Associating Remotely Sensed Seafloor Types with Groundfish Species in Hecate Strait

by

Christopher John Grandin, 2006

B. Sc., Computer Science, University of Victoria, 2002

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF SCIENCE
In the School of Earth and Ocean
Sciences

Associating Remotely Sensed Seafloor Types with Groundfish Species in Hecate Strait

by

Christopher John Grandin, 2006

B. Sc., Computer Science, University of Victoria, 2002

Supervisory Committee
Dr. Ross Chapman, (School of Earth and Ocean Sciences)
Supervisor

Dr. Stan Dosso, (School of Earth and Ocean Sciences)
Departmental Member

Dr. Adam Zielinski, (Department of Electrical Engineering)
Outside Member

Alan Sinclair, (Department of Fisheries and Oceans Canada)
Additional Member

Supervisory Committee

Dr. Ross Chapman, (School of Earth and Ocean Sciences)
Supervisor

Dr. Stan Dosso, (School of Earth and Ocean Sciences)
Departmental Member

Dr. Adam Zielinski, (Department of Electrical Engineering)
Outside Member

Alan Sinclair, (Department of Fisheries and Oceans Canada)
Additional Member

ABSTRACT

Traditional stock assessment methods do not incorporate remotely sensed ecosystem variables such as seafloor type, relief, and complexity. Incorporation of these and other ecosystem variables allows for targeting of species' optimal habitat during surveys. Recently, acoustic remote sensing methods have allowed us to gain insight into groundfish habitat.

In June 2002, a geophysical survey was performed in selected fishing areas of Hecate Strait. While underway, single beam acoustic data were collected along survey lines utilizing a 50 kHz echosounder coupled with Quester-Tangent's QTC VIEW 5; a Huntec seismic system, and a dual frequency sidescan sonar system.

Surficial sediment distribution and seabed features were mapped through examination of seismic, sidescan, and bottom grab data and compiled into a GIS. The surficial sediment classes were compared to bottom type classifications obtained from QTC single beam, with results showing the Gravel and Sand class from the surficial sediment data being classified best by the single beam system.

Catch data from the groundfish bottom trawl fishery for the areas of interest were made available by the Department of Fisheries and Oceans Canada (DFO). The distribution of groundfish aggregates and individual fish species were compared to surficial sediment classes using correspondence analysis to investigate habitat associations. Results show that the Rock Sole aggregate had a habitat preference of gravel and sand mixture and the Dover Sole and Arrowtooth Flounder aggregates had a habitat preference of sandy mud. Correspondence analysis allows for a 2-dimensional view of multivariate categorical data which are the norm for habitat-based biological studies.

Results suggest that the procedures developed in this work can improve stock assessment methodology and indicate that using various acoustic remote sensing techniques can be effective in characterizing seafloor habitats and ecological connections between groundfish species and seafloor types.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I would like to thank Alan Sinclair who made this whole study possible and gave me the opportunity to work on it, Ross Chapman whose patience and willingness to connect different fields of study was welcomed, the Centre for Marine Acoustic Remote Sensing (CMARS) for its support in the analysis of the single beam data, National Resources Canada (NRCan) for the surficial sediment maps, and to Lynne Yamanaka who supported me during much of the writing time of this thesis.

Chapter 1
Introduction

Traditional fisheries stock assessment methods do not formally incorporate ecosystem or habitat variables such as seafloor sediment type, bathymetry, or roughness. Instead, catch and effort data from scientific surveys and commercial, recreational, and aboriginal fisheries over many years are used to estimate current stock status. The time series generated from these data are used to estimate stock abundance for a particular species and recommend future harvests. The abundance index for a particular species usually applies to the whole west coast of British Columbia, whether or not the habitat is suited to the species. The underlying hypothesis in this thesis is that taking habitat into consideration in stock assessments will allow for improved abundance estimates and optimal sampling during surveys.

Associations between habitat and species distribution have been investigated by others (Romsos, 2004, DeLong and Collie, 2004, Whitmire, 2003, Greene *et al.*, 2004). However some of these studies do not incorporate catch data (Green *et al.*, 2004, Whitmire, 2003); instead they focus on bathymetric derivatives and reflectivity from backscatter to determine benthic habitat. Other studies (Romsos, 2004, DeLong and Collie, 2004) focus on single-species habitat associations and do not account for species assemblages. The Hecate Strait Ecosystem Project (HecStEP) was initiated in 2002 by the Department of Fisheries and Oceans Canada to expand knowledge of factors affecting the productivity of marine fish communities in Hecate Strait, and to use this information to develop new stock assessment techniques and incorporate them into the mainstream advisory process (Sinclair *et al.*, 2005). A part of this project was to determine seafloor type through use of single beam acoustic remote sensing, and to associate those types with species assemblages.

This study is designed to investigate multispecies habitat associations of groundfish in Hecate Strait that are found over relatively flat seafloor. Six test sites were selected based on species catch composition in the commercial bottom trawl fishery. The sites were grouped in three areas with pairs of sites with similar contrast in species composition. The principle hypothesis being tested is that variation in habitat among sites is related to the variation in species composition. The results of this work demonstrate that habitat can be determined efficiently and accurately by acoustic remote sensing.

Quester Tangent QTC View 5 single beam echosounder data were collected at the six sites during June 2002. The digital echo returns were checked for good bottom picks, and classified into several seafloor types. The processing, classification, and applicability of these data are detailed here and discussed. The single beam classifications were compared via correspondence analysis to surficial sediment maps supplied by Natural Resources Canada (NRCan); however the trackline densities for the single beam data at the sites were too sparse to allow for adequate spatial coverage. The original intent of the study was to examine correlations between single beam classes and trawl catch for the six areas of Hecate Strait. Due to the sparseness of single beam data, and due to the acquisition of surficial sediment maps, correlations between the catch data and surficial sediment were examined instead; these may more accurately represent the seafloor by including remotely sensed data such as sidescan sonar, HUNTEC seismic sub-bottom profile, and physical data from bottom grabs and cores. Correspondence Analysis is the analytical tool of choice for this study and is used extensively to show correlations between species and surficial sediment.

Chapter 2
Study Areas and HecStEP Data

The overall objective of this work is to find relationships between groundfish distributions and seafloor habitat type by acoustic remote sensing. This chapter outlines the study areas for the single beam acoustic survey, and the data recorded during the survey which includes HUNTEC deep-tow seismic sub-bottom profiler and Sidescan Sonar. Surficial sediment maps and groundfish catch data used in correspondence analyses later in the thesis are described here.

2.1 Study Areas

In June 2002 a hydroacoustic survey took place aboard the Canadian Coast Guard vessel *John P. Tully* in selected areas of Hecate Strait. The ship traversed the tracklines shown in Figure 1. The choice of locations was based on the distribution of species in the groundfish bottom trawl fishery during June of the years 1996-2000. The six trawl fishery areas included in this study are from north to south: Dundas, Two Peaks, White Rock, Shell Ground, and East & West Horseshoe. These areas were chosen for their significant differences in fish catch composition with the hypothesis that the contrast would be related to differences in benthic habitat.

The trackline density for some areas such as Dundas and the Horseshoes is sparse; ideally track line crossings should be present in the survey grid and the trackline density should look more like the density at the Shell Ground site.

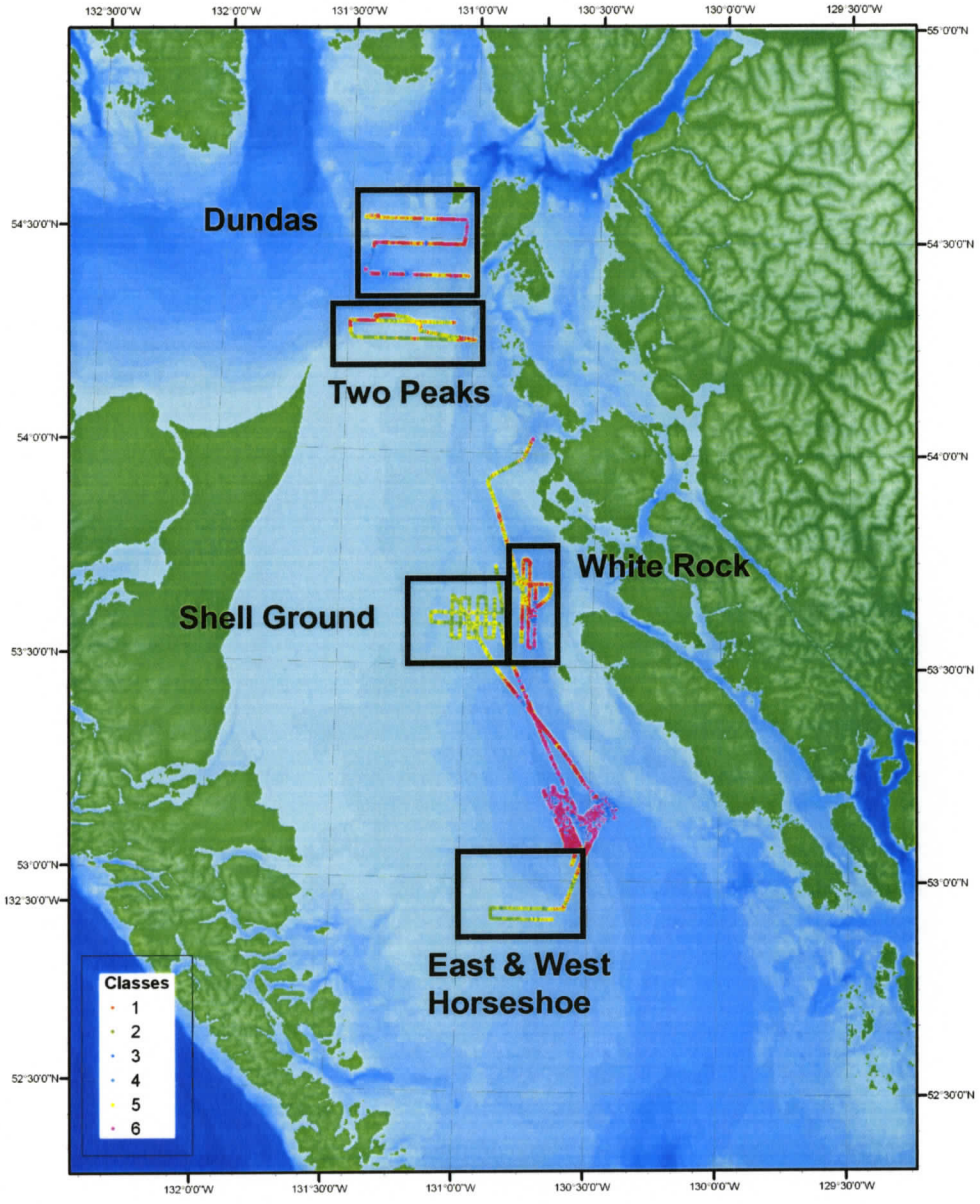

Figure 1 – Hecate Strait areas of study and acoustic survey tracklines. The colours on the track lines represent post-experiment bottom classes.

2.2 HecStEP Data

2.2.1 Single Beam Echosounder Data

Single beam acoustic data were recorded over the six sites shown in Figure 1. The figure shows the tracklines in post-classification colours. A 50 kHz Simrad echosounder was used in conjunction with the Quester Tangent Corporation QTC5 Single Beam System for acquisition of the returning echoes. Chapter 3 elaborates on these data and the classification procedures, and shows the comparison of the acoustic classes with the sediment ground truth maps.

2.2.2 HUNTEC High Resolution Seismic Data

A HUNTEC Deep-Tow System (DTS) with a low-frequency acoustic source in the band 1-10 kHz was towed behind the *C.C.G.S. John P. Tully*. The sound source is designed to be towed deep in the water, since being closer to the bottom increases resolution and reduces signal loss due to spreading and attenuation. The HUNTEC data were analyzed and integrated into an overall surficial sediment map by Conway *et al.* (2004). Output from the HUNTEC system can be seen in Figure 25 (Section 3.5.3.1).

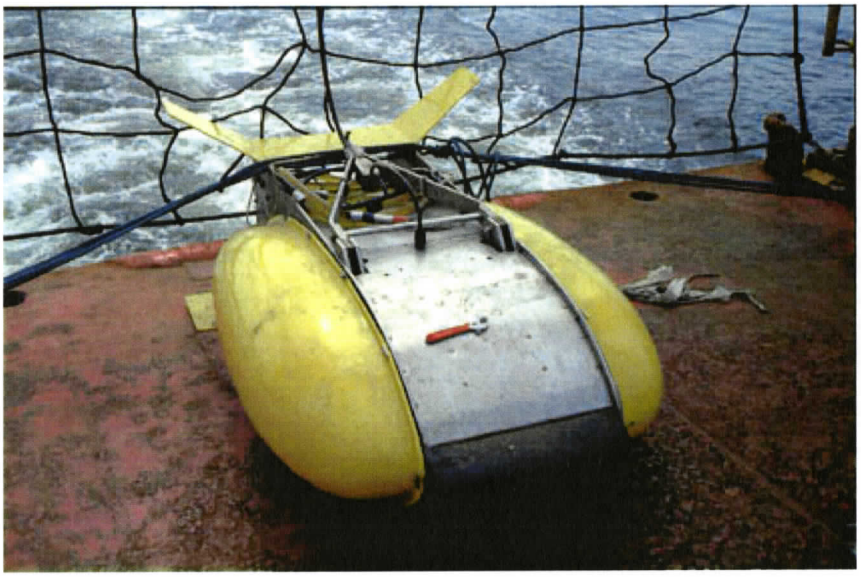

Figure 2 –The HUNTEC Deep-tow towfish. Credits: National Resources Canada.

2.2.3 Sidescan Sonar

Sidescan data were also acquired during the survey with a Simrad 992 dual frequency sidescan sonar (SSS) system using MUSE software for digital recording and display. These data were also integrated into the geophysical map post-survey by Conway *et al.* (2004). Output from the sidescan system can be seen in Figure 25 (Section 3.5.3.1).

2.2.4 Surficial Sediment

Surficial Sediment Maps were produced by Conway *et al.* (2004). These thematic maps incorporate Sidescan Sonar, HUNTEC sub-bottom, and grab sample data acquired throughout the past 30 years. The maps give a polygonal visualization of probable seafloor types, and also include features such as gas pockets, trawl marks, dunes, sand ribbons, and terraces. For this study, only the seafloor classes were used. Figure 3 shows the Shell Ground area as an example of the input layers for the GIS.

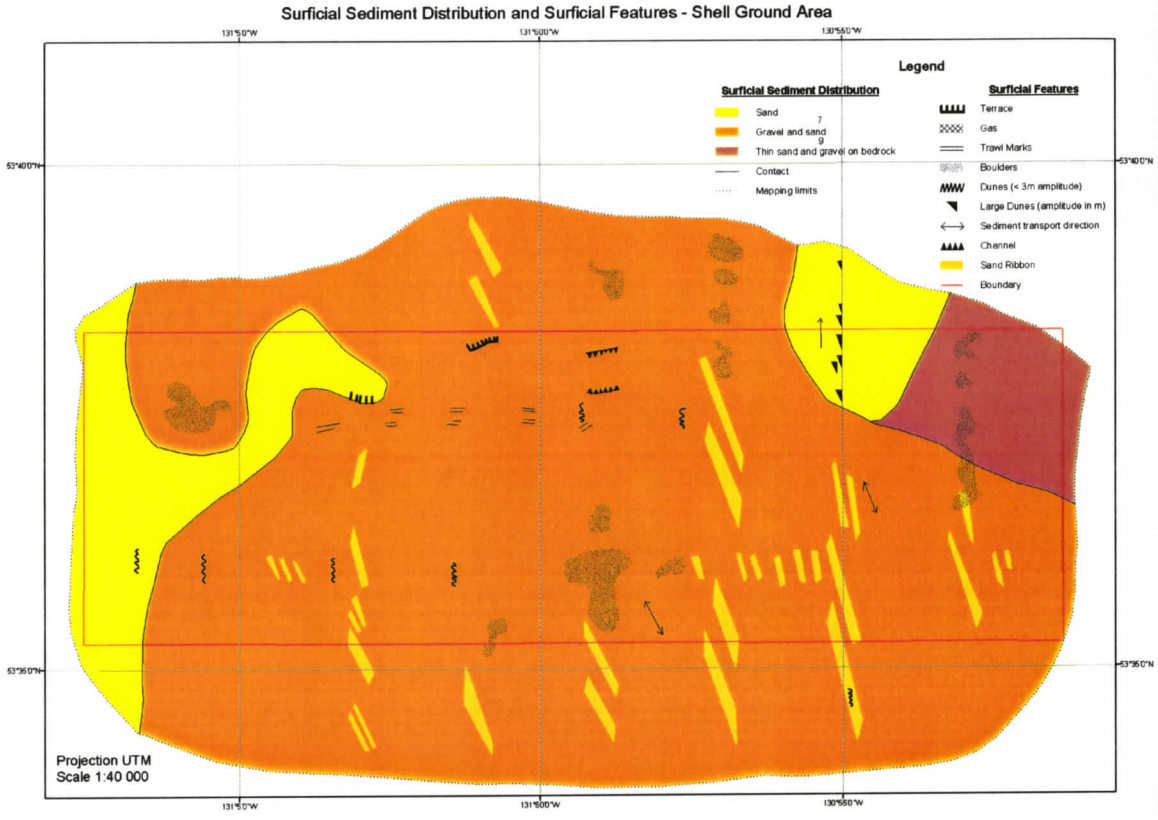

Figure 3 – Shell Ground surficial sediment distribution. Orange represents gravel and sand, yellow represents sand, and red represents thin sand and gravel over Bedrock. The area is 10.8 km from north to south and 19.7 km from east to west. Credits: Conway *et al.* (2004).

2.2.5 Groundfish Catch

Catch data were provided by the Department of Fisheries & Oceans Canada, Pacific Biological Station, Groundfish Section. The data were first queried to include only government-observed commercial trips and scientific surveys. All fishing sets or 'fishing events' falling inside the 6 areas defined by the surficial sediment layers as shown in Figure 1 were extracted for analysis by using a GIS (ArcMap 8.3). This ensured every fishing event had an associated bottom type for

the analysis. After this extraction process, the following datasets were used for the 5 fishing areas:

- 1996-2002 June only, to correspond to the single beam acquisition month.
- 1996-2002, All months, for contingency and correspondence analysis against surficial sediment.
- 1996-2002, Summer (April 1 - September 30), for contingency and correspondence analysis against surficial sediment.
- 1996-2002, Winter (October 1 – March 31), for contingency and correspondence analysis against surficial sediment.

The catch weight extracted was the total catch including the discarded catch for each trip/set combination; after grouping by species, a table with a total catch weight for every species caught for the dataset's time period was produced. From this table, the data for the fourteen most commonly caught species were extracted. Of the top fourteen, only the twelfth most caught species, a starfish (Asteriodea), was removed from the analysis because it is not an assessed groundfish species. Table 1 shows the species, catch weights, and proportion of catch for each of the analysis datasets. Round weight is defined as the total fresh weight of the fish, head on before any processing.

Species	Common Name	Weight	Weight (June)	% catch (June)	Weight (winter)	% catch (winter)	Weight (summer)	% catch (summer)
602	Arrowtooth Flounder	14318448	3264062	22.8	2479163	17.3	11839286	82.7
626	Dover Sole	5103895	1051738	20.6	987042	19.3	4116853	80.7
066	Spotted Ratfish	4866072	561975	11.5	2751366	56.5	2114706	43.5
628	English Sole	4325399	422581	9.8	2362044	54.6	1963355	45.4
056	Big Skate	4291051	284988	6.6	3008287	70.1	1282763	29.9
222	Pacific Cod	3407237	630605	18.5	829260	24.3	2577977	75.7
621	Rock Sole	2561749	701027	27.4	337697	13.2	2224053	86.8
044	Spiny Dogfish	2263722	194986	8.6	438592	19.4	1825130	80.6
610	Rex Sole	1789606	258807	14.5	786412	43.9	1003194	56.1
614	Pacific Halibut	1445793	303126	21.0	264010	18.3	1181783	81.7
228	Walleye Pollock	1166748	44262	3.8	868649	74.5	298099	25.5
059	Longnose Skate	963909	152896	15.9	485745	50.4	478164	49.6
467	Lingcod	644992	89512	13.9	145657	22.6	499335	77.4
455	Sablefish	605014	88999	14.7	260585	43.1	344429	56.9

Table 1 – Catch by season for 1996-2002. Winter is Oct. 1 – Mar. 31. Summer is Apr. 1 – Sept. 30. Weights are round weight in kg. These data are within the 5 fishing areas with surficial sediment coverage maps.

Chapter 3
Single Beam Acoustic Remote Sensing

Single beam acoustic remote sensing requires specialized hardware and software. This chapter describes the Quester Tangent QTC5 VIEW system in detail. The software for classification, QTC IMPACT, is also investigated and results of classifications are shown, including problems and pitfalls associated with single beam classification.

3.1 Acquisition Hardware

A 50 kHz Simrad echosounder was used in conjunction with the Quester Tangent Corporation (QTC) QT5 Single Beam System. The beam width was set to 8° and the pulse duration was 2 ms. A standard PC running the Microsoft Windows based QTC VIEW software was used for data acquisition.

3.2 QTC VIEW System Function

The QTC VIEW 5 system as described by Watt (1997) uses a conventional echo sounder to send acoustic pulses to the seafloor at a user-set ping rate that is dependent on the water depth and vessel speed. The transmit pulses are reflected off the seabed and returned to the echosounder transducer on the vessel. The shape and intensity of the returning echoes is a function of the roughness and composition of the seabed and other parameters such as transducer beam width and water depth. At shallow and moderate depths smooth, hard bottoms will tend to reflect most of the energy back, which results in an echo that has a narrow peak with almost no tail. If the same smooth bottom is composed of a muddy substrate, much of the energy will be absorbed, resulting in a lower

amplitude peak. If the bottom is complicated and rough, there will be more scattering of the signal and the echo will have a wider peak and longer tail. The roughness of the seabed and the density difference between the water and the seafloor are the main contributors to the return echo shape.

