

# APPLYING DISCRIMINATE ANALYSIS TO CHARACTERISE SHALLOW ROCKY REEF HABITAT

V. L. Lucieer

Marine Research Laboratories, University of Tasmania, Nubeena Crescent, Taroona Tasmania, Australia (v\_halley@utas.edu.au)

**KEY WORDS:** GIS, marine, AGDS, sidescan sonar, discriminate analysis, seabed habitat mapping

## ABSTRACT

A major goal of seafloor habitat mapping is to provide techniques that predict the distribution of species from physical and biotic parameters that define where species live. An area of seafloor comprising of rocky reef and sand habitats was surveyed with both a single beam acoustic ground discrimination system (AGDS) and sidescan sonar. Comparisons are made between predicted classes assessed from AGDS measurements and the digitised layer of the sidescan mosaic. The application of discriminate analysis to create a classification rule for the rocky reef areas revealed that 83% of the classification could be predicted by depth and hardness. These results can be used to improve classification of single beam acoustic data in the near-shore depth zone in the range 15-25 meters.

## 1. INTRODUCTION

One of the key endeavours of seafloor habitat mapping is to develop techniques, based on remote sampling, that predict the distribution and abundance of species and resources from physical and biotic parameters that define where species live (Bax & Williams, 2001). Quantification of reef area and complexity of shallow rocky reef substratum provides important information for describing the patterns and distribution of available habitat for many commercially fished species, such as abalone (sp. *Haliotis rubra*). The structure and size of reefs and the spacing of patches of reef habitat are important in determining abundances of local fish populations and their rates of change.

One approach to understanding and predicting reef pattern is to relate its distribution to physical descriptors collected using acoustic technology. However, application of this approach is limited because the gradients are locally defined. It may be possible to improve seabed habitat maps with simpler, more general biophysical data. The locations of reef systems are determined by exposure, geology, erosion and deposition and other factors that are largely unknown in detail. They all interact with a complexity that is difficult to untangle without detailed spatial and temporal measurements. If we are able to understand what influences the distribution of reef we may be able to estimate values at unsampled locations and predict uncertainty in our estimates. The scales of variability and distribution patterns must first be understood before any models for predicting reef distribution can be generated.

For many years, GIS in the marine environment has been driven by remote sensing (both aerial and satellite) to assess coastal terrestrial habitats and to delineate shallow seabed habitat boundaries (Finkbeiner *et al.*, 2001; Mumby *et al.*, 1997; Malthus and Karpouzli, 2003). Habitat boundary information can only be determined from a limited depth range in airborne sensors to 8-10 m limiting the ability to map habitats outside the immediate coastal strip. Therefore, the applicability of optical remote sensing for seabed mapping is restricted, because of factors such as depth limitations, wind disturbance and sun glint. While aerial remote sensing may provide continuous coverage in shallow depths, the numbers of identifiable habitats are fewer than can be defined through acoustic surveys. Acoustic techniques (single beam acoustics and sidescan sonar) are attracting considerable attention for habitat mapping because

they offer the potential to provide comparatively rapid and relatively accurate assessment of seabed properties (e.g. Greenstreet *et al.*, 1997; Collins and McConnaughey, 1998; Foster-Smith *et al.*, 1999).

Normal incidence acoustic ground discrimination systems (AGDS) can be used to obtain a variety of information about the reflective characteristics of the seabed. Such systems send pulses of sound at a particular frequency (usually within 30 – 200kHz) that reflects from the seabed and a transducer records the echo (Siwabessy *et al.*, 2004). This produces a continuous string of points along a transect, each point containing georeferenced information on the depth and derived parameters, such as the hardness and roughness of the seabed. A fundamental limitation of single beam acoustic sounder based systems arises from the limited spatial coverage provided by the sounder beam. While along-track coverage is essentially continuous, between-track coverage is limited by beam geometry to be a small fraction of the water depths requiring interpolation between transects to produce continuous maps.

Sidescan sonar sonifies a swath of the seafloor and produce a high-resolution georegistered image of the seabed (Brown *et al.*, 2005). The image provides information on the sediment texture and the distribution of patterns on the substrata. The image, whilst giving excellent information of topographic features has little measurable data on sediment characteristics or depth. There is no simple correlation between the reflected acoustic energy (or backscatter) recorded by sidescan sonar systems and bottom type, making interpretation of sidescan sonar images mainly descriptive and qualitative.

