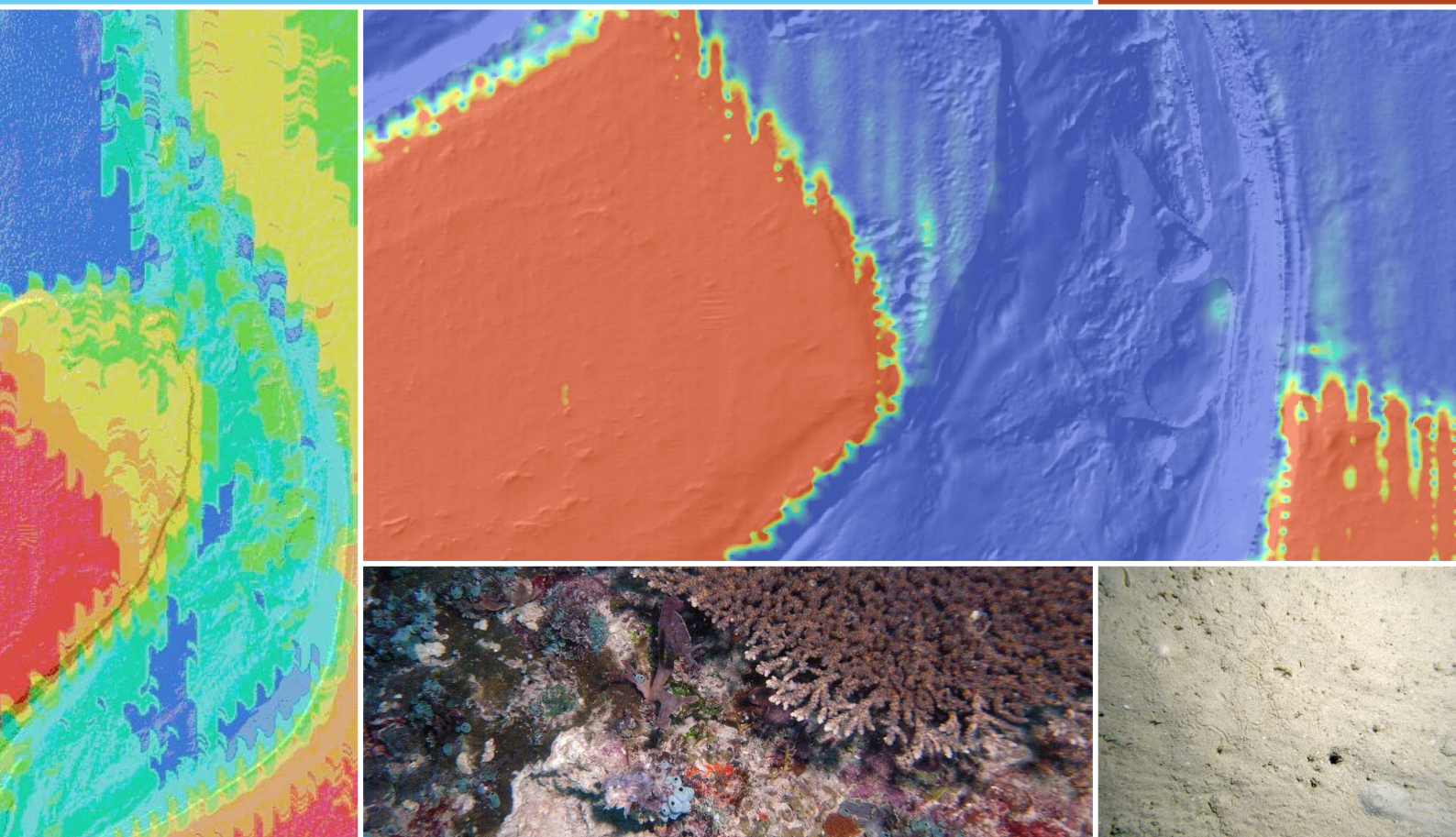




Australian Government
Geoscience Australia



Record 2013/11 | GeoCat 74092

Methodologies for seabed substrate characterisation using multibeam bathymetry, backscatter and video data

A case study from the carbonate banks of the Timor Sea, Northern Australia

P. Justy W. Siwabessy, James Daniell, Jin Li, Zhi Huang, Andrew D. Heap, Scott Nichol, Tara J. Anderson and Maggie Tran

Methodologies for seabed substrate characterisation using multibeam bathymetry, backscatter and video data

A case study from the carbonate banks of the Timor Sea, Northern Australia

GEOSCIENCE AUSTRALIA
RECORD 2013/11

P. Justy W. Siwabessy, James Daniell, Jin Li, Zhi Huang, Andrew D. Heap, Scott Nichol, Tara J. Anderson and Maggie Tran



Australian Government
Geoscience Australia

Department of Resources, Energy and Tourism

Minister for Resources and Energy: The Hon Gary Gray AO MP

Secretary: Mr Blair Comley, PSM

Geoscience Australia

Chief Executive Officer: Dr Chris Pigram

This paper is published with the permission of the CEO, Geoscience Australia



© Commonwealth of Australia (Geoscience Australia) 2013

With the exception of the Commonwealth Coat of Arms and where otherwise noted, all material in this publication is provided under a Creative Commons Attribution 3.0 Australia Licence.

(<http://www.creativecommons.org/licenses/by/3.0/au/>)

Geoscience Australia has tried to make the information in this product as accurate as possible. However, it does not guarantee that the information is totally accurate or complete. Therefore, you should not solely rely on this information when making a commercial decision.

Geoscience Australia is committed to providing web accessible content wherever possible. If you are having difficulties with accessing this document please contact feedback@ga.gov.au.

ISSN 1448-2177 (Print)

ISSN 2201-702X (PDF)

ISBN 978-1-922201-32-4 (Print)

ISBN 978-1-922201-33-1 (PDF)

GeoCat 74092

Bibliographic reference: Siwabessy, P.J.W., Daniell, J., Li, J., Huang, Z., Heap, A.D., Nichol S., Anderson, T.J. & Tran, M. 2013. *Methodologies for seabed substrate characterisation using multibeam bathymetry, backscatter and video data: A case study from the carbonate banks of the Timor Sea, Northern Australia*. Record 2013/11. Geoscience Australia: Canberra.

Contents

List of Figures	v
List of Tables	vii
Executive Summary	viii
Acknowledgements	ix
1. Introduction	1
1.1. Background	1
1.2. Study Objectives and Rationale	1
1.3. Study Area	2
2. Data Processing and Methodology	5
2.1. Multibeam Bathymetry Acquisition and Processing	5
2.2. Multibeam Backscatter Processing	5
2.3. Characterising Towed Video Data	7
3. Unsupervised Classification Case Study	8
3.1. Background and Theory	8
3.2. Methods	9
3.2.1. Angular Response Curves	9
3.2.2. AR Curve Clustering	10
3.2.2.1. Self Organising Map (SOM)	10
3.2.2.2. Hierarchical Clustering	11
3.2.3. Comparison of AR Curves and Video Classifications	12
3.3. Results and Discussion	12
3.3.1. Classification of AR Curves	12
3.3.2. Comparison of U-Matrix and Hierarchical Classification	14
3.3.3. Spatial Distribution of AR Classes	16
3.3.4. A refined Classification Scheme	19
3.3.5. Comparison of AR Classes, Video Observations and U-Matrix	22
3.4. Conclusion	25
4. Random Forest Prediction Case Study	27
4.1. Introduction	27
4.2. Methods	27
4.2.1. Generation of Grids	27
4.2.2. Angular Backscatter Response (AR) p-Value	28
4.2.3. Exploratory Analysis	30
4.3. Results and Discussion	30
4.3.1. Bathymetry, Backscatter and Derivative Grids	30
4.3.2. Angular Backscatter Response (AR) p-Value	52
4.3.3. Hardness Prediction	57
5. Summary	66
5.1. Approach Overview	66

5.2. Comparison of Methods	66
5.3. Limitations	67
5.4. Summary Comments and Recommendations	68
References	69
Appendices	74
Appendix A. Comparison of Primary Bottom Classification with Hierarchical Classification (Average Linkage with 12 classes).	74
Appendix B. Comparison of Relief Classification with Hierarchical Classification (Average Linkage with 12 classes).	75
Appendix C. Comparison of Biological Classification with Hierarchical Classification (Average Linkage with 12 classes).	76
Appendix D. Summary statistics describing sediment samples and predictor parameters at sediment sample locations.	77
Appendix E. The relationship of the hardness of seabed (i.e. the converted variable from soft and hard to 0 and 1) and relevant predictors.	81
Appendix F. The relationship among the predictors and hardness of seabed (i.e. the converted variable from soft and hard to 0 and 1).	82

List of Figures

Figure 1.1: Location of study areas on the Van Diemen Rise. Sample stations are represented and symbolised based on seabed hardness (soft or hard) observed in underwater towed video.	3
Figure 1.2: Bathymetry in meters and acoustic backscatter mosaic in dB units for the western sector of Area A. Locations of video transects are also shown.	4
Figure 3.1: a) Density map of seabed angular response within the study area, b) Greyscale density map with class means of the three main data classes (class A,B,C), c) Greyscale density map with class means of the twelve sub-classes shown.....	13
Figure 3.2: Dendrogram of hierarchical classification showing three main classes (labelled A-C) and twelve sub-classes (labelled 1-12) using average linkage method and Euclidean distance metric.....	14
Figure 3.3: Unified Distance Matrix (U-matrix) from SOM with results from hierarchical classification of SOM node averages. The main three classes are enclosed by polygons (Class A = red polygon, Class B = Yellow polygon, Class C = Blue polygon). The twelve sub-classes are numbered as per Figure 3.2).....	15
Figure 3.4: Van Diemen Rise survey area A1, showing (a) Three-class solution of angular response classification (A - red, B - yellow, C - blue); (b) Twelve-class solution, and (c). Geomorphic features adapted from Przeslawski et al. (2011).	17
Figure 3.5: Distributions of angular response sub-classes in the vicinity of the highest density of pockmarks.	19
Figure 3.6: Distributions of angular response classes reclassified as per Table 3.4.	21
Figure 3.7: Unified Distance Matrix (U-matrix) colour coded by percentage rock (primary) from video classification.	23
Figure 3.8: Unified Distance Matrix (U-matrix) colour coded by percentage low or moderate roughness (i.e. the sum of low or moderate roughness) from video classification.	23
Figure 3.9: Unified Distance Matrix (U-matrix) colour coded by percentage gardens (hard coral) from video classification.	24
Figure 3.10: Unified Distance Matrix (U-matrix) colour coded by percentage mixed patches from video classification.	25
Figure 4.1: (a)False colour backscatter mosaic of Area A showing video-derived classes at sampling stations and biology categories with representative photographs of different seabed types; (b) Angular backscatter response curves associated with different bottom types.	29
Figure 4.2: False colour bathymetry image of Area A with sampling stations indicated. All areas inside the black boxes were mapped during survey SOL5117, with remaining areas mapped during survey SOL4934.....	32
Figure 4.3: False colour bathymetry image of Area B with sampling stations indicated. All areas inside the black boxes mapped during survey SOL5117 and all others during survey SOL4934.	33
Figure 4.4: False colour bathymetry image of Area C with sampling stations indicated. All areas inside the black boxes mapped during survey SOL5117 and all others during survey SOL4934.	34
Figure 4.5: False colour bathymetry image of Area D with sampling stations indicated. All areas inside the black boxes mapped during survey SOL5117 and all others during survey SOL4934.	35

Figure 4.6: False colour backscatter image of Area A, showing video-derived classes at sampling stations. All areas inside the black boxes mapped during survey SOL5117 and all others during survey SOL4934.	36
Figure 4.7: False colour backscatter image of Area B, showing video-derived classes at sampling stations. All areas inside the black boxes mapped during survey SOL5117 and all others during survey SOL4934.	37
Figure 4.8: False colour backscatter image of Area C, showing video-derived classes at sampling stations. All areas inside the black boxes mapped during survey SOL5117 and all others during survey SOL4934.	38
Figure 4.9: False colour backscatter image of Area D, showing video-derived classes at sampling stations. All areas inside the black boxes mapped during survey SOL5117 and all others during survey SOL4934.	39
Figure 4.10: Histograms of backscatter values for the four study areas.	40
Figure 4.11: Boxplots of backscatter values grouped by geomorphic feature type.	40
Figure 4.12: Boxplots of backscatter values grouped into video-derived biology category.	41
Figure 4.13: Boxplots of backscatter values grouped into video-derived seabed type.	41
Figure 4.14: Seabed slope maps of the four study areas, Van Diemen Rise.	42
Figure 4.15: Seabed relief maps of the four study areas, Van Diemen Rise.	43
Figure 4.16: Surface area maps of the four study area, Van Diemen Rise.	44
Figure 4.17: Topographic Position Index maps of the four study areas, Van Diemen Rise.	45
Figure 4.18: Planar curvature maps of the four study areas, Van Diemen Rise.	46
Figure 4.19: Profile curvature maps of the four study areas, Van Diemen Rise.	47
Figure 4.20: Bathymetry Local Moran I of the four study areas, Van Diemen Rise.	48
Figure 4.21: Backscatter homogeneity maps of the four study areas, Van Diemen Rise.	49
Figure 4.22: Backscatter variance maps of the four study areas, Van Diemen Rise.	50
Figure 4.23: Backscatter Local Moran I maps of the four study areas, Van Diemen Rise.	51
Figure 4.24: False colour p-value map of seabed hardness for Area A, with biology categories indicated. All areas inside the black boxes mapped during survey SOL5117 and all others during survey SOL4934.	53
Figure 4.25: False colour p-value map of seabed hardness for Area B, with biology categories indicated.	54
Figure 4.26: False colour p-value map of seabed hardness for Area C, with biology categories indicated.	55
Figure 4.27: False colour p-value map of seabed hardness for Area D, with biology categories indicated.	56
Figure 4.28: The importance of the predictors for modelling the hardness using the Random Forest method.	60
Figure 4.29: Seabed hardness prediction map of Area A.	61
Figure 4.30: Seabed hardness prediction map of Area B.	62
Figure 4.31: Seabed hardness prediction map of Area C.	63
Figure 4.32: Seabed hardness prediction map of Area D.	64
Figure 4.33: Representative examples of comparison between video observations and prediction map with (a) & (b) showing good alignment with video track; and (c) & (d) showing misalignment with video due to the layback issue.	65

List of Tables

Table 3.1. Input variables to the SOMPak Self Organising Map routine	11
Table 3.2. Representation of original AR data by SOM Nodes.	15
Table 3.3. Summary of the three-class solution of angular backscatter response classification.	16
Table 3.4. Re-classification of twelve sub-classes based on underwater video observations. 1HC = hard corals.....	20
Table 4.1. Description of derived parameters from multibeam data presented in Figures 4.14 to 4.23.....	28
Table 4.2. Spearman correlation among the hardness and the predictors.	59
Table 4.3. Confusion matrix between predictions and observed values based on the average of 100 times of 10-fold cross validation.....	60
Table 5.1. Confusion matrix between classification map and video class at sampling stations within Area A.....	67
Table 5.2. Confusion matrix between prediction map and video class at sampling stations within Area A.....	67

Executive Summary

This report presents the results of a study on the use of multibeam sonar data to objectively classify and predict seabed substrate, with a focus upon delineating areas of hard and soft seabed. The analysis utilises multibeam sonar data, seabed samples and underwater video observations from four study areas on the Van Diemen Rise in the Timor Sea. This region of the continental shelf is characterised by a complex geomorphology comprising carbonate banks, shoals and reefs separated by incised valleys and surrounded by soft sediment plains. It is also a region where multiple stakeholders require accurate information on the seabed, including the offshore energy industry (through the need to place infrastructure), Government authorities responsible for managing the marine estate and conserving biodiversity in particular, and the fishing industry. For all these stakeholders, the development and testing of new techniques is a key component towards delivering robust and objective information on the biophysical properties of the seabed.

Two independent approaches to the analysis of multibeam data are tested: First, a two-stage classification-based clustering method, based solely on acoustic backscatter angular response curves, is used to derive a substrate type map. As such, this is the first documented case study to utilise angular backscatter response-derived products as a variable to predict and map continuous seabed substrate. Second, a prediction-based classification is produced using the Random Forest method based on bathymetry, backscatter data and their derivatives, with support from video and sediment data.

Results from the clustering method of hard and soft seabed types yielded classification accuracies of 78 - 87% when evaluated against seabed types as observed in underwater video. The prediction-based approach achieved a classification accuracy of 92% based on 10-fold cross-validation. For this result, the p -value of hard substrate was the most important predictor, highlighting the significance of the angular backscatter response curves in predicting seabed hardness. The Random Forest method identified seabed slope as a factor influencing hardness. Based on the strong performance of the predictive model, the Random Forest was also used to predict the distribution of hard and soft seabed types across the four study areas.

For the Van Diemen Rise study area, “acoustically hard” substrates are associated with areas of carbonate hard ground and relatively coarse seabed sediments. These areas are mostly found on banks and are associated with mixed sponge and coral gardens. In contrast, “acoustically soft” substrates are associated with finer sediments that form the plains and valley floors, with pockmarks observed in some areas.

These results demonstrate the utility of acoustic data to broadly and objectively characterise the seabed substrate and thereby inform on the distribution of key habitat types. The choice between the two approaches depends upon the availability and the reliability of the *a-priori* knowledge of the seabed for a given area. The classification-based approach uses only the angular backscatter response data but requires several steps to produce a continuous map because of the type of input data used. The prediction-based approach provides a continuous map over an area of interest but does require independent information on predictors (e.g. from sediment samples and video). The prediction approach is recommended for future work if continuous predictions of seabed hardness are required and if the information of relevant predictors is available.

Acknowledgements

This study was undertaken as an activity of the Seabed Mapping and Characterisation Project within the Coastal, Marine and Climate Change Group at Geoscience Australia. Data used for the study were collected in 2009 and 2010 on collaborative surveys undertaken between Geoscience Australia and the Australian Institute of Marine Science (AIMS) on board RV *Solander*. We thank the master and crew of the RV *Solander* and Marcus Stowar (AIMS) for support with the survey and towed-video and still photography acquisition. We also thank the Field and Engineering Support staff at Geoscience Australia. Sediment samples were analysed by the Palaeontology and Sedimentology Laboratory, Geoscience Australia. Tony Nicholas and Andrew Carroll (GA) provided constructive and valuable reviews. We thank Kim Picard for proof-reading the final report. This record is published with the permission of the Chief Executive Officer, Geoscience Australia.

1. Introduction

1.1. Background

Geoscience Australia provides spatial information about the physical and biological character of the seabed to support the decision-making of the Australian Government, specifically in relation to marine environmental management and responsible resource development. Central to this work is the prediction of Australia's seabed habitats from spatially continuous geophysical data, including multibeam sonar bathymetry and acoustic backscatter data as well as information from *in situ* sampling and observations.

Geomorphic features on the seabed provide benthic biota with habitats that are characterised by varying degrees of shelter, food supply, breeding grounds and refuge from predators. Physical properties that may influence the form and distribution of these seabed habitats include: substrate hardness, sediment texture and composition, water depth and seabed slope, rugosity, and aspect. Using geophysical datasets such as multibeam sonar bathymetry and backscatter to determine and map these seabed properties can provide useful insights into the availability of potential benthic habitats. For instance, finer-grained (muddy) sediments generally produce low backscatter intensity whereas coarser (sandy) sediments are more likely to produce higher backscatter intensity. This effect is due to higher density and sound speed properties, larger scale of apparent roughness, and porosity (APL, 1994; Briggs *et al.*, 2001, 2002; De Falco *et al.* 2010; Ferrini & Flood 2006; Goff *et al.* 2000; Sutherland *et al.*, 2007). Backscatter intensity is also dependent on the angle of incidence. For example, the backscatter strength resulting from small incidence angles is generally higher than larger incidence angles. This is due to the respective coherent versus incoherent returns arising from signal scattering (de Moustier & Alexandrou, 1991). This property is generally unique to each seabed type. For instance, sand-covered and mud-covered seabed show a rapid decrease in the backscatter strength toward the outer swath angles, whereas dense seagrass beds produces backscatter strength almost independent of the incidence angle (Siwabessy *et al.*, 2006).

The relationship between the physical environment and benthic biodiversity is complex, and substratum character is one of many physical variables that delineate habitats and influence benthic ecology (Newell *et al.*, 2001; Warwick & Davies, 1977; Hensley, 1996; Mackie *et al.*, 1995; Simpson *et al.*, 1986; McArthur *et al.*, 2010). Substrate hardness largely influences the nature of attachment of the organism to the seabed (sessile versus mobile versus in-fauna) (Williams & Leach, 1999), and affects the composition and abundance of benthic organisms. Prominent factors include: the sediment texture and composition, bed stress and the frequency and magnitude of seabed disturbance, water depth, and geomorphology.

1.2. Study Objectives and Rationale

The objective of this study is to develop and test new methods for classifying the seabed on the basis of the relative hardness of the substrate. Using multibeam sonar bathymetry and backscatter data, supported by information from seabed samples and underwater video, two approaches to the classification and prediction of seabed type are developed and evaluated. The test dataset is from the

Van Diemen Rise in the Timor Sea, an area of northern Australia where detailed information on the biophysical characteristics of the seabed is required to support the ongoing management and resource development of the region.

1.3. Study Area

The Van Diemen Rise in the Timor Sea comprises a spatially complex suite of geomorphic features. These include flat-topped carbonate banks that rise to within 10 to 40 m of the water surface, terraces, deep valleys that incise to depths up to 200 m and soft sediment plains ([Figure 1.1](#)) (Przeslawski *et al.*, 2011). This geomorphic diversity presents an opportunity to develop a methodology for distinguishing between hard and soft seabed types using multibeam and backscatter data. To achieve this, Geoscience Australia (GA) and the Australian Institute of Marine Science (AIMS) completed seabed mapping surveys on the Van Diemen Rise in 2009 (Heap *et al.*, 2010) and 2010 (Anderson *et al.*, 2011). These surveys provided high-resolution multibeam maps and seabed samples from four study areas (Areas A - D) located along a cross-shelf transect ([Figure 1.1](#)). The study areas encompass various geomorphic features (carbonate banks, shoals, reefs, valleys and plains), with Areas A and C characterised by a wider depth range than Areas B and D. Apart from the shallow water depths over a shoal located in the southeast corner of the study area, Area B generally exhibited relatively flat seabed with few banks and valleys throughout the centre. Area D contained the most homogenous terrain compared to other study areas. It mainly comprised a relatively flat bottom type, indicated by the narrow depth range.

Initial interpretation of multibeam backscatter ([Figure 1.2](#)) indicates that the shallow tops of the carbonate banks have the highest backscatter (i.e. acoustically hard). Terraces also show a strong acoustic return. In contrast, sediment deposits between the banks and terraces have the weakest acoustic returns.

