



Fisheries and Oceans Pêches et Océans
Canada Canada

Canadian Stock Assessment Secretariat
Research Document 98/60

Secrétariat canadien pour l'évaluation des stocks
Document de recherche 98/60

Not to be cited without
permission of the authors¹

Ne pas citer sans
autorisation des auteurs¹

SPANdex

SPANS GEOGRAPHIC INFORMATION SYSTEM
PROCESS MANUAL FOR
CREATION OF BIOMASS INDICES AND DISTRIBUTIONS
USING POTENTIAL MAPPING

by

D. W. Kulka
Department of Fisheries and Oceans
P. O. Box 5667
St. John's, Newfoundland, Canada A1C 5X1

¹ This series documents the scientific basis for the evaluation of fisheries resources in Canada. As such, it addresses the issues of the day in the time frames required and the documents it contains are not intended as definitive statements on the subjects addressed but rather as progress reports on ongoing investigations.

¹ La présente série documente les bases scientifiques des évaluations des ressources halieutiques du Canada. Elle traite des problèmes courants selon les échéanciers dictés. Les documents qu'elle contient ne doivent pas être considérés comme des énoncés définitifs sur les sujets traités, mais plutôt comme des rapports d'étape sur les études en cours.

Research documents are produced in the official language in which they are provided to the Secretariat.

Les documents de recherche sont publiés dans la langue officielle utilisée dans le manuscrit envoyé au secrétariat.

ISSN 1480-4883

Ottawa, 1998

Canada

Abstract

Indices of abundance based on fishing success have been used for calibrating Sequential Population Analysis (SPA) models to project future yields and provide a basis for advice on fishery management. Incorporating such abundance indices in SPA assumes that catch rate (catch per standardized tow in the case of research vessel data or catch per unit effort for commercial data) is directly proportional to fish-stock abundance. To satisfy this assumption, the stock area must be stratified into sub-areas within which density could be considered constant. Potential mapping, a spatial analysis technique in SPANS (Tydac Inc.) is a versatile tool for analysing both commercial and research survey catch rate data. It can be used to derive point and variance estimates of biomass of a population. Commercial catch rates alone without a spatial dimension, even if standardized using a multiplicative model are simply a measure of local fish density, not biomass. Much like conventional methods for analysing survey data such as STRAP, and unlike previous methods for deriving commercial catch rate indices, the SPANdex method elaborated here accounts not only for variation in stock density but also for changes in spatial distribution of the stock. However, it differs from the conventional approach to analysis of survey data in that strata are derived (post-stratified) from fish density rather than depth. The technique can be used at the same time to study the distribution of the population. It provides the methods, not possible with more conventional statistical analyses to spatially analyze changes in distribution of marine species to the finest scale of the data and spatially relate these changes to environmental and other variables that might affect the population. This paper is a procedure manual providing step by step details of the method used to derive estimates of biomass and distributional analyses using potential mapping in SPANS.

Résumé

Des indices d'abondance fondés sur la réussite de la pêche ont été utilisés pour étalonner des modèles d'analyse séquentielle de population (ASP) des rendements futurs et formuler des avis pour la gestion des pêches. L'ajout de tels indices d'abondance à l'ASP suppose que le taux de capture (captures par trait de chalut normalisé pour les données des navires de recherche ou prises par unité d'effort pour la pêche commerciale) est directement proportionnel à l'abondance du stock. Pour respecter cette hypothèse, la zone du stock doit être stratifiée en sous-zones au sein desquelles la densité peut être considérée constante. La cartographie potentielle, une technique d'analyse spatiale du système SPANS (Tydac Inc.) est un outil polyvalent d'analyse des données des taux de capture des pêches commerciales et des relevés de recherche. Elle peut servir à obtenir des estimations ponctuelles de la biomasse d'une population ou de sa variance. Les taux de capture commerciaux utilisés seuls, sans dimension spatiale, même normalisés par modèle multiplicatif ne constituent qu'une mesure de la densité locale du poisson, non de la biomasse. Tout comme les méthodes conventionnelles d'analyse des données de relevés, comme la méthode STRAP, et au contraire des méthodes antérieures d'obtention d'indices de taux de capture commerciaux, la méthode SPANdex dont il est traité dans l'article tient compte, en plus de la variation de la densité du stock, de celle de sa répartition spatiale. Elle diffère cependant de l'approche classique à l'analyse des données de relevés en ce que les strates sont obtenues (post-stratification) à partir de la densité du poisson et non de la profondeur. Cette technique peut être appliquée au même moment à l'étude de la distribution de la population. Elle permet, ce qui est impossible avec les analyses statistiques plus conventionnelles, d'effectuer une analyse spatiale de la variation de la distribution des espèces marines à l'échelle la plus fine des données et de relier dans l'espace ces variations à des variables environnementales, ou autres, pouvant influencer sur la population. L'article décrit la procédure, étape par étape, utilisée pour obtenir des estimations de la biomasse et effectuer des analyses de distribution à l'aide de la méthode de cartographie potentielle du SPANS.

Introduction

The Spatial Dimension of Abundance

Catch per unit effort from commercial fisheries data or catch per standard tow from research vessel surveys has been used as indicators of fish stock abundance. Incorporating such abundance indices into the Sequential Population Analysis models assumes that the commercial or survey catch rate is directly proportional to stock abundance. However, catch rate, even though it reflects density of the resource in the vicinity of the fishing gear, can be misleading. As long as fish are densely gathered, fishing fleets can experience high catch rates, regardless of how small in extent the concentration of fish may be. In fact, the areal extent of the concentrations may change dramatically over time while catch rates remain stable or even increase. Catchability has been reported to vary as a function of stock abundance in both groundfish and pelagic fish stocks as summarized in Kulka et al. (1996). It may also be affected by oceanographic fluctuations (e.g., Fréchet and Gagnon 1993) that may result in spatial variation in the stock.

Catch rate data from fishing sets simply on their own reflect only local fish density unless spatial components i.e. extent of the stock and density variations within the stock boundaries are constant. This is rarely the case for marine species. Thus, a spatial as well as a density component is required to derive reliable estimates of abundance. This is possible if the data (fishing sets) containing catch rate data are geo-referenced. Such catch rate observations with associated latitude and longitude can be converted to a continuous surface that depicts fish density over the stock area. Spatial techniques such as potential mapping, contouring, aggregation or kriging yield surface (continuous) interpretations of distribution from the point data and facilitate analyses of spatial patterns and biomass. The suitability of these various techniques for the types of analyses addressed in this paper is discussed.

To satisfy the assumption of the proportionality of catch rates to stock abundance, Gulland (1955) proposed compartmentalization of the total area of stock distribution into sub-areas within which density could be considered constant. For research surveys, a random stratified design referred to as STRAP (Smith and Somerton, 1981) is commonly used to stratify the survey area by depth. However, fish density within depth strata would be homogeneous only if the fish distribute by depth. That is, the assumption in this model is that density distribution of the fish is related to depth. This may not necessarily be the case since fish concentrations are often found to be discontinuous along lines of depth.

An alternate method elaborated in this paper, using post-stratification, makes use of the distribution of the fish to create density strata covering the survey or fishing area. This approach accommodates inter-annual changes in distribution both in terms of extent of the stock and any density variations that may occur. Using potential mapping (a GIS, Geographical Information Systems technique in SPANS, TYDAC Research Inc., Anon. 1997) to create density surfaces, this paper describes in detail the method used to derive biomass and analyze distribution. Potential mapping is most appropriate

for the analysis of interval or ratio point data that represents a non-continuous phenomenon. Non-continuous phenomena are data typified by a high degree of variance and uneven distribution (Anon 1997). Catch rate data are particularly suited for analysis by this method since they comprise ratio point data (catch/effort) that have a high degree of variance and an uneven distribution. Fish school behaviour results in dramatic variation from set to set. Kriging and contouring on the other hand assume that adjacent point values are correlated with each other spatially.

The method is well suited for spatially analyzing research survey data because it converts point estimates, in this case individual set catch rates, into continuous surfaces (density subareas) that perform as survey strata. Because observed fish density is used as the stratifying variable, it potentially reduces within strata variability of density. Extent and location of density constant subareas is allowed to vary according to distributional changes of the fish. That is, the technique makes use of the geo-referenced catch rate data to define location and extent of the density strata. The strata vary over time taking into account stock distributional shifts with a resulting lower within strata variability. In the case of research survey data, number of sets per strata, 32 in the case of NAFO Div. fall surveys, is greater and less variable than for STRAP (2-26). This likely reduces the estimate of variation.