Cost and logistics limit the quantity of sidescan data that can be collected for shallow water habitats. AGDS would become a more valuable tool for habitat mapping if it were able to be used to accurately interpolate the distribution of reef pattern or conversely small sample areas of AGDS could be taken with sidescan to provide a source of information for pattern validation. Although sidescan methods are not theoretically limited to a given depth range several practical considerations generally preclude boat operations in the very near shore (0-15 m). Wave height, submerged rocks, kelp canopy all present serious obstacles to data collection. Although airborne techniques can provide information on this near shore zone, the

technique lacks the ability to collect information of depth and substrate information. Laser and multispectral sensors can provide this information but the techniques and data are costly and difficult for small research centres to collect and process.

There are many factors affecting reef pattern and distribution, with the dominant influences due to wave energy, exposure and geology. Therefore, our predictions do not attempt to explain all spatial variability. In this study, we combined the general nature of applied multivariate statistics to test the ability of the physical parameters available to us to map the spatial pattern of reef in one coastal area. The objectives of this study were to (1) understand the relationships between hardness, roughness and depth with reef distribution, (2) identify the physical parameters that define the reef boundary using applied multivariate statistics and (3) estimate the effectiveness of the prediction by comparing the accuracy of the map to a digitized layer from the sidescan imagery collected within the study area.

## 2. METHODS

### 2.1 Study area

The Tasmanian commercial fishery for blacklip abalone (*Haliotis rubra*) and greenlip abalone (*H. laevigata*) contributes a significant component of the total Australian abalone catch, with annual landings of around 2590 tonnes in 2003 (Tarbath *et al.* 2003). The catch consists primarily of blacklip abalone (approximately 95%), which is fished throughout the State. One study site was selected for this research on the east coast of Tasmania at Friendly Beaches, which is typical blacklip abalone habitat (Figure 1). The area is characterised by nearshore patchy reef distribution with large offshore reef system. A feature of the reef morphology within the study area is the presence of a broad area of sand between the intertidal zone and the inner edge of the reef. This reflects the geomorphology of the region, which is dominated by granite headlands (Sharples, 2000) mixed with long sandy beaches.

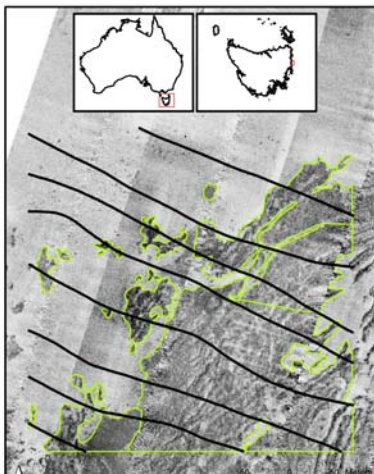


Figure 1. Location of the study site in Tasmania Australia, showing acoustic transects and interpreted polygon on sidescan sonar image

### 2.2 Sidescan Survey

A GeoAcoustics Dual Frequency sidescan sonar was used to collect information for this study from the *FRV Challenger*. Transects were run perpendicular to shore with 25% overlap between parallel beam footprints. Image scale is defined by the pixel dimensions (eg 1x1m) and is ultimately dictated by sounding density. Sounding density is related to the size of the beam footprint on the seafloor and the motion of the vessel (Bennell, 2005). As distance between the sonar transducer and the seafloor (range) increases, beam width also increases, resulting in increased footprint size and decreased horizontal resolution. The images were mosaiced using SonarWeb software. The sidescan mosaic was imported into ArcGIS 8.3 as a geotiff file. The data were interpreted by eye and two discrete habitat classes' reef and sand were digitised into shape files.