The acoustic pressure of the seafloor echo pulse is converted by the echosounder transducer to a time varying voltage. An example of the echo signal is shown in Figure 5. The voltage is then digitally sampled by an Analog-to-Digital (ADC) converter and stored as an envelope of the echo signal to save space. If appropriately chosen, the ADC and echosounder transducer in general contribute negligibly to noise in the data.

3.3 Echo Data Structure

The echo returns and auxiliary data from the single beam acoustic system are digitized and stored in ISAH-DB format, a data acquisition structure developed by Quester Tangent. To view individual echo returns (traces), the ISAH database class is accessed via code (MATLAB in this case), which consists of 6 folders, each containing several scripts and software classes. These ISAH-DB scripts and software classes are licensed from Quester Tangent Corporation. The user places the scripts into the MATLAB/work directory; this directory should look like Figure 4, including the two MATLAB user scripts used to access the traces, plot_wiggles.m and plot_traces.m. These scripts and software classes can be obtained from Quester Tangent or the author.

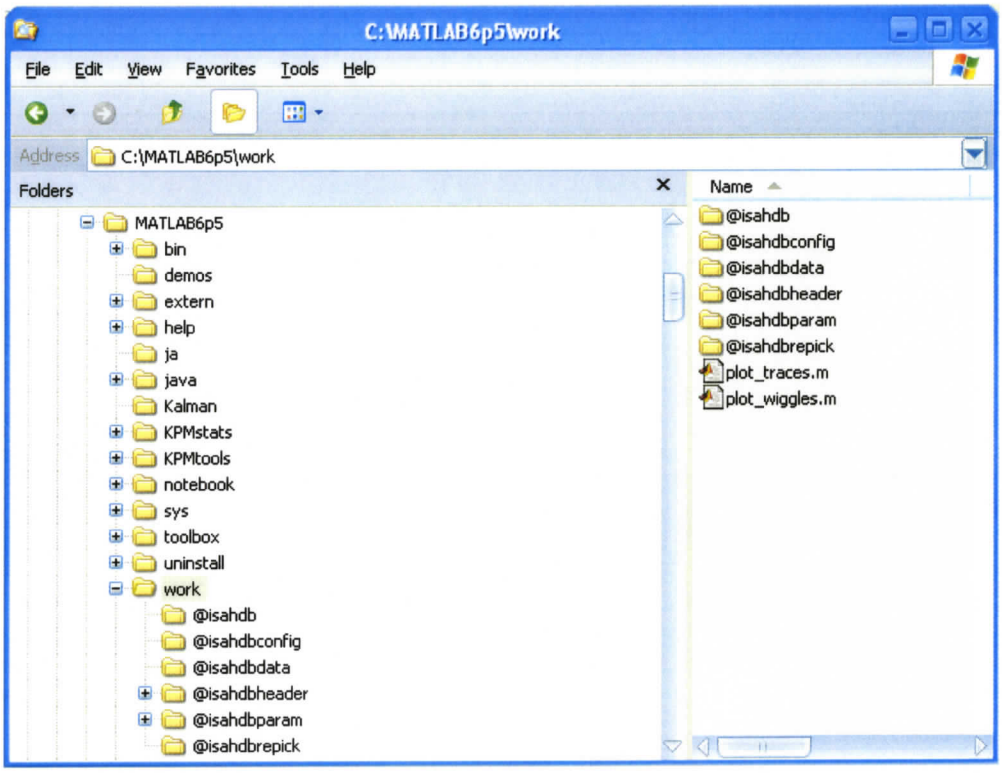

Figure 4 – MATLAB work directory structure for ISAH-DB trace extraction.

The two MATLAB scripts allow the extraction of any of the traces captured by the Quester Tangent single beam system for viewing. The y-axis can be either two-way travel time or depth; depth is calculated based on sample rate and a 1500 m/s speed of sound in seawater. An example plot is shown in Figure 5.

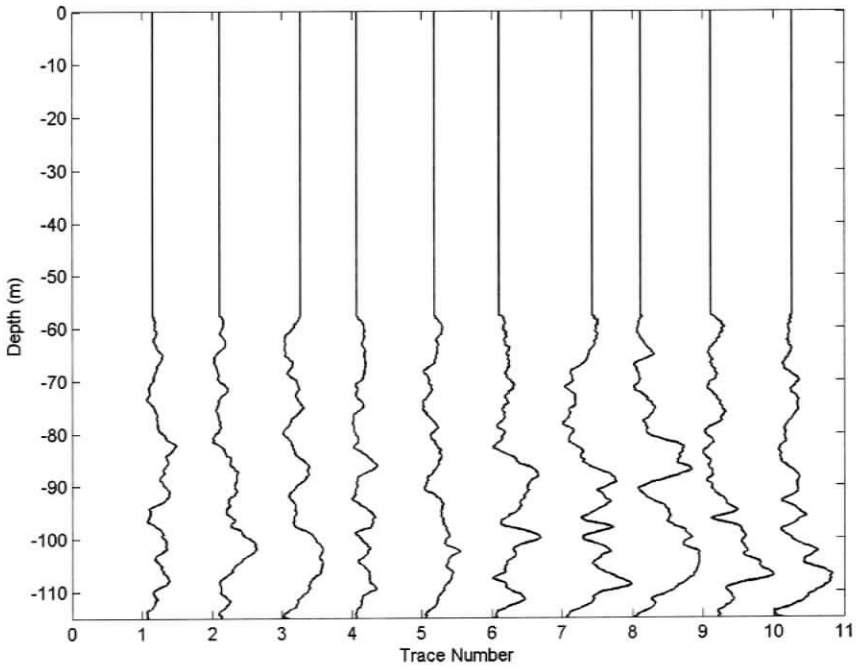

Figure 5 – An example of 10 consecutive echo traces collected on June 21, 2002 in the White Rock area. The amplitudes of the traces are voltages representing pressures of the return signals. The water depth is -60m for these data. There were 378,874 traces in the survey.

The single beam returns were cleaned, stacked, and classified. The details of these steps are found in Section 3.5.1.

3.4 QTC IMPACT Software

The Quester Tangent QTC IMPACT software is necessary to read and analyze the echoes returned by the QTC VIEW system. There are several important steps in the process:

1. Each echo must be checked for correct bottom pick (the depth at which the bottom is located in the echo), and corrected when needed. Automatic bottom picking is performed by the software, and is based on the largest return in the echo trace, usually the first return. IMPACT contains a utility called the waveforms editor, which is a window showing the individual traces in user-configurable sized groups. The repicking step can be quite tiresome, as any given survey can have hundreds of thousands of returns and each must be scrutinized individually. Figure 6 shows the waveforms editor. The red lines are the bottom picks and can be moved by the user if the software has incorrectly picked the bottom.

2. Figure 7 shows the trace viewer which displays only individual traces. Again, the red pick line can be adjusted by the user.

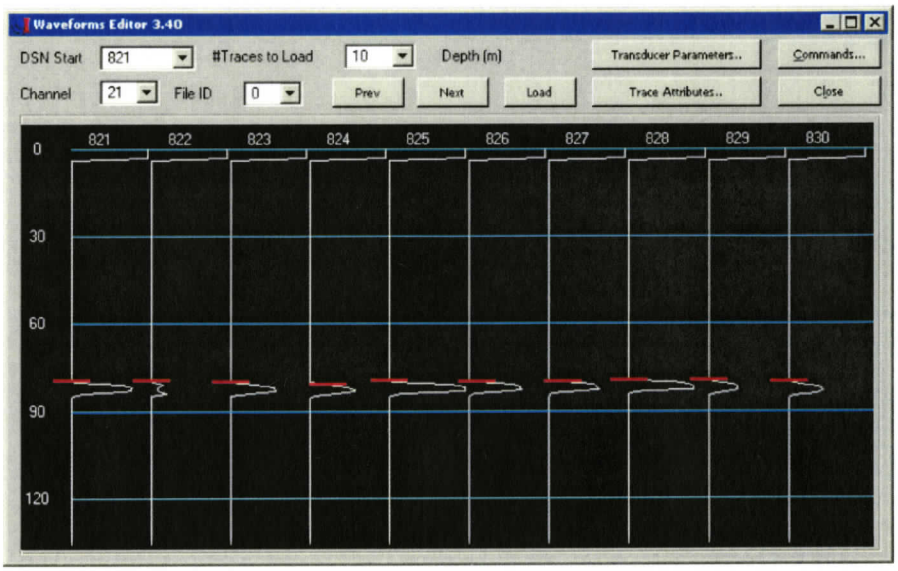

Figure 6 – The waveforms editor (QTC IMPACT User Manual, 2003). The red lines represent the pick for each echo. The Y-axis is the water depth (range from transducer).

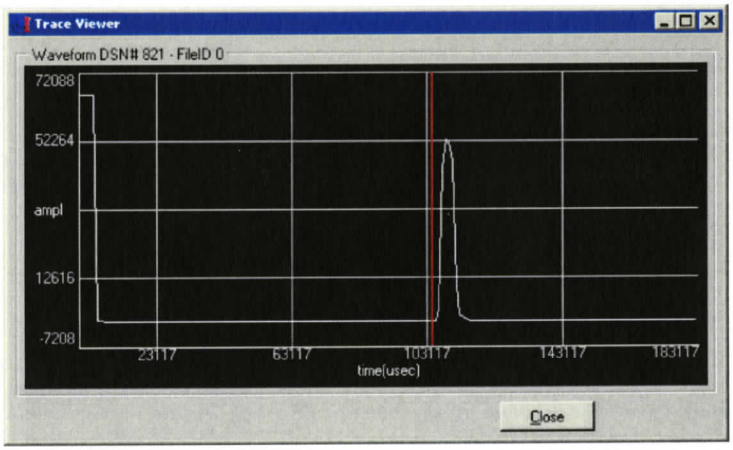

Figure 7 – The trace viewer (QTC IMPACT User Manual, 2003). The red line is the pick line for the echo. The X-axis is the water depth (range from transducer).

3. Each echo is analyzed using multiple proprietary algorithms which incorporate echo shape information in the signal to give a digital

description consisting of discrete elements, descriptors or features. These are combined into a Full Feature Vector (FFV). If traces are unpicked in the cleaning phase described above, they will be ignored in the FFV creation phase. This can pose a problem if there are too many consecutive traces which have been unpicked, since the vectors are stacks (means of consecutive picked traces); if the stacked traces are too far apart the FFV may not be representative of any actual type but a mixture of 2 or more types. QTC is addressing the issue of trace stacking and whether or not to expose the number of traces stacked as a variable. This would allow users to obtain a non-stacked classification by choosing 1 instead of 5 for consecutive traces to stack. Also it would allow a moving window averaging scheme with overlap, since the user could choose a negative stacking number.

4. The navigation data which were acquired through use of a GPS system must be merged to the FFV data. Each FFV is assigned a polynomial fit position based on time. The navigation files contain National Marine Electronics Association (NMEA) strings. For this analysis, the strings recorded by the QTC VIEW system from the GPS were of the wrong type used by IMPACT and needed preprocessing with a scripting language. The conversions were from $GPRMC strings to $GPGLL strings. The Perl code to do the conversion was written by and is available from the author.

5. The FFVs are checked in a similar manner as the raw traces; an FFV bathymetry editor window is used to unpick FFVs manually from the sequence (Figure 8). These unpicked FFVs stand out graphically from the others as they will have a much different depth or show unlikely slopes.

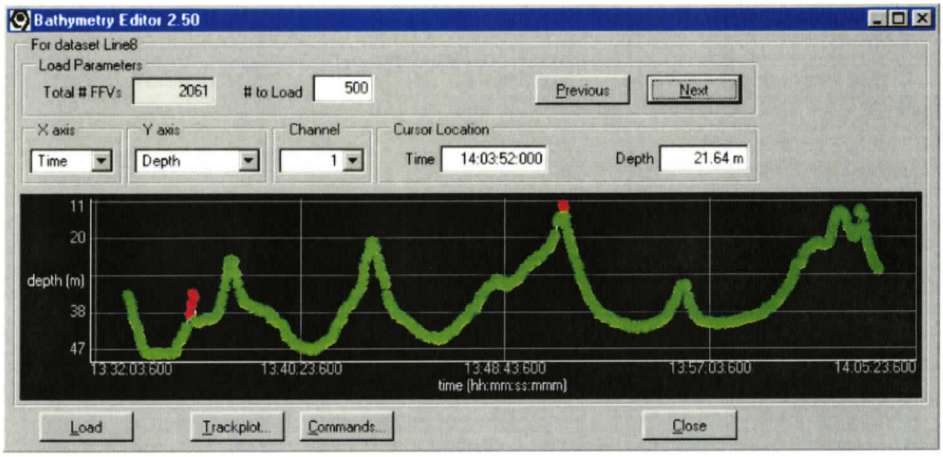

Figure 8 –The FFV bathymetry editor (QTC IMPACT User Manual). The green dots show picked bottom depths and form a bathymetry profile. The red dots show unpicked bottom depths. Only FFVs with picked bottom depths will be used.

6. The FFVs are reduced from 166 variables to 3 by Principal Component Analysis to enable 3-dimensional viewing during clustering by human analysts. These variables are dubbed Q1, Q2, and Q3 by Quester Tangent.

7. The Q variables are placed into a 3-dimensional view port coded in MATLAB. The first two principal axes are plotted in Figure 14, Section 3.5.2, and should appear to the eye to be clustered into several distinct groups. These groups represent acoustically similar echoes, although bottom type cannot be identified from the clusters alone. Ground truthing is required to determine what types the clusters represent.

8. A subjective method of assigning a class to each echo is implemented in IMPACT 3.20; the user splits a cluster by one of the three axes, primary, secondary, or tertiary. These three axes are defined as the axes of most

variability, secondary variability, and least variability respectively. In IMPACT 3.40 Quester Tangent added objective/automatic clustering, but the analysis here was done prior to the 3.40 release. The splitting process in IMPACT 3.20 is done manually, and iteratively. The recommended method as described in the QTC Impact Manual (2003) is to examine the class scores (Chi-square) and split the class with the largest value. A K-means algorithm is used to do the splitting, and new class scores are calculated. This process is repeated until the analyst is happy with the classes.

To stay organized during the processing and classification steps, a relational database containing all pertinent information was constructed. Figure 9 shows the relationships in the database created for this dataset, with primary keys boldface.

Figure 10 through Figure 13 show the tables' field definitions and datatypes as seen in Microsoft Access XP.

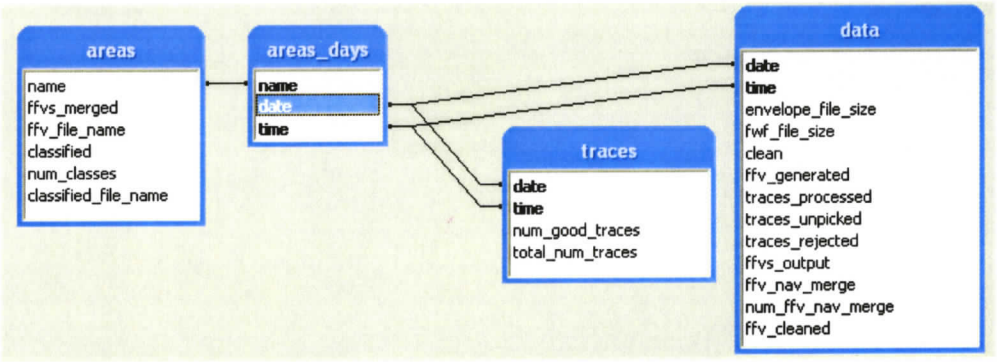

Figure 9 –Relationship diagram for the single beam classification procedure.

Field Name	Data Type	Description
name	Text	fishing area common name
date	Date/Time	day the data was collected
time	Date/Time	time the data was collected

Figure 10 – The areas_days table from the single beam processing database.

Field Name	Data Type	Description
name	Text	fishing area common name
ffvs_merged	Yes/No	indicates if all the ffvs from the area have been merged
ffv_file_name	Text	the path of the merged file
classified	Yes/No	has the area been classified?
num_classes	Number	number of acoustic classes
classified_file_name	Text	the path of the classified DAT file

Figure 11 – The areas table from the single beam processing database.

Field Name	Data Type	Description
date	Date/Time	date the data was collected
time	Date/Time	time the data started to be collected
envelope_file_size	Number	qtc5env_raw file size (KB)
fwf_file_size	Number	qtc5fwf_raw file size (KB)
clean	Yes/No	has the raw wave data been cleaned?
ffv_generated	Yes/No	has the ffv been generated?
traces_processed	Number	the total number of traces in the dataset
traces_unpicked	Number	the number of raw wave traces that were manually unpicked (cleaned)
traces_rejected	Number	the number of raw wave traces that were automatically rejected by IMPACT
ffvs_output	Number	the number of FFVs generated from the input dataset
ffv_nav_merge	Yes/No	has the ffv been merged with the navigation data?
num_ffv_nav_merge	Number	how many FFVs were successfully merged
ffv_cleaned	Yes/No	has the ffv data file been cleaned (edited) with the ffv editor?

Figure 12 – The data table from the single beam processing database.

Field Name	Data Type	Description
date	Date/Time	date the data was collected
time	Date/Time	time the data started to be collected
num_good_traces	Number	the number of traces accepted after cleaning
total_num_traces	Number	total number of traces

Figure 13 – The traces table from the single beam processing database.

3.5 Classification of Single Beam Acoustic Echoes

3.5.1 Pre-analysis Data Filtering

Cleaning the data involved checking each individual trace for proper bottom picks and noise (see Section 3.4). These data consisted of 378,874 individual traces before cleaning. After cleaning 279,739 traces remained. This step of the classification takes the most time as every trace needs to be checked. Figure 5 shows 10 consecutive echo traces collected on June 21, 2002 in a vertical arrangement. This is the same way the echoes are viewed during cleaning. Note that traces 1 through 5 are very similar in shape. This implies a very similar bottom type by visual inspection. The algorithms employed by QTC attempt to quantify these similarities and classify the traces accordingly.

In the typical acoustic survey, many bad traces will occur. This survey had 99,135 bad traces out of 378,874 (26%). These were caused by one of or a combination of the following factors (Collins and Rhynas, 1997):

- The vessel's hull can produce cavitations across the transducer face, resulting in a noisy signal and therefore a bad return or no return. This is exacerbated by high vessel speed or too small a turning radius.

- Electrical noise caused by the vessels engines, generators, and other electrical motors on board can interfere with the QTC5 system, causing incorrect interpretations of return signals and ultimately incorrect classification.

- Since the transducer is mounted rigidly to the vessel, rough seas will move the beam off nadir resulting in a poorer overall classification. This is due to the transducer facing away from the echo return at the time of

return. If there are many bad bottom picks, some periods of no returns and the survey is not too deep (<300m) then the sea state may play a role in data degradation.

- Water column stratifications such as thermoclines result in a change to the speed of sound and a distortion of the echo or a false indication of bottom.

Generally the noisy or non-returned echoes are very obviously different in shape and/or depth than the surrounding ones. For example if there is a set of consecutive echoes which all have a first return at similar depth, but the first return for one echo in the group is much deeper, that trace should be removed. Care needs to be taken while doing this; the depth and vessel speed should be taken into consideration. If the survey is deep and/or the vessel speed is fast enough to move the footprint significantly, then it is possible that the seafloor can change significantly from trace to trace. For example, an 8° beam in 20 m of water will have a 2.8 m diameter footprint. An 8° beam in 400 m of water will have a 56 m diameter footprint. It is possible that in a 56 m space, a deep but narrow chasm could be encountered by only one trace.

3.5.2 Clustering Feature Vectors

The classification was performed on the FFVs using QTC IMPACT 3.20 software which performs Principal Component Analysis (PCA) first to reduce the number of features from 166 to 3 so that human analysts can more easily interpret the results in 3-dimensional space. The classification was unsupervised, since the analysis did not include reference to known seafloor types.

The 3 principal components are designated Q1, Q2, and Q3 and when plotted in 3-dimensional space, acoustically similar FFVs will form clusters which represent

acoustically similar seafloors. IMPACT's cluster analysis tool identifies various data clusters and labels them as classes. It has several outputs; colour-coded bathymetry plot, Q-space plot, colour-coded track plot, and class statistics. The clustering method is a K-means based algorithm using Mahalanobis distances (Preston and Kirlin, 2003). The Mahalanobis distance is QTC's alternative to the Euclidean distance measure usually used in K-means clustering. The Euclidean distance does not take into account the covariance of the distance of the point from each centroid whereas the Mahalanobis distance does. The Mahalanobis distance is defined as:

$$d_i(\mathbf{x}) = (\mathbf{x} - \mathbf{m}_i)^T \mathbf{C}_i^{-1} (\mathbf{x} - \mathbf{m}_i)$$

where vector $\mathbf{x}$ is the observation, $\mathbf{m}_i$ and $\mathbf{C}_i^{-1}$ are the estimated mean and covariance of the cluster ω_i. This distance measure tends to produce clusters which are elongated ellipses (Figure 14). The major drawback of using Mahalanobis distances is that it is computationally intensive due to the calculation of covariance matrices. Since the covariance matrices are symmetric, their sizes increase quadratically by the number of observations. This may be very impractical for the large datasets gathered in acoustic classification systems. As Legendre (2003) points out, there is no particular reason why acoustic classification data should be structured in any particular way in multivariate space. It may be better to use a Euclidean distance for this application to reduce computation resources and avoid the need to catalogue a subset of data, and then apply that catalogue to the rest of the data as has been done here.