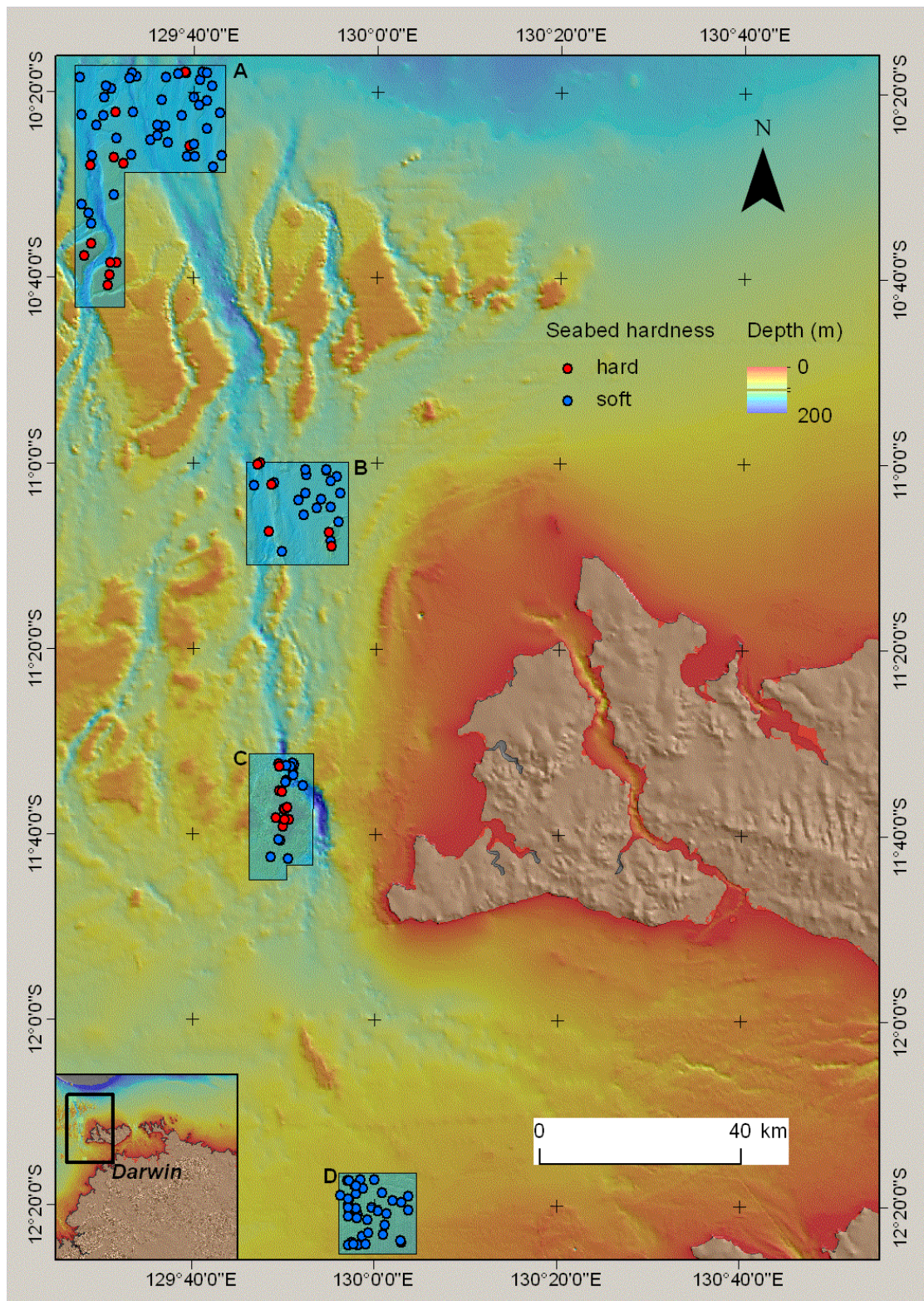


Figure 1.1: Location of study areas on the Van Diemen Rise. Sample stations are represented and symbolised based on seabed hardness (soft or hard) observed in underwater towed video.

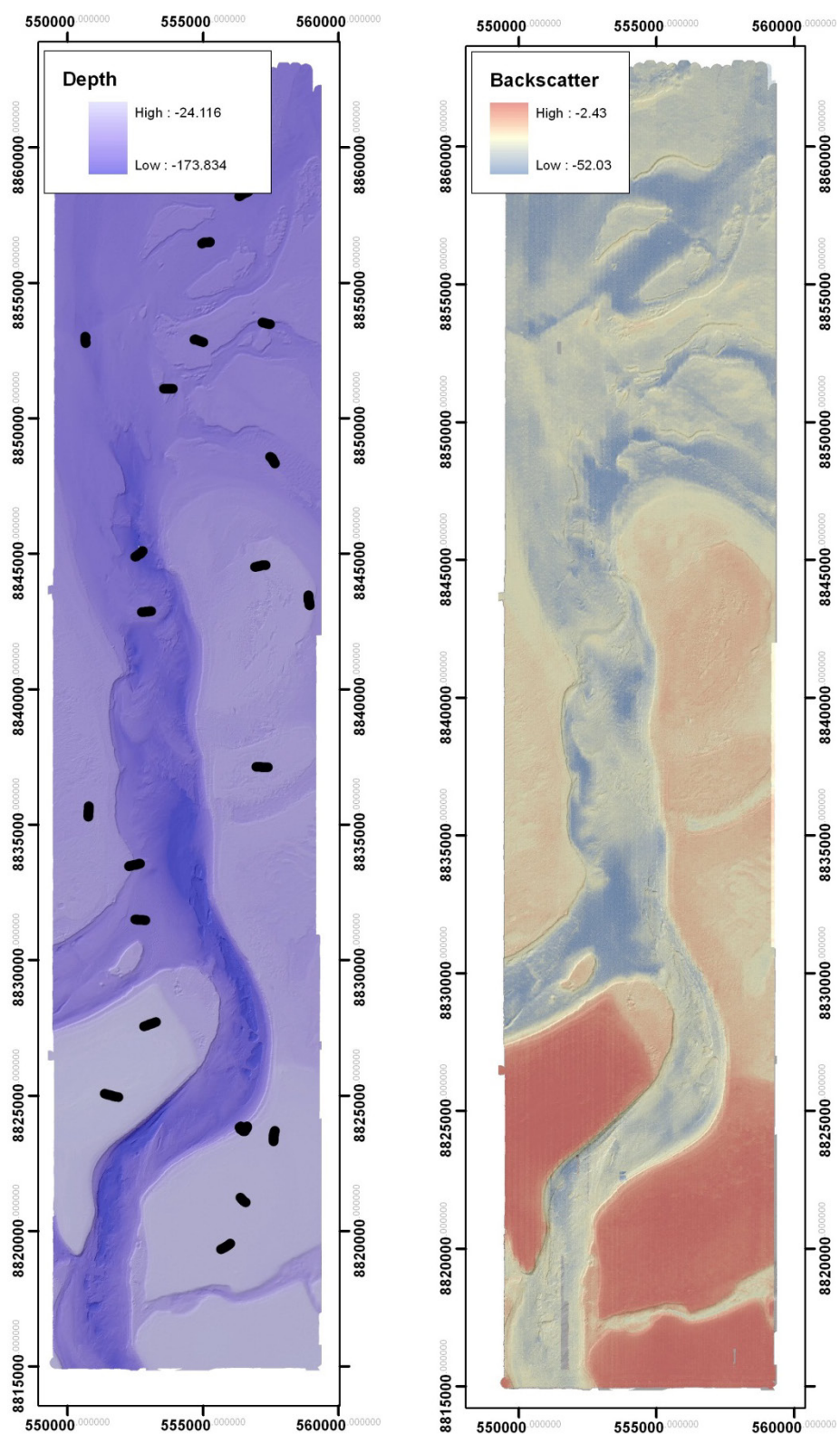


Figure 1.2: Bathymetry in meters and acoustic backscatter mosaic in dB units for the western sector of Area A. Locations of video transects are also shown.

2. Data Processing and Methodology

2.1. Multibeam Bathymetry Acquisition and Processing

A Simrad EM3002 300 kHz multibeam sonar (MBS) system was used on board the RV *Solander* to acquire multibeam data across four survey grids over two surveys (Heap *et al.*, 2010; Anderson *et al.*, 2011). Motion referencing and navigation data were collected with an Applanix Position and Orientation system and a C-Nav GPS system, respectively. Multibeam data were acquired using Kongsberg's *Seabed Information System* (SIS) software.

The multibeam bathymetry data were first processed using *Caris*TM *HIPS & SIPS* v6.1 software, and included: i) running algorithms that corrected for tide and vessel pitch, roll and heave, and ii) software filters and visual inspection of each swath line to remove any remaining artefacts and noisy data (e.g. nadir noise and data outliers). A bathymetry editing process was then conducted during the survey, using IVS Fledermaus 3D visualisation software (v.7.2). The HDCS files from *Caris*TM were imported into *IVS3D DMagic* to produce a Combined Uncertainty Bathymetry Estimator (CUBE) surface. This surface was edited manually within the *IVS3D* Editor tool. Final bathymetry surfaces were created within *Caris*TM using the edits from Fledermaus and then exported as a surface grid (bathymetric map) for display and further analysis. Final processing to minimise tidal bursts was completed after the survey using a co-tidal solution in *Caris*TM.

2.2. Multibeam Backscatter Processing

Multibeam backscatter data were processed using the *CMST-GA MB Process* v10.10.17.0 toolbox software co-developed by the Centre for Marine Science and Technology (CMST) at Curtin University of Technology and GA (described in Gavrilov *et al.*, 2005a, 2005b; Parnum, 2007). The fully processed backscatter strengths were corrected for transmission loss and insonification area, and were normalised to the transmitted pulse length. The process within the toolbox also involved: removal of the system transmission loss; removal of the system model; calculation of the incidence angle; correction of the beam pattern; calculation of the angular backscatter response within a sliding window with a 50% overlap in a 1° bin and; removal of the angular dependence and restoration to the backscatter strength at an angle of 25° (Daniell *et al.*, 2010).

The toolbox calculates the backscatter coefficient corrected for transmission loss and insonification areas. Calculation of the insonification area is based on the equation given in Talukdar *et al.* (1995). With these measurements, the corresponding incidence angle and coordinates on the seabed (X-Y) and depth (Z) are calculated. The full processing algorithm was developed into the computer software program *Matlab*[®], involving the following steps:

1. conversion from the Simrad raw ALL data format into *Matlab*[®] data format;
2. calculation of the absolute X, Y, Z position and the incidence angle θ for each beam and ping;
3. removal of the system transmission loss;
4. removal of the system-implemented correction;

5. calculation of the surface backscattering strength, which involves correction for transmission loss and area;

The Simrad system incorporates a correction that makes best use of the limited dynamic range of the electronics used (Hammerstad, 2000). The system-implemented correction includes Lambert's law and the scattering area. In addition, the Simrad system uses different expressions in three different angular domains of the incidence angle. The first domain relates to the normal incidence angle (i.e., $\theta = 0^\circ$); the second domain pertains to the incidence angle defined by $0^\circ < \theta < \theta_{crossover}$; and the third domain involves the oblique incidence angle $\theta > \theta_{crossover}$, where $\theta_{crossover}$ is the crossover angle which is between 5° and 30° as selected by the operator (Hammerstad, 2000). It is necessary to remove these manufacturer attempted corrections to obtain the true angular response of the backscatter.

The surface scattering coefficient can be determined by correcting the backscatter intensity for the actual transmission loss and normalising the values for the area under investigation (A_{ins}):

$$\sigma_{bs} = \frac{I(t_r)R^4(t_r)}{I_s R_0^2 \exp(2\alpha R) A_{ins}(t_r)} \quad (2.1)$$

where $I(t_r)$ is the instantaneous (received) backscatter intensity, I_s is the transmitted signal intensity, R_0 is 1m from the sonarhead, R is the range, α is the seawater acoustic absorption coefficient in neper per unit distance, σ_{bs} is the surface backscatter coefficient and (t_r) is the time usually chosen so that it corresponds to the two-way travel time of the sonar signal to the centre of the insonification area, A_{ins} . The transmission loss is the energy lost due to spherical spreading of acoustic energy and acoustic absorption in the water column. Spherical spreading loss is a function of range R and for two-way travel is equal to $40\log_{10}R$ (in dB). Absorption loss is also a function of range and the acoustic absorption coefficient (a) in dB, so for two-way travel this is $2aR$ (dB) (see Parnum, 2007; Parnum & Gavrilov, 2011).

The area that controls the peak backscatter intensity is called the insonified area. The more accurate insonified area is calculated as:

$$A_{ins} = \varphi R^2 \min \left\{ \cos(\theta) \tan \left[\cos^{-1} \left(\frac{\cos(\theta)}{1 + \frac{cT_w}{2R}} \right) \right] - \sin(\theta), \sin(\theta_t) \cos(\theta) \right\} \quad (2.2)$$

where T_w is the transmitted pulse length, φ is the along track width of the receive beam and θ_t is the transverse (across track) width of the receive beam.

In practice, the backscatter measure is commonly presented in logarithmic scale (dB) known as seabed backscatter strength, and follows (Parnum, 2007; Parnum & Gavrilov, 2011):

$$BS = 10 \log_{10}(\sigma_{bs}) \quad (2.3)$$

A sliding window approach was used for calculation of the angular backscatter response and removal of the angular dependence. It adopted a 50% overlap in a 1° bin of incidence angle (Gavrilov *et al.*, 2005a, 2005b; Parnum, 2007).

$$BScor(X, Y, \theta) = BS(X, Y, \theta) - \overline{BS}(X, Y, \theta) + \overline{BS}(X, Y, 25^\circ) \quad (2.4)$$

where $BS(X, Y, \theta)$ is all the backscatter data within the sampling window (X, Y) at angle θ , $\overline{BS}(X, Y, \theta)$ is the mean backscatter strength within the sampling window (X, Y) at angle θ , and $\overline{BS}(X, Y, 25^\circ)$ is the mean backscatter strength measured within the sampling window at the reference angle of 25° (Daniell *et al.*, 2010). The average angular backscatter within a sliding window is in fact the average angular backscatter response curve.

2.3. Characterising Towed Video Data

Seabed habitats and benthic macro-organisms of the study area were identified in real-time using the 3-tiered C-BED (Characterisation of the Benthos and Ecological Diversity) characterisation scheme described in Anderson *et al.* (2008) and Nichol *et al.* (2009). Further video analysis was required and conducted at the Ecology Laboratory within Geoscience Australia, adopting a similar method described in Mortensen & Buhl- Mortensen (2004). The video footage was broken up into a 15 second sequence of seabed sections at 30 second intervals along each transect to classify substratum composition and geomorphology (e.g. bedform and relief). Substratum composition (i.e. rock, rubble, boulder, cobble, pebble, gravel, sand, and mud as defined by the Wentworth scale) was categorised by primary (>50% of the area viewed) and secondary (>20% of the area viewed) percent-cover using the protocol of Stein *et al.* (1992), estimated to a precision of 10% (0, 10, 20...100%) (Wentworth, 1992). Geomorphology was defined as the local “vertical relief” (e.g. flat [0-0.3 m], low [0.3-1 m], moderate [1-3 m]). Four classes of biotic composition were identified; bioturbated, mixed patches, mixed gardens, and mixed gardens (hard corals).

C-BED characterizations were entered into “GNav Real-time GIS Tracker” software (© Gerry Hatcher, 2002) using a 142 key Cherry programmable keyboard (© Cherry, 2008) (details described in Anderson *et al.*, 2008). USBL navigation (UTC date, time, latitude, and longitude) was captured for each C-BED data-entry and was also logged continuously (1-2 second fixes) to provide navigational tracks for all video transects. All video footage was recorded to miniDV tapes and .avi digital files with a visual date/time stamp and audio-encoded 1-2 sec navigation (GPS date, time, latitude, longitude, and USBL depth). The system also allowed automatic logging of ship navigation. This is of extreme importance when the USBL navigation feed becomes unavailable (e.g. data collected in 2009).

The identified substratum composition was then used to form two broad video classes; “hard” and “soft”. Rock, boulders and cobbles identified in either primary or secondary substrata were merged to form the “hard” class, whereas gravel, sand and mud found in both primary and secondary substrata were combined to form the “soft” class. This process unavoidably introduced a mixed class in which the primary and secondary types were different. In this case, the class was merged into the “hard” class.

3. Unsupervised Classification Case Study

3.1. Background and Theory

Multibeam sonar instruments acquire seabed bathymetry and backscatter data over a port and starboard orientated swath, with a width of 120-150 degrees (Hughes-Clarke, 1994). This data, after correction for seabed slope, beam pattern, time varying and angle varying gains, and area of insonification, provide an estimate of acoustic backscatter strength (Fonseca & Mayer, 2007). Theoretical models and experimental observations demonstrate that acoustic backscatter from the seabed is a complex function of many factors, such as incidence angle, acoustic frequency, roughness scales, grain size distribution, presence of fauna and flora, biological reworking, and volume reverberation (Jackson *et al.*, 1986; de Moustier & Alexandrou, 1991; APL, 1994; Hughes-Clarke, 1994; Lyons *et al.*, 1994; Talukdar *et al.*, 1995; Jackson *et al.*, 1996; Novarini & Caruthers, 1998; Williams *et al.*, 2002; Siwabessy *et al.*, 2006b; Parnum, 2007; Fonseca & Mayer, 2007; De Falco *et al.*, 2010; Gavrilov & Parnum, 2010; Hamilton & Parnum, 2011; Hasan *et al.*, 2012). Of these, the incidence angle is of primary importance, with backscatter strength near the nadir (i.e. small incidence angles) generally higher than values recorded in the outer swath because of differences between reflection (near nadir) and scattering (outer swath) (de Moustier & Alexandrou, 1991; Ferrini & Flood, 2006). The variation of (discrete) measurements of backscatter, over a range of incidence angles, is referred to as an angular response and is considered an intrinsic property of the seabed (Hughes-Clarke *et al.*, 1997; Parnum *et al.*, 2004, 2006; Fonseca & Mayer, 2007; Gavrilov & Parnum, 2010; Hamilton & Parnum, 2011). Angular response (AR) analysis utilises the full backscatter angular response curve to segment the seabed into regions with similar acoustic properties (de Moustier & Matsumoto, 1993; Hughes-Clarke *et al.*, 1997; Fonseca & Mayer, 2007). Unlike pixel-based classification methods, angular response analysis is limited to two rectangular footprints on either side of the multibeam swath. Therefore, AR analysis is most effective where substrate variability is small relative to the half-swath width (Hughes-Clarke *et al.*, 1997).

Several morphological parameters can be extracted from multibeam sonar angular response curves and compared to formal mathematical models that link acoustic backscatter observations to seabed properties (Fonseca & Mayer, 2007; Fonseca *et al.*, 2009; Lamarche *et al.*, 2011). These models can then produce estimates of various seabed geotechnical properties, which can be used to predict the substratum properties of the seabed. This approach shows promise for seabed habitat mapping applications (Fonseca *et al.*, 2009; Lamarche *et al.*, 2011), although is yet to be tested over a range of benthic marine environments and acoustically complex sediments such as coarse-grained carbonates (Brown *et al.*, 2011).

The angular backscatter response curve is an inherent property of the seabed and can therefore be used as the primary means of seabed characterisation. Hamilton & Parnum (2011) argue that the direct clustering of angular backscatter response curves forms a standalone, independent map of acoustic properties. The virtues of the direct clustering method are that it uses the actual angular backscatter response curve, and is computationally simple and rapid.

Currently, there are few published studies on clustering AR curves. This study introduces a data mining approach to cluster AR curves. The approach involves two stages; an initial classification using

a Self Organising Map (SOM), and a second stage utilising hierarchical clustering based on the method of Vesanto & Alhoniemi (2000). This approach represents a modification to the method of Hamilton & Parnum (2011) as it provides visualisation of cluster relationships at both stages of processing.

In this study, there are two estimates of the number of clusters in a dataset. One estimate will be a gross over-estimation. Clustering results will then be refined by using video characterisations to identify regions of similar seabed substrates. A simple segmentation based on only three classes is also presented. Mapping of video observation data onto the SOM unified distance matrix (U-matrix) is novel and depicts environmental gradients within the data set and aids the understanding of cluster properties.

3.2. Methods

The seabed mapping process presented here comprises three steps. In the first step, the angular response curves were prepared (Section 3.2.1). The second step clustered the angular response using a two-stage clustering approach which involved an initial SOM classification and a hierarchical clustering using a dendrogram. This resulted in a gross over-estimation of the number of classes in the dataset and a simple segmentation of three main classes (Section 3.2.2). The last step used the video data to refine class solutions into a more practical number (Section 3.2.3).

3.2.1. Angular Response Curves

Stemming from the angular response curves, Fonseca & Mayer (2007) introduced a technique called Angular Range Analysis (ARA). It begins with the beam-by-beam time-series of acoustic backscatter provided by the multibeam sonar and then corrects the backscatter for seabed slope, beam pattern, time varying and angle varying gains, and area of insonification (Fonseca & Mayer, 2007). With accurate estimates of backscatter strength, the acoustic backscatter values from different acquisition lines can be reduced to a near calibrated scale of scattering strength. In the absence of an absolute calibration of the sonar, these estimated measurements are only relative values, but in certain conditions they can yield reliable near-absolute results (Fonseca & Mayer, 2007).

The extraction of angular response curves from the raw Simrad multibeam data was achieved using the multibeam backscatter *CMST-GA MB Process* v10.10.17.0 toolbox software co-developed by the Centre for Marine Science and Technology (CMST) at Curtin University of Technology and Geoscience Australia (described in Gavrilov *et al.*, 2005a, 2005b; Parnum, 2007). A number of corrections were introduced to the data and the angular response curves were produced as the average response curve within the adopted sliding windows in which port and starboard swath were processed separately as part of the process of the removal of the backscatter angular dependence, given as $\overline{BS}(X, Y, \theta)$ in equation 2.4.

Over 200,000 angular response curves were produced from the Area A dataset by averaging every 30 consecutive pings. Due to the ranges of depths encountered within the survey, a 45 degree angular sector at one-degree intervals was used for the AR analysis. A 2-dimensional histogram of angular response and grazing angle (for each 1 degree interval) was produced to aid the identification of clusters within the AR dataset.

3.2.2. AR Curve Clustering

The purpose of data mining is to gain insight into the structures within complex datasets. Efficient methods for data visualisation and summaries (as clusters) are important. Vesanto & Alhoniemi (2000) proposed a combination of a Self Organising Map (SOM) and a hierarchical clustering to produce summaries and identify structures within complex, multi-dimensional datasets. When a SOM has a large numbers of output units, it produces detailed summaries (or proto-clusters) of the input data which then allows for computationally efficient analysis of the SOM output by hierarchical clustering. The technique has been found to perform well when compared to direct clustering of the data and reduces the computation time needed. Another advantage of using both the SOM and hierarchical clustering is capacity to visualise cluster properties at both classification stages to aid the identification of clusters (Vesanto & Alhoniemi, 2000).