### 2.3 Single Beam Acoustic Survey

An ES60 echo-sounder with a 120 KHz single beam transducer was mounted on the side of a vessel underneath the GPS receiver in a survey seven days prior to the sidescan data collection. This equipment was connected to a laptop running ArcMap 6.0 GIS software and Simrad acoustic logging software. Geographic positioning data collected with a RACAL differential reference station and a Trimble Z Surveyor station onshore provided a correction message, with the estimated horizontal positional error of 3-5 metres. The corrected GPS message was combined using the time stamp from the acoustics data. Transects were performed across 15-25 m depth range in parallel (onshore-offshore) transects spaced at 50 m intervals resulting in 15 transects with an average profile length of 1000 m. Ten transects were sub-sampled for this study comprising of one thousand four hundred and eighty three points. Vessel speed was maintained at an average of 12km per hour (6 knots). Where possible transects were extended beyond the actual sampling area to avoid disturbances in the acoustic signal associated with boat manoeuvring within the survey area. ArcPad (ESRI) was employed in the field to display a transect spacing plan and using real-time GPS, ensured that the transect lines were being followed as accurately as possible. The single beam acoustic echogram was displayed in *Echoview* software (*SonarData Pty Ltd*). Data were imported into WEKA (Waikato Environment for Knowledge Analysis) for visualisation.

### 2.4 Acoustic variables

The acoustic system sends out a pulse of directed sound ensonifying a region of the seafloor that scatters sound back to the source to infer seabed depth, roughness and hardness (Kloser *et al.*, 2001). Depth is calculated by knowing the time of echo return with associate angle of arrival whilst hardness and roughness are inferred from the strength and variability in time and space of the backscattered signal as a function of angle (Basu and Saxena, 1999). The resolution of the sampling is related to the sounding range, frequency, reception and processing method. The sample interval was two pings per second although a GPS point was recorded only every second. Using ArcGIS 8.3 a spatial join was performed linking the acoustic points with the interpreted polygon and assigning to the acoustic point the class value of the polygon (1-reef, 2-sand). The acoustic data set was then subsetted into two parts, the training and test data sets. The model is created using the training data set and by applying the model to the test data set it can be used as an evaluation on how good the predictions are relative to the known values in the test dataset.

### 2.5 Discriminant analysis

Discriminant analysis is a method of predicting a classification based on known values of the variables. The technique is based on how close a set of measurement variables are to the multivariate means of the levels being predicted (Hastie et al, 2001). Based on the discriminant analysis of the training data set, the mahalanobis distance to each class cluster is computed. Based on this distance a probability can be calculated providing the likelihood that the sample is classified with a class label. The quality of model fit was assessed by comparing the accuracy of predicted classes with those interpreted from the sidescan image by a process of stepwise variable addition. The contribution of individual variables to the accuracy of prediction was assessed using the magnitude of the ratio of variances between consecutive stepwise additions to the model (F-ratio statistic).

### 2.6 Interpolation of predicted surface

To derive continuous hardness and depth surfaces at 2m cell resolution Ordinary kriging in the Geostatistical analyst of ArcGIS was applied. Spatial Analyst was used to apply each of the formula generated from the LDA to generate the probability surfaces. The raster calculator was used to generate a final predicted surface of rocky reef and sand.

## 3. RESULTS

The acoustic data were visually explored using WEKA (University of Waikato) software. It became clear which variables in the acoustic data showed a relationship with class distribution when coloured using the class label derived from the sidescan reference. Figure 2 shows a scatter plot of depth and hardness coloured according to the class label, showing that reef and sand form two clearly separable clusters.

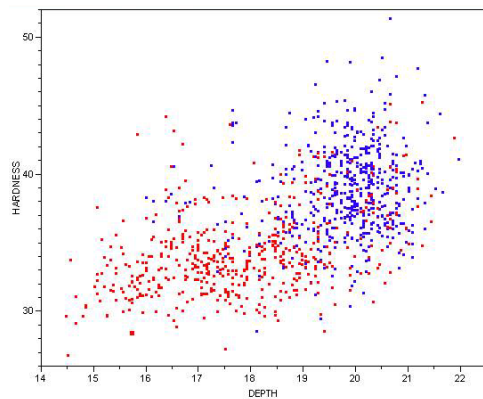


Figure 2. Scatter plot visually showing differences in two classes using hardness and depth (Class 1 [blue] rocky reef, Class 2 [red] sand).