The method of biomass estimation described here is updated from (Kulka *et al.* 1996) who used it to analyze commercial fisheries data. It is adapted to analyze research survey catch rate data. This tool also provided the opportunity, not possible with more conventional statistical analyses to analyze changes in distribution and biomass of marine species to the finest scale of the data and spatially relate these changes to variables such as temperature that might influence the populations. The examples in this manual utilize northern cod (*Gadus morhua*) data to demonstrate the procedures. However, the methods are applicable for any species where geo-referenced data exists. Past studies using potential mapping have been done for shrimp, crab, white hake, skate, monkfish, wolfish and witch among others (i.e. Kulka 1990, Kulka *et al.* 1995a, Kulka *et al.* 1995b, Kulka and Deblois 1996, Kulka and Power 1996, Kulka *et al.* 1996, Kulka 1997, Parsons and Veitch 1996, Parsons *et al.* 1996 and Wroblewski *et al.* 1995).

Methods

The Data Source

The extent to which a spatial dimension can be incorporated in analysis of commercial catch rate data and distributions depends on the geographical scale in the data. Gulland (1955), in his example of North Sea haddock fishing, was constrained by the data to using large (3000 km²) subareas. However, Paloheimo and Dickie (1964) proposed that the relationship between catch rate and stock abundance is a function of the number, size and density of fish schools. Thus, a spatial scale at the level of fish schools would be required to take full account of distributional changes in catch rate analysis. For groundfish, there is limited information on schooling behaviour. Rose (1993) measured schools of spawning and migrating cod off eastern Newfoundland that extended over about 2.5 km to 20 km, equivalent to a much smaller unit of area than used by Gulland. A typical commercial set by a trawler fishing cod in this area extends over 10-15 km and sets can be very densely spaced. In

contrast, research survey sets are much more widely spaced, between 20 and 70 km apart. Analysis of catch rates on a set by set basis seems necessary to provide spatial resolution at the level of fish schools. Thus, when data fisheries are geo-referenced, recorded set by set and approximate the scale of the fish distributions, a spatial dimension can be incorporated into the estimation of abundance and detailed distributional analyses are possible.

In order to create density surface maps, a geo-reference, generally latitude and longitude is required. Set by set geo-referenced fisheries data can be obtained from two sources, research vessel surveys and commercial catches as collected by fisheries observers. Both comprise point catch and effort data. Purchase slip information used to report landings do not include set by set details and fishing log data are not always accessible. The following sections describe the two available sources of detailed catch and effort.

Research - Catch, effort and bottom temperatures along with other data have routinely been collected during bottom trawl research surveys for the various areas around Newfoundland. For this study, fall research trawl set data collected from 1983 to 1996 in NAFO Divisions 2J, 3K and 3L, were used as the example in the application of the method used to estimate abundance and relate the distributional patterns observed to bottom temperature. Doubleday (1981) provides a summary of the stratified-random survey design adopted after 1970 to survey the waters of the Atlantic regions.

Commercial - All foreign and many domestic vessels fishing in Canadian waters carry fishery observers. Their job is to record catch and bycatch, collect scientific samples, and monitor fishing practices and compliance with regulations (Kulka and Firth, 1987). This work generates voluminous data on the catch and effort. Observers' reports include details of the catch, bycatch, discards; descriptions of the vessel's gear among other information related to the fishing operation. Observers also take diverse scientific data and samples including length measurements from the catch that can be analysed spatially. Such data has wide application in science and management, and in the fishery itself. The current example illustrates one such application using catch per hour recorded from thousands of fishing sets.

The Application of GIS

The association of a geo-reference with variables such as catch rate or temperature in a data set facilitates spatial analysis. The strength of the SPANS GIS is the ability to convert point data to surfaces that allow mapping and spatial modelling of the data. Values at a location can be averaged or totalled, or more complex calculations such as variance, surface complexity, rate of change over an area, and other calculations are possible. The data can also be overlaid and analysed in relation to other spatial data sets or mapped information, such as temperature, depth, salinity, or bottom type and in relation to digitized features such as shoreline and bathymetry.

The following sections provide a description of one particular application. They provide step by step procedures for conversion of point data to density surfaces and subsequent distributional analyses using potential mapping. Many other functions including surface modelling are also available in SPANS (Anon. 1997).

Importing the Data

This section demonstrates the conversion of point files of fisheries catch data to geo-referenced layers in SPANS. Fishing set data are the raw material for creating density surfaces in the form of potential maps. They can be used to illustrate stock distribution and are the basis for spatial and density estimates for calculations of biomass. Each line of the data to be imported into SPANS must contain at a minimum, a latitude and longitude value, plus catch rate. Each observation is a fishing set. The file is typically ASCII, with fields separated by a space. SPANS can also accept other common geographic formats as described in Anon (1997). An example of (a portion of) a typical geo-referenced input ASCII data file is shown in Table I.

Table 1 - Point file in ASCII as input into SPANS. The first two columns, latitude and longitude, define geographic location. Subsequent columns define the data to be analysed, in this example catch per hour and depth. The file must not contain headers.

53.2833	-52.2833	6.4500	423
53.2333	-52.2167	4.0000	437
53.2500	-52.3167	0.5680	423
53.3833	-52.3833	3.6961	427
53.2833	-52.4000	6.1500	432
53.1833	-52.2000	2.7000	440
53.2667	-52.3167	6.5945	432
53.3333	-52.4500	1.2306	432
53.2667	-52.2000	4.1038	486

The number of observations in a data set may be few to many thousand depending on the source. Research survey data generally contains hundreds of observations, commercial fishing data in the thousands. SPANS can accommodate tens of thousands of observations in a single data set. Latitude and longitude may be expressed in decimal degrees as shown in the example or can be imported as degrees and minutes, in several variations. Longitude must be negative to set the geo-reference to the western hemisphere.

Before analysis in SPANS, the file must first be prepared for importation into SPANS. This process involves creating a header file (with .tba extension). This is done externally of SPANS using the PNTBA command (included with the software) as follows:

```
PNTBA CR8705.PNT
```

or by hand in a text editor. PNTBA, invoked at the C:> prompt in a DOS window, external of SPANS, creates a table header from the ASCII point file. The result is a new file (Table 2):

Table 2- Creation of a header file with .tba extension for the importation of a data point file into SPANS

```

ID CR8705
TITLE May 1987 Cod Catch Rate.
TABTYPE 4
FTYPE free
KEYFIELD 0
DATAFILE CR8705.DAT
1 1 7.4 0 LAT Latitude
2 1 7.4 0 LONG Longitude
3 3 7.4 0 CHR Catch Rate (tonnes/hour)
4 3 3.0 0 DEPTH Depth (m)
DATA
  
```

This new file is assigned the same name as the original point file but with a .tbb extension. It can then be edited manually to change the TITLE, and each of the ATTRIBUTE lines before importing the data into SPANS or this can be done at a later stage from within SPANS in the open layer file. The procedure automatically assigns the name of the original point file to the ID, TITLE and DATAFILE lines. Attr0 through Attrxx are the default variable names. These can be renamed manually to reflect the variable names i.e. Lat for Attr0, Lon for Attr1, CHR for Attr2, Depth for Attr3 as shown in Table 2 or they can be renamed later from within SPANS.

Once the PNTBA routine has been run and after invoking SPANS, the study area (subdirectory) containing the data is opened by the following command from the dropdown menu:

FILE / OPEN / STUDY AREA

The point file import procedure is carried out within SPANS by the following command:

FILE / IMPORT / SPANS POINTS

A file selection box appears and the appropriate file, in this example CR8705 is selected. The following box appears:

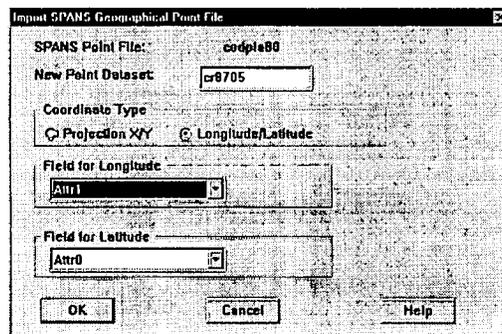


Figure 1- Selection of input data type and fields for latitude and longitude.

Select the **Longitude/Latitude** radio button and put Attr1 as Longitude and ATTR0 as Latitude in the **Field for pick boxes** (or Long. and Lat. if already renamed in the .tba header file). This causes SPANS to bring the point data (file pair .pnt as the data, .tba as the matching header) into SPANS by creating a combination binary .tbb file. This step assigns to each record an internal system coordinate in the form of a new variable called a Morton number using the latitude/longitude values. The new, internally geo-referenced .tbb file is referred to as a *point layer* and can now be displayed on a map or analysed in association with any other information in the SPANS subdirectory (geographical *universe*). Further details on importing point and other types of files such as vectors and rasters into SPANS and the various types of geographical layers can be found in Anon (1997).