3.2.2.1. Self Organising Map (SOM)

The SOM is a particular Neural Network algorithm that converts complex non-linear statistical relationships between high-dimensional data items into simple geometric relationships to produce a low-dimensional representation of the input space (Kohonen, 1995). A typical SOM consists of two-layer architecture, an input layer and a competitive layer (output layer). SOMs use an unsupervised iterative procedure to group similar vectors from the input data into nearby points on a 2-dimensional lattice of nodes (Marsh & Brown, 2008). Because there are more vectors than nodes, each node in the SOM summarises multiple vectors. The shape and number of nodes within the output layer is determined by the user. The arrangement of these nodes can be rectangular or hexagonal.

The output grid can be used as a convenient visualisation surface showing the features of the SOM (and thus the data) results in either a Sammon Map (Sammon, 1969) or a Unified Distance Matrix (Ultsch & Siemon, 1990). The Unified Distance Matrix (or U-Matrix), which will be featured in this report, presents the Euclidean distances between adjacent units in the SOM. Using a hexagonal topology, all nodes have six neighbours (apart from those occurring at the edge of the SOM lattice). The visualisation of the distances between nodes thus allows visualisation of the cluster structure within the SOM.

In this study, the freely available SOMPak software was used to run the SOM (Kohonen *et al.*, 1996), the variables used as input to the SOM routine are summarised in [Table 3.1](#). The AR curves that represent the acoustic response of the seabed at one degree intervals between 0 and 45 degrees are used as input to the SOM. These curves are mapped onto a 20x20 node output layer, resulting in 400 reference vectors to be used in the subsequent hierarchical classification.

Table 3.1. Input variables to the SOMPak Self Organising Map routine

SOM Variable	Input
Trials	100
Topology Type	Hexagonal
Neighbourhood Type	Bubble
Nodes in X-dimension	20
Nodes in Y-dimension	20
Training length of first part	10000
Training rate of first part	0.05
Radius on first part	15
Training length in second part	100000
Training rate of second part	0.01
Radius of second part	2

3.2.2.2. Hierarchical Clustering

The two main methods to cluster data (make partitions) are hierarchical and partitive approaches. Partitive clustering algorithms such as k-means give a single set of clusters with no particular organisation of structure within them. Hierarchical clustering is a method of cluster analysis which seeks to build a hierarchy of clusters (Kaufman & Rousseeuw, 1990). Strategies for hierarchical clustering generally fall into two types:

- Agglomerative: This is a “bottom up” approach: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.
- Divisive: This is a “top down” approach: all observations start in one cluster, and splits are performed recursively as one moves down the hierarchy.

In general, the merges and splits are determined in a “greedy” manner. A greedy algorithm is any algorithm that follows the problem-solving heuristic of making the locally optimal choice at each stage with the hope of finding the global optimum. The results of hierarchical clustering are usually presented in a dendrogram. Agglomerative clustering methods are the most common and will be considered here. Agglomerative clustering routines usually have the following steps:

1. Initialise: assign each vector to its own cluster
2. Compute distances between all clusters
3. Merge the two clusters that are closest to each other
4. Return to step 2 until there is only one cluster left.

The resulting dendrogram (clustering tree) is then used to interpret the data structure and determine the number of clusters. The dendrogram does not provide a unique clustering. Rather, a partitioning can be achieved by cutting the dendrogram at certain levels.

The hierarchical clustering was performed using R (R Development Core Team, 2011) employing the Average Linkage method with a Euclidean distance metric. The Average Linkage method uses the average of the distances between all pairs of objects, (where one member of the pair is from each of the cluster) as the distance between two clusters. An advantage of the Average Linkage method is it

considers all members within a cluster rather than just single points and therefore ensures the method is less influenced by extreme values than the single or complete linkage methods (Kaufman & Rousseeuw, 1990). The initial number of classes chosen was 12, with the intention of reducing this number based on a comparison of habitat and geomorphology from seabed video observations for each class.

3.2.3. Comparison of AR Curves and Video Classifications

The geometric centre of each AR curve was calculated based on the spatial location of the central beam within each side of the swath. Every video classification that occurred within $2.5 \times$ water depth (approximately half one swath width) of an AR curve was used to construct a set of tables indicating the occurrence of seabed types for each AR class. A comparison table was produced for each of the four types of video observation (Primary Bottom Type, Secondary Bottom Type, Relief, and Biology).

3.3. Results and Discussion

3.3.1. Classification of AR Curves

Histograms of the angular response for each 1 degree interval were combined to create a density plot (Figure 3.1). This plot is used to observed structure within the dataset and aid the classification procedure. Peaks in the density plot indicate clusters within the dataset. Three distinct clusters are observed (Figure 3.1a), with an upper-cluster (higher acoustic reflectivity indicating a hard substrate) well separated from two lower clusters (inferred softer substrates).

Using hierarchical clustering, the 400 SOM nodes were split into three classes to represent a simple segmentation based on cluster structure provided by the density plot. These three classes were labelled A, B, and C (Figure 3.1b). The SOM nodes were also split into twelve sub-classes (numbered 1-12) to provide an over-segmented dataset for the analysis of geomorphology and habitat variance within the three main classes (Figure 3.1c). The relationships between all classes are represented as a dendrogram in Figure 3.2. The “acoustically hard” cluster is represented by only two sub-classes. These two classes are quite distinct from the others, as observed by the height (or separation) within the dendrogram. The other two classes are each represented by five sub-classes. Within classes B and C, sub-classes 4 and 12 are the most distinct (based on height within the dendrogram).

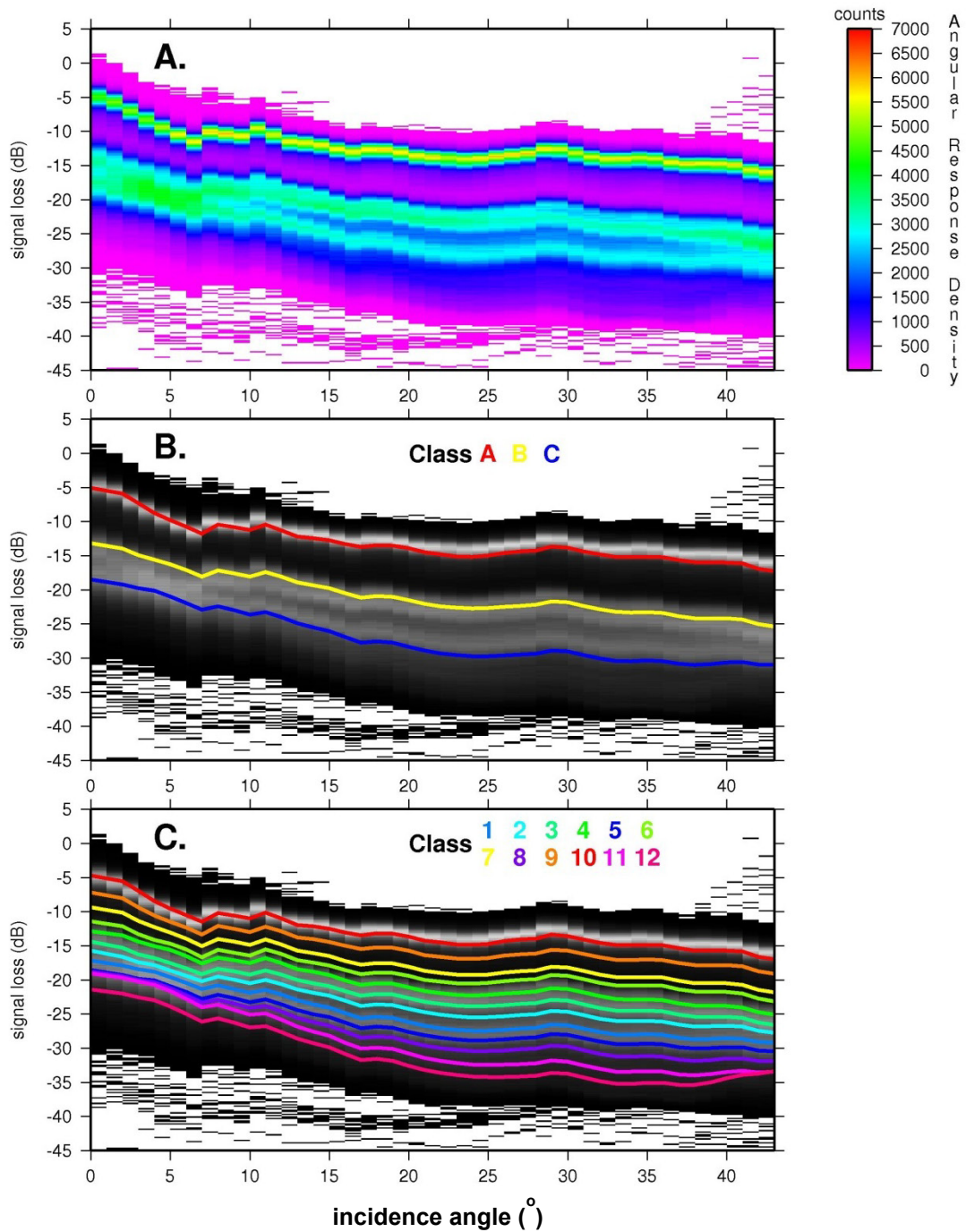


Figure 3.1: a) Density map of seabed angular response within the study area, b) Greyscale density map with class means of the three main data classes (class A,B,C), c) Greyscale density map with class means of the twelve sub-classes shown.

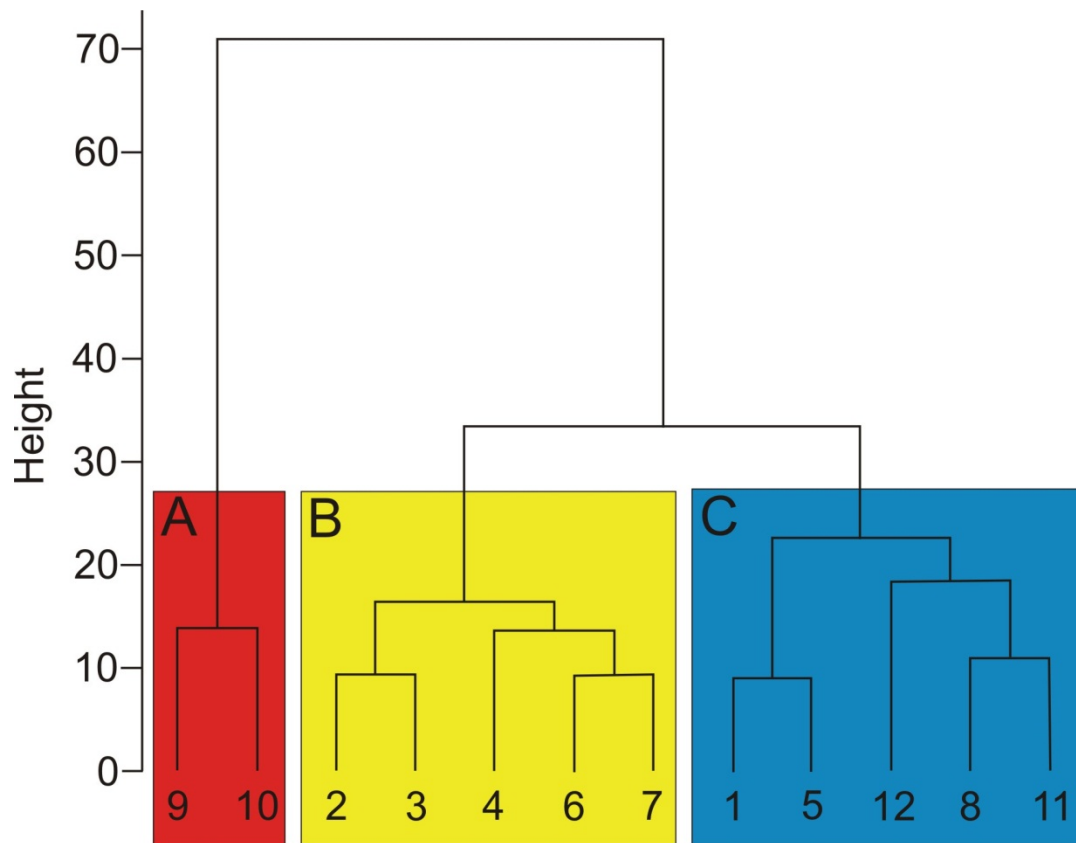


Figure 3.2: Dendrogram of hierarchical classification showing three main classes (labelled A-C) and twelve sub-classes (labelled 1-12) using average linkage method and Euclidean distance metric.

3.3.2. Comparison of U-Matrix and Hierarchical Classification

The results of the hierarchical clustering (both the three main classes and twelve sub-classes) have been remapped on a SOM Unified Distance Matrix (Figure 3.3). The U-Matrix presents the Euclidean distances between adjacent units in the SOM. Distances between the SOM nodes within classes A, B and C are typically in the range of 0-20 over the U-Matrix. The U-Matrix does not show a discernable boundary between classes B and C, however, distances between classes A and B are on the order of 50-100. This result shows much stronger separation between classes A and B than between B and C (a result also inferred from the initial density plot).

Table 3.2 shows a summary of the number of SOM nodes and AR curves which are mapped onto each class and sub-class. All the classes consist of approximately the same number of AR curves (61842, 72084, and 78489 for classes A, B, and C, respectively). However, the 61842 AR curves of class A (which consists of only 2 sub-classes) are mapped on to only 15 SOM nodes indicating an unusually tight cluster within this study.

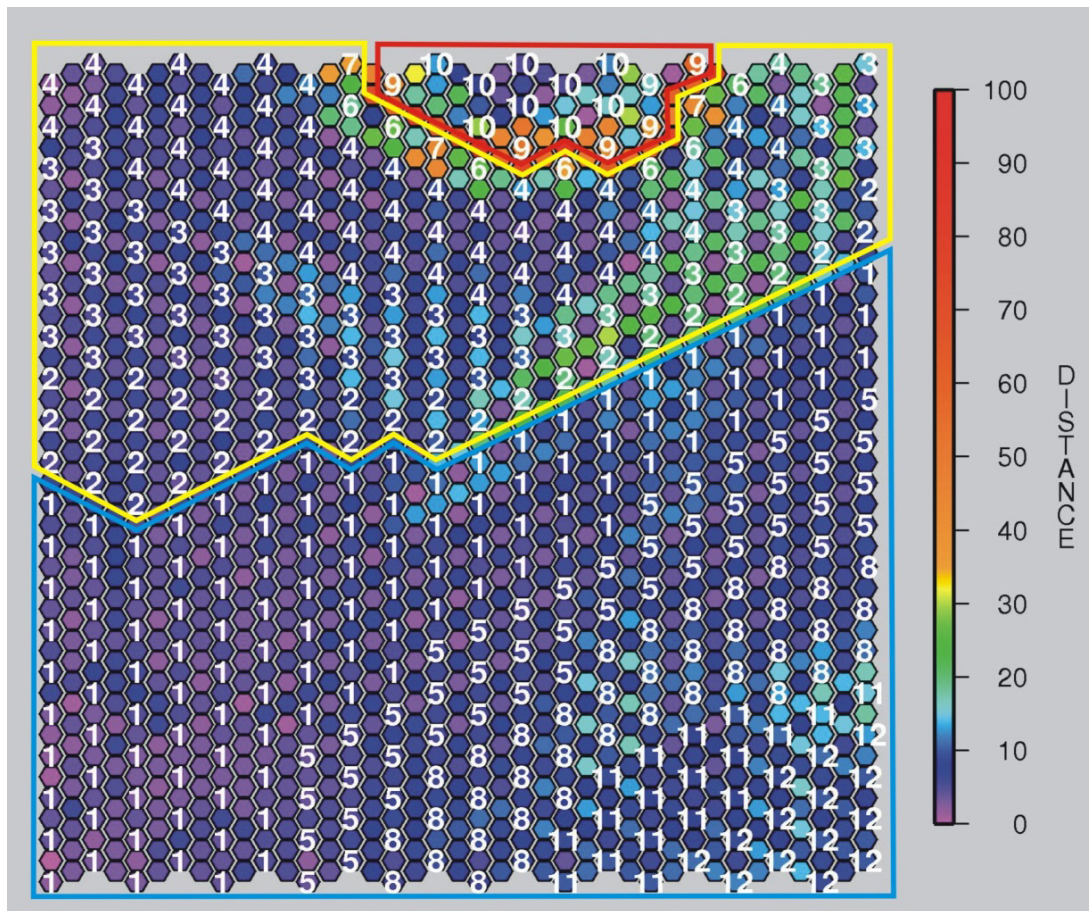


Figure 3.3: Unified Distance Matrix (U-matrix) from SOM with results from hierarchical classification of SOM node averages. The main three classes are enclosed by polygons (Class A = red polygon, Class B = Yellow polygon, Class C = Blue polygon). The twelve sub-classes are numbered as per [Figure 3.2](#).

Table 3.2. Representation of original AR data by SOM Nodes.

Class		SOM nodes	AR Curves	AR Curves per SOM node
A	9	6	8489	1414.8
	10	9	53353	5928.1
B	2	34	12380	364.1
	3	55	29171	530.4
	4	56	25140	448.9
	6	7	3119	445.6
	7	3	2274	758
C	1	114	30354	266.3
	5	45	14472	321.6
	12	14	11344	810.3
	8	38	14498	381.5
	11	18	7821	434.5
Summary	A	15	61842	7342.9
	B	155	72084	2547
	C	229	78489	2214.2

3.3.3. Spatial Distribution of AR Classes

The geomorphology of the survey area was previously classified into five geomorphic features by Przeslawski *et al.* (2011) based on definitions from Heap & Harris (2008): 1) Banks – local or regional areas of elevated seabed with one or more steep sides; 2) Terraces – relatively flat or gently sloping seabed with a moderately steep rise on one side and a moderately steep drop on the other side; 3) Valleys – tapered depressions on the shelf characterised by laterally converging contours of increasing depth; 4) Ridges – long, narrow elevation with steep sides; and 5) Scarps – elongated and comparatively steep slope separating more gently sloping areas (Figure 3.4).

Table 3.3. Summary of the three-class solution of angular backscatter response classification.

Class	Sub-classes	Acoustic hardness	Geomorphic facies	Relief	Primary Substrate	Habitats
A	9, 10	High	Banks	Low-Flat	Rocky	Mixed Patches Mixed Gardens Mixed Gardens (HC)
B	4, 2, 3, 6, 7	Medium	Terraces Ridges Scraps	Flat-Low	Sandy-Rock	Mixed Patches
C	1, 5, 8, 11, 12	Low	Valleys Ridges	Flat	Muddy-Sand	Bioturbated

Table 3.3 provides a summary of a simplified three-class solution of angular backscatter response classification. This represents the results of a two stage unsupervised classification, shown as A - red, B - yellow and C - blue in Figure 3.2 together with their corresponding sub-classes.

Class A represents sub-classes 9 and 10 mostly found in banks where the substrate is dominantly (>80%) rock (Appendix A), with low or flat relief (Figure 3.4; Appendix B). The dominant habitats are mixed patches and mixed gardens for sub-class 9 and mixed gardens and mixed gardens with hard corals for sub-class 10 (Appendix C). The presence of hard corals in sub-class 10 increases the acoustic hardness and roughness of the substrate, which separates these two sub-classes. This makes sub-class 10 the highest in acoustic hardness.

Class B represents sub-classes 2, 3, 4, 6 and 7 and is generally found on terraces and scarps but also in smaller patches on ridges. This class has medium acoustic hardness due to marginally dominant hard substrate (~60%). While sub-classes 2, 3 and 4 have balanced proportions of rocky and sandy substrate, sub-classes 6 and 7 have a higher proportion of rocky substrate (>80%). The dominant habitats are mixed patches, with flat or low relief. The dendrogram (Figure 3.2) suggests that sub-class 4 is different from other sub-classes. Sub-class 4 has the highest proportion of moderate relief in comparison to all the sub-classes and is characterised by mixed gardens and mixed gardens with hard corals (~40%), where the other sub-classes are dominantly mixed patches (>90%).

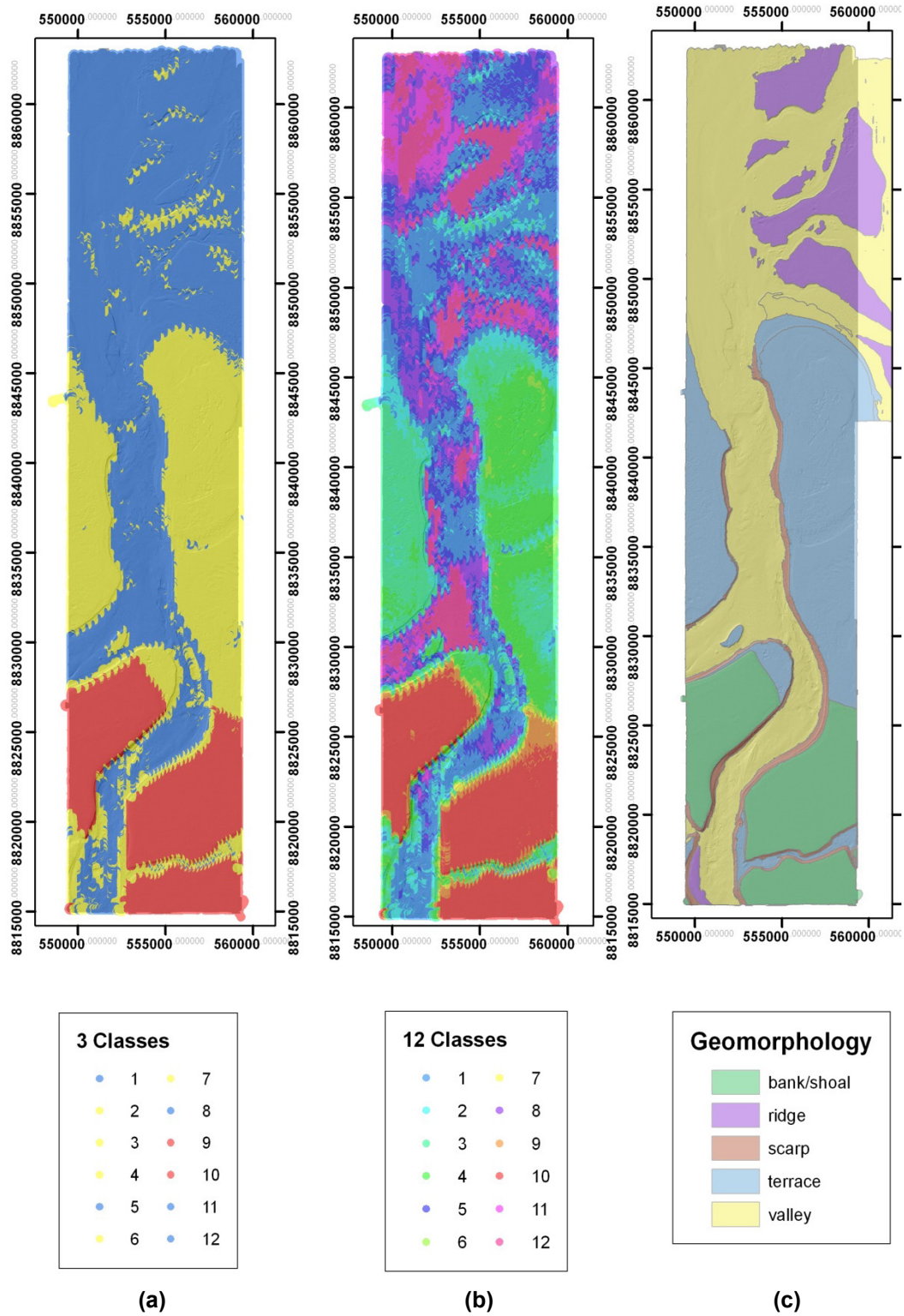


Figure 3.4: Van Diemen Rise survey area A1, showing (a) Three-class solution of angular response classification (A - red, B - yellow, C - blue); (b) Twelve-class solution, and (c). Geomorphic features adapted from Przeslawski et al. (2011).

