary between 0, when the entire echogram segment is covered by a single patch, and 1, when there is a single patch, one cell in area (McGarigal and Marks, 1995).

Patch area—10th, 50th, and 90th percentiles (A10, A50, and A90): percentiles of the patch-area distribution.

Relative patch area—10th, 50th, and 90th percentiles (RA10, RA50, and RA90): percentiles of the patch-area distribution, expressed as a fraction of the total patch area.

Biomass depth—10th, 50th, and 90th percentiles (BD10, BD50, and BD90): percentiles of weighted pixel depths. Weighted pixel depths (wpd) are defined as wpd = pd × S_v/sbd, where pd is the pixel depth, measured from the water surface, S_v the pixel volume-backscatter coefficient, and sbd the mean bottom depth in the current echogram segment.

Median distance to nearest neighbour (medNND): the median distance from each patch to its nearest neighbour, measured from patch edge to patch edge.

Volume-backscattering strength (S_v mean): volume-backscattering strength of the echogram segment (MacLennan *et al.*, 2002).

Acoustic density—10th, 50th, and 90th percentiles (S_v10, S_v50, and S_v90): percentiles of the distribution of the volume-backscattering strength (S_v; MacLennan *et al.*, 2002) of individual pixels. Only pixels above the threshold are considered.

Index of aggregation (Iagg): metric proposed by Bez and Rivoirard (2001) to measure aggregation using density data. It is calculated as

$$\mathrm{Iagg} = \frac{\sum z}{S(\sum z)^2},$$

where z is the density of individuals in each sample, in our case each echogram pixel, and S the area of the sampling unit (20 m^2).

Inertia in the horizontal and vertical directions (inx and iny): measurement of the variability of the spatial distribution of individuals. As proposed by Bez and Rivoirard (2001), inertia is the square deviation of the location of individuals from its mean. Inertia in the horizontal and vertical directions is computed as:

$$\mathrm{inx} = \frac{\sum\left(x - (\sum xz/\sum z)\right)^2 z}{\sum z} \quad \text{and}$$

$$\mathrm{iny} = \frac{\sum\left(y - (\sum yz/\sum z)\right)^2 z}{\sum z},$$

where z is the density of individuals, x the horizontal location of the sample, and y its vertical location.

doi:10.1093/icesjms/fsn087