Using ArcGIS the acoustic samples were divided into two data sets using the create subset command; the training data set contained 60% of all samples to build a predictive model and the test data set contained 40% of all samples to validate the model. Linear discriminant analysis (LDA) was applied on the training data set using depth, roughness and hardness as continuous variables to describe the nominal class variable. A stepwise variable selection was conducted to review the contribution of each variable to the class distribution. The results of the discriminant analysis show the inclusion of the variables hardness and depth resulted in an accuracy prediction

of 83.4% (F-ratios of 168.220 and 209.159 respectively) of the class values for reef (class 1) and sand (class 2). The LDA model was applied to the test data set, calculating class labels based on depth and hardness. This validation process provided an overall accuracy of 83.6%. Stepwise addition of the variable roughness (F-ratio 19.604) improved the accuracy of fit by 0.2% but decreased the overall prediction accuracy to 80.6% and was therefore excluded from the model.

Based on the LDA model probability surfaces were created. From these probability surfaces a class map was generated. The predicted class map can be seen in figure 3. The figure shows that general agreement between predicted reef and sidescan interpretation is adequate.

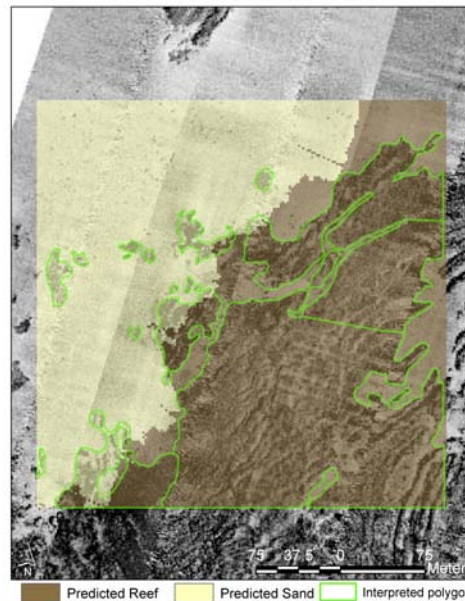


Figure 3. Predicted class map result using the model to generate classes 1 (rocky reef) and 2 (sand) overlaid on sidescan image.

## 4. DISCUSSION

Linear discriminate analysis seeks to find a way to predict a classification (X; class) variable based on known responses (Y; depth, roughness and hardness) (Hastie et al, 2001). The technique goes further to show how close a set of measurements variables are to the multivariate means for the levels being predicted. Decision theory for classification requires class posteriors for optimal classification; here we take the class values from the digitised layer of the sidescan image and compare these to the modelled classes. LDA relies on two main assumptions; that the data has a normal distribution and that the variables have the same variance and covariance. When plotted, both variables of depth and hardness displayed close to a normal distribution with a skewness of -0.608 for depth and -0.32 for hardness, therefore indicating that the data are appropriate for this type of analysis.

In some cases, the physical features of habitat are sufficient to explain their distribution and pattern. This model shows that classes can be described by hardness and depth with an accuracy of 83% when compared to the classified sidescan layer. The predicted map of reef and sand distributions shows that the model is able to determine the inner boundary of the reef edge to within an average of 30 meters. Given the density of data and the distance between transects (50 meters) this prediction

appears to work quite successfully. The small reef areas to the east of the image were not predicted but this may occur due to low data density. Within the depth range of 25-30 m this model can aid in explaining the distribution of reef.

## 5. CONCLUSION

The strength of combining the two techniques relies on using the AGDS to provide a more accurate classification of the sediment type and the sidescan to provide information of the patterns of seabed distribution. It is important to be able to describe quantitatively how rocky reef areas vary spatially and to gather knowledge about the uncertainty in interpolation from acoustic ground discrimination (AGDS) point data. From this study area we have been able to clearly determine which parameters in the acoustic signal have the greatest effect on predicting reef presence.

Categorical map analysis involves mapping the system property of interest by identifying patches that are relatively homogenous with respect to that property at a particular scale and that exhibit a relatively abrupt transition (boundary) to adjacent areas (patches) that have a different intensity (or quality). By identifying the mahalanobis distance of each point from the multivariate mean (centroid) we can take into account the correlation structure of the data (variance and covariance) and be able to estimate the probability that data belong to one of two habitat classes. The multivariate distance calculation is useful for spotting outliers in many dimensions. Even if the data are correlated, it is possible for a point to be unremarkable when seen along one or two axes but still be an outlier by violating the correlation.

If we can identify the differences in hardness and depth that identify a boundary between the two classes we can predict the amount of variability (patchiness) from the transect data and give an overall idea how heterogenous the seabed is in areas where the level of variability is smaller than the resolution of the sidescan imagery. This technique can be applied to other areas along the east coast where the geology and exposure are similar in spatial scale.