Class C represents sub-classes 1, 5, 8, 11, 12 found dominantly throughout the ridge and valley geomorphic feature types. Sub-classes 1 and 5 are typically found on ridges and within channels where sediment cover is patchy and/or of mixed grain size (mud to sand). Sub-classes 11 and 12 are common within the valleys where sediment has accumulated. Where a continuous cover of sediment deposits occur within the survey area, pockmarks are also observed (Przeslawski *et al.*, 2011). Unlike classes A and B, class C is dominantly sand and mud with minor amounts of rocky substrate (~11%). It is also the only class with significant mud content (33.4 % overall). Class C is also dominantly flat and comparatively little relief compared to other two classes. Sub-classes 1 and 5 are more bioturbated and have less mixed patches than sub-group 2 and 3. The dendrogram ([Figure 3.2](#)) suggests that sub-class 12 is different from other sub-classes. Sub-class 12 has the lowest overall acoustic hardness of all the angular backscatter response classes associated with the highest densities and is dominantly bioturbated mud. The areas encompassed by sub-class 12 also have the highest densities of pockmarks within the survey area ([Figure 3.5](#)).

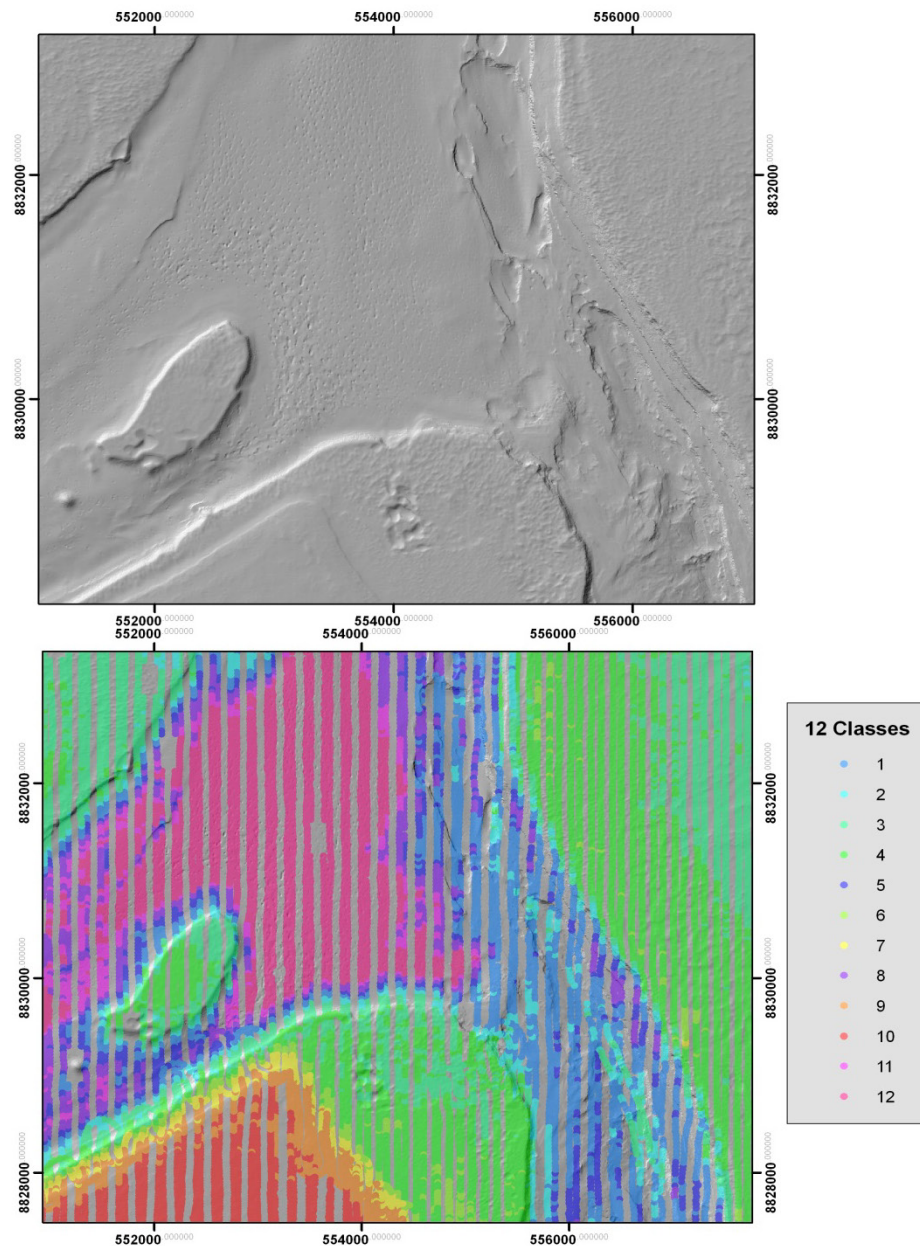


Figure 3.5: Distributions of angular response sub-classes in the vicinity of the highest density of pockmarks.

3.3.4. A refined Classification Scheme

A comparison of habitat and geomorphology from seabed video observations for each of the 12 classes derived from the unsupervised technique has led to a reduction to a possible six-class solution of seabed type. Table 3.4 and Figure 3.6 shows the resultant six-class re-classification based on substrate, relief, and habitat observations from the underwater video.

Sub-classes 10 and 4 are placed in new sub-classes based on habitats (they are the only sub-classes with significant hard-coral habitats). Sub-class 4 is the only sub-class with a significant degree of moderate relief making it unique in the study area.

Sub-classes 6, 7, and 9 are placed together into a single new sub-class. They retain high proportions of rocky habitat but only support mixed patch habitats. Spatially and statistically (based on the U-matrix), the sub-classes are transitional between sub-classes 10 and 4. Sub-class 9, while being most similar to sub-class 10 in the dendrogram (Figure 3.2), is considered more similar to sub-classes 6 and 7 based on video observations (as opposed to Euclidean distances between AR curves).

The three remaining revised sub-classes consist of sub-classes 2 and 3, 1 and 5, and 8, 11, and 12. These three revised sub-classes are structurally consistent with the original dendrogram (Figure 3.2), with original sub-classes combined based on the video observations.

Table 3.4. Re-classification of twelve sub-classes based on underwater video observations. 1HC = hard corals.

Revised Sub-Class	Old Sub-classes	Primary Substrate	Relief	Habitats/Biological Community
1	10	Rocky	Flat-Low	Mixed Gardens Mixed Gardens (HC ¹)
2	6, 7,9	Rocky	Flat-Low	Mixed Patches
3	4	Sandy-Rock	Low-Mod	Mixed Gardens (HC)
4	2, 3	Sand-Mud	Flat-Low	Mixed Patches
5	1, 5	Sand-Mud	Flat-Low	Mixed Patches Bioturbated
6	8,11,12	Mud-Sand	Flat	Bioturbated

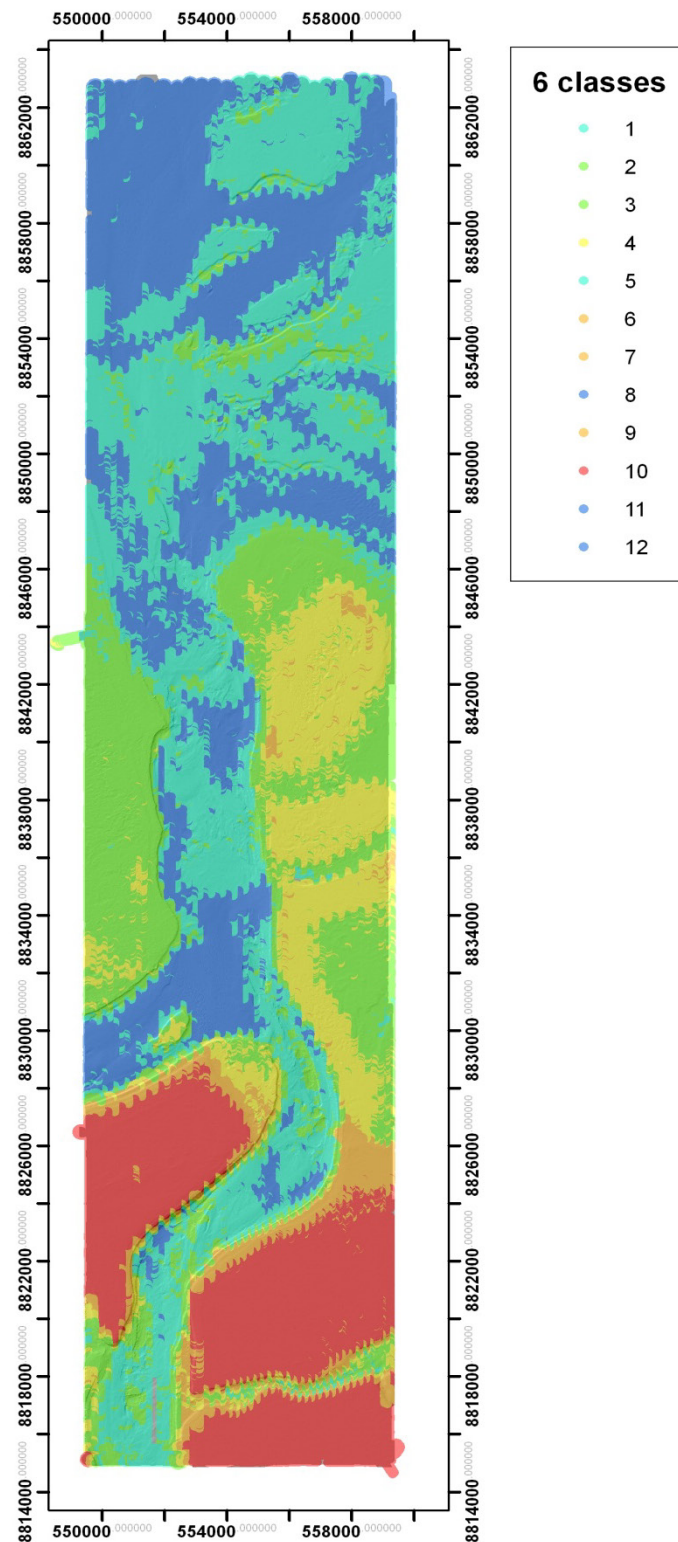


Figure 3.6: Distributions of angular response classes reclassified as per [Table 3.4](#).

3.3.5. Comparison of AR Classes, Video Observations and U-Matrix

The sub-class averages for percent primary rock, percent low or moderate relief (i.e. the sum of both), percent hard coral and percent mixed patches were remapped onto the SOM U-matrix to observe how the variables changed within a multivariate space (i.e. Euclidean distance between nodes). Percent primary rock showed a near-continuous trend across the U-matrix (Figure 3.7). This trend is consistent with the observation that average acoustic hardness within the classes is likely to increase with the presence of consolidated substrate or bedrock (i.e. substrates with elevated acoustic impedances).

The presence of mud in this study is inversely correlated with the presence of rock. This is because the presence of mud works to lower the acoustic impedance of the sediment through increased porosity and permeability, and consequently lowering sediment density.

Distribution of percent coral and relief mapped onto the U-Matrix however, shows a more complex relationship when mapped (Figures 3.8 & 3.9). Sub-classes 4 and 10 both have high percentages of coral and relief (percentage mixed gardens is also similarly correlated but not shown) but substantially different proportions of rock and sediment cover. The U-matrix shows that these sub-classes were separated by sub-classes 6, 7, and 9 and by relatively large Euclidean distances (Figures 3.8 & 3.9).

Przeslawski *et al.* (2011) consider that the terraces within the survey area were once reefs and/or carbonate banks that were sub-aerially exposed and eroded during a previous sea-level lowstand. During this time, the sub-aerially exposed carbonate would have been subject to chemical (i.e. karst) weathering, producing a comparatively rough and “pitted” surface. The presence of karst terrain in the near sub-surface would still result in a rough seabed due to the relatively thin sediment cover over much of the terrace. However, it is not possible to conclude whether the presence of mixed gardens (hard coral) are responding to the presence of a rocky substrate in the subsurface or whether roughness may be the key driver for the presence of mixed gardens (hard corals) in the study area.

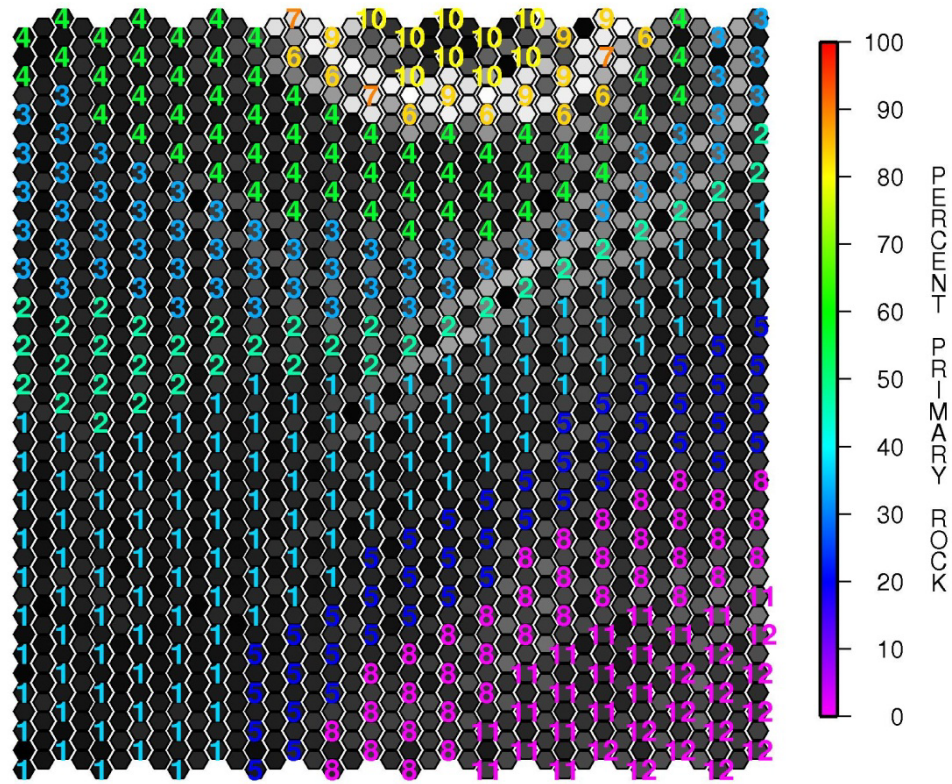


Figure 3.7: Unified Distance Matrix (U-matrix) colour coded by percentage rock (primary) from video classification.

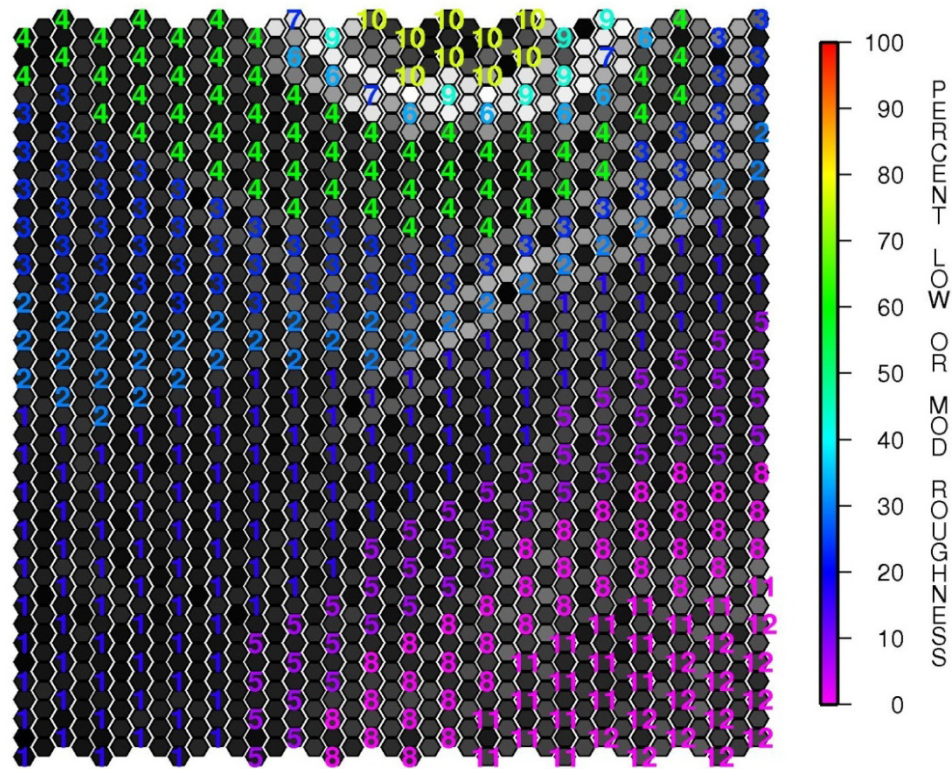


Figure 3.8: Unified Distance Matrix (U-matrix) colour coded by percentage low or moderate roughness (i.e. the sum of low or moderate roughness) from video classification.

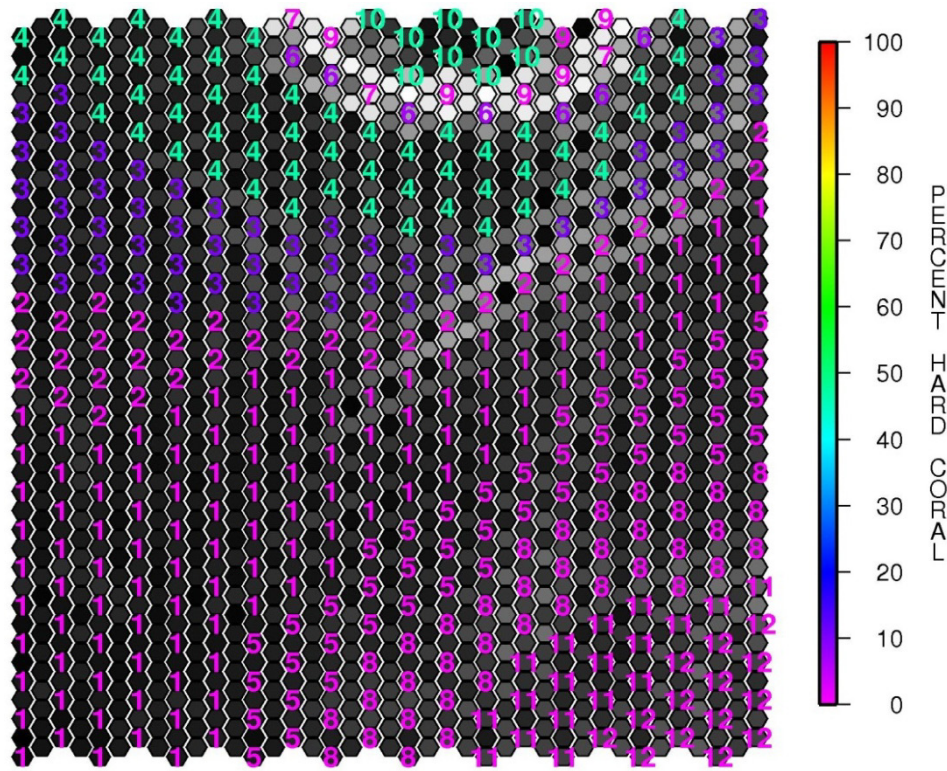


Figure 3.9: Unified Distance Matrix (U-matrix) colour coded by percentage gardens (hard coral) from video classification.

The presence of mixed patches through the survey area (Figure 3.10) is presented as a comparison to percentage hard corals (Figure 3.9). Mixed patch habitats were found adjacent to mixed garden (hard coral) habitats (sub-classes 4 and 10). The presence of mixed patches decreases away from hard coral habitats and are uncommon within sub-classes 8, 11, and 12. The driver for this may be depth, presence of mud, or potentially the presence of gas in the subsurface.

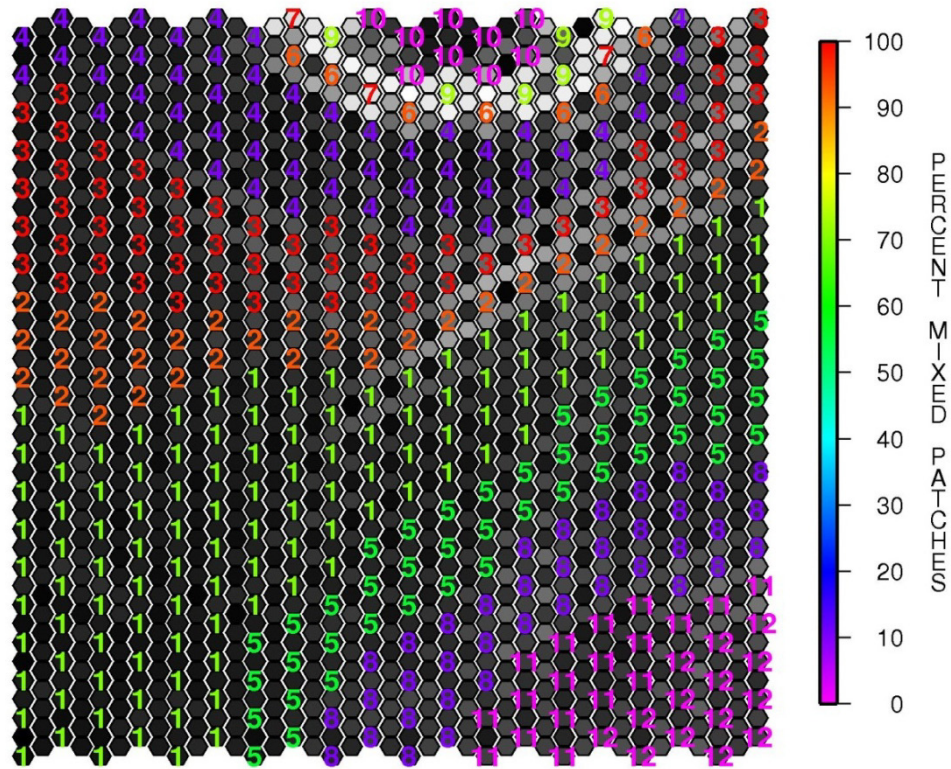


Figure 3.10: Unified Distance Matrix (U-matrix) colour coded by percentage mixed patches from video classification.