Using the command:

FILE / OPEN / LAYER

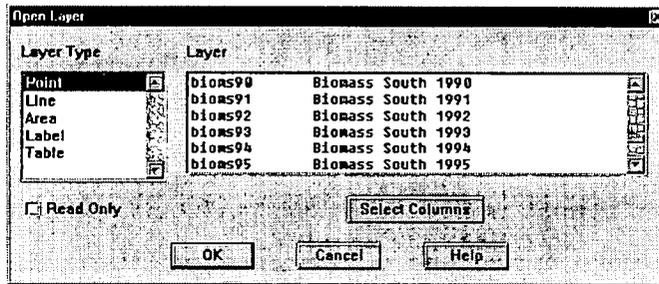


Figure 2- Selection of layer type, in this case points and file containing those points.

opens a layer selection box (Fig. 2). Under **Layer Type Point**, choosing the appropriate **Layer** name (.tbb file) displays the newly imported points on the screen in their proper position along with a geo-referenced spreadsheet (Fig. 3):

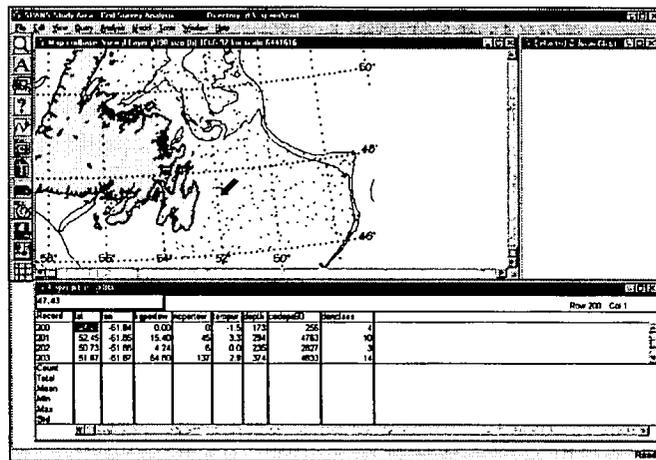


Figure 3- Opened point layer showing a spreadsheet of the data, a map with the point locations and a chart of selected variables.

Query by example, using equations defining variable relationships will be explained in greater detail later as it is useful in sub-setting the data points into the various density categories. Further details of these functions can be found in Anon, 1997.

Layers and maps

There are various layer types in SPANS. They are point, line, area, label, table, raster and quadtree. The first five can be opened in a spreadsheet (Fig. 3 and 4 illustrate opened point layers) and have attributes that can be edited. The last two, rasters and quadtrees form continuous surfaces whereas points and vectors occupy a location, but have no area. Analyses conducted on fisheries data transform the point layer composed of fishing sets to a quadtree comprising a fish density surface. Refer to Appendix 1 (quadtrees) and Anon. 1997 for detailed descriptions of the various layer types.

Layers can be combined as maps. In SPANS, a "map" is considered to be an image that contains one to several layers such as point (set locations), vector (lines such as depth contours) or continuous surfaces i.e. raster (satellite images) or quadtrees. Various annotations such as titles, legends, or lat./long. grids, can also be added to the maps. Completed maps can then be printed directly or exported as Windows metafiles (.wmf extension) to other graphic applications for further editing.

Potential mapping

The following is an edited extract from the SPANS manual (Anon, 1997).

The SPANS potential mapping surface function converts a point data layer to a surface representative of a selected attribute, referred to as a *Z-value*, of the points. The surface derivative models available are the weighted average of Z-values (mean of the attribute), slope, aspect, X- and Y-derivatives, angle of incidence and sum. This function uses an averaging technique and does not create any new values greater than the maximum value of the input data nor less than the minimum value of the input data.

The potential mapping surface derivative function generates a surface by applying a sampling radius to each point in the data layer. No output values are calculated for areas lying outside of any sampling radius. Thus the user controls the surface area through selection of sampling radius size. This function generates output in the quadtree data format (App. 1) and during the enquadding process a "grid" representative of the quad level is applied to the surface described by the study area view. The value for each cell in the output quadtree is calculated based on the value of each point whose sampling radius overlaps the centre of that quad cell.

This function is most appropriate for interval or ratio point data that represents a non-continuous phenomenon. Non-continuous phenomena are data typified by a high degree of variance and uneven distribution.

Catch rate data are ratio point data that have a high degree of variance and an uneven distribution.

The following function generates the potential map surface model for weighted average of z-values (mean of the attribute):

$$\sum (f_i w_i(d) z_i) / \sum (f_i w_i(d)) \quad \text{Equation 1}$$

Where:

z = Z-value

I = each point at the centre of the sampling radii that overlap the quad cell

f = frequency

$w(d)$ = distance weighting

Creating a potential map - Once the point data are accessible, potential mapping is used to transform the point data (catch rate by set) to a surface depicting differential densities of fish (catch rate subareas). The resulting surface is in the form of a "quadtree". Refer to Appendix 1 for a definition of a quadtree.

The following command:

MODEL / POTENTIAL MAPPING / SURFACE - invokes the following input screen:

Figure 5- Input screen for creating a potential map.

Input - From the first category, select a point layer containing the catch rate data. In the *Z-value Field*, choose the variable that describes catch rate. For this example using survey data, **kgpertos** is selected as the point measure of fish density.

Distance and weighting functions affect the manner in which the point data used in the calculation, point interactions and distance of influence of the points from the *Point Layer* jkl85.

Selecting a function - Spatial analysis *Functions* are set under Distance and Weighting Options box. **Weighted Average of Z-values** (i.e. mean of **kgpertos** inside the circle) is the function required to fish density surfaces (other available functions such as **Sum**, **Slope** and **Aspect** are generally applicable only to land surface analyses).

Selecting a circle function - Potential mapping performs the point to surface conversion by placing a circle around each point and averaging the values of all points that fall within the circle. This is

illustrated in the following Fig. 6.

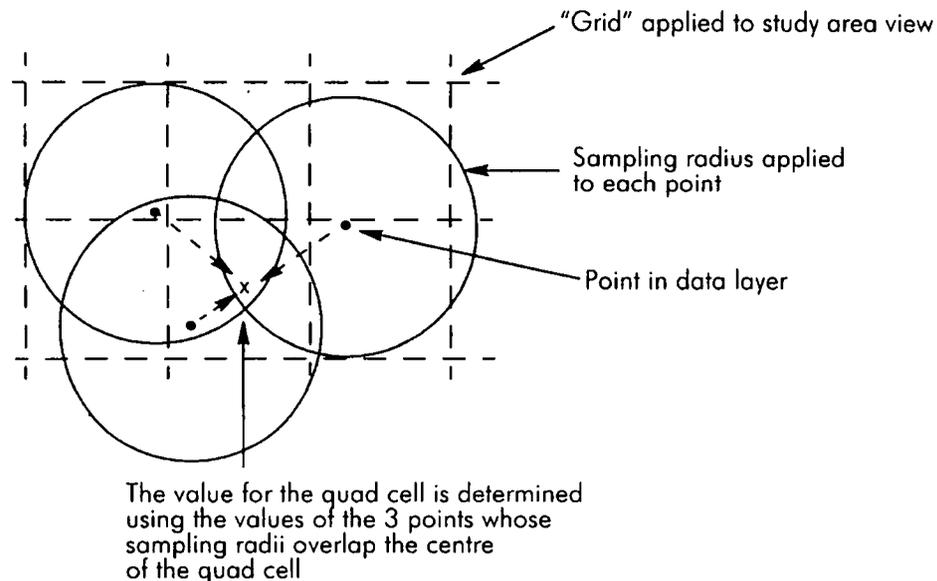


Figure 6- Sampling radius applied to each point (fishing set kg per tow) (after Anon. 1997).

A circle is placed around each point. This process performed on all of the points effectively smoothes the data at its maximum resolution, i.e. at the spatial distance of survey fishing sets, by averaging the catch rates of all fishing sets that occurred within the chosen radius of the scanning circle. A further averaging takes place where the circles overlap and the resulting values are assigned to the underlying quadcell (a quadcell being a variable sized raster dividing the study area into a fine grid, App. 1). Fig. 6 shows circles that are slightly larger than the quadtree. For fisheries applications the circle size is usually much larger than the quadtree. Using the smallest possible quadcell size (quad level 15, Fig. 5, yields a minimum quadcell size of 0.04 km^2 in the 2J3K1 survey area) means that the circle is much larger than the quad cell. For example a circle with radius 31 km as used for fall survey analysis is more than 70,000 times larger than the quad cell. Thus, hundreds or thousands of quadcells comprise each circle segment. An integration of classified quadcells forms a density surface analogous to Gulland's subareas. Refer to App. 1 and Anon. (1997) for a detailed technical description of the quadtree structure. How the quadcells are classified is described below.