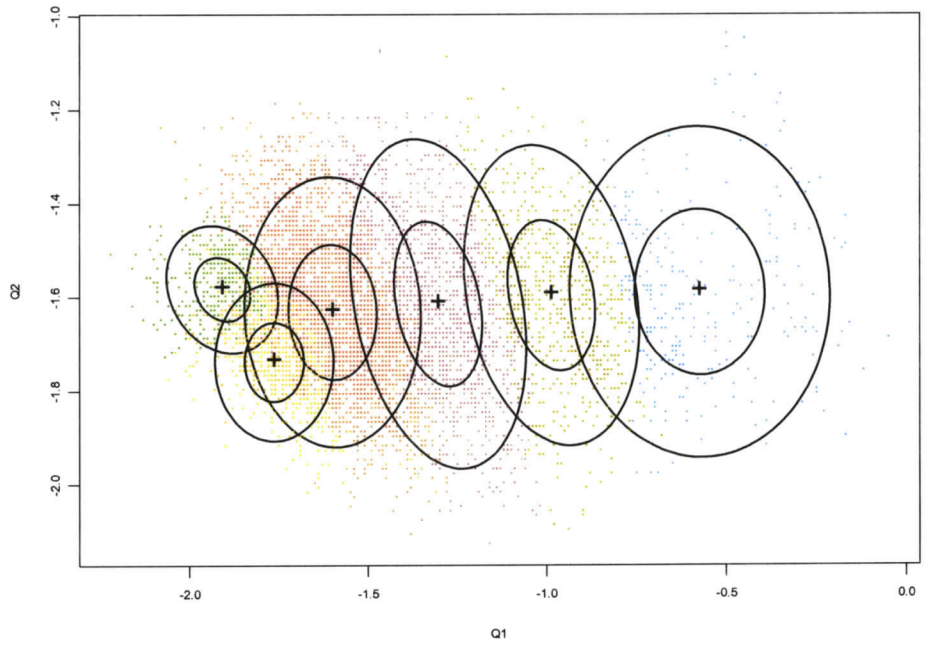

Figure 14 – Dundas Q-space plot (first 2 principal components) for catalogued single beam. There are 6 classes. The crosses are the means, the small ellipses represent one standard deviation from the means, and the large ellipses represent two standard deviations from the mean.

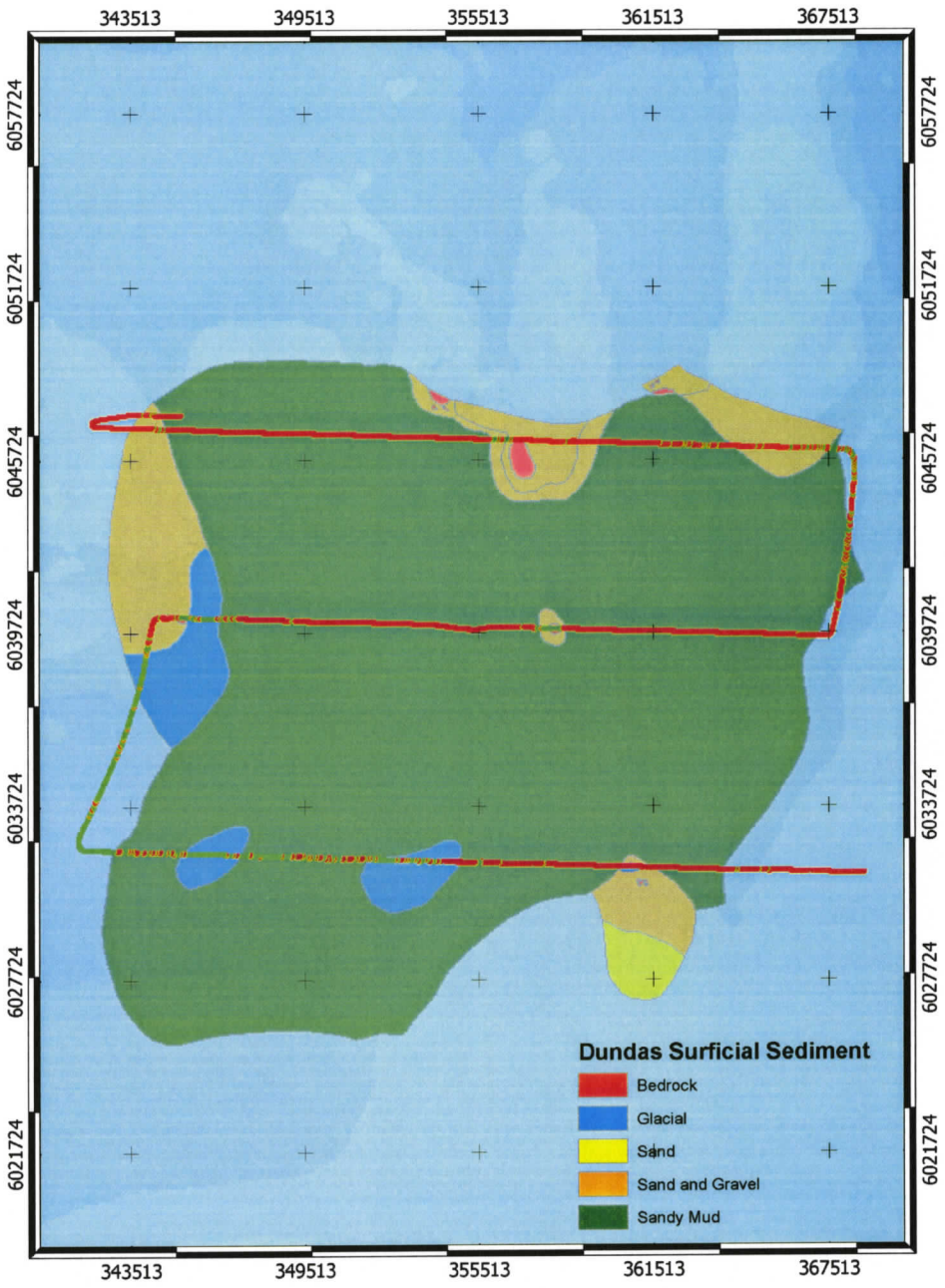

Figure 15 – Dundas surficial sediment map overlaid with non-catalogued single beam. Note that only 2 classes were found using this method.

3.5.2.1 First attempt – separate non-catalogued single beam

In the first attempt at classification, the six areas were classified separately and the results compared with the surficial sediment maps supplied by Conway *et al.* It was discovered that a lack of dissimilar bottom types would result in a poor classification. For example, looking at the Dundas area (Figure 15), the overlaid single beam classification shown was generated by using only the Dundas area echoes. Since Dundas is almost all sandy mud, the clusters are not as distinct and only two unique classes were found. In this case, sandy mud and sand & gravel appear to be clustered together as if they were the same type.

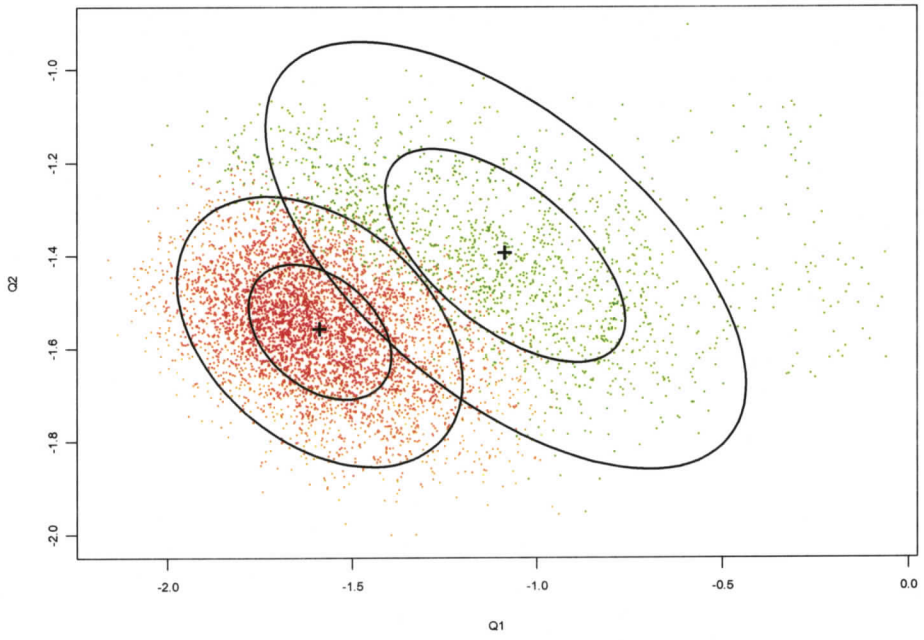

Figure 16 – Dundas Q-space plot (first 2 principal components) for non-catalogued single beam. There are 2 classes. The smaller ellipses show 1-standard deviation, the larger show 2 standard deviations from the mean of each class.

The Q-space scatterplot shows how the clustering occurred (Figure 16). The inner ellipse represents one standard deviation from the mean (centre) of the cluster (QTC IMPACT Manual, p96), the outer ellipse represents the two standard deviations from the mean of the cluster. Only the first two principal components are shown. Clearly the far right side of the figure has some values that could be split into another cluster, and the smaller red ellipsoid appears too small; many of the points belonging to the red cluster are more than two standard deviations from the center of the cluster.

3.5.2.2 Second attempt – catalogued single beam

The method used to correct the problem of too few clusters was to create a catalogue of bottom types from an area or combination of areas with as much benthic diversity as possible so the clustering algorithm will find more unique and well defined clusters. The catalogue is then used to identify the bottom types of other areas based on return echo features seen from the original diverse site(s). This is accomplished by QTC's use of a catalogue file, which contains the class numbers or names along with the FFV records. The algorithm to analyze the FFVs is proprietary as mentioned previously, all of the 166 FFV variables are used to determine class, and this record can be used for other datasets. Two sites which encompassed the majority of bottom types were chosen, and an unsupervised classification was performed on their data, to create the best possible catalogue for our entire area of study.

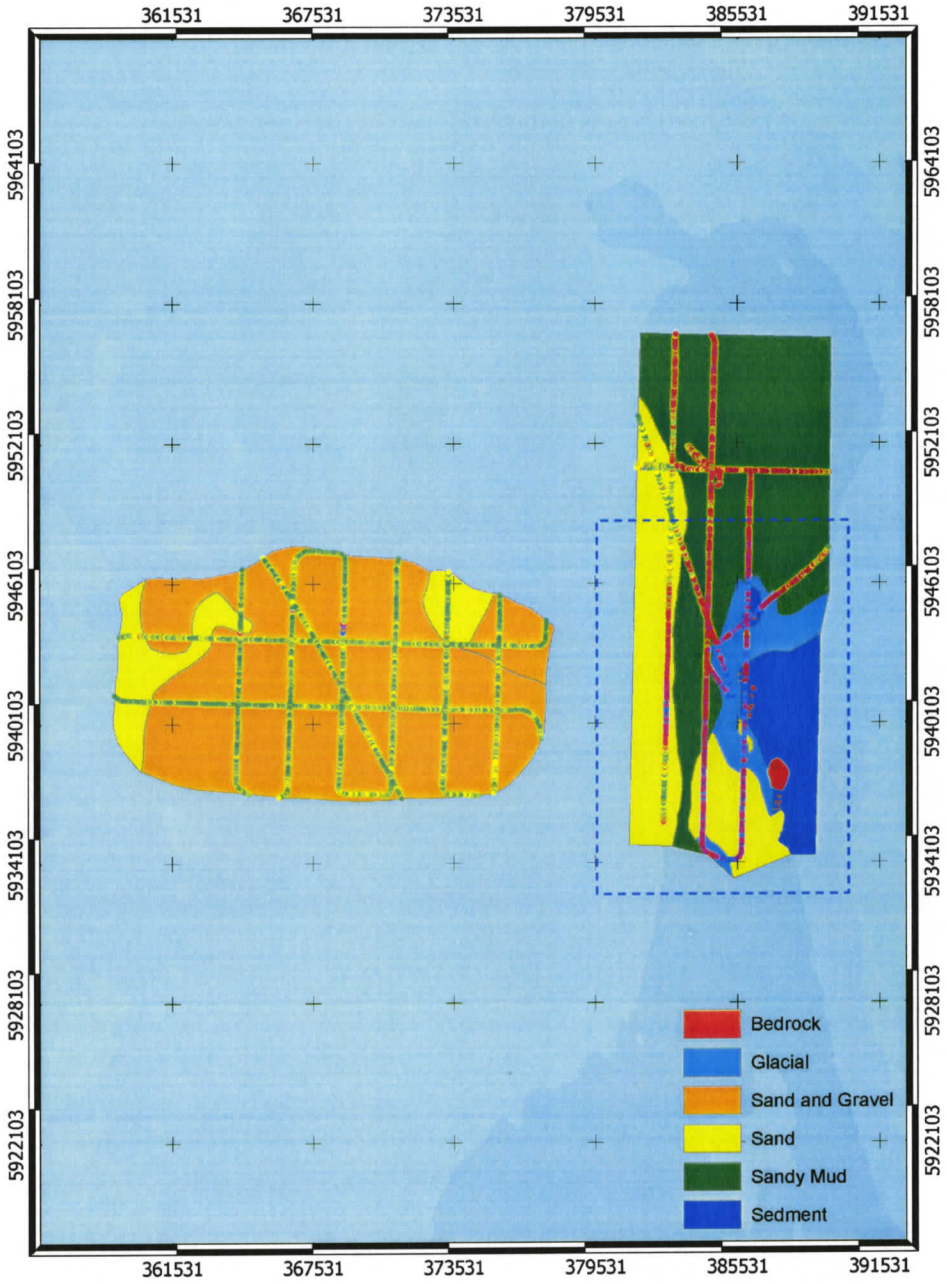

Figure 17 – Shell Ground and White Rock classified together. The dashed rectangle can be seen in close-up in Figure 18.

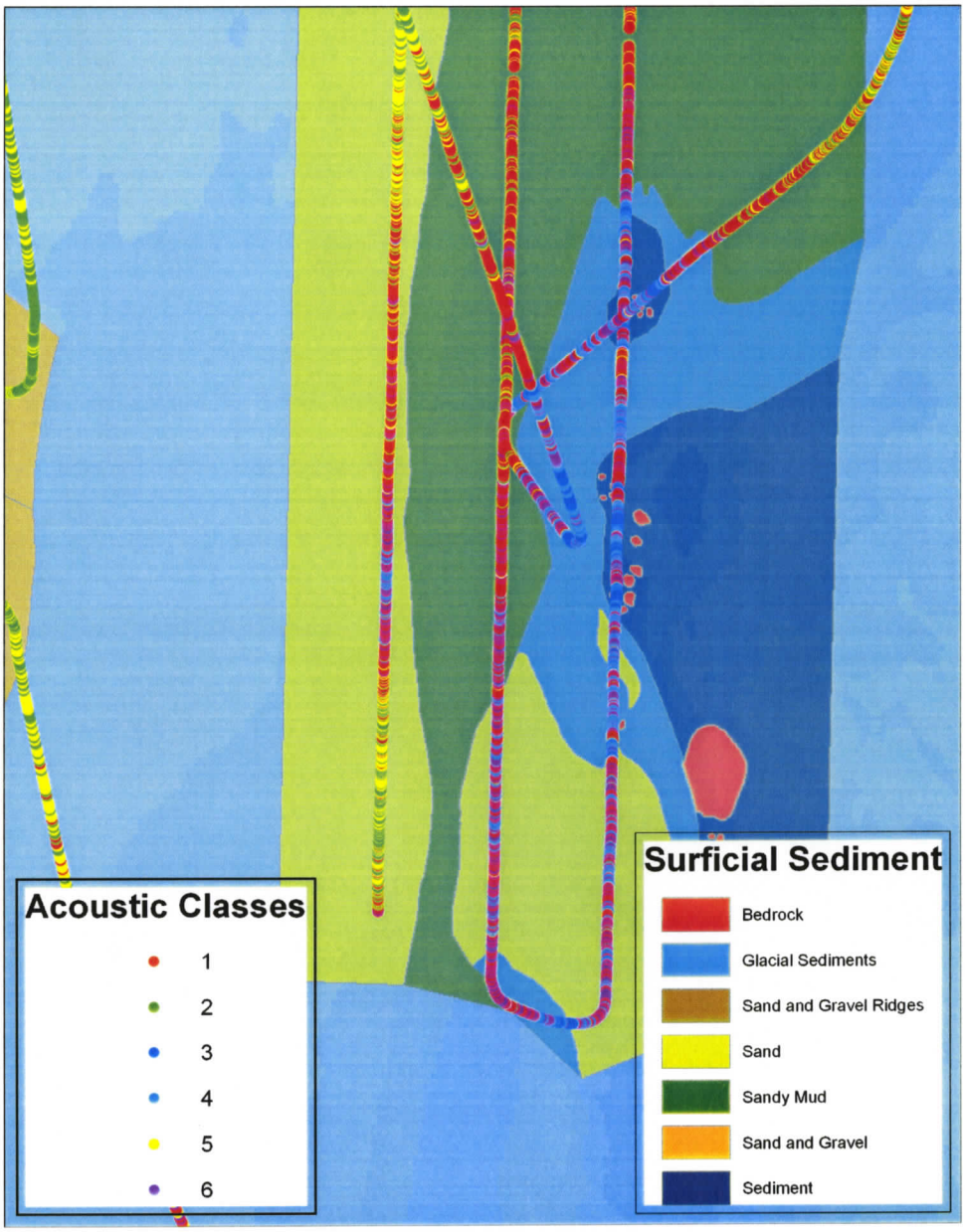

Figure 18 – Close-up of White Rock. This is the area in the rectangle in Figure 17.

In this case, Shell ground (flat plain) and White rock (high relief trench) together meet that criteria (Figure 17 and Figure 18). Figure 19 and Figure 20 show the Dundas area classified using a catalogue derived from the Shell Ground and

White Rock areas. The improvements between the catalogued and non-catalogued clustering can be seen by comparing Figure 14 and Figure 16 respectively. The catalogued clustering produced 6 classes while the non-catalogued clustering produced only 2. The catalogued method has an added benefit; avoidance of the tedious splitting process for future single beam data in areas with similar bottom types as the Shell Ground and White Rock areas. Catalogues also allow for the comparison of different areas since they will share the same acoustic class definitions.

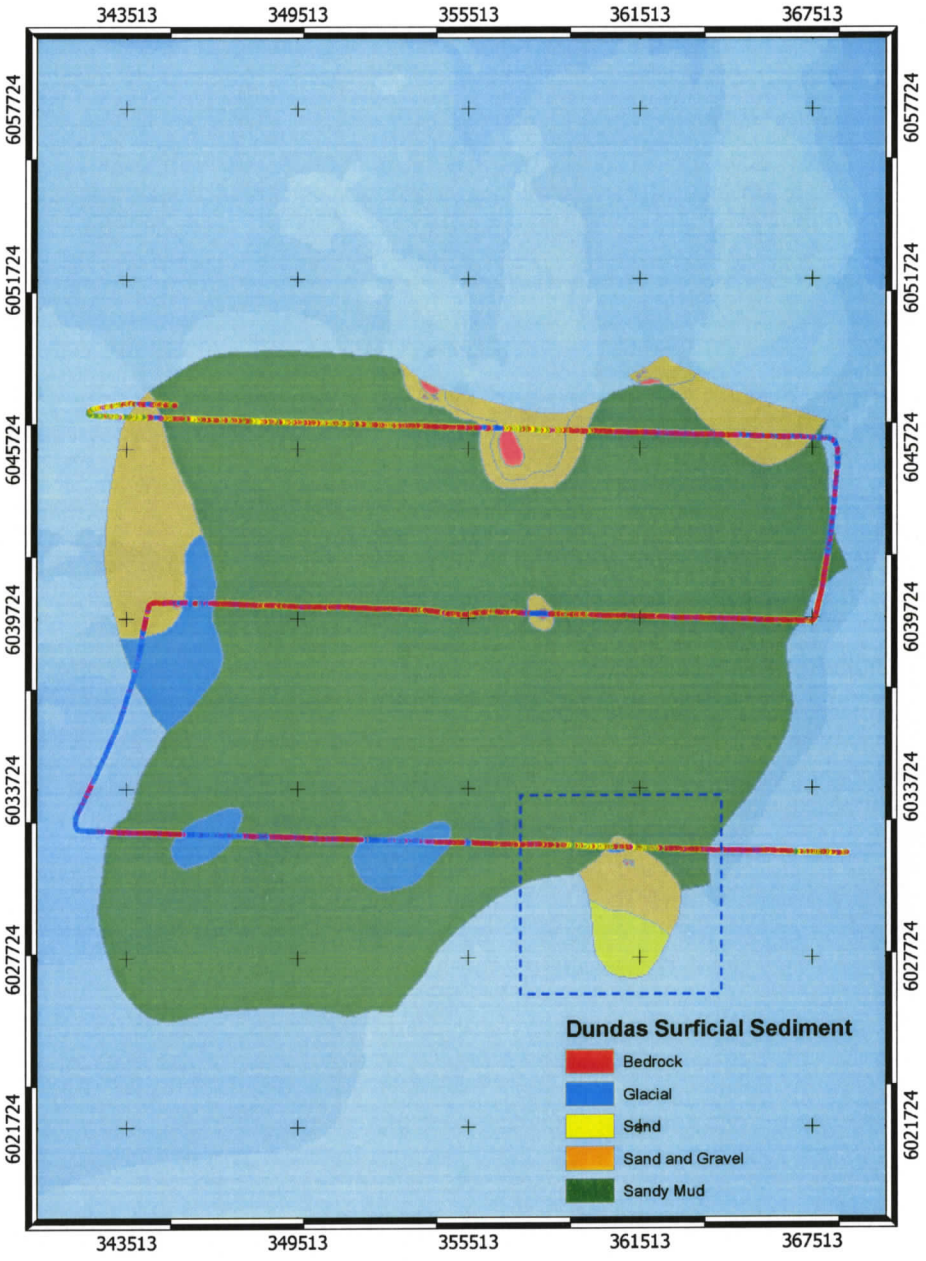

Figure 19 – Dundas surficial sediment map overlaid with catalogued single beam. The dashed rectangle can be seen in close-up in Figure 20.