3.4. Conclusion

The combination of a SOM and hierarchical clustering appears to be a useful method to cluster multibeam sonar angular response curves. The methods provide a high degree of flexibility as it has the capacity to reduce quite large datasets (100,000's input vectors but potentially significantly more) into a summary of the input dataset. This summary can then be treated more rigorously or with numerous other statistical methods better suited to a smaller dataset (such as hierarchical clustering). The unified distance matrix provided a tool for observing cluster relationships within a multivariate space. Many of these relationships were also observable in the angular response density plot.

Consistent with the current state of knowledge on geoacoustics, the most “acoustically hard” substrates were dominated by hard substrates. The least “acoustically hard” substrates were muddy sediments with high densities of pockmarks. Interstitial gas in the sediment is likely to contribute strongly to the low reflectivity of the sediment. A surprising result from the distribution of substrates was the similarity of benthic habitats between sub-classes 4 and 10, both of which are predominantly hard coral communities; however the acoustic response of each of the sub-classes was dramatically different. It is hypothesised that the presence of hard substrates in the shallow sub-surface may assist the hard corals in sub-class 4 but further work is required to understand this relationship.

Within the data used here, it is possible to observe cluster structure using both the dendrogram and SOM U-Matrix. The dendrogram, U-matrix, and underwater video observations were useful tools for reclassifying the twelve sub-classes into a final “user-guided” seabed classification. The final classification represents six distinct facies based on seabed type, relief, and benthic habitats.

Remapping video observation data onto the U-Matrix is innovative and enabled the depiction of environmental gradients within the dataset. This was particularly useful and illustrated by the separation of hard coral habitats, which leads to an updated understanding of substrates and the capacity for similar hard coral habitats to occupy acoustically distinct substrates.

A number of future recommendations are noted here: research and develop methods to statistically determine the numbers of clusters within the datasets; compare outputs of the Geocoder software (Fonseca *et al.*, 2009) for the delineation of substrates; determine if there is a difference in species composition or morphology between habitats of sub-classes 4 and 10; apply the two stage classification approach on a more complex dataset to assess its performance; and determine if the AR classification corresponds to bathymetric or backscatter mosaic textures.

4. Random Forest Prediction Case Study

4.1. Introduction

The aim of this second case study is to delineate “hard” and “soft” substrates for the carbonate banks of the Van Diemen Rise study area using the Random Forest method. A number of predictors are used to generate a prediction map of seabed substrate type. The Random Forest method has been used for generating continuous layers from point samples which involves two essential steps (Li *et al.*, 2010). First, a model is established from the relationship between a set of explanatory variables and a target variable from samples. It supports both linear and non-linear relationships, and high-dimensional inputs. Second, the derived model is used to predict the value of the target variable at a location for which values of the explanatory variables are known.

4.2. Methods

4.2.1. Generation of Grids

After a number of corrections for vessel attitude, refraction and tide, and removal of artefacts and noise for bathymetry data (see [Section 2.1](#)), bathymetry grids of the four study areas were produced at 10 m resolution using *CarisTM HIPS & SIPS* v6.1 software. Backscatter grids were generated using *CMST-GA MB Process* v8.11.02.1 software at 10 m resolution after removal of the system transmission loss and system-implemented correction, and correction for transmission loss and insonified area to calculate the surface backscatter strength (see [Section 2.1](#)).

A common way to extract and calculate statistics and derivatives (textures) from images is to use a “kernel” filter. A kernel filter typically uses a square or rectangular matrix of values (the kernel) to calculate a new value for the centre pixel based on a function of the cells within the kernel. Kernel dimensions are often odd integer values to maintain symmetry.

A total of seven terrain and morphometric bathymetry-derived parameters and three textural backscatter-derived measures were extracted using *ArcGIS* v9.3 and *ENVI* v4.4 software, following Huang *et al.* (2011) and Daniell *et al.* (2007) ([Table 4.1](#)). The seven bathymetry-derived variables include slope, relief, surface area, topographic (benthic) position index (TPI), planar curvature, profile curvature and local Moran I as described in Lundblad *et al.* (2006), Lanier *et al.* (2007), Wilson *et al.* (2007), Erdey-Heydorn (2008), Holmes *et al.* (2008), Dunn & Halpin (2009), Verfaillie *et al.* (2009), and Zieger *et al.* (2009). The three backscatter-derived parameters are homogeneity, variance and local Moran I as discussed in Haralick *et al.* (1973), Cochrane & Lafferty (2002), and Lucieer (2008). Daniell *et al.* (2007) provides a summary as to how bathymetric derivatives are calculated.

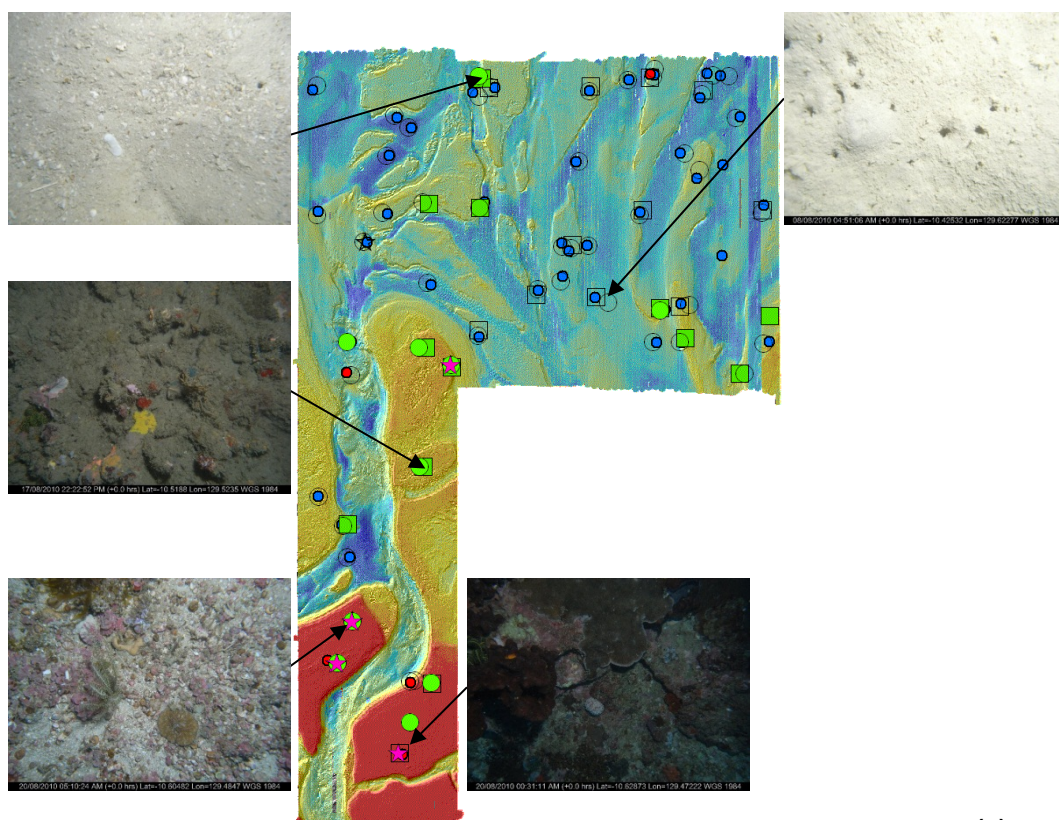
Daniell *et al.* (2007) examined few different kernel sizes using the correlation coefficients and finally adopted a 7x7 kernel for their study. A 7x7 kernel was also adopted here in this study. This 7x7 kernel creates a spatial scale of 70x70 m for a 10 m grid resolution.

Table 4.1. Description of derived parameters from multibeam data presented in [Figures 4.14 to 4.23](#).

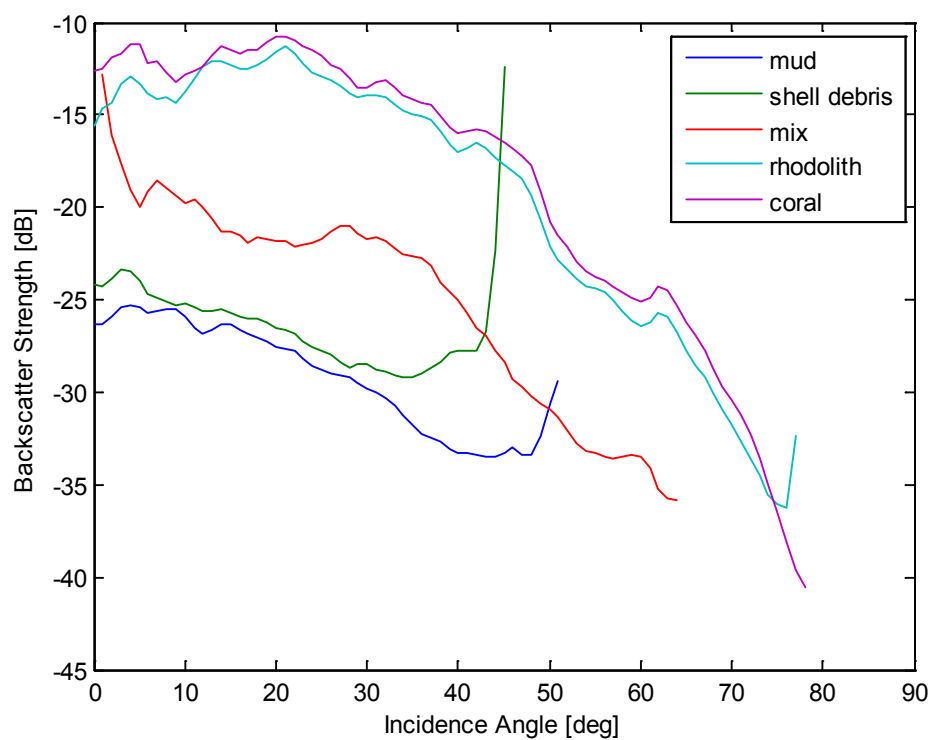
Parameter	Description
<i>Bathymetry-derived parameter:</i>	
Slope	Slope gradient (first derivative of bathymetry).
Relief	Topographic relief.
Surface area	Ratio of the cell surface area grid and its planimetric area (i.e. the square of the grid cell size).
Topographic position index (TPI)	Difference between a cell elevation and the average of the elevation values in the surrounding cells. Positive values mean the cell is higher than its surroundings while negative values means it is lower.
Planar curvature	The curvature of the surface perpendicular to the slope direction (second derivative of bathymetry).
Profile curvature	The curvature of the surface in the direction of slope (second derivative of bathymetry).
Local Moran I	Measure of global spatial autocorrelation which is characterized by a correlation in data among nearby locations in space.
<i>Backscatter-derived parameter:</i>	
Homogeneity	Measure of closeness of the distribution of elements in the Gray-Level Co-occurrence Matrix (GLCM) to the GLCM diagonal.
Variance	Measure of the dispersion of the values around the mean within the GLCM.
Local Moran I	Measure of global spatial autocorrelation which is characterized by a correlation in data among nearby locations in space.

4.2.2. Angular Backscatter Response (AR) p-Value

Video-derived classes were used to locate the hard-bottomed seabed areas for which the angular backscatter response curves were extracted ([Figure 4.1](#)). Of those response curves, the lowest angular backscatter response curve was associated with the “mixed patches/mixed gardens” class and the highest response curve with the “mixed gardens (hard coral)” class. Of these lowest angular backscatter response curves, one for each study area was extracted and compared to all other angular backscatter response curves within its corresponding area for incidence angles between 0° and 60° using the Kolmogorov-Smirnov goodness of fit to estimate the probability (*p*-value) of hard bottom. Finally, the IDW interpolation technique was used to produce a continuous layer of the *p*-value of hard bottom for each study area.



(a)



(b)

Figure 4.1: (a) False colour backscatter mosaic of Area A showing video-derived classes at sampling stations and biology categories with representative photographs of different seabed types; (b) Angular backscatter response curves associated with different bottom types.

4.2.3. Exploratory Analysis

To validate the predictability of the Random Forest model, a 10-fold cross validation was adopted and the correct classification rate (Fielding & Bell, 1997; Li & Hilbert, 2008) and kappa (Cohen, 1960) were used to measure the accuracy of the model developed. To reduce the random error associated with each 10-fold cross validation, the cross validation procedure was repeated 100 times. The final results were based on the average of the 100 cross validations. All the data modelling work was implemented in R (R Development Core Team, 2011).

Given that the hardness of seabed is measured in terms of “hard” and “soft”, the response variable is a categorical variable and needed to be transformed into a continuous variable. To explore the relationships of hardness and the predictive variables, the hardness variable was converted into the numeric variables 0 (“soft”) and 1 (“hard”).

4.3. Results and Discussion

4.3.1. Bathymetry, Backscatter and Derivative Grids

Figures 4.2 to 4.5 show the bathymetry of the four study areas (Areas A - D). Areas A and C presented wider depth range and more complex geomorphology than Areas B and D. Geomorphic features identified in the four study areas included banks, terraces, valleys and plains. Geomorphic units superimposed on the identified geomorphic features included floors, scraps, ridges, flanks, platforms, depressions and crests. Areas A and C comprised more complex morphology compared to Areas B and D and included the maximum number of geomorphic features/units. Pockmarks were identified on the soft sediment sections of valley in Areas A and D and also within bedform fields (Area A).

Backscatter values varied between the study areas and the associated geomorphic features (Figures 4.2 to 4.9). Histograms of the backscatter values reveal distinct differences between each of the areas surveyed (Figure 4.10). Areas A and C had the greatest range in backscatter values whereas Areas B and D had narrower ranges of backscatter values, indicating a smaller variation in substratum types across these areas (Figure 4.10). Comparisons between backscatter values were dependant on the geomorphic feature types present. The shift in backscatter values for each geomorphic feature suggests that there is higher backscatter association with bank/shoal features in backscatter response (Figure 4.11). Bank environments had the highest backscatter values (mean -14.38 dB; range -22.16 to -5.63 dB from sampling locations), which along with video observations indicated areas of hard substratum (Figure 4.11). In contrast, valleys had the widest range of backscatter values (mean -26.39 dB; range -46.79 to -16.76 dB from sampling locations) representing a mix of soft and hard substrata (Figure 4.11).

When grouping the backscatter areas according to video categories and benthic habitat type, mixed gardens (HC) had the highest backscatter values with a mean of -14.24 dB and a range from -21.97 to -5.63 dB from sampling locations whereas bioturbated seabed habitat had the lowest value and widest range of backscatter with a mean of -24.89 dB and a range from -46.79 to -12.03 dB from sampling locations (Figure 4.12). With regard to the specific grouping of benthic habitats into “soft” and “hard” substratum, we correctly identified that high backscatter values with the video-derived bottom type were “hard” whereas low backscatter values were associated with the video-derived bottom type “soft” (Figure 4.13). At all sampling stations, 84% of the video-derived “hard” bottom were associated with %gravel>10% and backscatter values mostly greater than -22 dB (84%) and median grain size of

365 μm . For the video-derived “soft” bottom at sampling locations, 72% were associated with %gravel<10% and backscatter values mostly less than -22 dB (70%) and smaller median grain size (155 μm).

Results of these bathymetry and backscatter derivatives of slope, relief, surface area, TPI, planar and profile curvature, homogeneity, variance and local moran I of backscatter and bathymetry using a 7×7 kernel size are shown in [Figures 4.14 to 4.23](#).

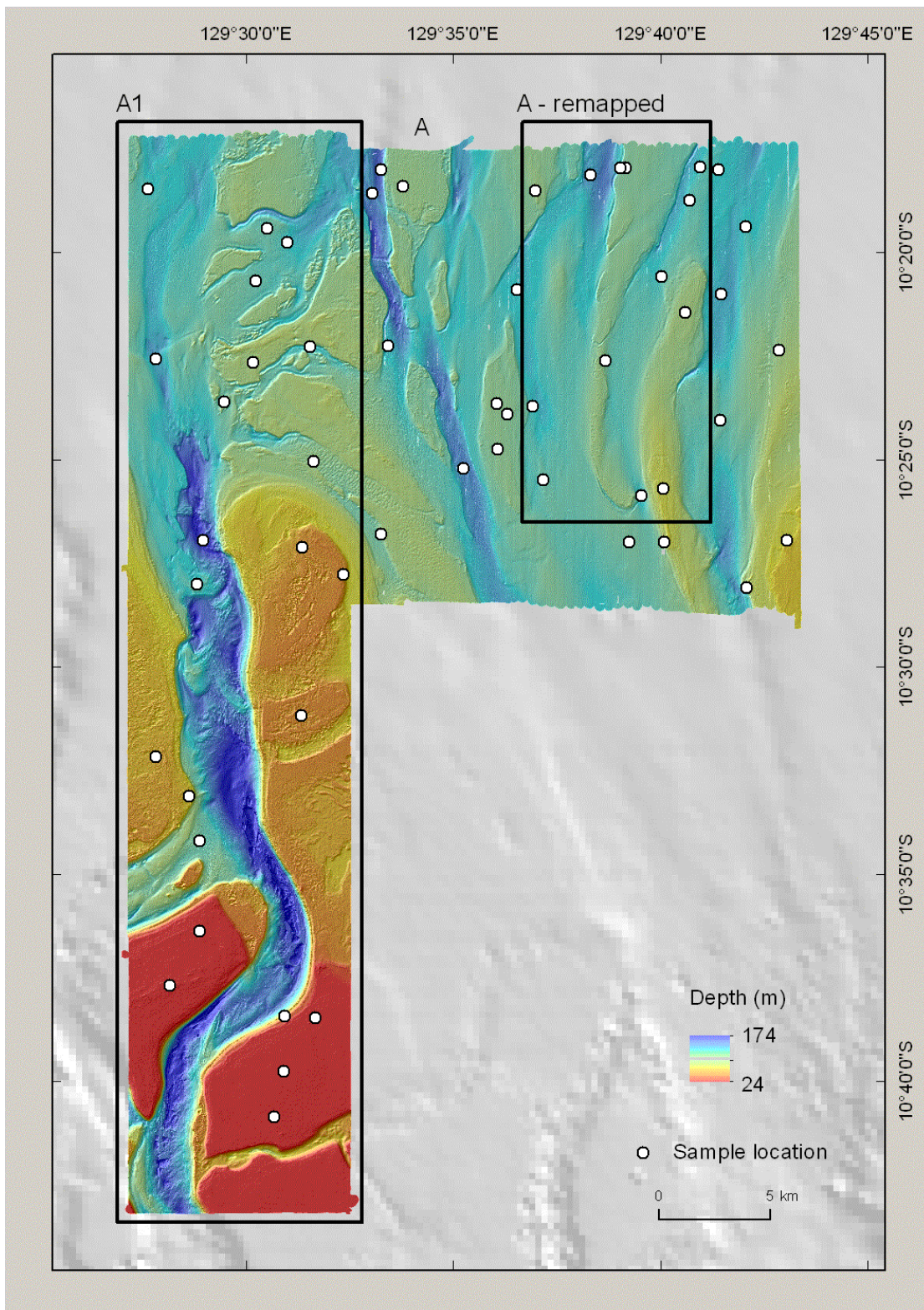


Figure 4.2: False colour bathymetry image of Area A with sampling stations indicated. All areas inside the black boxes were mapped during survey SOL5117, with remaining areas mapped during survey SOL4934.

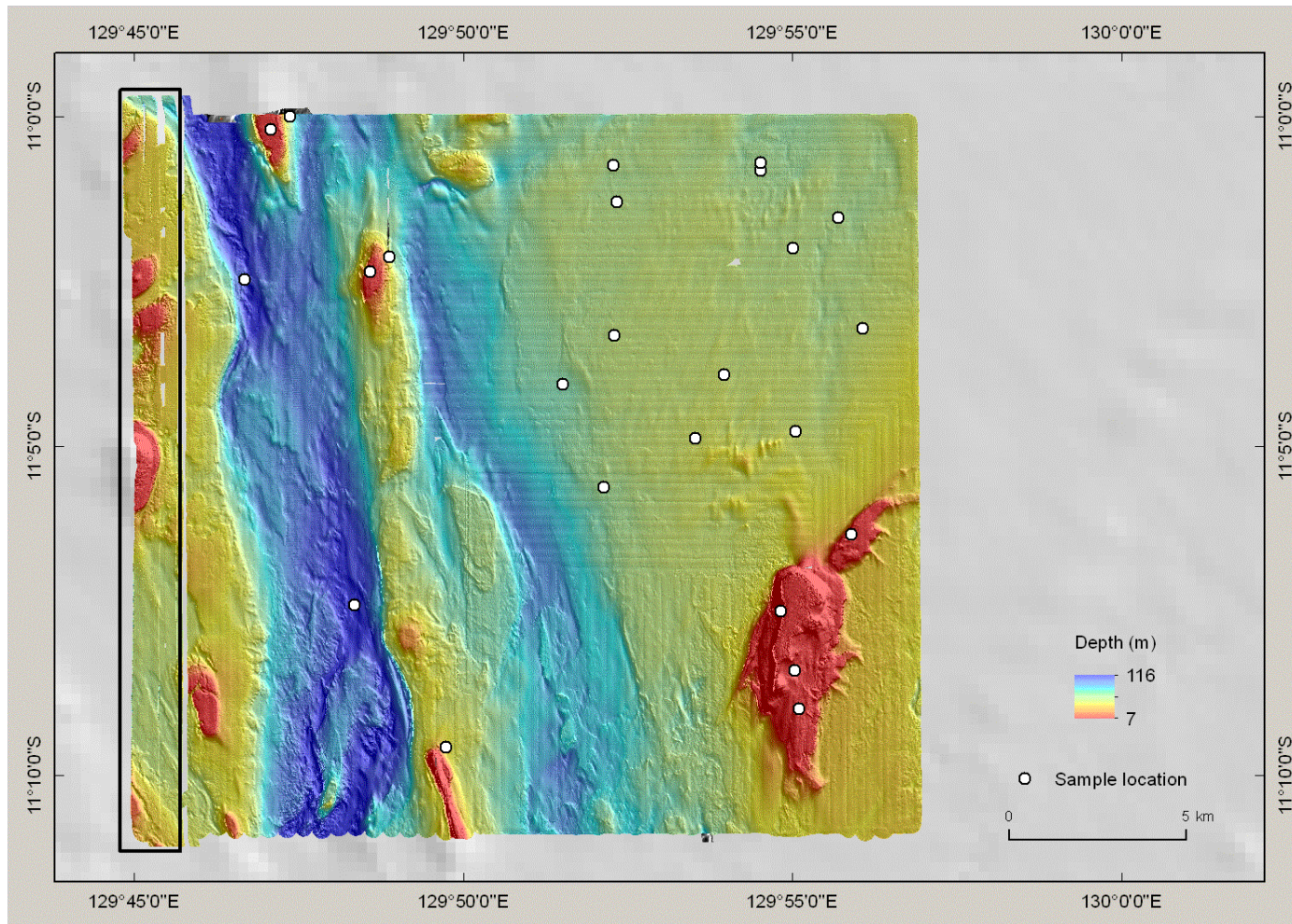


Figure 4.3: False colour bathymetry image of Area B with sampling stations indicated. All areas inside the black boxes mapped during survey SOL5117 and all others during survey SOL4934.