Inner and *Outer Radius* has the greatest effect on the outcome both in terms of biomass calculations and distribution as illustrated later. These functions define the size of the scanning circle and how the point data in the circles are to be weighted from the centre out. Inner radius plus outer radius equals the total sampling radius. Selecting a larger circle radius for the *Outer Radius* and a decay rate different from 1 causes the points in the circle to be given different weights. Fig. 7 explain how this works:

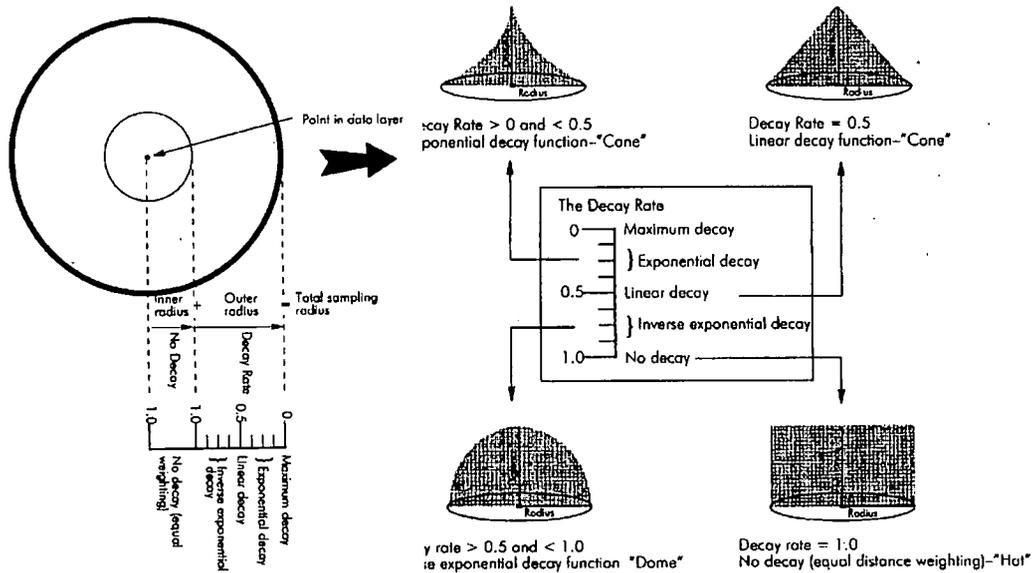


Figure 7- Selecting weighting and circle size using inner and outer circle radii (after Anon, 1997).

For fisheries data, a Decay Rate of 1 is normally used such that all points within the circle are given equal weight. Weighting is generally used where circle size is large, points are dense and there is a reason to give points on the circle periphery less weight when calculating the mean. The decay rate is only applied to the outer radius. Thus, if a sampling radius of 31 km with no weighting of point values within the circle is required then outer plus inner radius must add to 31 and Decay Rate must be set at 1. Refer to Anon (1997) for further details.

Selecting a scanning circle size – The optimum size of the scanning circle is one that matches the scale of the data. Kulka *et al.* (1996) from an analysis of commercial catch rate data found that increasing radius size increased the resulting estimate of biomass because of the increased surface size due to an edge effect (areas surrounding the fishing area increased with larger circle size).

For survey data, the distance between sampling points (20 to 70 km) exceeds the size of fish schools and length of tow. In this case, distance between sets determines the circle radius, such that circle size is increased until there were no gaps in the resulting surface and the entire survey area is covered. This is done using several runs where increasingly larger circle sizes are tested. Size should be increased until the total surface area created stabilizes. The exception where there are known areas not sampled such as near the coast in the earlier years of the survey in NAFO Div. 2J3KL. In the case of random stratified research surveys, circle size is set such that total area approximates the total area of the depth strata. This would be less than the point of stabilization because of unsurveyed areas near the coast. The affect of varying circle size on area and biomass estimates was tested. Potential maps were created from a series of circle sizes for 1985, namely circles with 2 to 42 km. radii by 2 km intervals. A scanning circle with radius 31 km. was chosen as the smallest that would create a surface with minimal gaps. Because the surveyed area was very similar from year to year, the study area periphery was set by a 'cookie cut' (referred to as a basemap cut in SPANS) such that the resulting density surface was constrained on all sides by the land/water boundary, the 1000 m

depth contour and the 2J and 3L NAFO Division lines. Within the bounds of NAFO Divs. 2J, 3K and 3L, this resulted in relatively constant surface size of the study area. For 1985, it was 354,912 km². Area varied slightly from year to year because of variation in extent of un-surveyed grounds close to the coast. This 31 km circle was chosen to yield an survey area size that approximates the area of the depth strata (354,000 km²) used in the random stratified design.

An area and biomass curve for the various circle sizes is shown in Fig. 8.

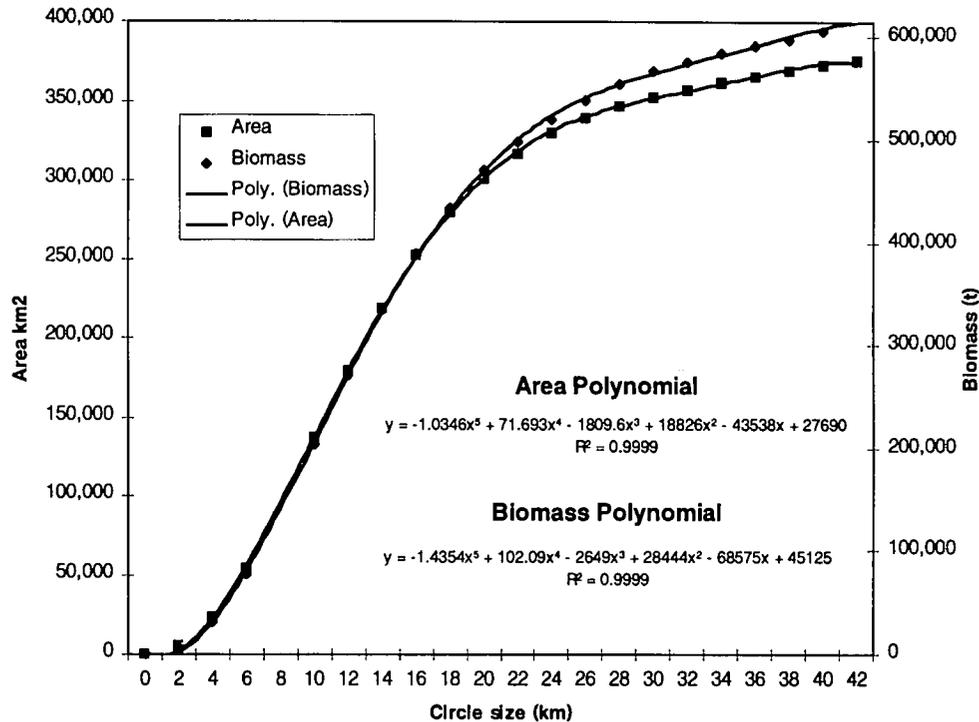


Figure 8- Affect of increasing circle size on area of the map and biomass estimate produced from survey data.

For fall survey data in 2J3KL, a circle of about 31 km. creates a surface with no gaps except close to the coast where there were no survey sets. These variations in set locations along the fringe of the survey area cause minor fluctuations in total area from year to year depending on how close to the coastal the sets were. However, an adjustment factor can be applied to standarize areas across years. These adjustments were found to make very little difference to the biomass calculated.

Densities of cod adjacent to the non-surveyed areas were generally found to be low. The exception for the fall cod surveys in NAFO Divs. 2J3KL is along the coast of Labrador where the North concentration (Kulka, 1998) has not yet completed its offshore migration. Increasing circle size under this circumstance can cause the biomass estimate to increase proportionately faster than the area added as can be seen in Fig. 8. This because high density values are extrapolated to non-surveyed areas. If density is low at the edge of the surveyed grounds, the shape of the area and biomass curves will match more closely.