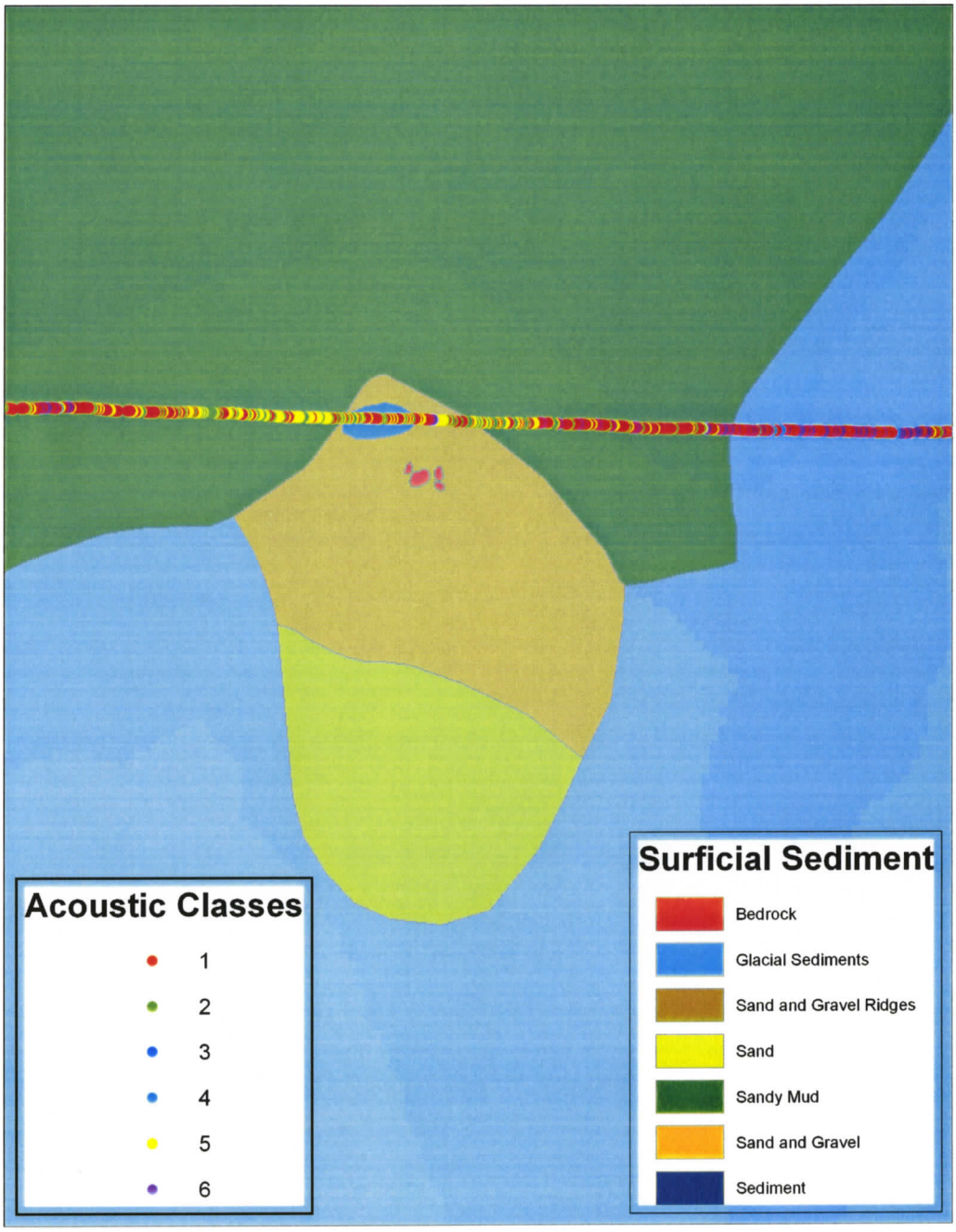

Figure 20 – Close-up of southeast corner of Dundas. This is the area in the rectangle in Figure 19.

By comparing Figure 17 and Figure 21, we can see that White Rock, when classified alone, produces a very poor classification (Figure 21). The classification is highly homogeneous, even over differing surficial sediment. When classified along with Shell Ground however, the classification improves dramatically (Figure 17). The single beam classes closely follow the surficial sediment. It was mysterious at first why the data from White Rock was classified so poorly, even after cleaning, but after Shell ground was included in the classification, it was apparent that the lack of unique bottom types in the single beam data was to blame. The improvement of the Dundas classification can be seen by comparing Figure 15, Figure 19 and Figure 20. Before classification with the catalogue, the southeast corner of Dundas was classified as only one type. After applying the catalogue, several types can be seen that correlate reasonably well with the surficial sediment type. Small differences between the surficial sediment maps in and the single beam classification (Figure 18 and Figure 20) are likely due to the low resolution of the surficial sediment map, and uncertainties in navigation data for the various surveys used in construction of the surficial sediment maps.

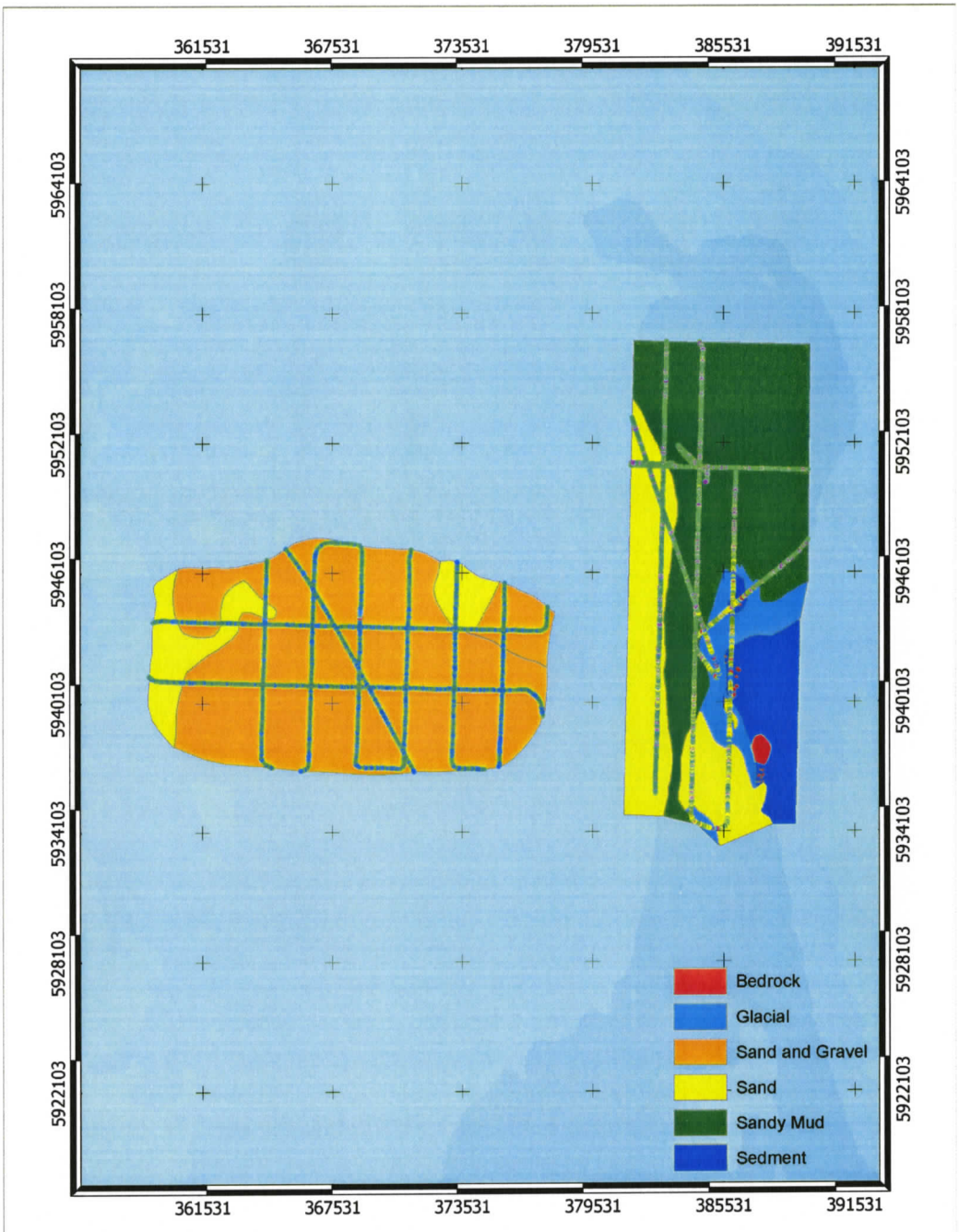

Figure 21 – Shell Ground and White Rock classified separately. Note the lack of class separation on White Rock (Right).

3.5.3 Issues of the single beam classification

A scatter plot shows the relationship between all three Q-values, depth, and the acoustic class (colour) by catalogued cluster analysis (Figure 22). The principal component Q3 contains little of the variance in the data, and is nearly insignificant in the interpretation. The striping we see in the Q1-Q2 and Q1-Q3 relationships is caused by the QTC modified K-means clustering algorithm. The algorithm creates elongated (hyper-ellipsoidal) clusters instead of the hyper-spherical clusters that the general K-means algorithm produces. Legendre (2002) points out that the structure of acoustic class clusters in multivariate space is not as important as the choice of sonar backscatter variables and the number of principal components. Therefore, this striping is not necessarily a bad thing; it is just an artifact of the analysis process.

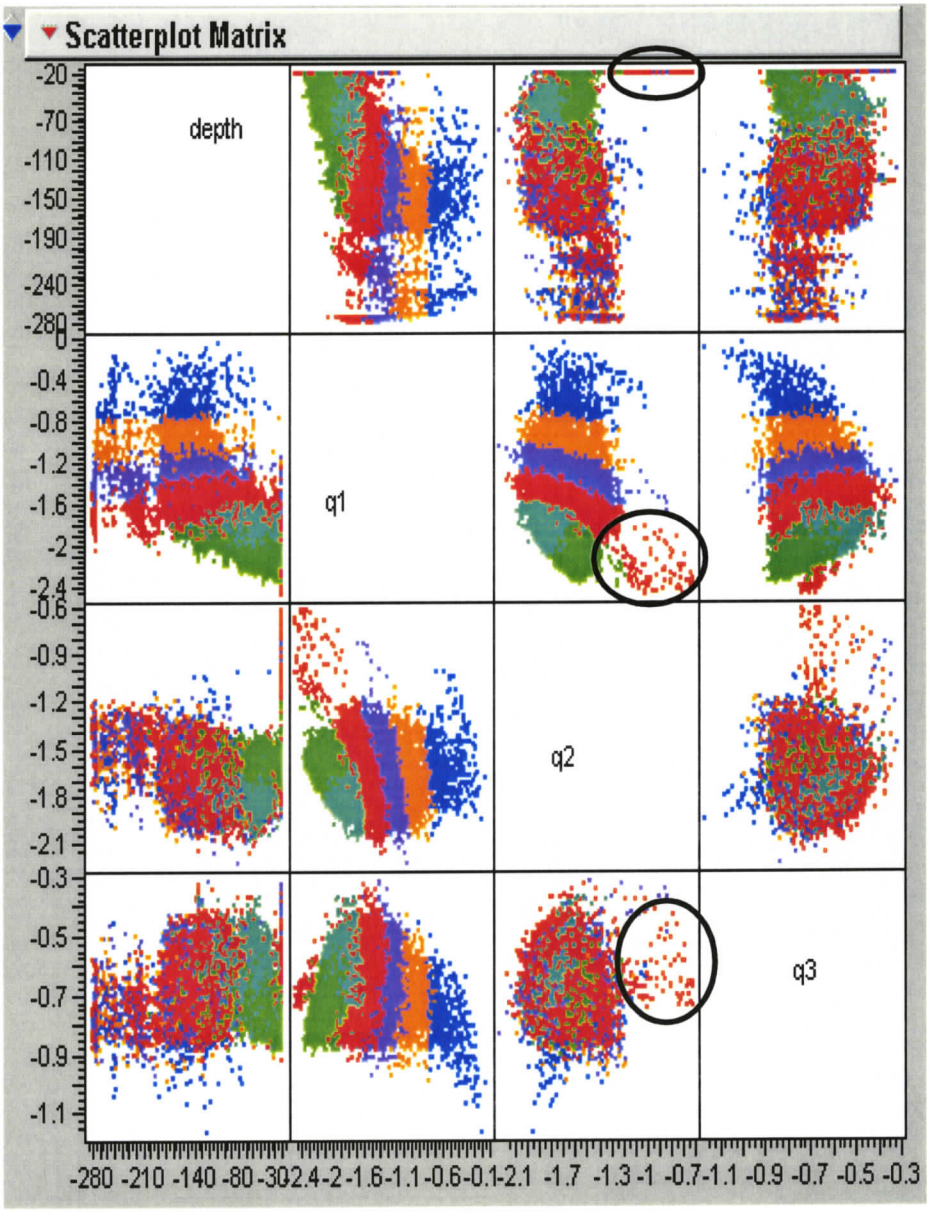

Figure 22 – Relationship between Q-values and Depth from all areas after classification. The black ellipses show a group of observations which appear at shallow depth but may be a separate class. The points are coloured to represent the bottom class assigned by catalogued clustering. Output from JMP 4.0.4 statistical software.

3.5.3.1 Slope Anomalies

The black ellipses in Figure 22 show outlier points that all occur at the same depth. Plotting these points in the GIS revealed that they were all located in the Sand & Gravel ridges region of Two Peaks Figure 23 and Figure 24.

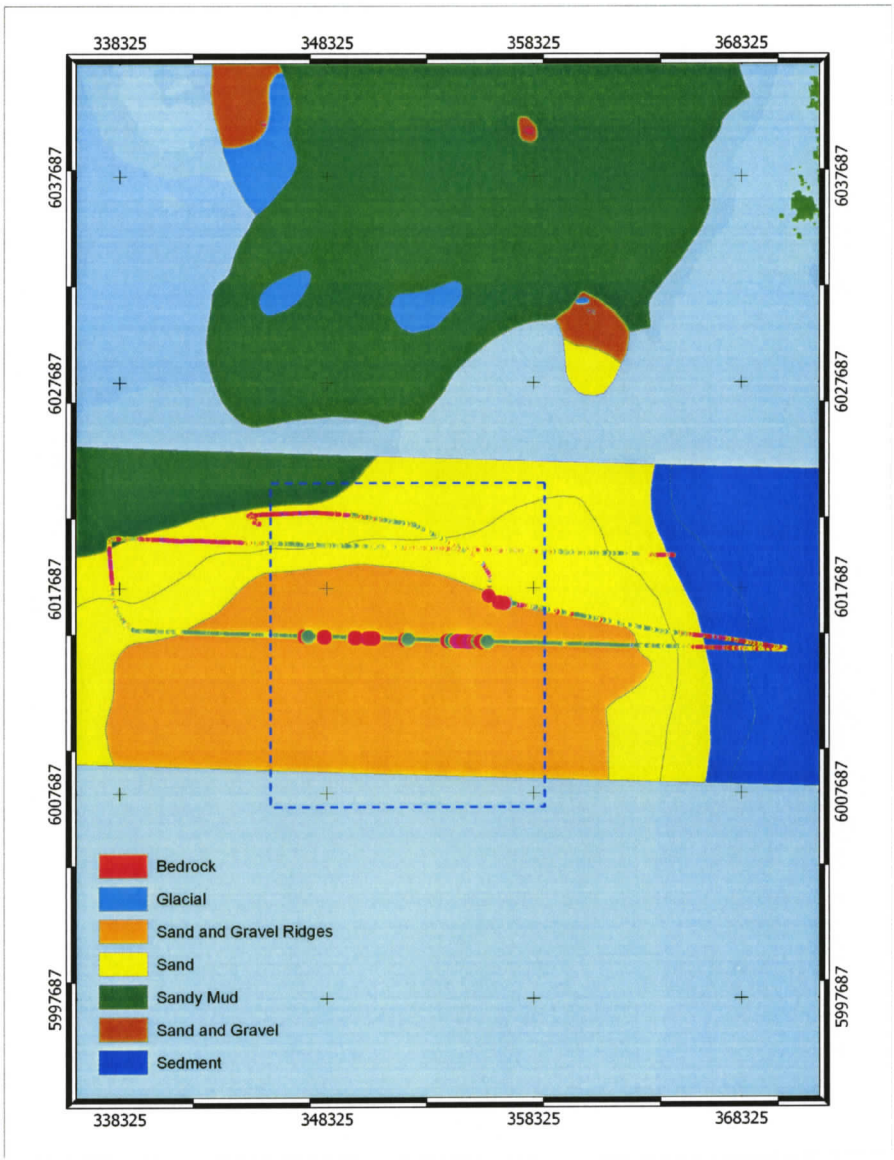

Figure 23 – Outlier points located in the Two Peaks region. The larger dots represent the same points that are inside the ellipses in Figure 22. The area inside the dashed rectangle can be seen close up in Figure 24.

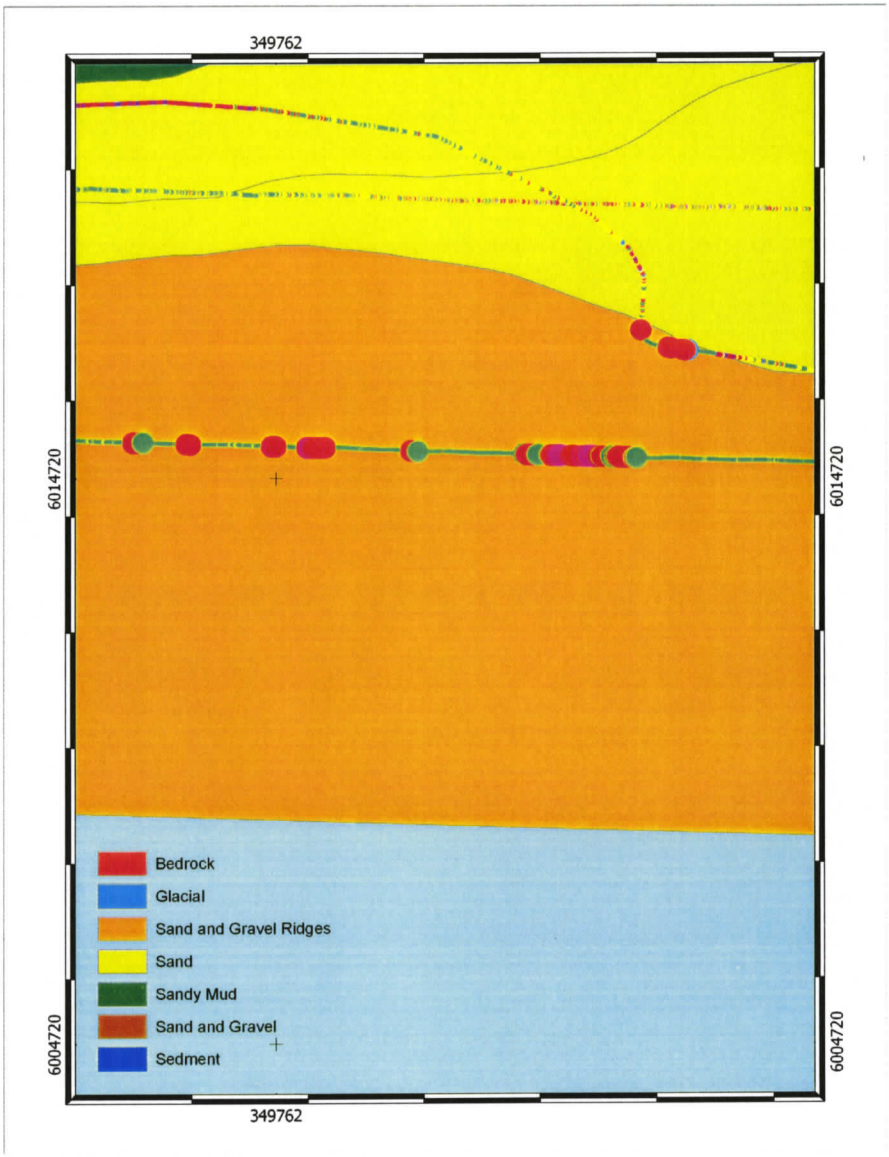

Figure 24 – Close-up of points located in the Two Peaks region. The orange area is Sand & Gravel Ridges which could explain the erroneous single beam classifications for these points. This is the area in the dashed rectangle in Figure 23.

The most likely causes for the misclassifications are sea floor slopes/ridges. It is known that the QTC single beam system can and does misclassify on ridged or sloped terrain (Biffard et al 2005); i.e. it will classify one bottom type on an up slope and a different bottom type on a down slope even though the actual bottom type remains unchanged. Without compensation for slope, single-beam echosounder seabed classification is not possible for slopes greater than one-half the beamwidth. The sidescan and Huntec data showing the mobile sand dunes in question in the Two Peaks region can be seen in Figure 25.