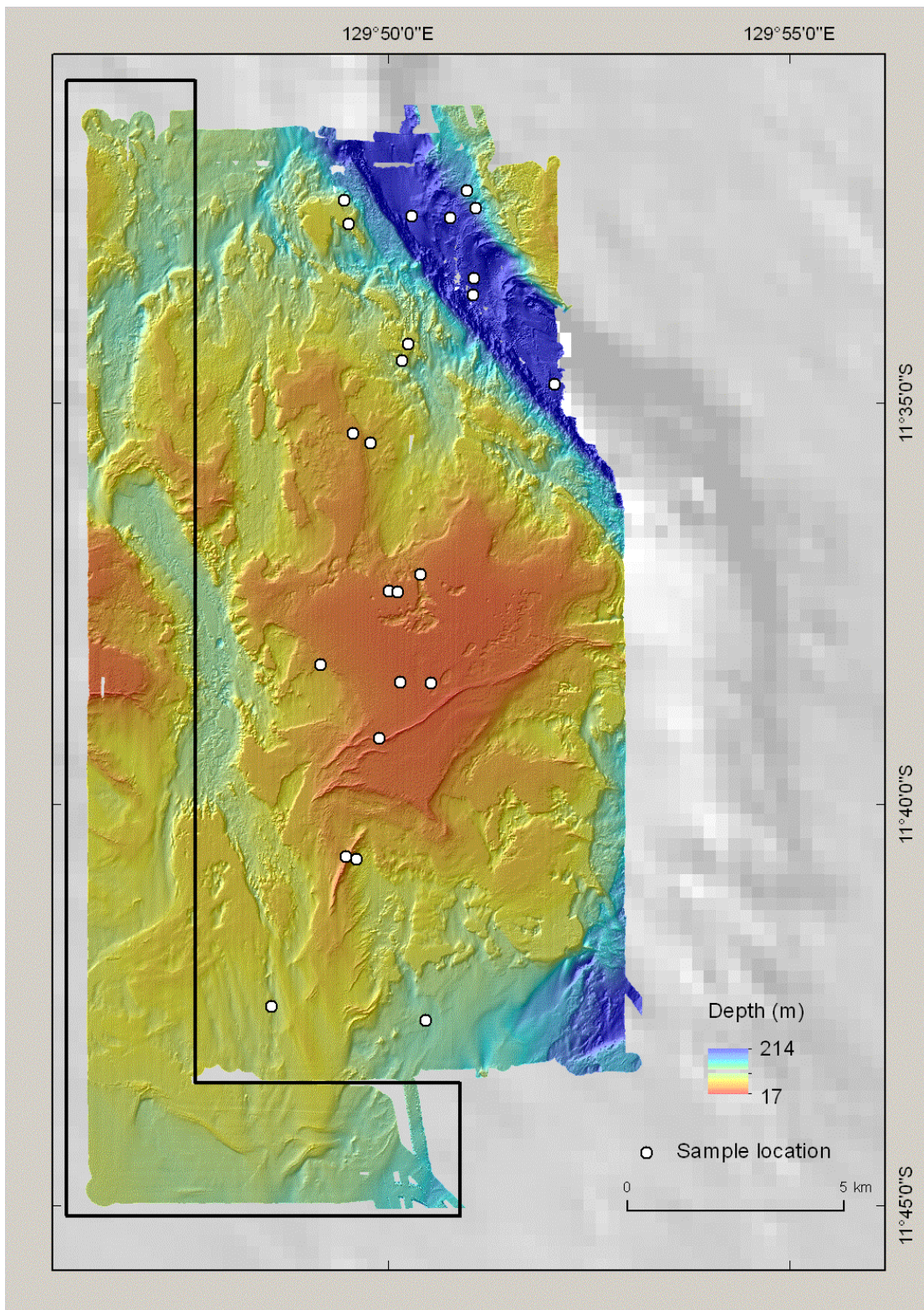


Figure 4.4: False colour bathymetry image of Area C with sampling stations indicated. All areas inside the black boxes mapped during survey SOL5117 and all others during survey SOL4934.

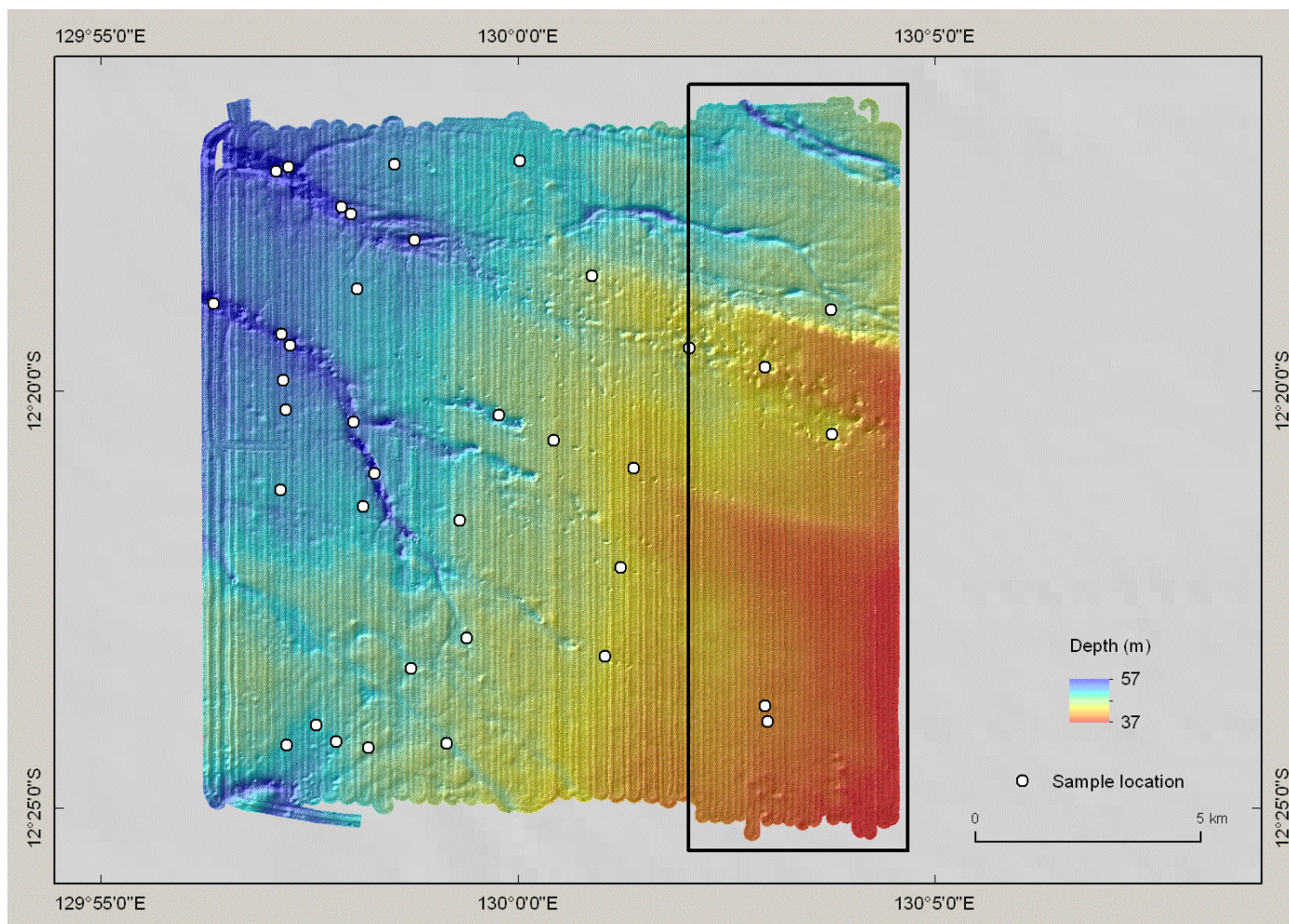


Figure 4.5: False colour bathymetry image of Area D with sampling stations indicated. All areas inside the black boxes mapped during survey SOL5117 and all others during survey SOL4934.

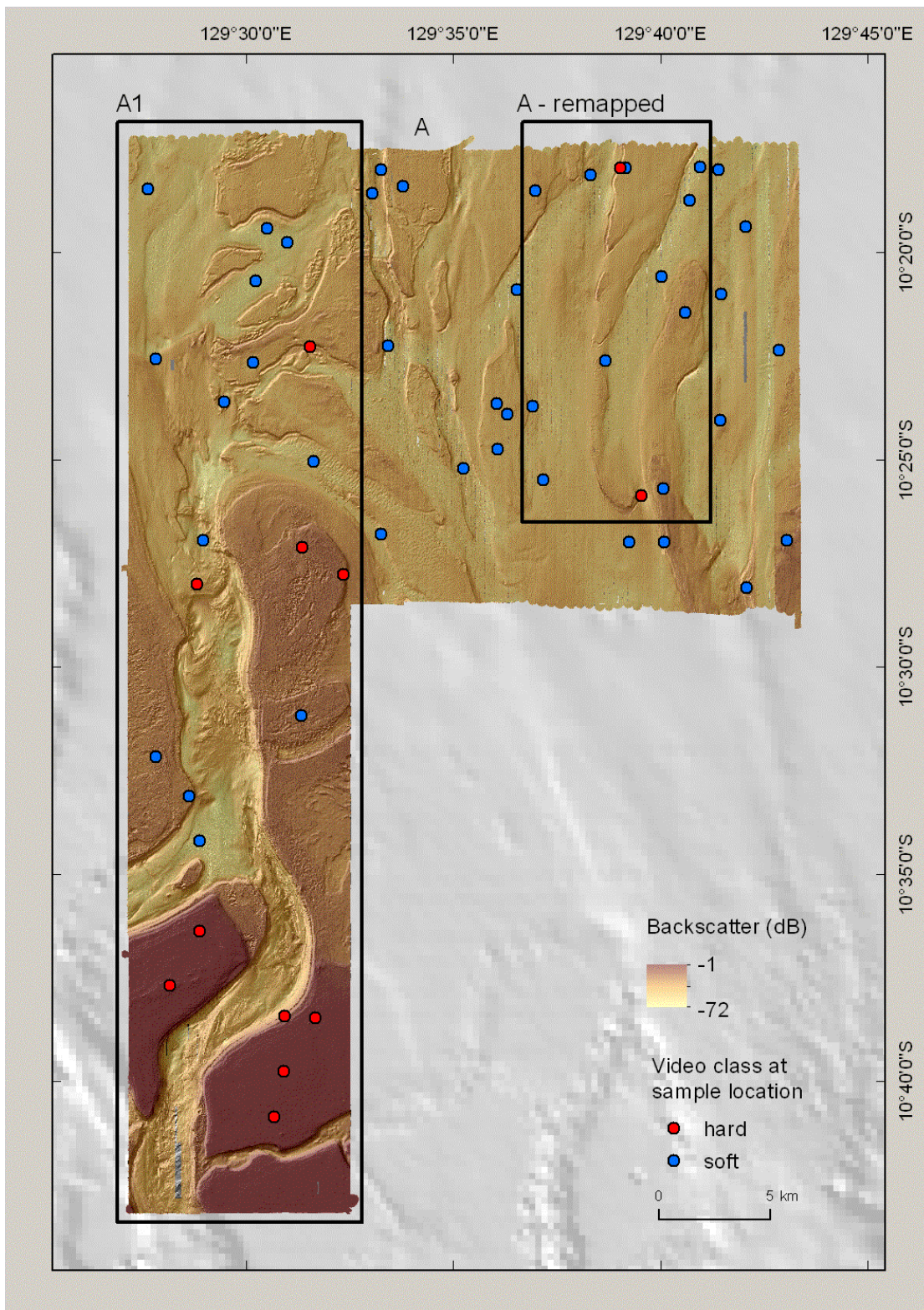


Figure 4.6: False colour backscatter image of Area A, showing video-derived classes at sampling stations. All areas inside the black boxes mapped during survey SOL5117 and all others during survey SOL4934.

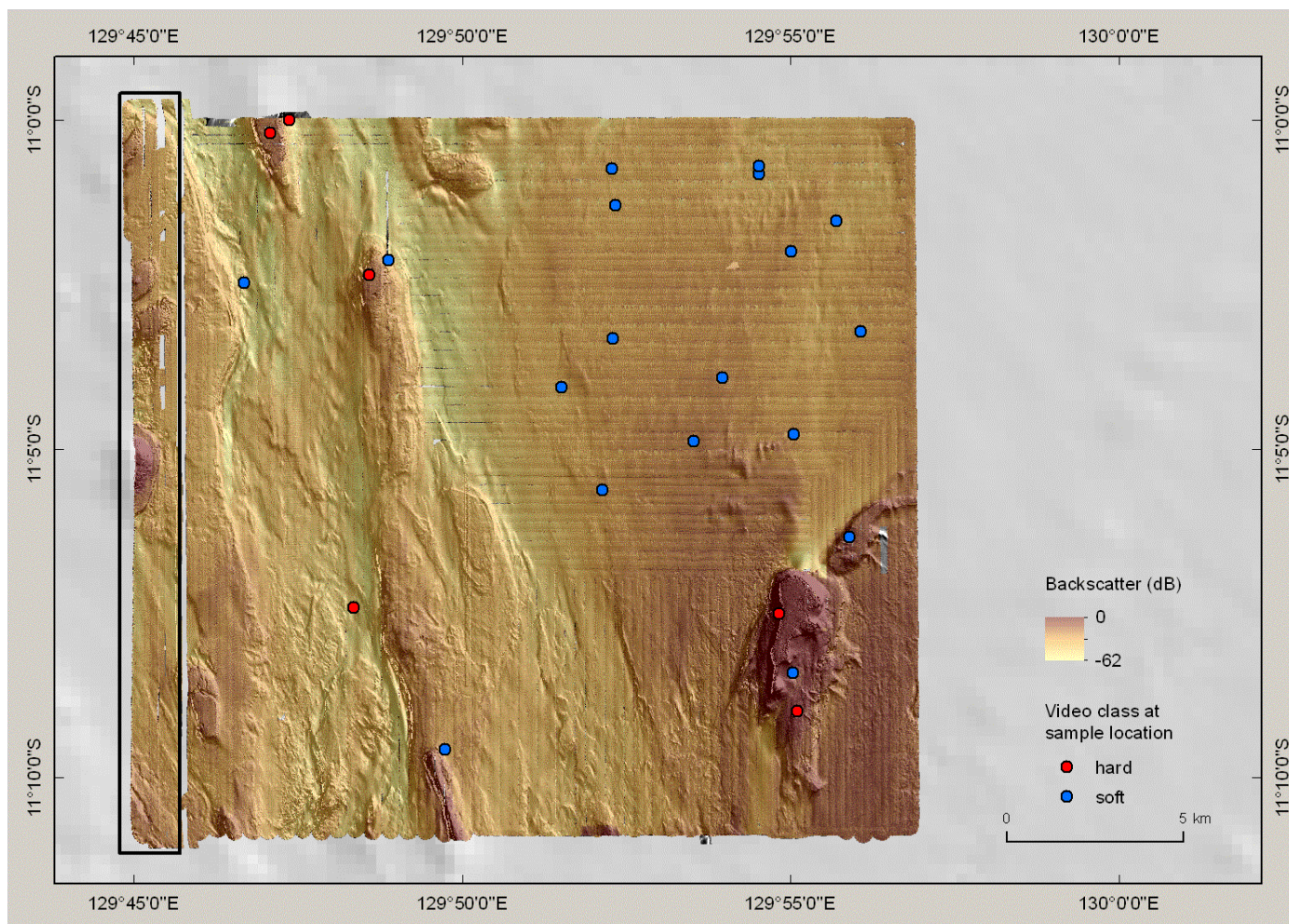


Figure 4.7: False colour backscatter image of Area B, showing video-derived classes at sampling stations. All areas inside the black boxes mapped during survey SOL5117 and all others during survey SOL4934.

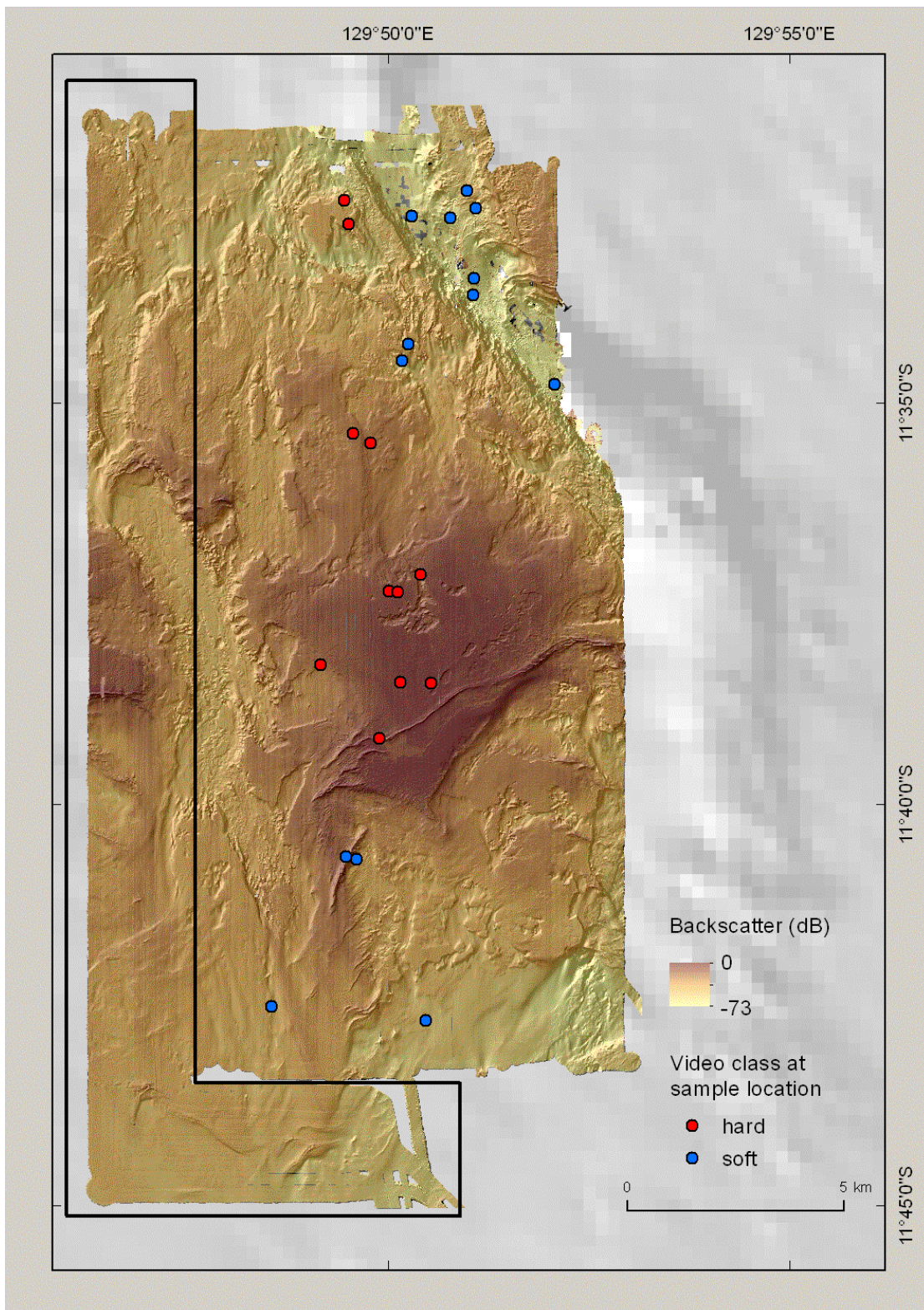


Figure 4.8: False colour backscatter image of Area C, showing video-derived classes at sampling stations. All areas inside the black boxes mapped during survey SOL5117 and all others during survey SOL4934.

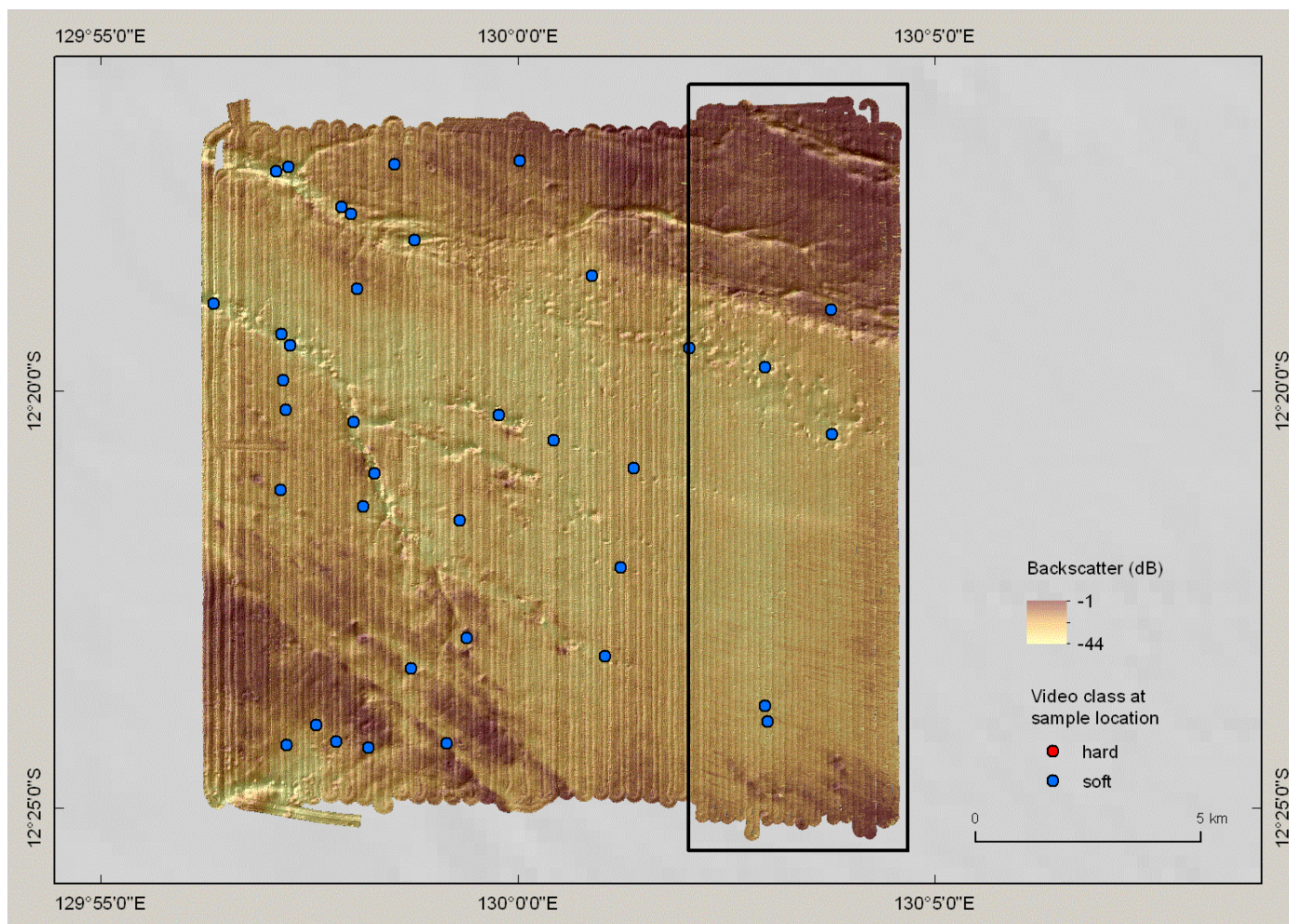


Figure 4.9: False colour backscatter image of Area D, showing video-derived classes at sampling stations. All areas inside the black boxes mapped during survey SOL5117 and all others during survey SOL4934.