If an estimate of biomass is required to include non-surveyed areas such as near the coast and in the bays, a larger circle size can be used. The assumption here is that the density of fish in the non-surveyed areas is the same as the adjacent surveyed areas. For example, in the case of the 1985 fall survey, if it can be assumed that cod density in the non-surveyed areas is similar to the adjacent surveyed area then the biomass estimate of 616,662 t derived from a potential mapping procedure using a 42 km circle is reasonable for the 375,739 km² area covered. The 42 km circle creates a surface that is 6 % bigger than the 354,912 km² area created by the 31 km circle used to approximate the random stratified survey area and a biomass estimate 13 % larger.

For commercial fisheries data, different criteria can be used to determine circle size. The fishing sets are usually much more closely spaced allowing for a smaller scanning circle, if required. For example, Kulka et al. 1996 in an analysis of 2J3KL cod winter fishery data chose a sampling circle of 7 km to match the scale of the data. The circle diameter chosen was equivalent to the average length of a fishing set and was on the scale of the size of cod schools. Thus, for that analysis the potential map could resolve fish schools. This is not possible with the much more widely spaced survey sets. Both the data and the application must be considered when choosing the best circle size. Changing circle size changes magnitude of the estimate of biomass. However, a circle of 31 km and one of 42 km applied to survey data for example yield biomass estimates of different magnitude but with the same trend.

It was noted earlier that from an analysis of commercial catch rate data Kulka *et al.* (1996) found that increasing radius increased the resulting estimate of biomass because of the increased surface size due to an edge effect (areas surrounding the fishing sets increased with larger circle size). However, this problem can be avoided by a 'cookie cut' (referred to as a basemap cut in SPANS) such that the resulting density surface is constrained on all sides. For research surveys, because the surveyed area is very similar from year to year, the study area periphery can set to a standard area that approximates the area of the depth strata used in the random stratified design. In the case of 2J3KL surveys, the basemap perimeter would be the land/water boundary, the 1000 m depth contour and the 2J and 3L NAFO Division lines. In the case of commercial fisheries data the basemap might be set as the extent of the fishing grounds. Creating and applying a Basemap to the view is described in Anon (1997).

Further increasing circle size beyond where surface size increases (constrained by the basemap cut) increasingly smoothes the data to a point where a single circle encompassing all points results in a single stratum with an average catch rate (density estimate) of all points. This removes any spatial definition of density. Thus, the smallest circle that can provide complete coverage of the survey area, the better the definition of spatial variation of fish density within the survey ground.

Maximum neighbours can be set at a high number so that all points in the circle are included in the calculation. *View* is user defined and in this case refers to the survey area i.e. NAFO Divs. 2J3KL. The entire universe is the Canadian Atlantic. The view *jl* (previously user defined), for example, includes only 2J3KL. Views are defined by:

VIEW / DEFINE NEW (or using the Spyglass button seen in upper left of Fig.4)

Finally, *New Quadtree* under **Output** is used to define the name of the quadtree file (layer) being created. This can be any 8-digit name. *Quadlevel* should be set at 15 (top level) to provide the smoothest graphical output and *Minimum Quadlevel at 12*. The quad level determines the resolution of quadtree output. The higher the quad level, the smaller the cell size and, therefore, the better the resolution. Higher quadcell size results in longer computer processing time. This not as big an issue with today's high speed computer. Refer to App. 2 and Anon (1997) for further details.

Classification of fish density - Once all the categories on the potential mapping surface input screen have been completed, pressing **OK** brings up the screen illustrated in Fig.9. This screen allows the user to categorise density into levels or strata. A classified histogram allows the user to visualize area proportionality in the area of the strata.

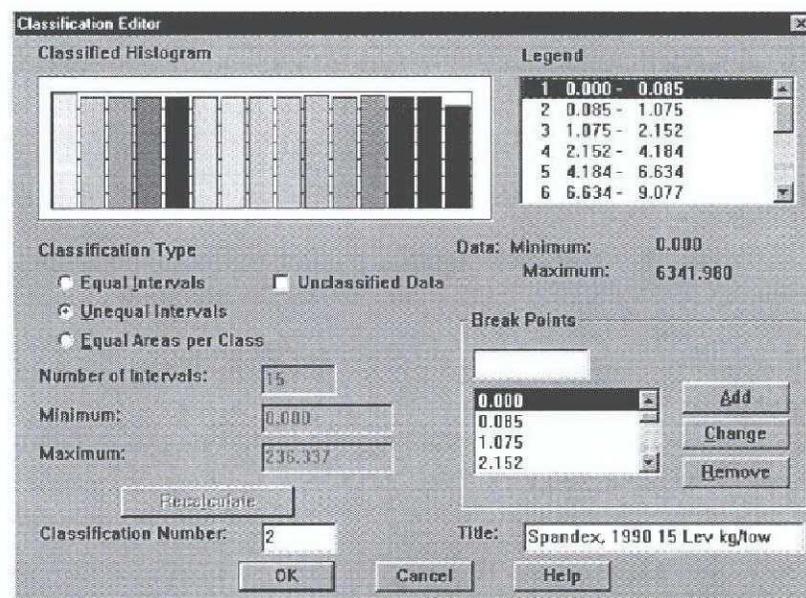


Figure 9- Classifying density subareas with a user-defined legend.

To set up density strata for SPANdex, under **Classification Type**, select *Equal Areas per Class* and specify 15 as the *Number of Intervals*. This sets up 15 density strata of (approximately) equal area. SPANS automatically set the legend break points. These are the density categories that will be used to calculate biomass. A different classification (legend) is used for each year such that each year has the same number (15) of equal area strata. For distribution maps, a different approach is required. A single legend set for a baseline year is used across all years to show inter-annual variation (elaborated later). The user has complete control over how the legend is structured and can even custom design break points using the add, change and remove functions.

Once the legend is set, and **OK** is pressed, in the map view. Set point locations appear and then a series of circles builds around each point as shown in Fig. 10. Finally the quadtree classified by selections made on the Classification Editor starts to build on screen using the default colours. These can be modified later when building a map if the quadtree is being produced for display purposes.

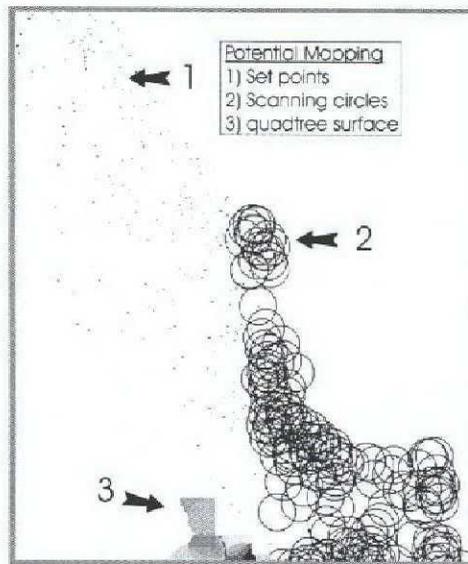


Figure 10- Formation of the potential map showing set point locations, scanning circles and the surface being created.

There is an intermediate step, explained here to assist in understanding potential mapping method. There is creation of unclassified circle segments not seen by the user since the procedure automatically groups the many areas created according to the classification scheme set up by the user (Fig. 7) while it creates the map. This intermediate stage is shown in Fig. 11.

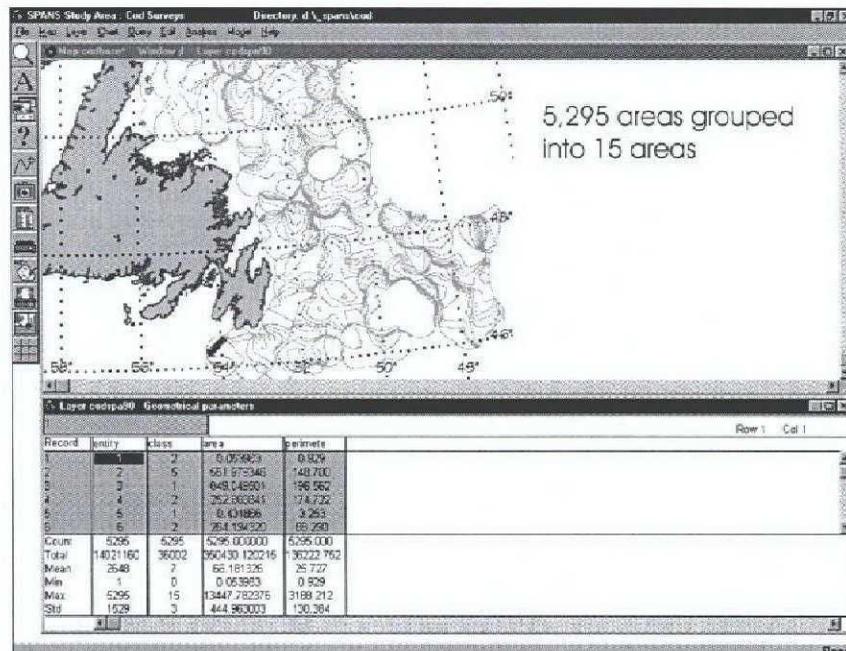


Figure 11- Individual areas created by the potential mapping procedure.