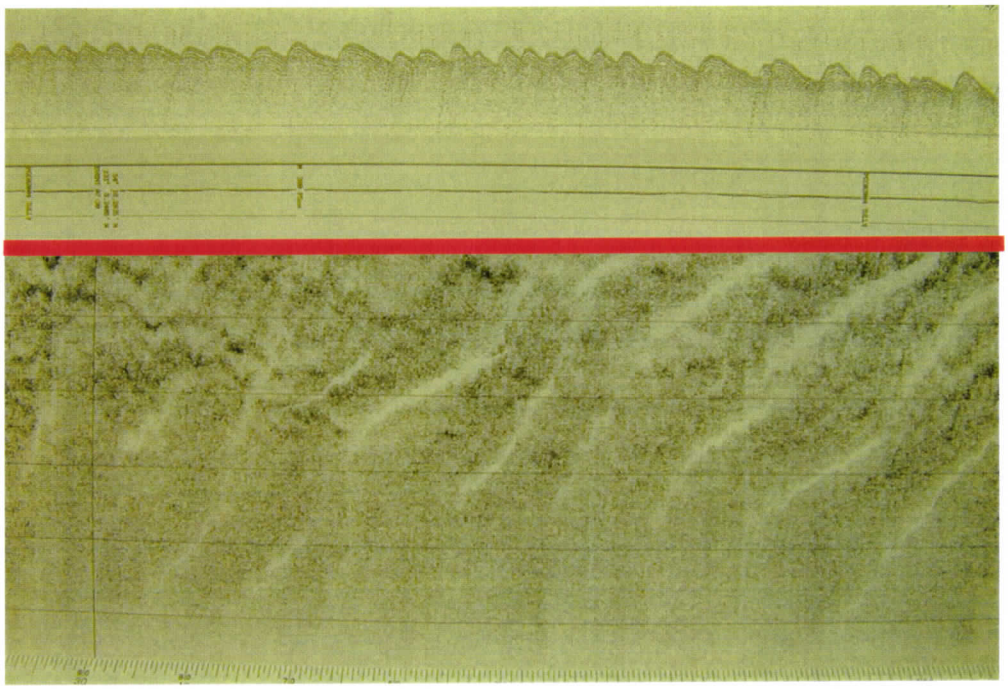

Figure 25 – Mobile sand dunes in the Two Peaks region. Sidescan sonar is shown below the red line and HUNTEC sub-bottom profile above it. The windward slope is gentle, the leeward slope is steep. Vertical scale is 1m, horizontal scale is 10m.

3.5.3.2 Non-discrete Boundaries

Note that the surficial sediment maps in Figure 15, Figure 17 - Figure 21, and Figure 23 - Figure 24 show discrete lines between bottom types. In most cases

this will not be the situation; these lines in reality are more likely bands of intermediate types. This would also explain some of the overlap issues seen in the single beam classifications. For example, in Figure 19 (inside the rectangle) and Figure 20 (close-up view) the southeast corner of Dundas single beam shows a sand & gravel (yellow & green) type where the surficial sediment map shows sandy mud (green). If a merging zone were shown, we would observe that the single beam classification is more correct in that area. This occurrence can also be seen in Figure 17 and Figure 18 on the west side of White Rock (inside the rectangle) where the sand and sandy mud meet. These intermediate zones will be wider for areas of low relief since the bottom type is correlated to depth.

Chapter 4
Correspondence Analysis

Correspondence Analysis is a multidimensional data compression technique used to show correlation between row and column profiles (counts) in a frequency table, in this case between surficial sediment type and either single beam classification or catch composition. A frequency table is defined here as a table with rows representing some dependent variable and columns representing some independent variable, with the frequency of the independent variable for each dependent variable in the cells that intersect these. Figure 26 shows an example of a frequency table that could be used in a correspondence analysis.

		Groundfish Catch			
		Arrowtooth Flounder	Dover Sole	English Sole	Rock Sole
Surficial sediment	Bedrock				
	Glacial				
	Gravel and Sand				
	Sand				
	Sandy mud				

Figure 26 – Example of a frequency table as used in correspondence analysis. The empty cells in this table will be frequencies of occurrences of the species over the respective bottom types.

The output from a correspondence analysis is a two or three-dimensional plot representing the majority of the variability in multidimensional data. This lower dimension plot is created through the Singular Value Decomposition of the Chi-square distance matrix of the frequency table, the details of which are described in the following two sections.

4.1 Inertia - Distance from Independence

To say something meaningful about the frequency tables, the observed distribution must be compared with the expected values. Chi-square tests for independence are performed to test the null hypothesis that the frequency in the cells is what would be expected, given the row and column sums. This χ^2 statistic is used as a measure of distance that separates this distribution from independence; $\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

where O_{ij} is the observed frequency in a contingency table for the cell located at row i and column j, E_{ij} is the expected frequency for the cell. In correspondence analysis, the χ^2 must be divided by the grand total of the frequency table, to avoid the effect of a relative increase in frequencies increasing the χ^2 statistic. For example, if each frequency in the table were doubled, the relative distances to the mean will not increase but the χ^2 value will. This statistic, χ^2 divided by frequency total, is called the *Inertia*, $\phi^2 = \frac{\chi^2}{N}$. Since total table frequency N is a factor of inertia, the inertia is a normalized value and will fall between zero and one with zero representing no association and one representing perfect association. In a correspondence analysis there exists an Inertia value for each of the singular values calculated by Singular Value Decomposition. The word Inertia is an analogy to the applied mathematics *moment of inertia* which is the integral mass times the squared distance to the centroid. This is known as Pearson's mean square contingency coefficient computed on the table of frequencies (Greenacre 1984, p35).

4.2 Singular Value Decomposition

The Singular Value Decomposition (SVD) of a matrix **A** is the decomposition of the matrix as a product of three matrices of simpler form and geometric interpretation (Greenacre, 1984, p341). Performing this decomposition allows the data to be transformed onto new axes, which are orthogonal to one another, thus avoiding correlation between axes. The decomposition of **A** is defined as follows:

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathbf{T}}$$

where:

A is a real m × n matrix (the frequency table).

U and $\mathbf{V}^{\mathbf{T}}$ are m × k and k × n matrices respectively.

The columns of **U** and the rows of $\mathbf{V}^{\mathbf{T}}$ are orthonormal

D is a k × k diagonal matrix of real nonnegative numbers (the singular values), in descending order.

The columns of **U** are left singular vectors, and are the eigenvectors of the square matrix $\mathbf{A}\mathbf{A}^{\mathbf{T}}$.

The rows $\mathbf{V}^{\mathbf{T}}$ are right singular vectors, and are the eigenvectors of the square matrix $\mathbf{A}^{\mathbf{T}}\mathbf{A}$.

For an n × n matrix **W**, a nonzero vector **x** is the eigenvector of **W** if $\mathbf{W}\mathbf{x} = \lambda\mathbf{x}$ or $(\mathbf{W} - \lambda\mathbf{I})\mathbf{x} = 0$ for some scalar λ. The scalar λ is the eigenvalue of **W**. Since the vector **x** is nonzero, the matrix $\mathbf{W} - \lambda\mathbf{I}$ must be singular (non-invertible), meaning its determinant is zero; $|\mathbf{W} - \lambda\mathbf{I}| = 0$. The characteristic

polynomial is used to solve this equation exactly for small matrices, but for larger matrices numerical methods are employed for an estimate. These methods vary and are generally iterative with convergence to the solution. JMP 4.0.4 (2001) software returns:

- Singular values - which are the diagonal entries in the matrix $\mathbf{D}$, their squared values are the Inertias of the principal axes.

- Singular vectors – which are the left and right orthonormal vectors $\mathbf{U}$ and $\mathbf{V}^{\mathbf{T}}$. These vectors define the orthogonal principal axes of the column and row points respectively and are used in the biplot routine.

4.3 Analysis Details

Correspondence Analyses between the following datasets were performed for this study:

1. Single beam classes and the surficial sediment.

2. Groundfish catch data and surficial sediment, where the catch data were organized as follows:

    a. Single species data, no clustering or grouping.

    b. Multi-species assemblages based on hierarchical clusters.

Correspondence analysis was not performed between single beam classes and groundfish catch for reasons explained in Section 4.3.2.

4.3.1 Single Beam and Surficial Sediment

To prepare the data, a layer-based GIS was used (ArcMap 8.3) where the single beam classes were represented as a point layer with a class attribute for each point and the surficial sediment a polygon layer with polygons for each bottom type. Only the single beam points contained within the surficial sediment polygons were used; a frequency table of single beam class over surficial sediment polygon was generated. The resulting frequency table was imported into a statistical package (JMP 4.0.4) where a contingency table and correspondence analysis plot were produced to show the relationships (Figure 27 and Figure 28).

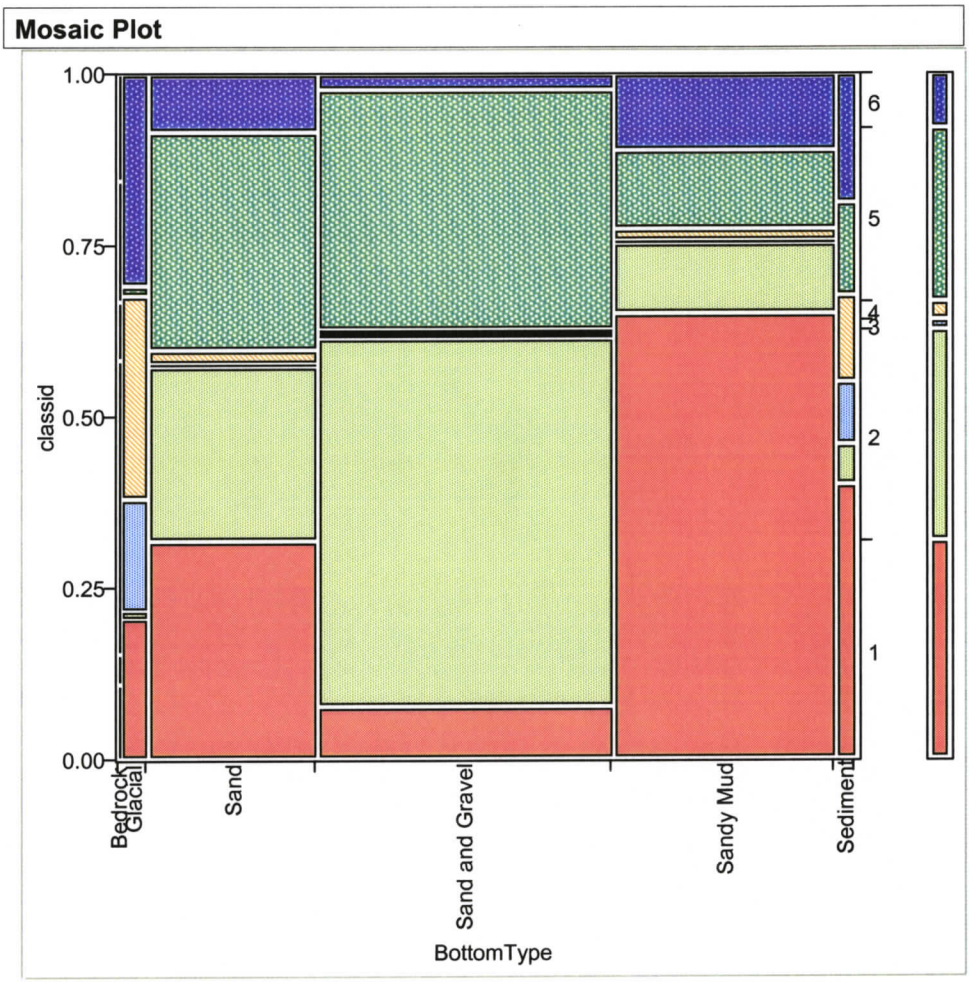

Figure 27 – Contingency plot - Relationship between surficial sediment (X-axis) and single beam acoustic classes (Y-axis, labeled 1-6).

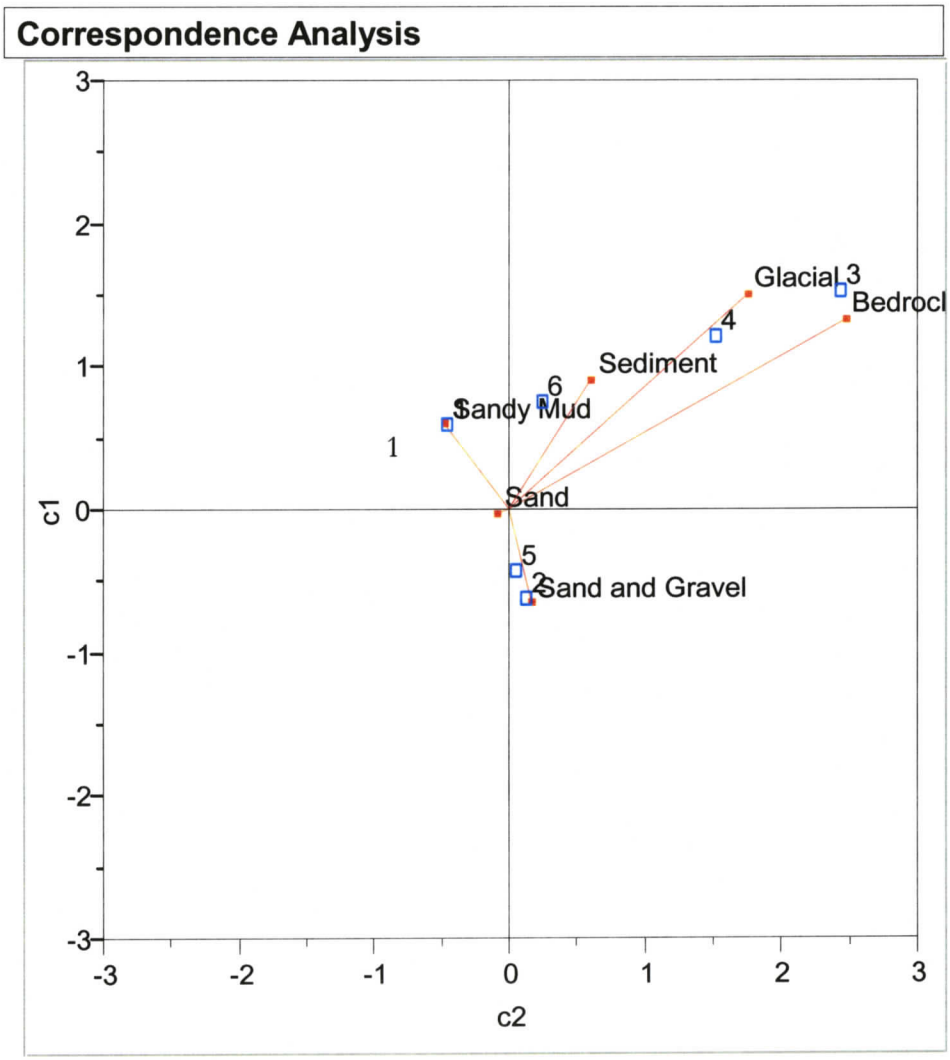

Figure 28 – Correspondence biplot - between surficial sediment (red lines) and single beam classes (blue squares, labeled 1-6). This plot shows the correspondence analysis results for the contingency plot seen in Figure 27.

The areas of the rectangles in the contingency plot can be used as a visual guide to determine correlation. For example, in Figure 27, single beam class 1 can be said to correlate with the 'Sandy Mud' surficial sediment type because the

rectangular area in the mosaic is large. This observation is reinforced by the correspondence biplot (Figure 28), where the blue square of single beam class 1 is directly underneath the end of the red ray of the 'Sandy Mud' surficial sediment class. To elaborate on these visual cues, an examination of numerical outputs is warranted. Some other notes on the contingency plots such as that in Figure 27:

- The colour 'thermometer' at the right hand side represents the proportion of each single beam class in the input data.

- The horizontal width of the surficial sediment bottom types represents the proportion of that type to the total for all types shown.

- The previous two points imply: that the area of each rectangle represents the correlation between surficial sediment bottom types and single beam classes. The larger the area, the higher the correlation.

			Axis c1					Axis c2					Axis c3				
Variable	**QLT**	**mass**	**c1**	**CTR**	**CTR(%)**	**COR**	***α***	**c2**	**CTR**	**CTR(%)**	**COR**	***β***	**c3**	**CTR**	**CTR(%)**	**COR**	***ζ***
Bedrock	599	0.0029	1.329	5	1.3	134	68.6	2.481	18	7.9	465	47.0	2.303	15	85.2	401	50.7
Glacial	990	0.0388	1.515	89	22.6	421	49.5	1.76	120	53.2	569	41.0	-0.23	2	11.4	10	84.3
Sand	858	0.2300	-0.027	0	0.0	90	72.6	-0.079	1	0.6	768	28.8	-0.034	0	1.5	142	67.8
Gravel and Sand	1000	0.4011	-0.651	170	43.1	938	14.4	0.167	11	5.0	62	75.6	0.004	0	0.0	0	89.7
Sandy mud	999	0.3011	0.601	109	27.6	624	37.8	-0.466	65	28.9	375	52.2	0.021	0	0.7	1	88.4
Sediment	996	0.0261	0.913	22	5.5	687	34.0	0.612	10	4.3	309	56.2	0.068	0	0.7	4	86.5
Single Beam Class 1	998	0.3192	0.587	110	27.9	613	38.5	-0.465	69	30.5	385	51.7	0.0336	0	2.0	2	87.4
Single Beam Class 2	1000	0.3086	-0.626	121	30.6	960	11.5	0.127	5	2.2	40	78.5	-0.0011	0	0.0	0	89.9
Single Beam Class 3	919	0.0136	1.521	31	8.0	258	59.5	2.434	81	35.7	661	35.6	0.8537	10	54.9	81	73.4
Single Beam Class 4	942	0.0284	1.212	42	10.6	365	52.8	1.523	66	29.2	577	40.6	-0.4842	7	36.9	58	76.0
Single Beam Class 5	1000	0.2516	-0.43	47	11.8	985	7.1	0.053	1	0.3	15	83.0	0.0045	0	0.0	0	89.4
Single Beam Class 6	978	0.0786	0.749	44	11.2	880	20.2	0.249	5	2.2	97	71.8	-0.1192	1	6.2	22	81.4

Table 2 – Numerical Output from correspondence Analysis of Single Beam classes with Surficial Sediment types.

Numerical output of the correspondence analysis is shown in Table 2. This table was built using Microsoft Excel based on an example given in Greenacre (1984, p75). The table is made up of three groups of columns, one group per axis of correspondence; these are labeled Axis c1, Axis c2, and Axis c3. Black text signifies surficial sediment bottom type and red text signifies single beam class. Each point on the CA biplot (Figure 28) is defined by the axis values c1 and c2. For each of c1, c2, and c3:

- c1, c2, or c3 is the point value along the primary axis; these are obtained from the left and right singular vectors of the SVD.

- CTR is the contribution to the first principal inertia by the point. This field is calculated: $CTR = rc_n^2$ where r is the mass of the point (row or column sum from the mass column) and c_n is the mass of the point on axis n (column c1, c2, or c3 in the table). This value has been multiplied by 1000 for ease of examination.

- CTR(%) is the contribution to the principal inertia by the point as a percentage. This is calculated as $CTR(\%) = \frac{rc_n^2}{\phi_n^2}$ where c_n is the axis mass; c1, c2, or c3 from the table, r is proportion of the row or column sum that the point accounts for, and ϕ_n^2 is the inertia of axis n.

- α, β, and ζ are the angles from each axis to the given point. They were calculated using the direction cosines: $\cos\alpha = \frac{c_1}{\sqrt{c_1^2 + c_2^2 + c_3^2}}$, $\cos\beta = \frac{c_2}{\sqrt{c_1^2 + c_2^2 + c_3^2}}$, $\cos\zeta = \frac{c_3}{\sqrt{c_1^2 + c_2^2 + c_3^2}}$. The angle to the axis is

shown in each case. The angles are required to find the contribution of inertia to the point (COR) and also to allow the analyst to visualize all angles from axes, including the third axis which is not shown in the biplot (Figure 28). Greenacre uses another, equivalent method for finding the angles, by examining the appropriate submatrices of the generalized Singular Value Decomposition (SVD), and multiplying the mass of the point by the square of the position on the axis (Greenacre, p61 & 67).

- COR is the contribution of the axis to the inertia of the point. $COR = \cos^2 \alpha$, $\cos^2 \beta$, and $\cos^2 \zeta$ for axes c1, c2, and c3 respectively. Comparison of all three of these correlations for each point show which axes represent the point well. This value has been multiplied by 1000 for ease of examination.

- Mass is the proportion of the row or column sum that the point accounts for.

- QLT is the quality of the point in a two dimensional plot. It is the squared correlation (cosine squared of α and β) of the point to the plane defined by the axes c1 and c2. This value is a simple sum of the first two axes' COR fields, which represent the individual cosine squared correlations. It is possible to take the sum of the correlations in this case because $\cos^2 \alpha + \cos^2 \beta = \cos^2 \theta$ where θ is the angle between the point and the plane defined by the c1 and c2 axes.

Some other observations from Figure 27, Figure 28 and Table 2:

- Since sand is located at the origin of the correspondence plot, it does not correspond to any class and therefore does not have its own acoustic class. This can also be seen in the contingency plot (Figure 27) as an equal

distribution across 4 of 6 single beam acoustic classes. Sand is located at the origin because its contribution to the inertia of those axes (CTR value) is zero (Table 2).

- Gravel and Sand and both single beam classes 2 and 5 are highly correlated to the first primary axis. These can be seen as values close to 1000 for axis c1 COR in Table 2. Similarly, Glacial appears to correspond well with single beam classes 3 and 4, although not as strongly on the first primary axis, which accounts for most of the variability.

The lack of trackline crossings and sparseness of tracklines precluded using single beam to classify all habitats encapsulated within the fishing areas chosen for the study.

4.3.2 Groundfish Catch and Surficial Sediment

Catch data for individual species were extracted as explained in section 2.2.5, and correspondence with surficial sediment was analyzed in the same way as the single beam classes. Inertia only was examined at this stage, and compared with the clustered species' Inertias to determine which method should by explored further. The clustered species data had greater Inertias for the first two primary axes and therefore that group was more exhaustively analyzed.