The results of this study greatly contribute to a larger research project that aims to quantify uncertainty in seabed mapping although further research should be conducted in other areas using differing data density and transect spacings to see if the results of predicting reef occurrence can be improved.

## REFERENCES:

- Basu, A. and Saxena, N.K., 1999. A Review of Shallow-Water Mapping Systems. *Marine Geodesy*, 22, pp. 249-257.
- Bax, N. J. and Williams, A., 2001. Seabed habitat on the south-eastern Australian continental shelf: context, vulnerability and monitoring. *Marine and Freshwater Research*, 52, pp. 491-512.
- Bennell, J. D., 2005. Procedural Guideline No. 1-5 Mosaicing of sidescan sonar images to map seabed features. *Marine Monitoring Handbook*, pp 405
- Brown, C.J., Mitchell, A., Limpenny, D.S., Robertson, M. R., Service, M and Golding, N., 2005. Mapping seabed habitats in the Firth of Lorn off the west coast of Scotland; evaluation and comparison of habitat maps produced using the acoustic ground discrimination system, *RoxAnn*, and sidescan sonar. *ICES Journal of Marine Science*, 62, pp. 790-802.

Collins, W.T. and McConnaughey, R.A. 1998. Acoustic classification of the seafloor to address essential fish habitat and marine protected area requirements, Canadian Hydrographic Conference 1998, CHS, Victoria, British Columbia, Canada.

Finkbeiner, M., B. Stevenson, A., (2001). Guidance for benthic habitat mapping, NOAA Coastal Services Center, 53.

Foster-Smith, R.L., Davies. J. and Sotheran. I., 1999. Broad scale remote survey and mapping of sublittoral habitats and biota, Final Technical Report of SeaMap Research Group, University of Newcastle-upon-Tyne, UK.

Greenstreet, S.P.R., Tuck, I.D., Grewar, G.N., Armstrong, E., Reid, D.G. and Wright, P.J., 1997. An assessment of the acoustic survey technique, *RoxAnn*, as a means of mapping seabed habitat *ICES Journal of Marine Science*, 545, pp. 939-959.

Hastie, T., Tibshirani, R., Friedman, J., 2001. *The Elements of Statistical Learning; Data mining, Inference and Prediction*. Springer, New York, pp. 84-94.

Kloser, R.J., Bax, N.J., Ryan, T.E., Williams, A. and Barker, B.A., 2001, Remote sensing of seabed types- development and application of normal incident acoustic techniques and associated ground truthing. *Marine and Freshwater Research*, v52, pp. 475-89.

Malthus, T. and Karpouzli, E., 2003. Integrating field and high spatial resolution satellite-based methods for monitoring shallow submersed aquatic habitats in the Sound of Eriskay, Scotland, UK. *International Journal of Remote Sensing*, 24.(13), pp.2585-2593.

Mumby, P. J., Green, E.P., Edwards, A.J. and Clark, C.D., 1997. Coral reef habitat mapping; how much detail can remote sensing provide? *Marine Biology*, 130, pp.193-202.

Sharples, C. 2000. Tasmanian Shoreline Geomorphic Types and Oil Spill Response Types, Report and Data Dictionary to accompany Tasmanian Shoreline Geomorphic Types v1.0 (2000), GIS Data Set, Department of Primary Industries, Water & Environment, Tasmania.

Siwabessey, P.J.W., Tseng, Y, Gavrolov, A.N. 2004. Seabed habitat mapping coastal waters using a normal incident acoustic technique. *Proceedings of Acoustics 2004*. Gold Coast, Australia.

Tarbatch, D., Mundy, C. and Haddon. M. 2003. Fishery Assessment Report: Tasmanian abalone fishery 2002. Marine Research Laboratories, Tasmanian Aquaculture and Fisheries Institute, pp. 144.

## ACKNOWLEDGEMENTS

The author would like to acknowledge The Tasmanian Aquaculture and Fisheries Institute at the University of Tasmania for providing support and financial assistance for this research. Andy Bickers at the University of Western Australia for processing and mosaicing the sidescan imagery. Richard Coleman and Alan Jordan for their encouragement and guidance and Arko Lucieer and Hugh Pederson for their advice on multivariate statistics.