An unclassified potential map (of 1990 data) comprises a total of 5,295 areas (different for each data set) is temporarily created, each area with a different value of density. These circle segments are the result of hundreds of overlapping circles, each segment having a unique averaged value of fish density. These many areas are reduced to 15 levels (or however many levels of classification were specified by the user) in the final product.

All classified areas are not necessarily contiguous since the fish may form concentrations at several places within the survey area. For example, in 1983 (typical of all years of fall survey data), there several areas of high densities of cod separated by lower densities as shown in Fig. 12. This comprises the final product, the classified potential map. A 7 level grey shade map of fish density expressed in kg. per standard tow for is used to illustrate the density layer created by potential mapping.

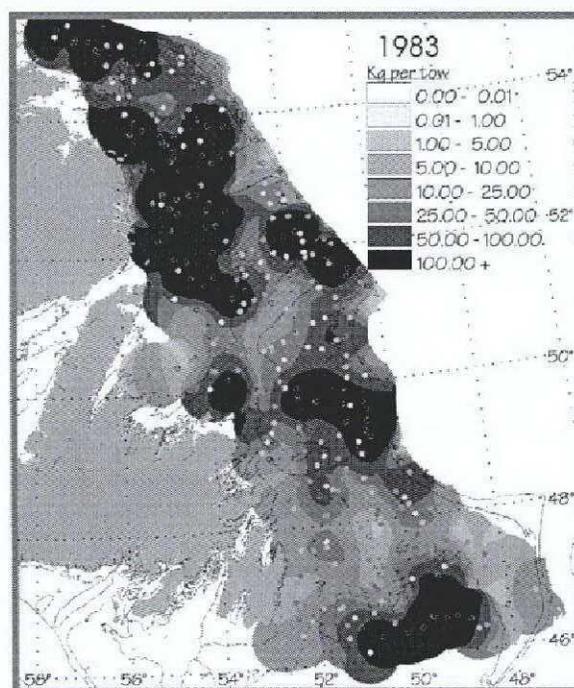


Figure 12 - Potential map of cod density for 1983. Darker areas denote denser areas of dense cod.

SPANdex -For the purpose of biomass estimation using the SPANdex (SPANS index) method (adapted from Kulka *et al.* 1996), in the case of 2J3KL fall surveys, each stratum covered about 6.7% or 1/15th of the survey area across all years as produced from the potential mapping procedure described above. Being related to a complex distribution of fish, each stratum derived by the potential mapping procedure was not necessarily contiguous and each one varied from year to year in terms of shape and location, dependant on the distribution of the fish.

This approach to stratification of the survey area differs from the STRAP method (Smith and Somerton 1981) in several important ways. The STRAP stratified random survey design consists of many more strata (76 in fall NAFO Div. 2J, 3K and 3L) and they are fixed in space over time. They

are delineated primarily by bottom depth rather than fish density and each stratum is contiguous. Also, each SPANdex stratum contains an average of 32 survey points (total average of 487 points per year for fall surveys in 2J3KL) as compared to between 2 and 26 for the STRAP random stratified design.

Similar to STRAP, a mean catch rate is calculated for each SPANdex stratum from the all of the points (sets) that fall within the bounds of the stratum. The mean density and area of each class is used to estimate a biomass index using the formula from Kulka *et al.* (1996). The formula for biomass calculation, Equation 2 below, is an areal expansion based on the area of each stratum and the average catch rate within the stratum, much like STRAP:

$$B = n \sum \{(a \times c) / [(t \times w) / h]\} \quad \text{Equation 2}$$

where

B=biomass index

n=number of catch rate classes

a=area of catch rate class (km²)

c=mean (or midpoint) of catch rate class (t/hr)

t=average tow length (km)

w=wingspread of net (km)

h=average number of hours per tow.

For the fall research survey example, W, wingspread of net was set at 0.014 km for 1983 to 1994 and 0.01681 km. for 1995-96 when the gear was changed from Engel to Campelen. H, average number of hours per tow was constant at 0.334 given the standard 20 minute tows. As noted above, the resulting area of the survey as derived from SPANdex varies slightly from year to year due to changing distribution of survey sets. A biomass standardized to the random stratified survey area can be calculated by multiplying ratio of the random stratified area to the SPANdex area by the unadjusted SPANdex biomass. The same principles apply to commercial fisheries data although certain variables, constant across years for surveys vary. The following Table 4 shows an example of how fishing behaviour changed in the 2J3KL winter cod fishery. It is important that these changes are reflected in the biomass calculations.

Table 4 - Statistics for calculation of biomass for commercial data.

Year	Tow Length, t (km)	Duration, h (hrs)
1980	16.54	2.30
1981	14.69	2.10
1982	12.73	1.80
1983	13.17	1.90
1984	11.71	1.70
1985	9.28	1.30
1986	11.89	1.70
1987	10.47	1.50
1988	10.04	1.50
1989	10.88	1.60
1990	11.39	1.80

Biomass from Equation 2 can be calculated in a spreadsheet. The example in Table 3 illustrates the biomass calculation. B, a and c refer to the equation parameters.

Table 3- Biomass calculation for SPANdex using 1990 fall survey data in 2J3KL.

Density Class	Legend		c			a		B	Set Count	% of Biomass	Cum. % Biomass	% of Area
	Lower	Upper	Density Mean	Stdev	Area(%)	sq km	Biomass (t)					
1	0.0	0.2	0.0	0.0	9.0	31,473	0.00	22	0.0%	100.0%	100	
2	0.2	1.1	0.4	0.7	6.4	22,403	168	31	0.0%	100.0%	90.99	
3	1.1	2.3	1.4	1.5	6.4	22,464	660	21	0.1%	100.0%	84.58	
4	2.3	4.4	2.0	2.0	7.5	26,125	1,138	34	0.2%	99.9%	78.15	
5	4.4	7.0	3.3	3.8	7.2	24,982	1,763	35	0.3%	99.7%	70.68	
6	7.0	9.4	6.7	6.0	6.3	21,957	3,174	32	0.5%	99.4%	63.53	
7	9.4	12.6	8.4	5.5	6.3	21,924	3,955	26	0.6%	98.9%	57.25	
8	12.6	17.5	12.8	9.7	6.8	23,661	6,517	28	1.1%	98.2%	50.98	
9	17.5	25.0	18.8	11.5	6.6	23,076	9,291	33	1.5%	97.1%	44.21	
10	25.0	36.1	20.1	18.0	6.4	22,374	9,622	32	1.6%	95.6%	37.61	
11	36.1	57.1	46.3	26.5	6.5	22,730	22,554	22	3.7%	94.0%	31.21	
12	57.1	87.8	38.2	34.6	6.4	22,455	18,409	28	3.0%	90.3%	24.71	
13	87.8	127.4	68.5	70.3	5.7	19,885	29,209	33	4.8%	87.3%	18.29	
14	127.4	242.1	194.5	149.2	6.2	21,662	90,351	33	14.8%	82.5%	12.6	
15	242.1		858.0	1,251.1	6.4	22,387	412,018	28	67.7%	67.7%	6.4	
			100.0	349,558.0		349,558.0	608,830	438				

As shown in Equation 2, biomass is a function of fish density, in this example kg. per standardized tow and a spatial component, in the example, area in each density class expressed in km². Mean (c), the fourth column in Table 3 is the mean catch per standardized tow for all sets that fall within the specified density class (col. 1). The mean is derived by the following command in SPANS:

ANALYSIS / OVERLAY POINTS

The quadtree layer (*codspa90*), the classified fish density surface derived through potential mapping and the original point layer (*jkl90*) are selected as shown in Fig. 13:

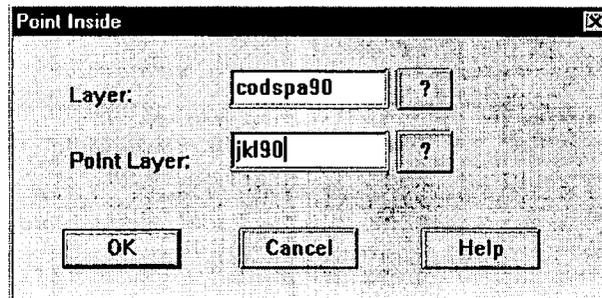


Figure 13 - Selection of the surface and point layer for matching points to classes in the surface density layer.