Hierarchical clustering on the catch data was performed using Ward's linkage, a minimum variance method (JMP statistical software manual). The distance between the clusters was calculated as the Analysis of Variance or ANOVA (JMP statistical software manual) sum of squares. This method starts with each data point being its own cluster, then links the pair of clusters which produce the smallest variance in the merged group. The smallest mean squared distance to the centroid for each pair is found, and those two clusters become one. This process

is iterative, ending when there is only one cluster left, containing all the original data points. The distance for Ward's linkage is given as (JMP manual, 2001):

$$D_{KL} = \frac{\left\| \overline{x_K} - \overline{x_L} \right\|^2}{\left(\frac{1}{N_K} + \frac{1}{N_L} \right)}$$

where:

$\overline{x_K}$ is the mean vector for cluster K

$\overline{x_L}$ is the mean vector for cluster L

$\left\| \overline{x_K} - \overline{x_L} \right\|$ is the square root of the sum of squares of the elements of $\overline{x_K} - \overline{x_L}$

N_K is the number of observations in cluster K

N_L is the number of observations in cluster L

A dendrogram is a tree diagram that lists each observation and shows which cluster it is in and the dissimilarity level the observation was at when it entered the cluster. With JMP software, a dendrogram is produced and the number of clusters is entered by the analyst; the software finds the proper dissimilarity level and returns the number of clusters requested. The correspondence analyses allow us to explore the clusters' affiliation for certain bottom types over a period of 7 years and come to some conclusions about what bottom types certain groups of fish prefer. A preliminary examination of the axis inertias of the clustered species catch and the single species catch showed that the clustered species data

performed better in the correspondence analysis, thus warranting a look at the numerical outputs for these data.

Chapter 5
Correspondence between Habitat and Species

Surficial sediment maps are associated with catch data in this chapter using correspondence analysis. First, a correspondence analysis between the 14 species with highest catch in Hecate Strait and surficial sediment is performed, then this is compared with the correspondence between multi-species assemblages and surficial sediment.

5.1 Single Species and Surficial Sediment

Correspondence analysis was performed between the single species (non-clustered) catch data and surficial sediment type. Figure 29–36 show the contingency tables and correspondence biplots for each year 1996-2002. The widths of the bottom types (X-axis) on the contingency plots represent the proportion of area of that bottom type with respect to the total area. The colour 'thermometer' on the right side of the contingency plot represents the proportion of each species caught with respect to the total caught. Larger area of the species-bottom type rectangles can therefore be interpreted as a higher correlation of the two variables. The species codes can be found in Table 1.

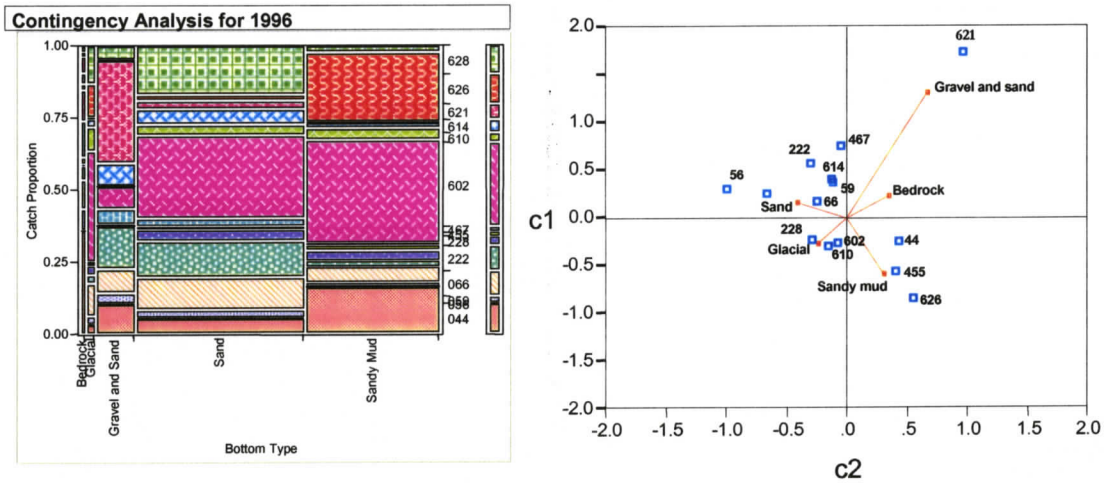

Figure 29 – Contingency plot and correspondence analysis for 1996 catch data and surficial sediment bottom type. Rays represent bottom type and squares represent species.

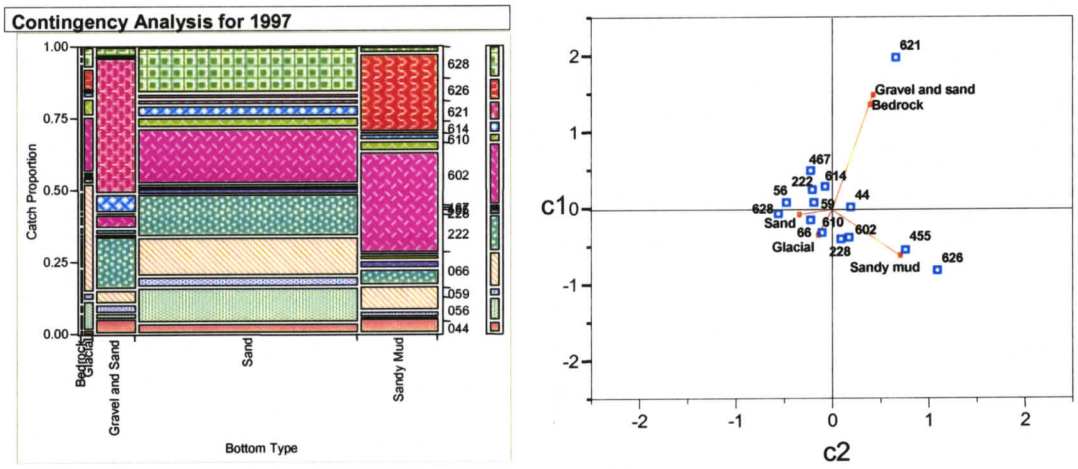

Figure 30 – Contingency plot and correspondence analysis for 1997 catch data and surficial sediment bottom type. Rays represent bottom type and squares represent species.

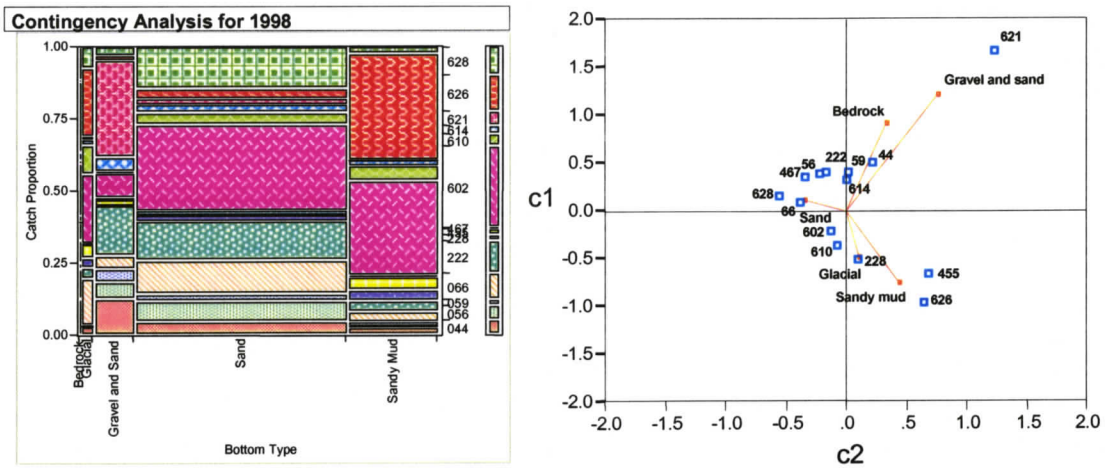

Figure 31 – Contingency plot and correspondence analysis for 1998 catch data and surficial sediment bottom type. Rays represent bottom type and squares represent species.

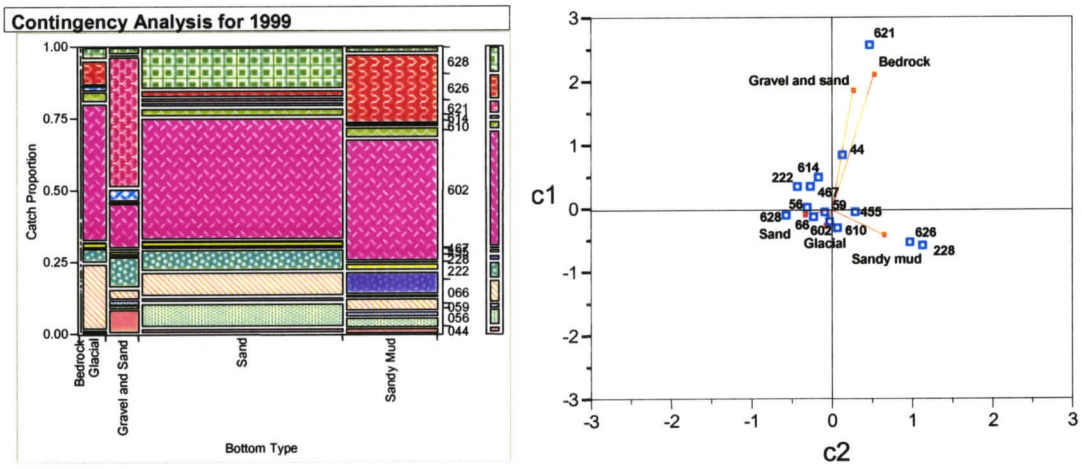

Figure 32 – Contingency plot and correspondence analysis for 1999 catch data and surficial sediment bottom type. Rays represent bottom type and squares represent species.

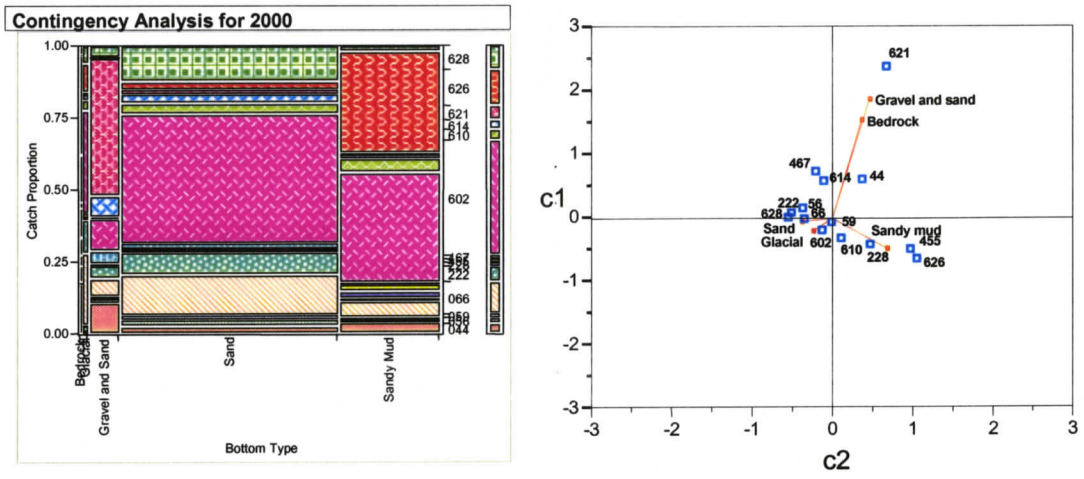

Figure 33 – Contingency plot and correspondence analysis for 2000 catch data and surficial sediment bottom type. Rays represent bottom type and squares represent species.

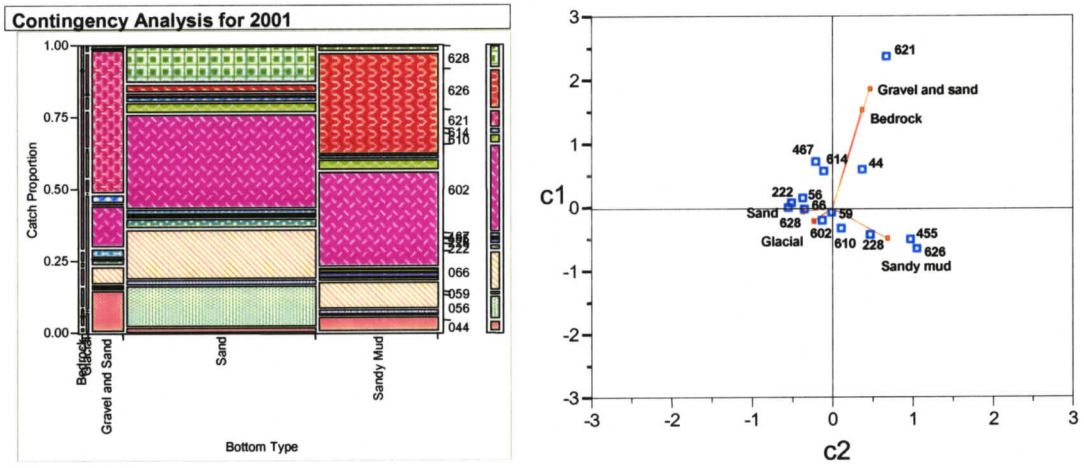

Figure 34 – Contingency plot and correspondence analysis for 2001 catch data and surficial sediment bottom type. Rays represent bottom type and squares represent species.

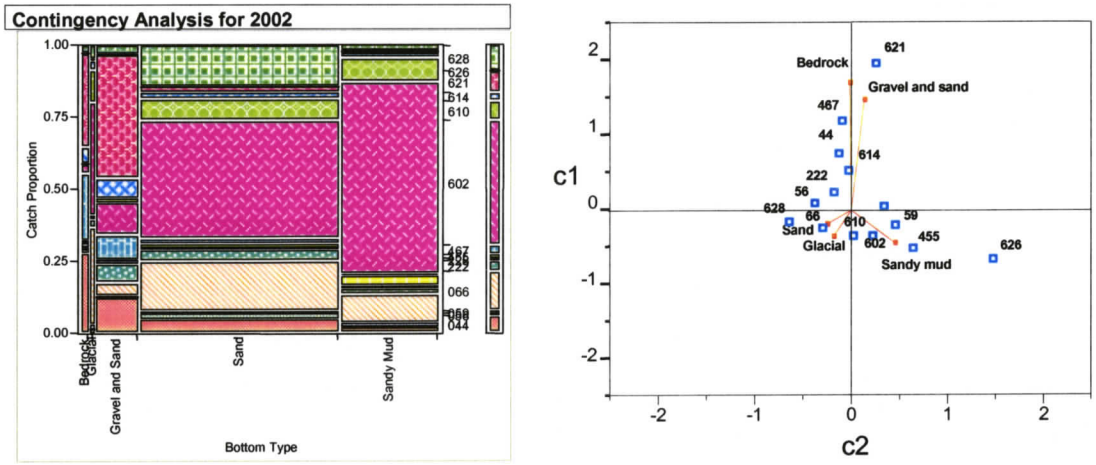

Figure 35 – Contingency plot and correspondence analysis for 2002 catch data and surficial sediment bottom type. Rays represent bottom type and squares represent species.

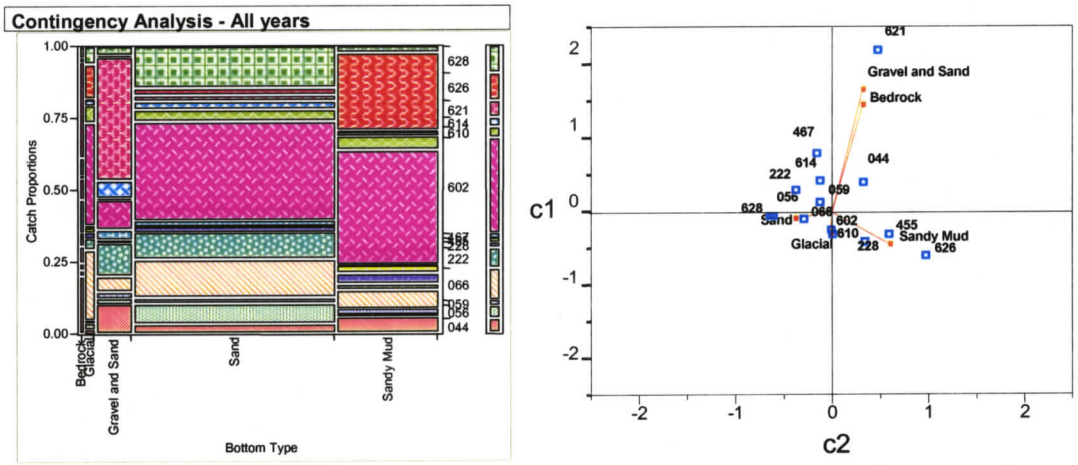

Figure 36 – Contingency plot and correspondence analysis for years 1996-2002 combined catch data and surficial sediment bottom type. Rays represent bottom type and squares represent species.

For Figure 29 - Figure 36, the blue squares represent the 14 species, and the red X's with rays represent the 5 surficial sediment bottom types. The linear (Euclidean) distance between the bottom types and species along each axis indicates the correspondence between them; the closer they are, the more they are associated. Species associations decrease near the origin of the plots and at the origin there is no association between bottom type and species. Only two axes are shown in the correspondence analysis biplots. The SVD of the contingency tables is summarized in Table 3. The Inertias are the squares of the singular values. The contribution percentage is calculated as the axis inertia divided by the total inertia. For these datasets, there were four axes used in the calculation, but the fourth axis contributed less than 1% of the inertia and was excluded for simplicity.

The principal and secondary axes (Axis 1 & 2) account for most of the inertia in all datasets, for example for the ALL column in Table 3, the principal axis accounts for 66.9% and the secondary axis accounts for 31.8%, giving a total of 66.9% + 31.8% = 98.7% of the inertia being described by the two dimensional plot of Figure 36. The individual years form a very similar pattern, with 60%-70% of the inertia being carried by the principal axis, and 20%-30% being carried by the secondary axis.

	1996	1997	1998	1999	2000	2001	2002	ALL	Summer	Winter
Axis 1 Singular Value	.62917	.64122	.64179	.70109	.69099	.77400	.71737	.66958	.68823	.68153
Axis 1 Inertia	.39585	.41116	.41189	.49152	.47746	.59908	.51461	.44834	.47367	.46448
Axis 1 (%)	**69.4**	**64.0**	**61.3**	**69.5**	**67.0**	**65.1**	**78.5**	**66.9**	**69.2**	**70.4**
Axis 2 Singular Value	.39056	.46665	.49952	.44802	.48002	.55641	.33205	.46158	.44862	.42933
Axis 2 Inertia	.15254	.21776	.24952	.20072	.23042	.30959	.11026	.21306	.20126	.18433
Axis 2 (%)	**26.8**	**33.9**	**37.4**	**28.4**	**32.3**	**33.7**	**16.8**	**31.8**	**29.4**	**27.9**
CUM (%)	**96.2**	**97.9**	**98.7**	**97.9**	**99.3**	**98.8**	**95.3**	**98.7**	**98.6**	**98.3**
Axis 3 Singular Value	.13740	.11586	.09950	.11694	.06314	.08858	.16302	.08116	.08084	.09312
Axis 3 Inertia	.01888	.01342	.00990	.01368	.00399	.00785	.02658	.00659	.00653	.00867
Axis 3 (%)	**3.3**	**2.1**	**1.5**	**1.9**	**0.6**	**0.9**	**4.1**	**1.0**	**1.0**	**1.3**

Table 3 – Correspondence Analysis Singular values, Inertias, and percent contribution for the first three primary axes. ALL is a combination of all years 1996-2002, Summer and Winter are for all years

To view seasonal effects, the data were extracted by season for the 7 years 1996-2002. Winter is defined as October 1 through March 31, and summer is the rest of the year. The results of these two datasets are seen in Figure 37 and Figure 38.

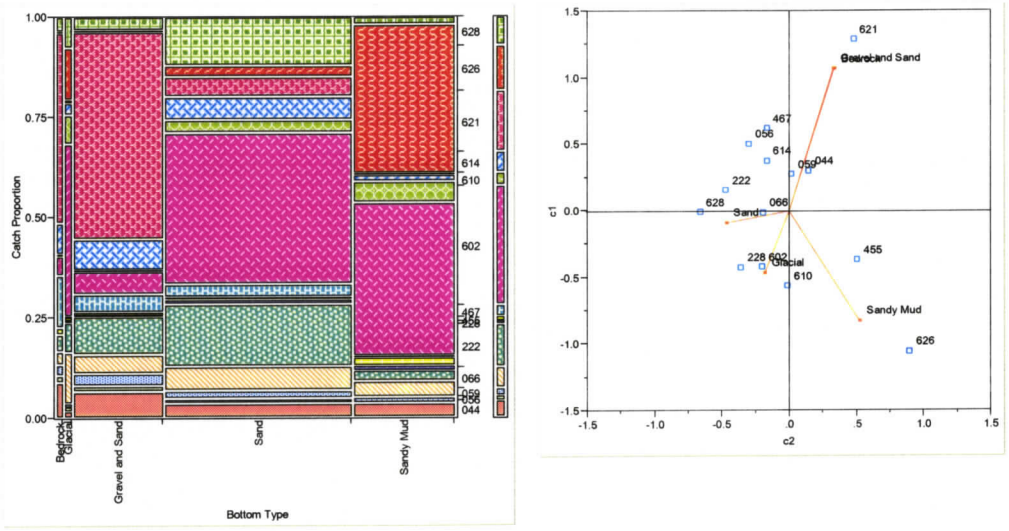

Figure 37 – Mosaic plot and correspondence analysis for the summers of 1996-2002 combined catch data and surficial sediment bottom type. Rays represent bottom type and squares represent species.