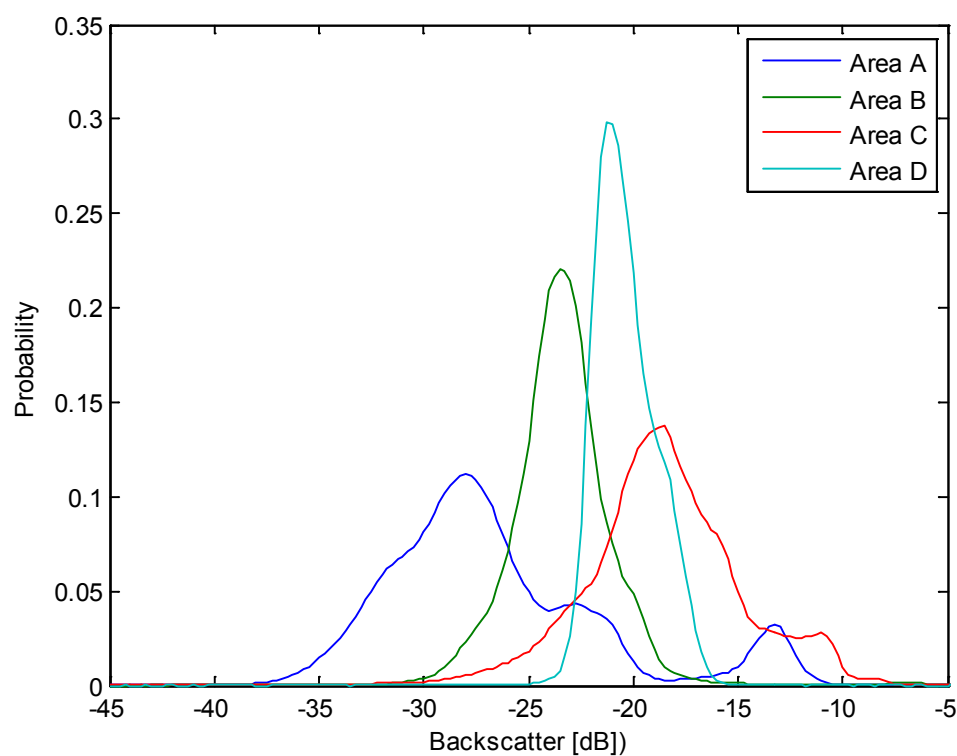


Figure 4.10: Histograms of backscatter values for the four study areas.

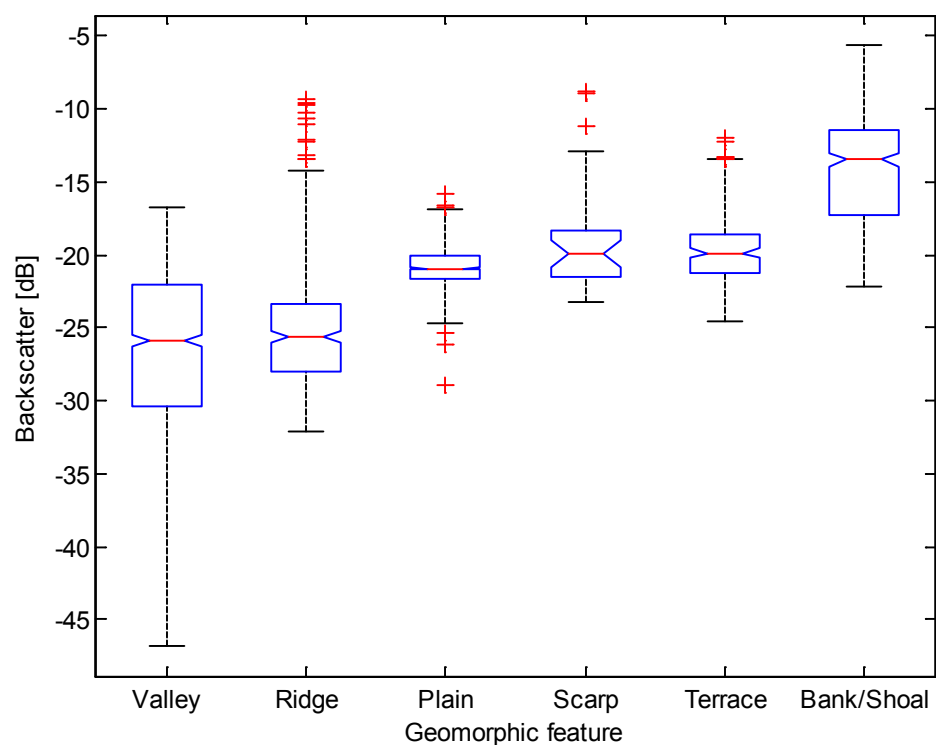


Figure 4.11: Boxplots of backscatter values grouped by geomorphic feature type.

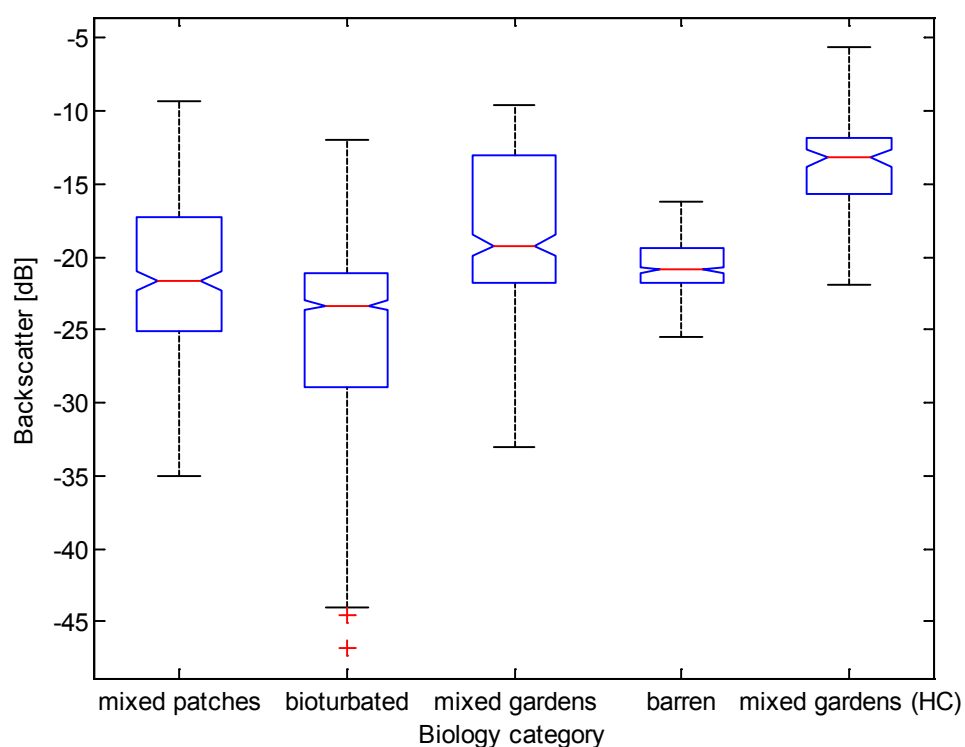


Figure 4.12: Boxplots of backscatter values grouped into video-derived biology category.

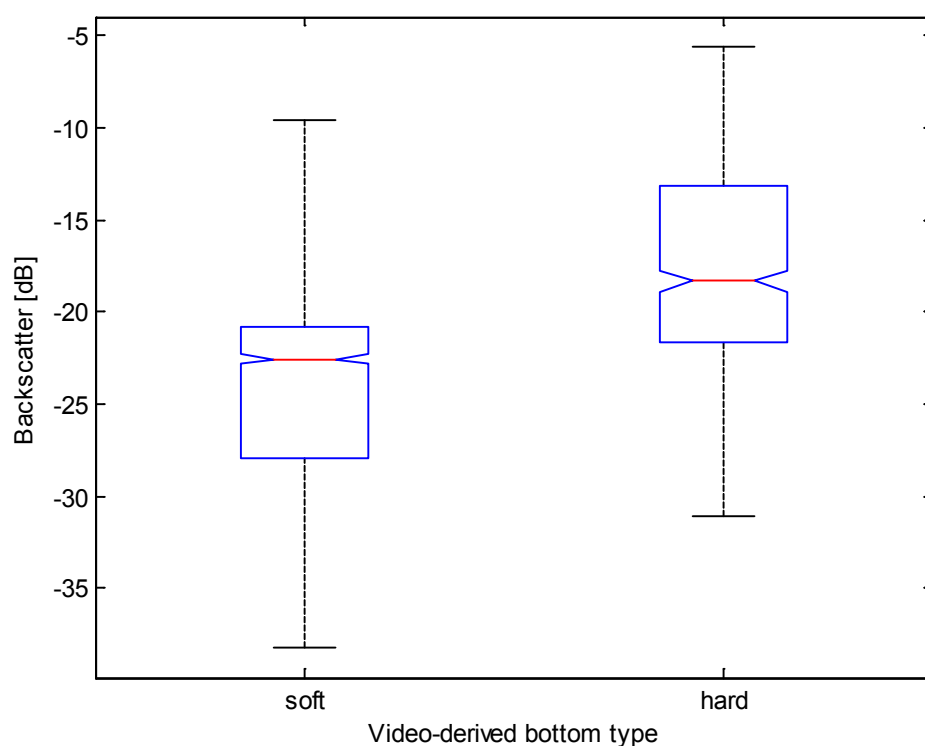


Figure 4.13: Boxplots of backscatter values grouped into video-derived seabed type.

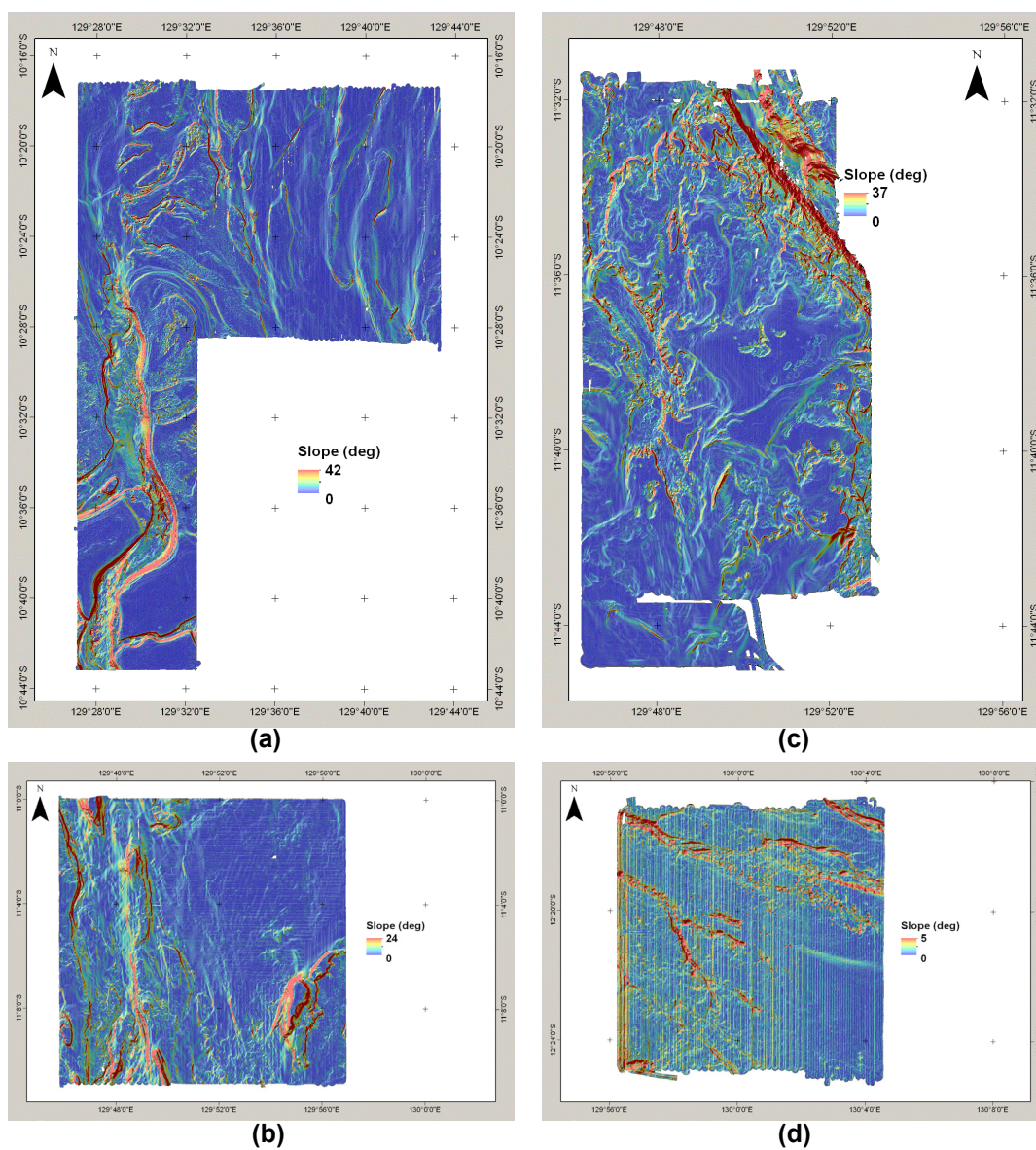


Figure 4.14: Seabed slope maps of the four study areas, Van Diemen Rise.

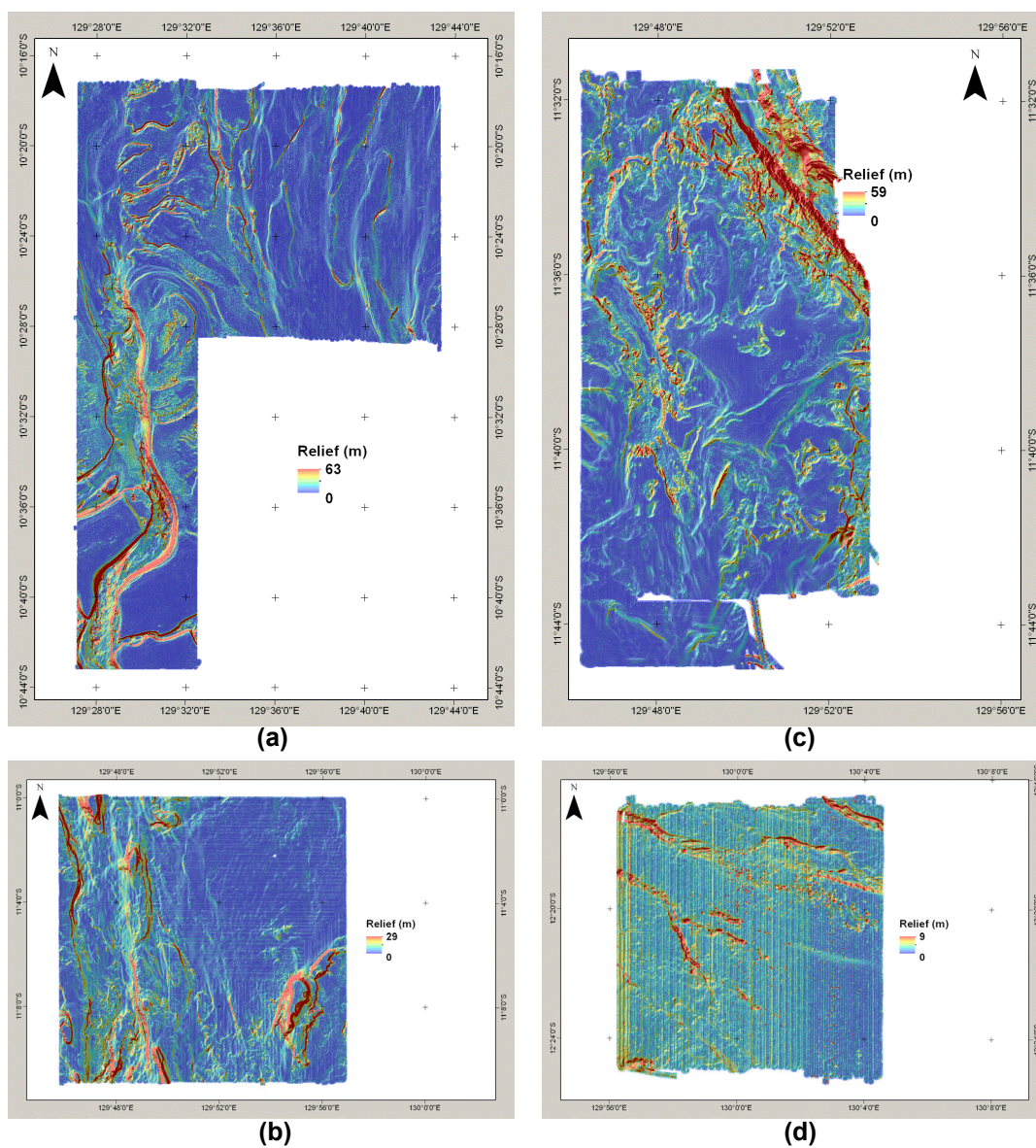


Figure 4.15: Seabed relief maps of the four study areas, Van Diemen Rise.

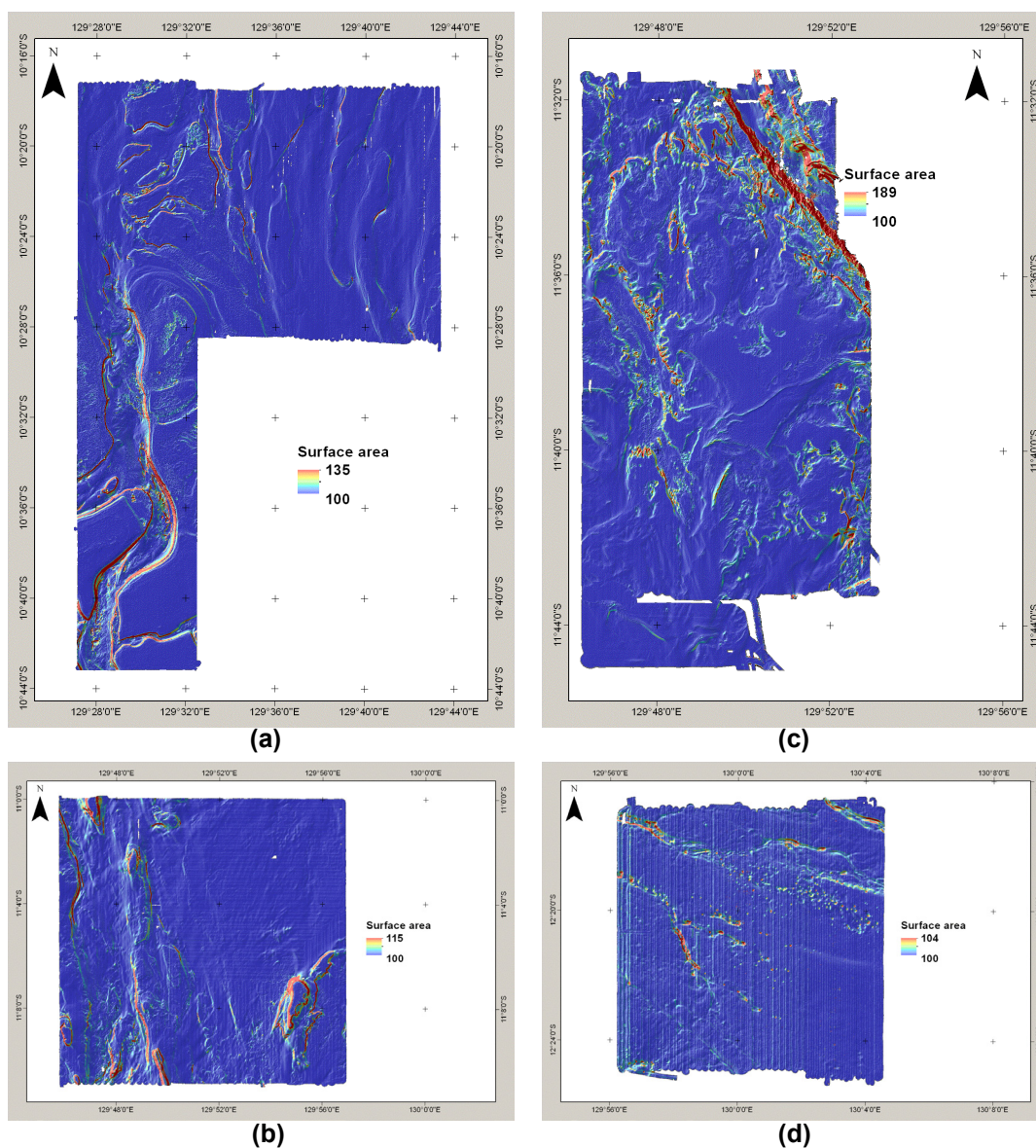


Figure 4.16: Surface area maps of the four study area, Van Diemen Rise.