This procedure overlays the quadtree data layer with a point data layer. The result is the creation of a new column (denclass) in the point layer containing the class of the quadtree within which each point falls. The new column is shown as the last column on the spreadsheet in Fig. 14:

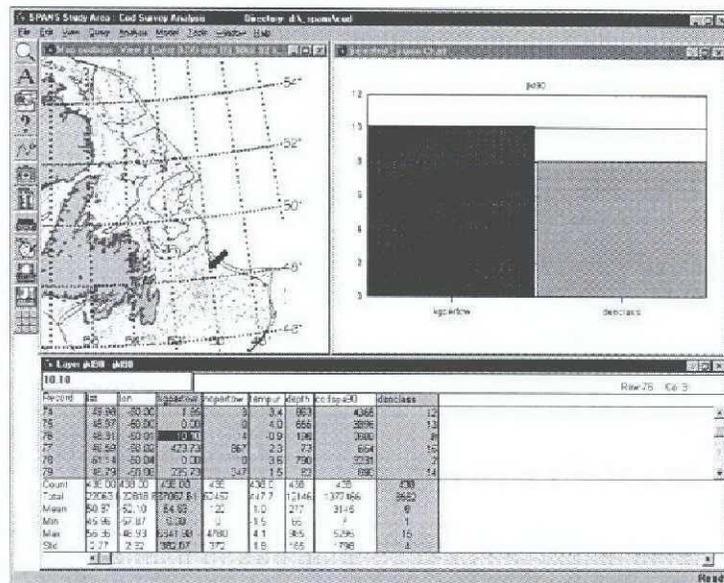


Figure 14 – Addition of a column denclass to the point layer file.

Denclass is a number between 1 and 15 referring to the fish density classifications and is appended to each line on the spreadsheet that is a point observation. The average catch rate can then be calculated using several methods. One way is to export the point layer file from SPANS using the following command:

FILE / EXPORT / LAYER

In the example, the file name jk190 would be selected for export. The resulting ACHII file can then be imported into a spreadsheet for the sorting and calculation of mean kg per tow values by density class. These values can then be placed into the SPANdex calculation as shown in Table 3, Col. 4. Alternately, the SPANS command:

QUERY / BY EXAMPLE

can be used to calculate class means as follows:

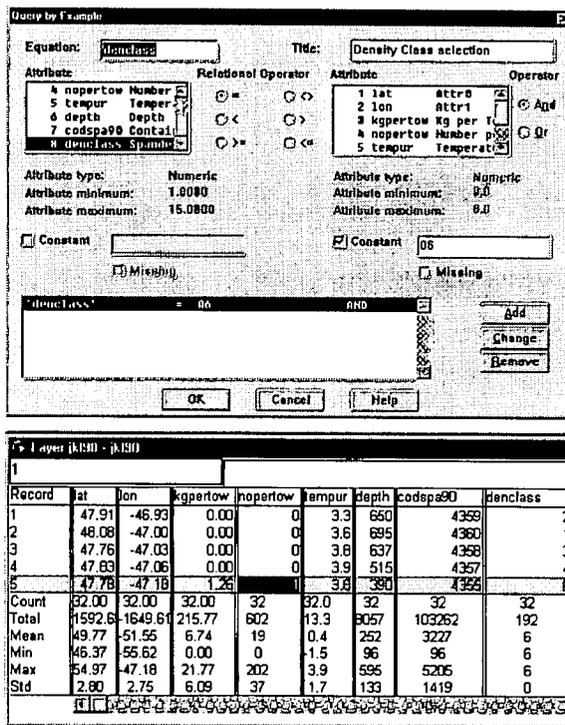


Figure 15 – Query by example selecting denclass = 6 and the resulting calculation in the point layer spreadsheet.

Each denclass, 1 to 15 must be calculated separately then placed into Table 3. Col. 4. This meyhod is slow and not recommended for large batches of calculations. Finally, perhaps the most efficient method of calculating class means is by using:

ANALYSIS / AGGREGATE / ATTRIBUTES

bringing up a screen in Fig. 16:

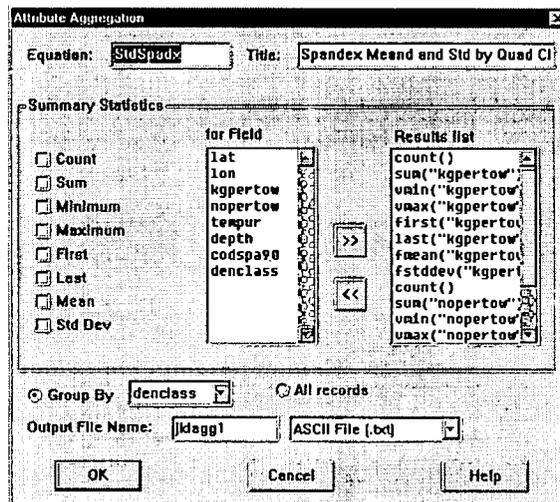


Figure 16 – Selecting aggregate statistics with respect to denclass.

This produces summary statistics by denclass (or other variables) for any of the other fields (variables) in the point layer. The numbers contained in the .txt files produced can be cut and pasted into the SPANdex spreadsheet.

The geographic extent of each class on the fish density map is used for biomass estimation. These density areas are analogous to random stratified survey strata. The area (col. 7) in Table 3 is derived by an area calculation function in SPANS invoked by:

ANALYSIS / STATISTICS / AREA CALCULATION

This will create an ASCII file with a .rep extension. This file lists the class numbers and corresponding area in km² of each class on the map. The class areas can be imported into the SPANdex spreadsheet (Col. 7).

Distribution maps - For creating distribution maps, the density strata are set differently. Usually, fewer strata may be considered for more interpretable grey shade display. Also, each stratum is set with equal area in the first (or other selected baseline) survey year (for example, 1983). The strata class bounds (catch per tow legend values) are then held constant across years (a single legend is used for all years) so that varying amounts of each grey shade displayed depicting a density level would vary, reflecting relative changes in density. Refer to Fig. 9 for a description of how to classify the potential map. In the case of distribution maps, the same *Classification Number* is used for all years. Similarly, bottom temperature grey shade maps can be created using seven strata of equal size varying from -1.6 to 3.36+ °C. If displaying in colour (red-yellow-green gradation), fifteen levels provides a very interpretable map product.

For commercial fisheries data, two types of potential maps can be created. One is the density distribution described above that uses catch rate to define density areas. This type of map illustrates density of fish schools, independent of the number of sets fished at any location. The other type of map uses:

MODEL / DENSITY

to shows the density of fishing activity expressed as number of sets per km² over the fishing grounds. This type of map is useful for fishing intensity studies such as analyses that define extent of grounds scoured by trawls or dredges or to help the fishing industry in developing new fisheries.

Functions other than mean and point density are also possible. The Statistics command selected as:

MODEL / POTENTIAL MAPPING / STATISTICS

produces surfaces depicting standard error of the mean, coefficient of variation and other measures of spatial variance of the selected variable. For example, high coefficient of variation areas on maps of survey kg. per tow closely match areas of high mean kg. per tow. That is, areas where the fish are most dense are also most variable. Areas of high density that correspond with areas of low variance

may be considered desirable fishing locations.

Batch Processing

Some of the operations in SPANS are repeated and may take considerable time to complete if manually. To save time, batch processing can be used to allow unattended processing. Underlying all processes in SPANS is a command language, lines of code that correspond to the operation. SPANS allows operations to be audited, recording the commands. When the following command is invoked, the system is set to record all processes:

TOOLS / COMMAND FILE / AUDIT

Once the set of operations is done, the audit function is turned off (same command). This produces an ACHII file, user named, with an .eas extension. This file containing command language for the processes run while in audit mode can then be edited to include multiple runs of the audited processes. Refer to Anon (1997) for details. Once this is done, the following command runs the audit file after selecting the correct command file name:

TOOLS / COMMAND FILE / EXECUTE

For example, a potential map could be produced for one year and an area calculation run while in audit mode. The .eas created would then be edited by copy/paste of the recorded functions (in command language) setting up a command file that will automatically run all years of data.

Discussion

Distribution studies of exploited marine species and in particular northern cod, have often employed either illustrative methods such as expanding symbol or point plots (i.e. Hutchings 1996, Lilly 1994 and Shelton *et al.* 1996) but this approach is limited with respect to analyzing spatial changes. These descriptive distributional studies cannot quantitatively examine the patterns of shifting fish concentrations over the extent of the stock over time. It is these types of changes in spatial structure of a population that may occur as stock destabilizes and begins to decline. GIS, in this case SPANS has several advantages over descriptive mapping. Various analyses at the scale of the data is possible. That is, the researcher can select any portion of a stock area to analyse and compare among years or with influencing factors. Biomass calculations for example need not be constrained to NAFO Divisions that very often do not conform to stock boundaries. Spatial modelling of changes in the species being studied with environmental variables such as bottom temperature, bottom types or salinity can help define how the environment affects the stock. In a practical application, species mixes in commercial catches can be spatially analysed to show where the most desirable mixes can be found. Changes in bottom type can be related to trawl and dredge activity. These are a few examples of possible fisheries applications of spatial analysis.