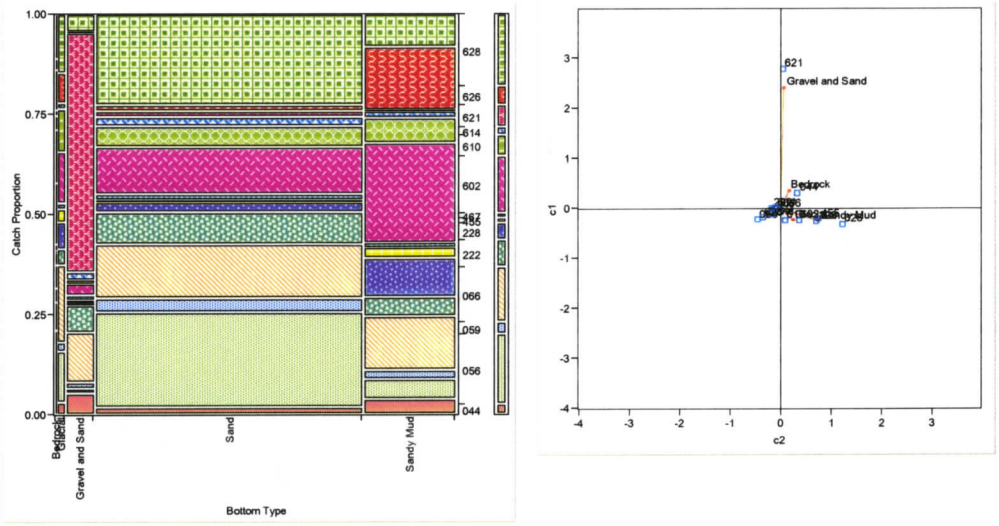

Figure 38 – Mosaic plot and correspondence analysis for the winters of 1996-2002 combined catch data and surficial sediment bottom type. Rays represent bottom type and squares represent species.

When compared to the all-seasons 1996-2002 plot (Figure 36), correspondence is elevated for summer data and reduced for winter data with the exception of Rock Sole (621) in the winter. This is not due to significantly greater catch in the winter; for Rock Sole 86.8% of the catch was in the summer months.

5.2 Clustered Species and Surficial Sediment

5.2.1 Cluster Analysis Distribution

The hierarchical clustering of the bottom trawl data differed by year; some species were not caught in any appreciable amount for some years, so the clusters were not always the same by year. The number of clusters chosen for inclusion in the JMP 4.0.4 software was 4, by viewing the dissimilarity levels on the dendrograms and choosing the number of clusters that encapsulated the most data. The species with the largest number in each cluster became the name of the

cluster; i.e. a cluster with the name 'Rock Sole' would be composed of mostly Rock Sole with other species aggregated. Figure 39 - Figure 45 show the clusters included for each analysis.

5.2.2 Correspondence Analysis

The same analysis methods used in the previous section on single-species data were also be used for clustered species data.

The clusters were used as inputs into a correspondence analysis along with surficial sediment bottom type. Figure 39 - Figure 45 show the contingency tables and results for the correspondence analysis.

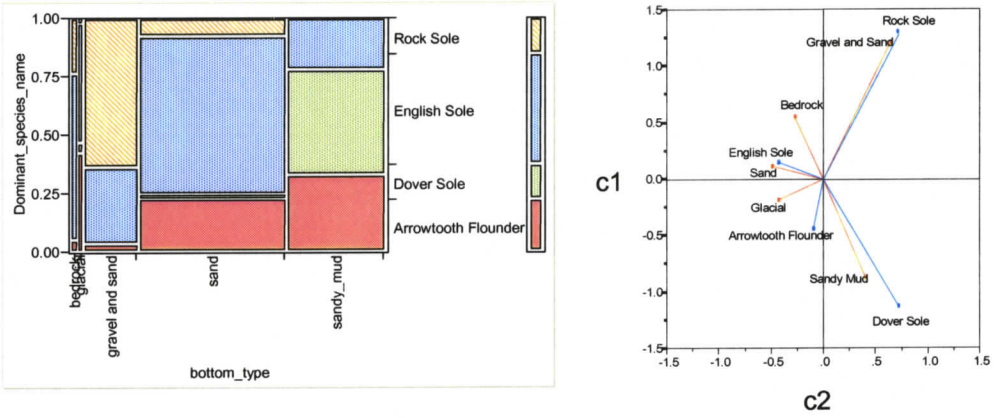

Figure 39 - Correspondence Analysis of dominant species by bottom type– 1996

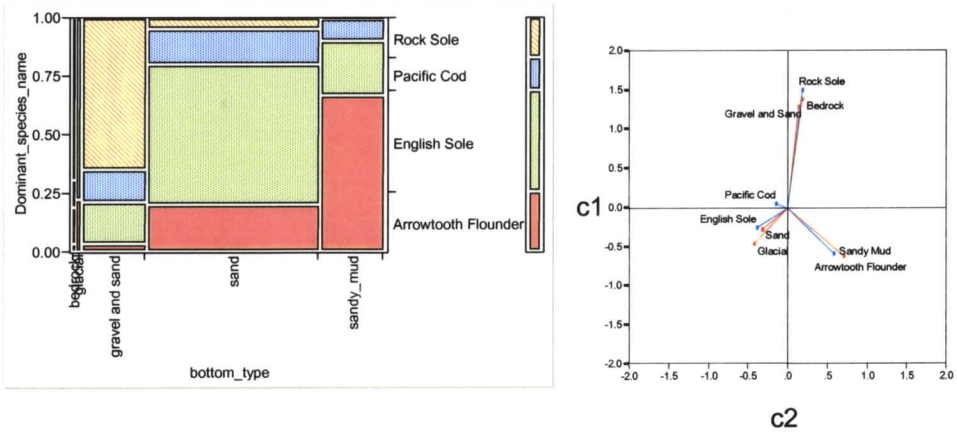

Figure 40 - Correspondence Analysis of dominant species by bottom type– 1997

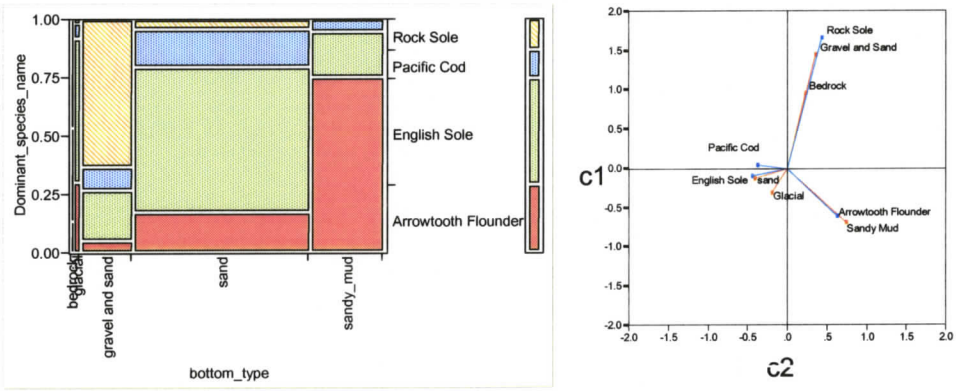

Figure 41 - Correspondence Analysis of dominant species by bottom type – 1998

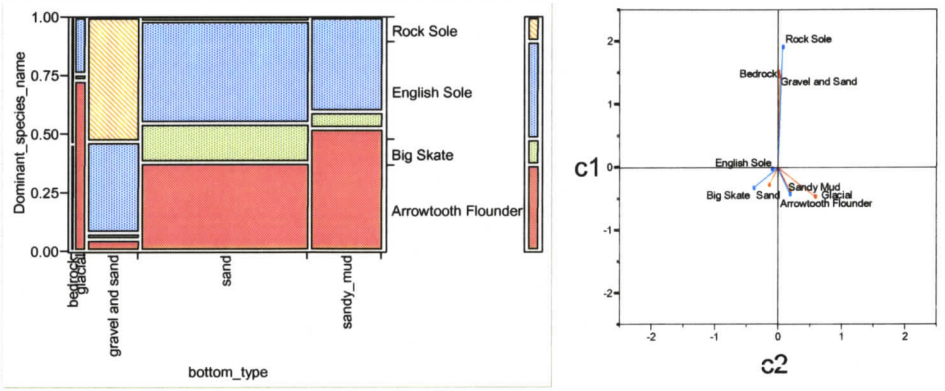

Figure 42 - Correspondence Analysis of dominant species by bottom type – 1999

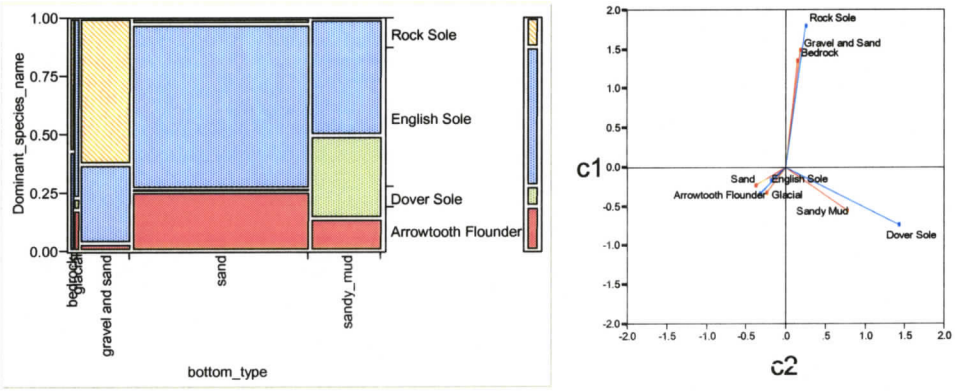

Figure 43 - Correspondence Analysis of dominant species by bottom type – 2000

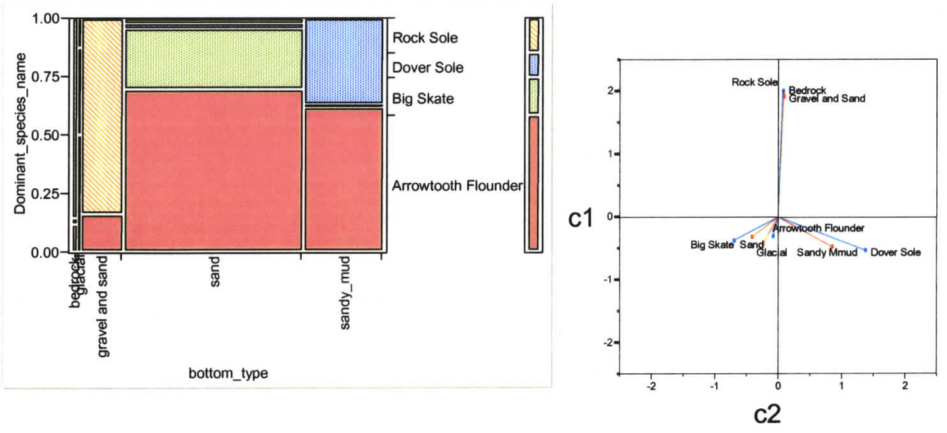

Figure 44 - Correspondence Analysis of dominant species by bottom type – 2001

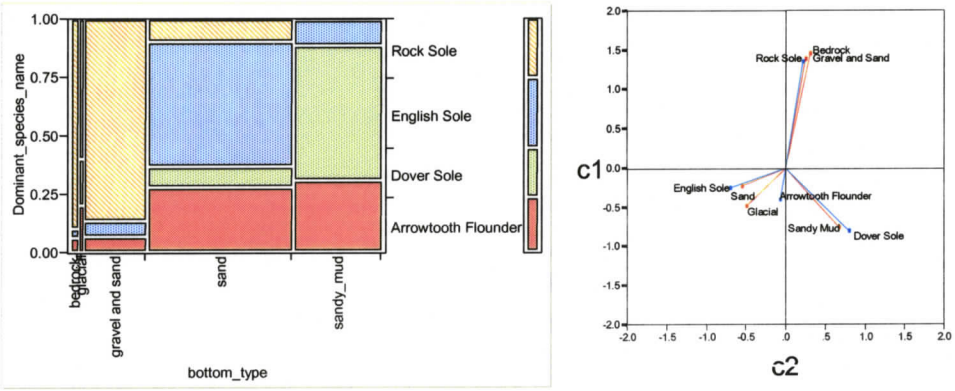

Figure 45 - Correspondence Analysis of dominant species by bottom type – 2002

Visually, the correspondence plots show a closer relationship between bottom type and species cluster than the individual species' correspondence analysis. Also, comparison of Table 3 and Table 4 shows that the cumulative inertia (CUM) is greater in all years for the clustered data meaning the two dimensional plots for the clustered data are a better representation than the single-species plots. Because of this, a more in-depth examination of the correspondence analysis is shown for the clustered data.

Axis inertias and contributions are given in Table 4. The axis singular values are output from a SVD of the frequency table. The singular values and therefore the inertias for the first two axes account for most of the variability in the data. For 1999, the primary axis (Axis 1) accounts for 92.8% of the variability, a departure from the 65%-70% for the other years. This can be seen on the plot for 1999, Figure 42; all the points appear closer to the c1 axis (vertical) and do not spread as much across the c2 axis. The singular values for the third axis are very low, as they should be, and therefore their overall contribution to the correspondence is low.

	1996	1997	1998	1999	2000	2001	2002
Axis 1 Singular Value	.71356	.70759	.68936	.68031	.70702	.84414	.81001
Axis 1 Inertia	.50917	.50069	.47522	.46283	.49988	.71257	.65612
Axis 1 (%)	**67.3**	**74.8**	**65.6**	**92.8**	**68.5**	**71.6**	**69.1**
Axis 2 Singular Value	.49277	.40660	.49720	.18556	.47820	.53020	.54115
Axis 2 Inertia	.24282	.16533	.24721	.03443	.22868	.28111	.29284
Axis 2 (%)	**32.1**	**24.7**	**34.1**	**6.9**	**31.4**	**28.2**	**30.8**
CUM (%)	**99.4**	**99.5**	**99.7**	**99.7**	**99.9**	**99.8**	**99.9**
Axis 3 Singular Value	.06770	.06008	.04744	.04172	.03008	.04550	.02666
Axis 3 Inertia	.00458	.00361	.00225	.00174	.00090	.00207	.00071
Axis 3 (%)	**0.6**	**0.5**	**0.3**	**0.3**	**0.1**	**0.2**	**0.1**

Table 4 – Correspondence Analysis Singular values, Inertias, and percent contribution to the axis by the point for the first three axes, clustered catch data.

For 1996, the first and second axes are determined almost exclusively by the by the bottom types 'Gravel and Sand' (51.0%, 30.4%) and 'Sandy Mud' (46.7%, 21.5%) and the species clusters 'Dover Sole' (37.8%, 32.3%) and 'Rock Sole' (51.4%, 32.5%) (Table 5, CTR(%)). These high contributions to the first and second primary axes and an examination of the biplot (Figure 39) show that the fish in the 'Dover Sole' cluster prefer 'Sandy Mud' as habitat and the fish in the 'Rock Sole' cluster prefer 'Gravel and Sand' as habitat. The bottom type 'Glacial' is not well represented in the figure because the quality in the biplot (QLT) value of 577 is much lower than the others. Since this value is the sum of the correlations to the first two axes (COR for the first two axes) only, the correlation to the third axis is 1000 – 577 = 423. Since the correlation is cosine squared, the closer the angle from an axis to the point is to 0°, the more that point correlates to that axis up to a maximum of 1000.

The 1997 results are similar in the first primary axis, with 'Gravel and Sand' (70.0%) and 'Sandy Mud' (16.1%) being the predominant bottom types contributing. 'Rock Sole' (76.5%) and 'Arrowtooth Flounder' (17.8%) contribute the most of the species clusters to the first primary axis (Table 6, CTR(%)). This year differs from 1996 in the second axis, where the highest-contributing bottom type is 'Sandy Mud' and the highest-contributing species cluster is 'Arrowtooth Flounder'. Again, the 'Glacial' bottom type is not represented well in two dimensions, with a quality value of 657. The contribution (CTR%) to the third axis by the 'Glacial' point is 94.2%, meaning this point contributes most to the third axis. Also, 'Pacific Cod' is not well represented with a quality value of 519, and an angle of 46.1°. Note that 'Bedrock' has a similar correlation to the primary axis as both 'Rock Sole' and 'Gravel and Sand'. This can be seen in both Figure 40 and the first COR column in Table 6. The figure shows their spatial correspondence, and the COR column shows that the three points have similar correlations (978, 988, and 984). This means that both bottom types correspond

to 'Rock Sole' but bedrock only contributes 4.7% to the inertia of the first primary axis (CTR%), making 'Gravel and Sand' the predominant type.

The 1998 analysis is similar to 1997, with 'Gravel and Sand' (72.8%) and 'Sandy Mud' (23.4%) being explained best in the first primary axis. 'Rock Sole' and 'Arrowtooth Flounder' again contribute most to the first primary axis for the species clusters. 'Gravel and Sand' and 'Rock Sole' share the closest correlation to the primary axis as seen in the first COR field of Table 7. Examination of the second axis fields in the table shows that 'Sand' and 'English Sole' are most correlated along the secondary axis with values of 903 and 950 respectively. 'Pacific Cod' also appears correlated to them with a value of 891 but further examination of the third axis shows that it has an angle of 71.6°, far from the angles of the other two (90° and 86.1°), meaning that 'Pacific Cod' is also correlated to the third axis whereas the other two species clusters are not.

The 1999 analysis is somewhat different than the others with most of the variability (92.8%) being explained by the first primary axis (Table 4), 'Gravel and Sand' and 'Rock Sole' are the main contributors to the first primary axis, with 'Sand', 'Sandy Mud', and 'Arrowtooth Flounder' having similar correlations to all axes and similar contributions of inertia for all axes. 'English Sole' has the lowest quality representation in the two-dimensional display with a QLT value of 799 (Table 8). Most of this value (764) is correlation with the secondary axis, but with a contribution of 9.4% to the secondary axis, it has negligible correspondence with other points. This can also be seen in Figure 42 as the point is very close to the origin, indicating low correspondence in the first two axes. Also, the third axis angle (63.3°) is closer to 45° making the correspondence to the third axis relatively high.

The year 2000 shows a similar pattern with 'Gravel and Sand' and 'Rock Sole' having the greatest contribution to the first axis. 'Sandy Mud' and 'Dover Sole' do

not contribute as much to the first axis as in other years, but they contribute much more to the second axis' inertia (62.7% and 80.7%). 'Glacial' is again correlated to the third axis, and cannot be said to correspond to any species cluster.

Table 10 shows the 2001 analysis. This year again shows a high correspondence between 'Gravel and Sand' and 'Rock Sole', but 'Bedrock' is contributing 12.2% of the inertia to the first axis. The plot (Figure 44) shows that those three points do line up and along with the numerical output we can say that the 'Rock Sole' cluster corresponds to both 'Gravel and Sand' and 'Bedrock' for this year. Bedrock's CTR value however is very low; at 12% it does not contribute much to the inertia. The mosaic plot for this year (Figure 44) also shows that 'Bedrock' catch consisted mainly of fish in the 'Rock Sole' cluster.

The 2002 output is very similar to 2001 output. 'Gravel and Sand' corresponds to 'Rock Sole' and 'Sandy Mud' corresponds to 'Dover Sole'. As in 2001, the 'Bedrock' point has an inflated contribution to the first axis' inertia. The mosaic plot (Figure 45) is very similar to the 2001 mosaic, with the 'Bedrock catch consisting mainly of fish in the 'Rock Sole' cluster.

			Axis c1					Axis c2					Axis c3				
Variable	**QLT**	**mass**	**c1**	**CTR**	**CTR(%)**	**COR**	***α***	**c2**	**CTR**	**CTR(%)**	**COR**	***β***	**c3**	**CTR**	**CTR(%)**	**COR**	***ζ***
Bedrock	832	0.0168	0.551	5	1.0	667	35.2	-0.274	1	0.5	165	66.0	-0.2761	1	28.0	168	65.8
Glacial	577	0.0208	-0.185	1	0.1	93	72.3	-0.422	4	1.5	484	45.9	0.3949	3	70.8	423	49.4
Gravel and sand	1000	0.1776	1.209	260	51.0	779	28.1	0.645	74	30.4	221	61.9	0.012	0	0.6	0	89.5
Sand	1000	0.4700	0.114	6	1.2	52	76.9	-0.488	112	46.1	948	13.2	-0.0075	0	0.6	0	89.1
Sandy mud	1000	0.3149	-0.869	238	46.7	820	25.1	0.407	52	21.5	180	64.9	-0.0069	0	0.3	0	89.6
Arrowtooth Flounder	937	0.2270	-0.442	44	8.7	899	18.5	-0.09	2	0.8	38	78.8	0.117	3	67.8	63	75.5
Dover Sole	998	0.1506	-1.131	193	37.8	709	32.7	0.722	78	32.3	289	57.5	-0.0671	1	14.8	2	87.1
English Sole	992	0.4691	0.15	11	2.1	111	70.5	-0.422	84	34.5	881	20.2	-0.0402	1	16.6	8	84.9
Rock Sole	1000	0.1533	1.307	262	51.4	769	28.8	0.717	79	32.5	231	61.2	0.0157	0	0.8	0	89.4

Table 5 – Correspondence analysis output and calculations for 1996 clustered species. See Figure 39 for the plot.

			Axis c1					Axis c2					Axis c3				
Variable	**QLT**	**mass**	**c1**	**CTR**	**CTR(%)**	**COR**	***α***	**c2**	**CTR**	**CTR(%)**	**COR**	***β***	**c3**	**CTR**	**CTR(%)**	**COR**	***ζ***
Bedrock	998	0.0122	1.386	23	4.7	978	8.4	0.194	0	0.3	19	82.0	0.0686	0	1.6	2	87.2
Glacial	657	0.0169	-0.462	4	0.7	364	52.9	-0.415	3	1.8	294	57.2	0.4485	3	94.2	343	54.2
Gravel and sand	1000	0.2114	1.288	351	70.0	988	6.2	0.139	4	2.5	12	83.8	0.0004	0	0.0	0	90.0
Sand	999	0.5541	-0.276	42	8.4	442	48.3	-0.309	53	32.1	556	41.8	-0.016	0	3.9	1	87.8
Sandy mud	1000	0.2053	-0.626	80	16.1	434	48.8	0.715	105	63.4	566	41.2	0.0016	0	0.0	0	89.9
Arrowtooth Flounder	1000	0.2608	-0.585	89	17.8	494	45.4	0.592	91	55.3	506	44.6	0.01	0	0.7	0	89.3
English Sole	995	0.4327	-0.253	28	5.5	299	56.9	-0.386	65	39.1	697	33.4	0.0318	0	12.1	5	86.1
Pacific Cod	519	0.1356	0.048	0	0.1	49	77.2	-0.148	3	1.8	469	46.8	-0.1501	3	84.6	481	46.1
Rock Sole	1000	0.1709	1.497	383	76.5	984	7.3	0.192	6	3.8	16	82.7	0.0233	0	2.6	0	89.1

Table 6 – Correspondence analysis output and calculations for 1997 clustered species. See Figure 40 for the plot.

			Axis c1					Axis c2					Axis c3				
Variable	**QLT**	**mass**	**c1**	**CTR**	**CTR(%)**	**COR**	***α***	**c2**	**CTR**	**CTR(%)**	**COR**	***β***	**c3**	**CTR**	**CTR(%)**	**COR**	***ζ***
Bedrock	892	0.0063	0.951	6	1.2	841	23.5	0.232	0	0.1	50	77.0	0.3415	1	32.7	108	70.8
Glacial	711	0.0264	-0.312	3	0.5	515	44.1	-0.193	1	0.4	196	63.7	0.2336	1	64.0	289	57.5
Gravel and sand	1000	0.1654	1.446	346	72.8	941	14.1	0.362	22	8.8	59	75.9	-0.0078	0	0.4	0	89.7
Sand	999	0.5680	-0.133	10	2.1	97	71.9	-0.406	94	37.9	903	18.2	-0.0098	0	2.4	1	88.7
Sandy mud	1000	0.2338	-0.689	111	23.4	460	47.3	0.747	130	52.7	540	42.7	-0.0063	0	0.4	0	89.6
Arrowtooth Flounder	1000	0.2926	-0.611	109	23.0	480	46.1	0.635	118	47.8	520	43.9	0.0002	0	0.0	0	90.0
English Sole	995	0.4547	-0.096	4	0.9	46	77.7	-0.438	87	35.3	950	12.9	0.0302	0	18.4	5	86.1
Pacific Cod	900	0.1217	0.038	0	0.0	10	84.4	-0.366	16	6.6	891	19.3	-0.1225	2	81.2	100	71.6
Rock Sole	1000	0.1310	1.662	362	76.1	934	14.9	0.442	26	10.3	66	75.1	0.0088	0	0.5	0	89.7

Table 7 – Correspondence analysis output and calculations for 1998 clustered species. See Figure 41 for the plot.

			Axis c1					Axis c2					Axis c3				
Variable	**QLT**	**mass**	**c1**	**CTR**	**CTR(%)**	**COR**	***α***	**c2**	**CTR**	**CTR(%)**	**COR**	***β***	**c3**	**CTR**	**CTR(%)**	**COR**	***ζ***
Bedrock	992	0.0099	1.523	23	5.0	992	5.3	0.004	0	0.0	0	89.9	-0.1408	0	11.3	8	84.7
Glacial	967	0.0413	-0.47	9	2.0	377	52.1	0.588	14	41.4	590	39.8	0.1388	1	45.7	33	79.5
Gravel and sand	1000	0.1698	1.446	355	76.7	1000	1.2	0.029	0	0.4	0	88.9	0.0086	0	0.7	0	89.7
Sand	999	0.5436	-0.284	44	9.5	803	26.3	-0.14	11	31.0	195	63.8	0.0122	0	4.6	1	87.8
Sandy mud	984	0.2354	-0.369	32	6.9	762	29.2	0.199	9	27.2	222	61.9	-0.0527	1	37.6	16	82.8
Arrowtooth Flounder	998	0.3705	-0.419	65	14.1	822	25.0	0.194	14	40.6	177	65.2	0.0198	0	8.3	2	87.5
Big Skate	978	0.1128	-0.328	12	2.6	413	50.0	-0.384	17	48.3	565	41.3	0.0764	1	37.8	22	81.4
English Sole	799	0.4131	-0.019	0	0.0	35	79.2	-0.089	3	9.4	764	29.1	-0.0455	1	49.2	201	63.3
Rock Sole	1000	0.1037	1.929	386	83.4	998	2.4	0.076	1	1.8	2	87.7	0.0277	0	4.6	0	89.2

Table 8 – Correspondence analysis output and calculations for 1999 clustered species. See Figure 42 for the plot.

			Axis c1					Axis c2					Axis c3				
Variable	**QLT**	**mass**	**c1**	**CTR**	**CTR(%)**	**COR**	***α***	**c2**	**CTR**	**CTR(%)**	**COR**	***β***	**c3**	**CTR**	**CTR(%)**	**COR**	***ζ***
Bedrock	990	0.0130	1.357	24	4.8	978	8.5	0.151	0	0.1	12	83.7	0.1364	0	26.9	10	84.3
Glacial	821	0.0181	-0.321	2	0.4	528	43.4	-0.239	1	0.5	293	57.2	0.1868	1	70.2	179	65.0
Gravel and sand	1000	0.1672	1.493	373	74.6	985	7.2	0.187	6	2.6	15	82.9	-0.0092	0	1.6	0	89.6
Sand	1000	0.5706	-0.235	32	6.3	288	57.6	-0.37	78	34.1	712	32.4	-0.0055	0	1.9	0	89.3
Sandy mud	1000	0.2311	-0.551	70	14.0	328	55.0	0.788	143	62.7	672	35.0	-0.0019	0	0.1	0	89.9
Arrowtooth Flounder	984	0.1890	-0.334	21	4.2	529	43.3	-0.31	18	7.9	456	47.5	-0.0574	1	69.2	16	82.8
Dover Sole	1000	0.0894	-0.741	49	9.8	210	62.7	1.437	185	80.7	790	27.3	-0.0072	0	0.5	0	89.7
English Sole	992	0.5947	-0.167	17	3.3	481	46.1	-0.172	18	7.7	511	44.4	0.0212	0	29.7	8	84.9
Rock Sole	1000	0.1269	1.804	413	82.6	980	8.1	0.256	8	3.6	20	81.9	-0.0087	0	1.1	0	89.7

Table 9 – Correspondence analysis output and calculations for 2000 clustered species. See Figure 43 for the plot.

			Axis c1					Axis c2					Axis c3				
Variable	**QLT**	**mass**	**c1**	**CTR**	**CTR(%)**	**COR**	***α***	**c2**	**CTR**	**CTR(%)**	**COR**	***β***	**c3**	**CTR**	**CTR(%)**	**COR**	***ζ***
Bedrock	999	0.0222	1.976	87	12.2	997	2.9	0.073	0	0.0	1	87.9	0.0693	0	5.2	1	88.0
Glacial	524	0.0086	-0.432	2	0.2	399	50.9	-0.243	1	0.2	126	69.2	0.4719	2	92.5	476	46.4
Gravel and sand	1000	0.1372	1.926	509	71.4	998	2.9	0.096	1	0.4	2	87.2	-0.0102	0	0.7	0	89.7
Sand	1000	0.5759	-0.319	59	8.2	387	51.5	-0.401	93	33.0	613	38.5	-0.0064	0	1.1	0	89.3
Sandy mud	1000	0.2561	-0.471	57	8.0	234	61.1	0.853	186	66.3	766	28.9	-0.0021	0	0.1	0	89.9
Arrowtooth Flounder	989	0.5889	-0.312	57	8.0	931	15.2	-0.078	4	1.3	58	76.0	-0.0334	1	31.7	11	84.1
Big Skate	989	0.1572	-0.375	22	3.1	221	62.0	-0.699	77	27.3	768	28.8	0.0842	1	53.8	11	83.9
Dover Sole	999	0.1048	-0.542	31	4.3	133	68.6	1.381	200	71.1	865	21.5	0.0528	0	14.1	1	88.0
Rock Sole	1000	0.1491	2.01	602	84.5	999	2.2	0.076	1	0.3	1	87.8	0.0061	0	0.3	0	89.8

Table 10 – Correspondence analysis output and calculations for 2001 clustered species. See Figure 44 for the plot.

			Axis c1					Axis c2					Axis c3				
Variable	**QLT**	**mass**	**c1**	**CTR**	**CTR(%)**	**COR**	***α***	**c2**	**CTR**	**CTR(%)**	**COR**	***β***	**c3**	**CTR**	**CTR(%)**	**COR**	***ζ***
Bedrock	1000	0.0319	1.459	68	10.3	956	12.1	0.313	3	1.1	44	77.9	-0.0071	0	0.2	0	89.7
Glacial	871	0.0098	-0.488	2	0.4	434	48.8	-0.49	2	0.8	437	48.6	0.2662	1	97.8	129	68.9
Gravel and sand	1000	0.2046	1.389	395	60.2	966	10.6	0.261	14	4.7	34	79.4	0.0027	0	0.2	0	89.9
Sand	1000	0.4664	-0.23	25	3.8	151	67.2	-0.546	139	47.5	849	22.8	-0.0056	0	2.1	0	89.5
Sandy mud	1000	0.2873	-0.761	166	25.4	553	41.9	0.684	134	45.8	447	48.1	-0.0011	0	0.0	0	89.9
Arrowtooth Flounder	987	0.2384	-0.4	38	5.8	966	10.6	-0.06	1	0.3	21	81.6	-0.0457	0	70.1	13	83.6
Dover Sole	1000	0.2072	-0.811	136	20.8	502	44.9	0.807	135	46.0	497	45.2	0.0207	0	12.5	0	89.0
English Sole	999	0.3029	-0.256	20	3.0	121	69.7	-0.691	145	49.4	879	20.4	0.0202	0	17.4	1	88.4
Rock Sole	1000	0.2515	1.355	462	70.4	973	9.4	0.224	13	4.3	27	80.6	0.002	0	0.1	0	89.9

Table 11 – Correspondence analysis output and calculations for 2002 clustered species. See Figure 45 for the plot.

Chapter 6
Conclusions

Six fishing areas in Hecate Strait were surveyed in June 2002 with single beam echosounder, sidescan, and Huntec acoustic systems to obtain bottom information and classification. The sidescan and Huntec data from the survey were integrated into surficial sediment maps of the fishing areas (Conway *et al.*, 2004). The raw data for the single beam classifications were digitized by Quester Tangent's QTC VIEW 5 single beam capture software and processed with IMPACT acoustic classification software. The steps for single beam classification including raw echo cleaning, bottom picking, feature extraction, database organization of the analysis and Quester Tangent's data storage methods and algorithms were discussed. The single beam data collected on the survey were quite sparse for some areas. Track crossings are required to ensure quality of data; two returns over the same area reinforce the classification over that area, and makes interpolation between ship tracks more plausible. This survey covered six areas, among these the Dundas area had the worst coverage (Figure 15) and Shell Ground had the best (Figure 17). Dundas had no crossings of tracklines, and Shell Ground had 17 crossings. An effort to interpolate the Shell Ground data to obtain a two-dimensional classification map was made but this is not included here because the sparseness of data lead to many algorithmic artifacts.

If good tracklines are maintained, single beam remote sensing appears to be a promising alternative to more expensive methods such as submersible or ROV work. There are several processing shortcomings. One of these is that the Quester Tangent k-means partitioning method uses a Mahalanobis distance measure which requires the computation of covariance matrices. This computation is an $O(n^2)$ operation which means that large datasets will cause the Random Access Memory (RAM) to fill up on many systems and the swap file will

be used excessively. Also, post-classification analysis is required to alleviate the misclassifications caused by mobile sand dunes (Figure 25) and other similar bottom features. The distribution of bottom types in the classification also appears to be an issue, if one type dominates over others the QTC algorithm will tend to assimilate them into the main type. The surficial sediment 'Sand' type is particularly interesting because it showed no correspondence with any single beam class in this analysis. This is probably due to the small amount of sand in the fishing areas; too few single beam echoes were returned from the bottom for the sand class to allow it to remain a distinctive class during the K-means clustering routine.

The surficial sediment maps of the same five fishing areas used in the single beam analysis were also used in the correspondence analyses for the seven years 1996-2002. Catch data, from fisheries-observed trips were used in a two-way frequency table with the surficial sediment bottom type. A full correspondence analysis required a plot for visualization, along with numerical output to determine correlations to axes and contributions to inertia of the axes. These things taken into consideration show that, over all the years the 'Gravel and Sand' bottom type is preferred by Rock Sole. In both the clustered correspondence and the single species correspondence the two are correlated. Another strong correspondence occurs between 'Sandy Mud' and both 'Arrowtooth Flounder' and 'Dover Sole'. These species clusters as well as the single species themselves have high correlation with that bottom type.

Some bottom types and/or species clusters may not be represented well in a two-dimensional correspondence plot. The only way to ascertain this is to examine the numerical output of the correspondence analysis, adding the two axis correlations of each point to determine if they make up most of the overall correlation of the point (QLT field in numerical output tables). For the bottom type 'Glacial', every

year was similar; the quality of the point in a 2-dimensional plot was shown to be poor. 'Glacial' appears correlated to the third axis, which is a very low contributor to overall inertia. This implies very little correspondence between the 'Glacial' bottom type and any species or cluster of similar species. Correspondence analysis of remotely sensed, acoustic data against species caught has exposed correspondences between several species and bottom types, and eliminated 'Glacial' as an accurate predictor of species.

Incorporation of habitat into stock assessment for fisheries is a burgeoning idea and the ultimate goal of this investigation. This work has shown that first steps in incorporating habitat into assessments could include acoustic seabed classification, visualization of seafloor maps and abundance and biomass estimates based on area of habitat instead of total area of fishing grounds.

Literature Cited

Anderson, J.T., *Classification of Marine Habitats Using Submersible and Acoustic Seabed Techniques*, Spatial Processes and Management of Marine Populations, AK-SG-01-02, pp 377-393, 2001.

Anderson, J.T., Gregory, R.S., and Collins, W.T., *Acoustic classification of marine habitats in coastal Newfoundland*, ICES Journal of Marine Science, 59, pp 156-157, 2002.

Barrie, J.V. and Bornhold, B.D., *Surficial Geology of Hecate Strait, British Columbia Continental Shelf*, Can. J. Fish. Aquat. Sci., 26, 1241-1254, 1989.

Biffard, B., Bloomer, S., Chapman, R., and Preston, J., *Single Beam Seabed Classification and the Problem of Slope,* Boundary Influences in High Frequency, Shallow Water Acoustics, pp 227-232, University of Bath, UK, September 5-9, 2005.

Bloomer, S.F., Mosher, D.C., Collins, W.T., Preston, J.M., and Rosenberger, A., *Subsurface Classification of High-resolution Seismic Data with Multivariate Statistical Techniques: A Case Study from the Strait of Georgia,*, Proceedings of MTS/IEEE Oceans '00, 3, pp 1677-1684, 2000.

Collins, W.T. and Rhynas, K.P, *Acoustic Seabed Classification using Echo Sounders: Operational Considerations and Strategies*, Proceedings of Canadian Hydrographic Conf., Victoria Canada, 1998.

Collins, W.T. and LaCroix, P., *Operational Philosophy of Acoustic Waveform Data Processing for Seabed Classification*, Proceedings of COSU '97, 1997.

Conway, K.W., Barrie, J.V., Picard, K., Hill, P.R., Yamanaka, K.L., and Sinclair, A., *Habitat Mapping for Ocean Management on the Western Canadian Continental Shelf,* GeoHab 5th International Symposium, May 5-7, Galway Ireland, 2004.

Davidson, I., *Understanding k-Means Non-Hierarchical Clustering*, SUNY Albany Technical Report 02-2, 2002.

DeLong, A.K., and Collie, J.S., *Defining Fish Habitat: A Model-Based Approach*, Rhode Island Sea Grant, Narragansett, R.I., 2004.

Elkan, C., *Using the Triangle Inequality to Accelerate k-Means*, Proceedings of the 20th International Conference on Machine Learning (ICML-2003), 2003.

Fader, G.B.J., Pickrill, R.A., Todd, B.J., Courtney, R.C., and Parrott, D.R., *The Emerging Role of Marine Geology in Benthic Ecology*, Bedford Institute of Oceanography Science Review, 1999.

Galloway, J.L. and Collins, W.T., *Dual Frequency Acoustic Classification of Seafloor Habitat using the QTC VIEW*, Proceedings of IEEE Oceans '98, Nice, France, 1998.

Greene, H.G., Barrie, V., Lopez, H.L., Palsson, W., Tilden, J., and Endris, C., *Potential Marine Benthic Habitats of an Inland Sea, San Juan Islands and the Transboundary Region of Canada and the US: A Successful International Cooperative Program*, Poster, GeoHab 2004, Galway, Ireland, 2004.

Greenacre, M.J., *Theory and Applications of Correspondence Analysis*, Academic Press (Harcourt Brace Jovanovich, Publishers), 1984.

JMP 4.04 Statistical Software, Copyright 2001 SAS Institute

Kenny, A.J., Todd, B.J., and Cooke, R.C., Procedural Guideline No. 1-4 *The Application of Sidescan Sonar for Seabed Habitat Mapping*, JNCC Marine Monitoring Handbook, pp 199-210, 2001.

Kostylev, V.E., Courtney, R.C., Robert, G., Todd, B.J., *Stock Evaluation of the Giant Scallop (Placopecten Magellanicus) Using High-resolution acoustics for seabed mapping*, Science Direct Fisheries Research, 60, pp 479-492, 2003.

Kostylev, V.E., Todd, B.J., Fader, G. B. J., Courtney, R.C., Cameron, G.D.M., and Pickrill, R.A., *Benthic habitat mapping on the Scotian Shelf based on multibeam bathymetry, surficial geology and sea floor photographs*, Mar. Ecol. Prog. Ser., 218, pp 121-137, 2001.

Legendre, P, Ellingsen, K.E., Bjornbom, E., and Casgrain, P., *Acoustic Seabed Classification: Improved statistical method*, Can. J. Fish. Aquat. Sci., 59, pp 1085-1089, 2002.

Legendre, P., *Reply to the comment by Preston and Kirlin on "Acoustic Seabed Classification: Improved statistical method"*, Can. J. Fish. Aquat. Sci., 60, pp 1301-1305, 2003.

Li, D., Tang, D., and Frisk, G., *Evaluation of Sound Propagation Models Used in Bottom Volume Scattering Studies*, J. Acoust. Soc. Am., 108(5), pp 2039-2052.

McConnaughey, R.A. and Smith, K.R., *Associations between flatfish abundance and surficial sediments in the eastern Bering Sea*, Can. J. Fish. Aquat. Sci., 57, pp 2410-2419, 2000.

Maushake, C. and Collins, W. T., *Acoustic Classification and Water Injection Dredging*, Hydro International, 6(2), pp 7-9, 2002.

Perry, R.I., Stocker, M. and Fargo, J., *Environmental Effects on the Distributions of Groundfish in Hecate strait, British Columbia*, Can. J. Fish. Aquat. Sci., 51, pp 1401-1409, 1994.

Pickrill, R.A. and Todd, B.J., *The Multiple Roles of Acoustic Mapping in Intergated Management, Scotion Shelf, Canada*, Proceedings of International conference on coastal and ocean space utilization, 1, pp 109-113, 2000

Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T., *Numerical Recipes in C The Art of Scientific Computing*, Cambridge University Press, 1988.

Preston, J.M. and Kirlin, R.L., *Comment on "Acoustic Seabed Classification: Improved statistical method"*, Can. J. Fish. Aquat. Sci., 60, pp 1299-1300, 2003.

Preston, J.M. and Collins, W.T., *Bottom Classification in Very Shallow Water by High-speed Data Acquisition*, Proceedings of MTS/IEEE Oceans '00, 2, pp1277-1282, 2000.

Preston, J.M., Christney, A.C., Bloomer, S.F., and Beaudet, I.L., *Seabed Classification of Multibeam Sonar Images*, Proceedings of MTS/IEEE Oceans '01, pp 2616-2623, 2001.

QTC IMPACT User Manual, 2003, DMN-MIMP-0001-R03C, Quester Tangent Corporation, 2003.

Romsos, C.G., *Mapping Surficial Geology Habitats of the Oregon Continental Margin Using Integrated Interpretive GIS Techniques*, Oregon State University Master's Thesis, 2004.

Sinclair, A., Conway, K.W., and Crawford, W., *Associations Between Bathymetric, Geologic, and Oceanographic Features and the Distribution of the British Columbia Groundfish Trawl Fishery*, ICES CM 2005/L:25, 2005.

Siwabessy, P.J.W., Penrose, J.D., Fox, D.R., and Kloser, R.J., *Bottom Classification in the Continental Shelf: A case study for the North-west and South-east Shelf of Australia*, Aus. Acoust. Soc. Conf., 2000.

Siwabessy, J., Penrose, J., Kloser, R., and Fox, D., *Seabed Habitat Classification*, Proceedings of Shallow Survey '99 – International Conference on High Resolution Surveys in Shallow water, 1999.

Todd, B.J., Kostylev, V.E., Fader, G.B.J., Courtney, R.C., and Pickrill, R.A., *New Approaches to Benthic Habitat Mapping Integrating Multibeam Bathymetry and Backscatter, Surficial Geology and Sea Floor Photographs: A Case Study from the Scotian Shelf, Atlantic Canada*, Proceedings of ICES Science Conference, 2000.

Tsemahman, A.S., Collins, W.T., and Prager, B.T., *Acoustic Seabed Classification and Correlation Analysis of Sediment Properties by QTC View*, Proceedings of Oceans '97, Halifax, Canada, 1997.

Watt, J.V., Seabed Classification – A New Layer for the Marine GIS, Quester Tangent Corporation, 1997.

Whitmire, C.E., Integration of High-Resolution Multibeam Sonar Imagery with Observational Data from Submersibles to Classify and Map Benthic Habitats at Hecata Bank, Oregon, Oregon State University Master's Thesis, 2003.