The functions described in this paper comprise only a small part of the mapping, analysis and

modelling capabilities of SPANS. Other surface generators such as kriging, contouring and point aggregation are available. Raster images such as ocean surface temperatures can be imported and spatially analysed. Built in modelling functions facilitate overlaying surface and point layers in many combinations that provide powerful spatial analysis tools. The available modelling language can introduce transformation and equations into surface analyses that allow a thorough analysis of spatial data.

References

- Anon, 1991. SPANS. Spatial analysis system. Vol.1. Reference Manual. Intera-Tydac Technologies Inc. 3rd Edition.
- Burrough, P.A. 1986. Principles of Geographical Information Systems for land resource assessment. Clarendon Press. Oxford.
- Doubleday, W.G. 1981. Manual on groundfish surveys in the Northwest Atlantic. NAFO Sci. Coun. Stud. No. 2. xxp.
- Fréchet, A. and P. Gagnon. 1993. Changes in distribution and failure of the winter fixed gear cod (*Gadus morhua*) fishery off southwestern Newfoundland. NAFO Sci. Coun. Stud. 18: 71-77.
- Gulland, J.A. 1955. Estimation of growth and mortality in commercial fish populations. Fish. Invest., Lond., (2) 18 (9), 47p.
- Hutchings, J. A. 1996. Spatial and temporal variation in the density of northern cod and a review of hypotheses for the stocks collapse. Can. J. Fish. Aquat. Sci., 53: 943-962.
- Kulka, D.W. 1990. Areal patterns in the 2J3KL offshore witch fishery, 1979-89. CAFSAC Res. Doc. 90/82, 9p.
- Kulka, D. W. 1997. Discarding of cod (*Gadus morhua*) in the Northern cod and Northern shrimp directed trawl fisheries, 1980-94. NAFO Sci. Coun. Studies, 29: 67-79.
- Kulka D. W., and J. R. Firth 1987. Observer Program training manual. Can. Tech. Rep. of Fish. & Aquat. Sci. No. 1335 (revised) 197 p.
- Kulka D. W., A. T. Pinhorn, R. Halliday, D. Stansbury and D. Pitcher 1995a. Accounting for Changes in Spatial Distribution of Groundfish When Estimating Abundance from Commercial Fishing Data. Fish Res.
- Kulka, D.W., Wroblewski, J.S., and Naryanan, S. 1995b. Inter-annual patterns in the winter distribution and recent changes and movements of northern Atlantic cod (*Gadus morhua* Linnaeus, 1758) on the Newfoundland-Labrador shelf. ICES J. mar. Sci., 52: 000-000.
- Kulka, D. W. and D. Power 1996. Bycatch in the NAFO Div. 3M shrimp fishery, 1993-1995. NAFO SCR Doc. 96/64 15 p.
- Kulka, D. W. and E. M Deblois 1996. Non-traditional groundfish species on the Labrador Shelf and Grand Banks – Wolfish, monkfish, white hake and winter (blackback) flounder. DFO Atl. Fisheries Res. Doc 96/97 49 p.

Kulka, D. W., E. M Deblois and D. B. Atkinson 1996. Non-traditional groundfish species on the Labrador Shelf and Grand Banks – Skate. DFO Atl. Fisheries Res. Doc 96/89 29 p.

Lilly, G.R. 1994. Predation by Atlantic cod on capelin on southern Labrador and Northeast Newfoundland shelves during a period of changing spatial distributions. ICES mar. Sci. Symp., 198: 600-611.

Paloheimo, J. E. and L. M. Dickie. 1964. Abundance and fishing success. Rapp. p.-v. Reun. Cons. Int. Explor. Mer, 155: 152-163.

Parsons, D. G. and P.J. Veitch 1997. Regional review of the status of Northern shrimp (*Pandalus borealis*) resources in areas off Newfoundland and Labrador (Divisions 0B to 3K). DFO Atl. Fisheries Res. Doc 97/05 38 p.

Parsons, D. G., D. W. Kulka and P. J. Veitch 1996. Mapping the distribution, abundance and biomass of shrimp (*Pandalus borealis*) on Flemish Cap (NAFO Division 3M) based on data obtained from a Canadian research vessel trawl survey, September-October 1996 NAFO SCR Doc. 96/47 12p.

Pope, J.G. and D. J. Garrod. 1975. Sources of error in catch and effort quota regulations with particular reference to variations in the catchability coefficients. ICNAF Res. Bull. 11: 17-30.

Rose, G.A. 1993. Cod spawning on a migration highway in the northwest Atlantic. Nature. 366: 458-461.

Shelton, P. A., D. E. Stansbury,, E. F. Murphy, G.R. Lilly and J. Bratley 1996. An assessment of the cod stock in NAFO Divisions 2J+3KL. DFO Atl Fisheries Res. Doc. 96/80, 65 p.

Smith, S. J. and G. D. Somerton 1981. STRAP: A user-oriented computer analysis system for groundfish research trawl survey data. Can. Tech. Rep. Fish. Aquat. Sci. 1030 65p.

Wroblewski, J. S., D. W. Kulka, S. Naryanan, A. Oake, a. Collier and B. McGrath 1995. Winter distribution and movements of northern Atlantic cod (*Gadus morhua*) along the Newfoundland-Labrador continental shelf edge derived from scientific observations on commercial trawlers. Fish. Oceanogr., 4:2 128-146.

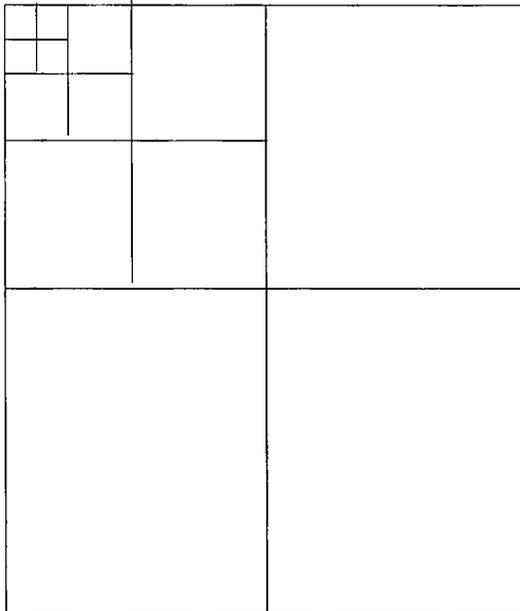
APPENDIX 1- Quadtree Data Structure

Data Storage Format

Graphic programs use two basic formats for storing and displaying information. A vector based program stores only the outline of images as a series of connected points, or as a mathematical formula. Spatial relationships, such as inside or outside of a circle are determined by the relationship to the "inside" or "outside" of the line boundary. This format is very efficient for the storage of information, but requires more processing time to derive spatial relationships.

A raster based program stores data as a continual series of points that each have a specific location and an attribute or a surface with associated extent and area. The attribute may identify the spatial relationship of the point, such as "inside" a circle, or it can contain a unique value, or "colour", that can be compared to the pixels on a television set. A raster has a set pixel size comprising a grid of equal size rectangles. This format requires a great deal of storage space, however it is very efficient and determining spatial relationships.

SPANS uses a variation of a raster format, which is referred to as a quadtree data structure. Using this format, the size of the pixel is variable, and can appear as a fine grid of squares as in Fig a. When a quadtree is being formed, the system divides the study area into four equal rectangles as shown in Fig. a.



If there is only one class of data in the rectangle (the data are homogeneous) then this area forms a single large pixel. Otherwise, the system subdivides the rectangle once again into 4 new rectangles. This process is repeated up to 15 times to form a grid of variable sized pixels, each homogeneous.

The minimum size of these grid blocks, or "quadcells" is set by the "quadlevel" of a map that is selected by the user. The quadlevel varies can be set from 1 to 15, with each whole number increment resulting in 4x change in the size of the quadcell. The absolute size of the quadcell also varies with the size of the universe. The key to a quadcell structure is that neighbouring quadcells that have a common value are combined into a single block. This decreases the storage space required for the map, and increases processing speed.

Fig. a – Quadtree structure.