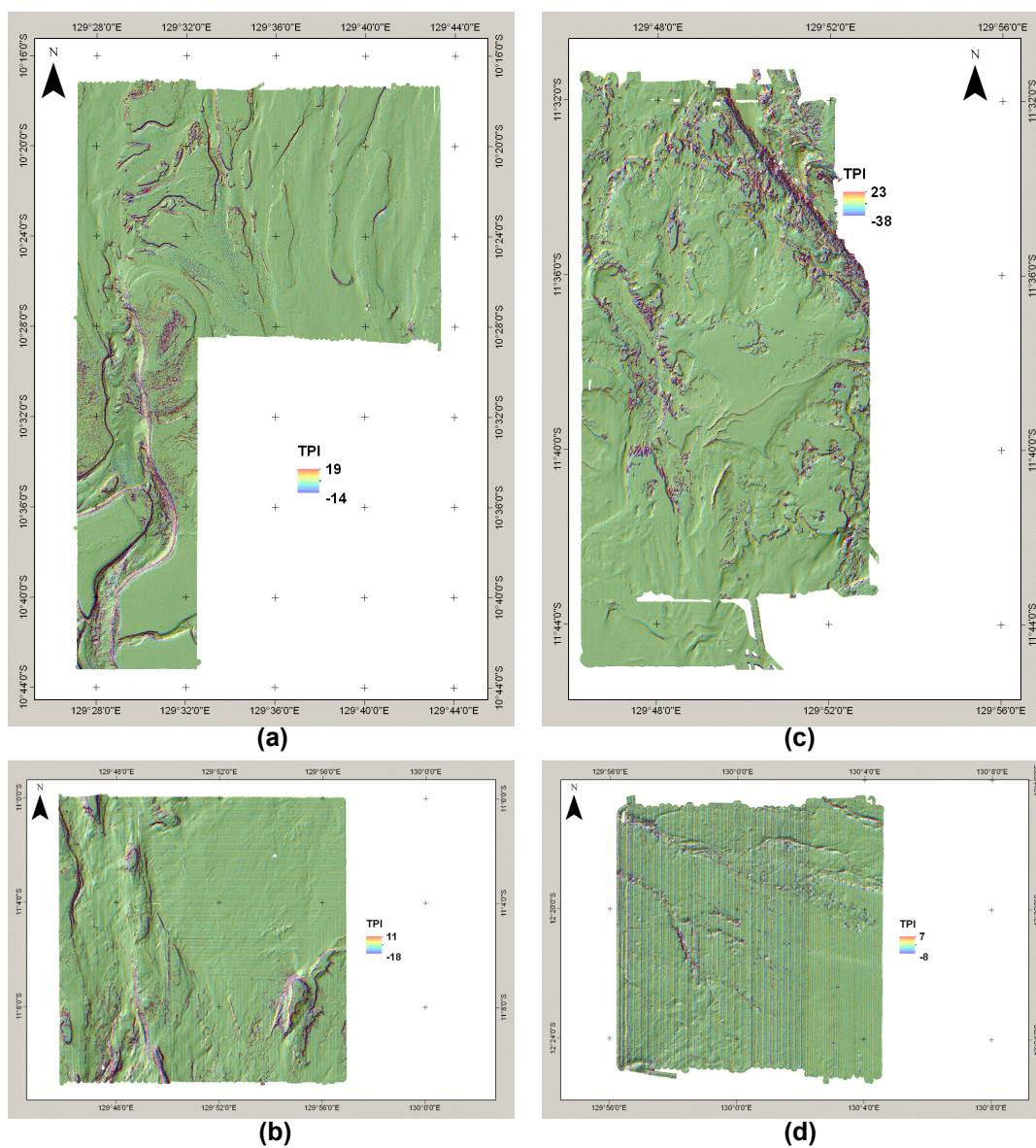


Figure 4.17: Topographic Position Index maps of the four study areas, Van Diemen Rise.

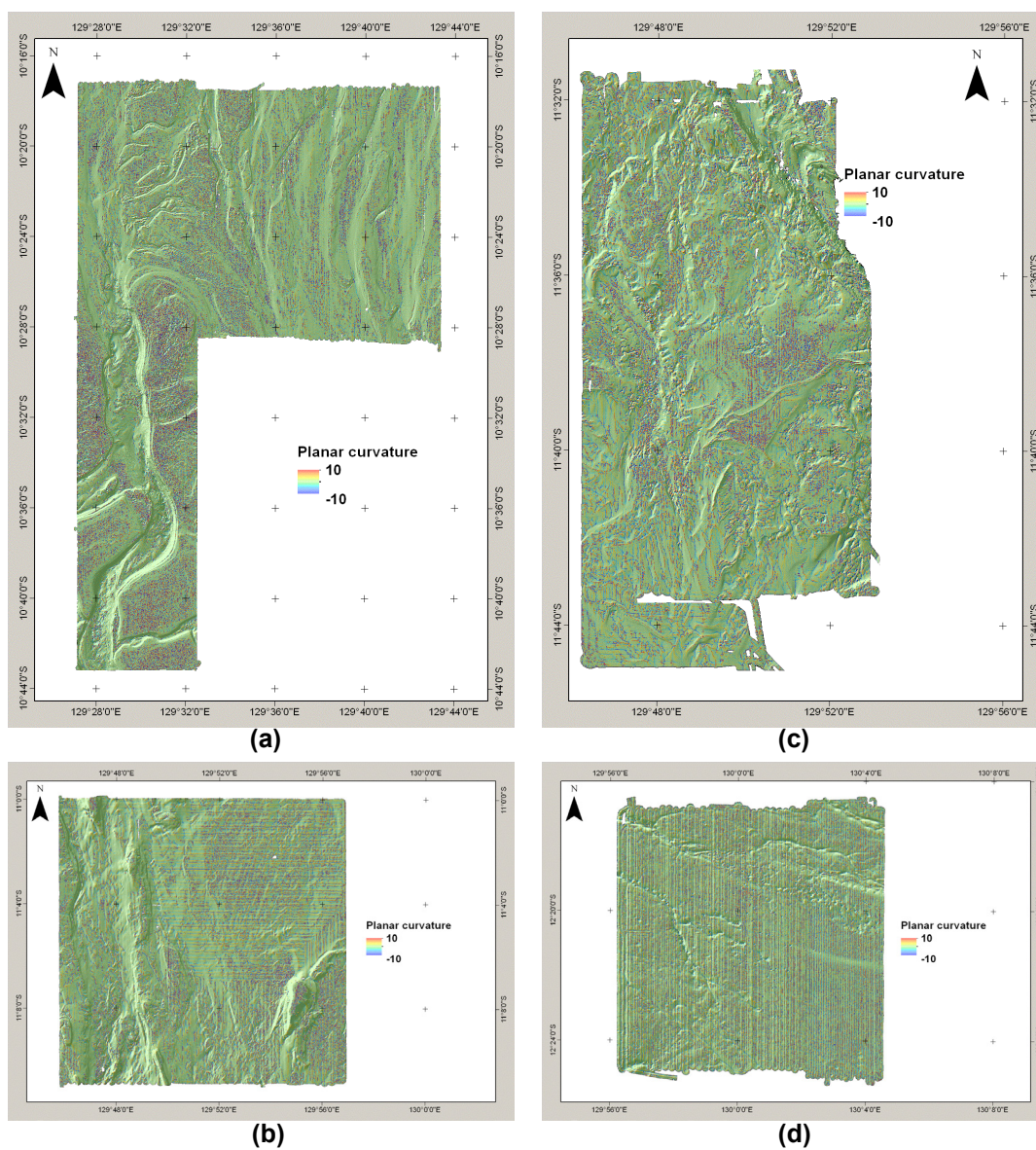


Figure 4.18: Planar curvature maps of the four study areas, Van Diemen Rise.

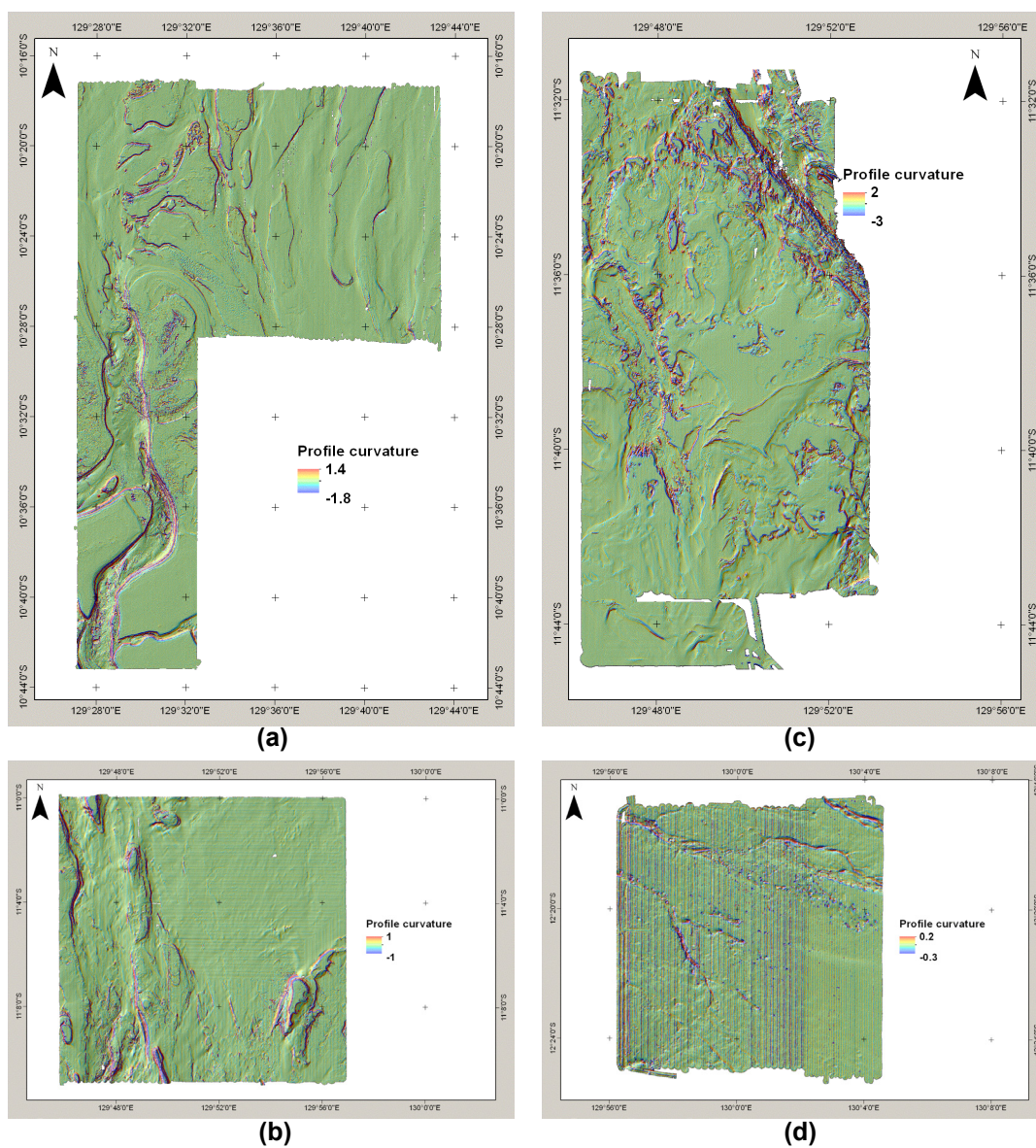


Figure 4.19: Profile curvature maps of the four study areas, Van Diemen Rise.

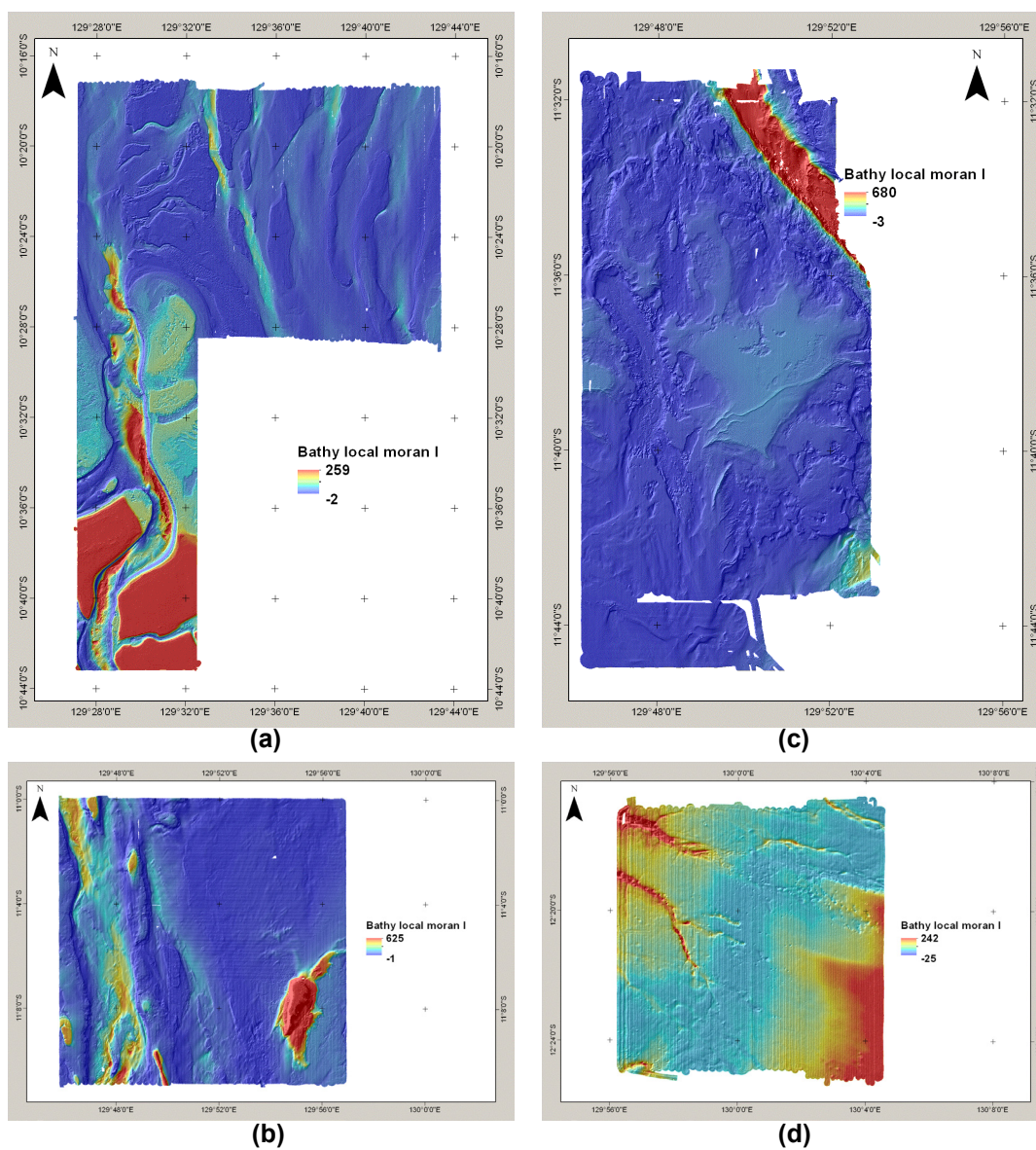


Figure 4.20: Bathymetry Local Moran I of the four study areas, Van Diemen Rise.

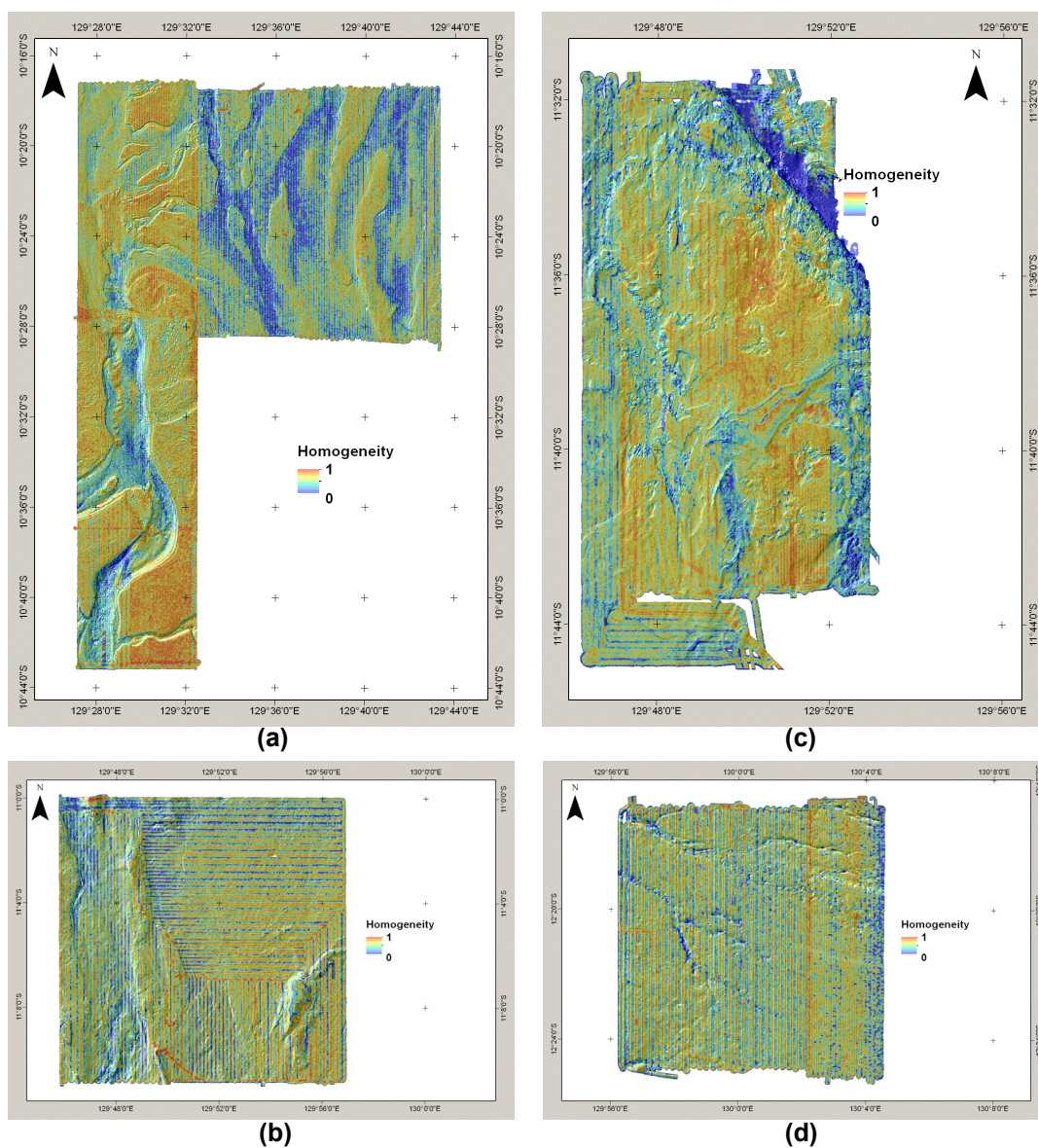


Figure 4.21: Backscatter homogeneity maps of the four study areas, Van Diemen Rise.

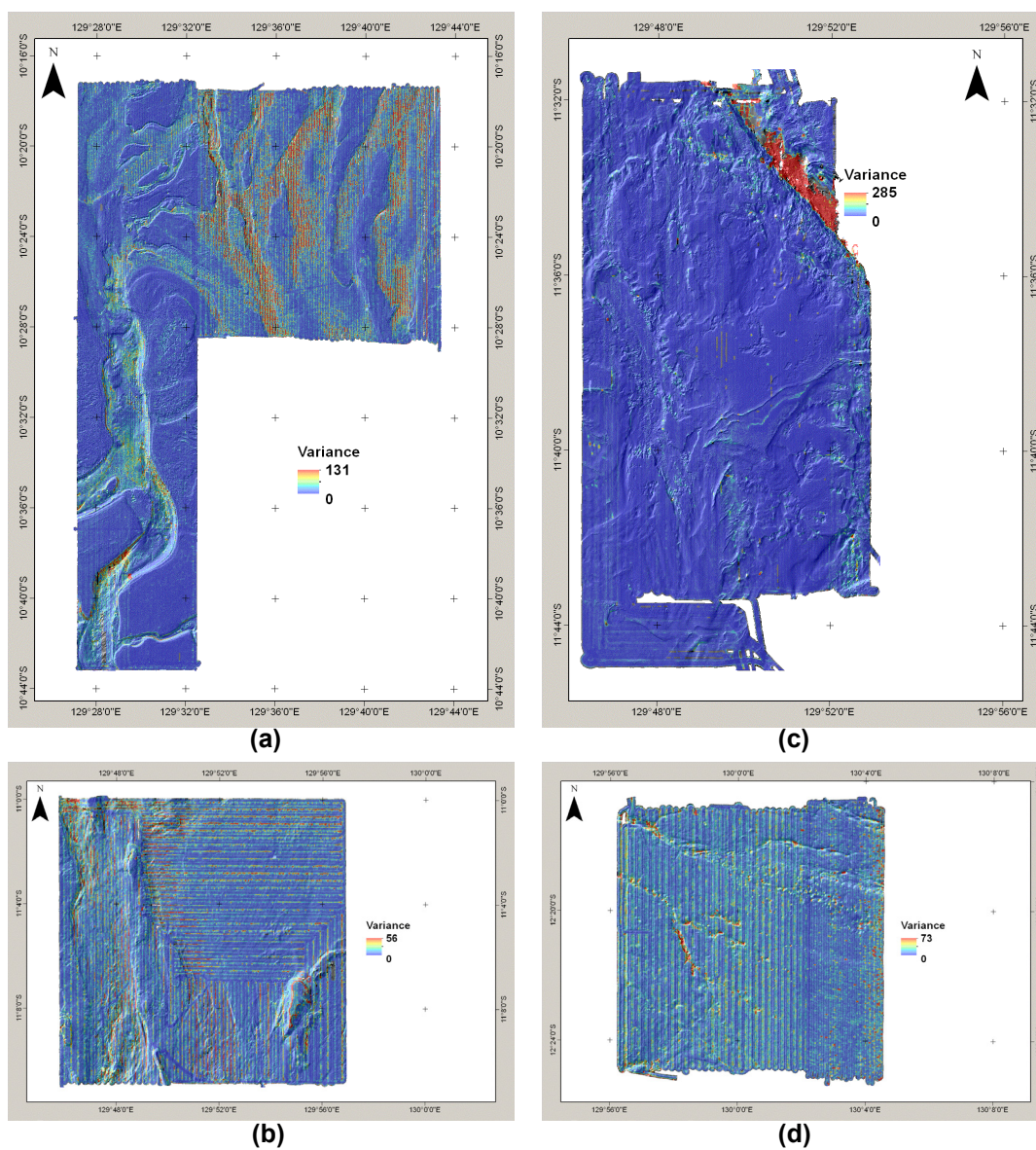


Figure 4.22: Backscatter variance maps of the four study areas, Van Diemen Rise.

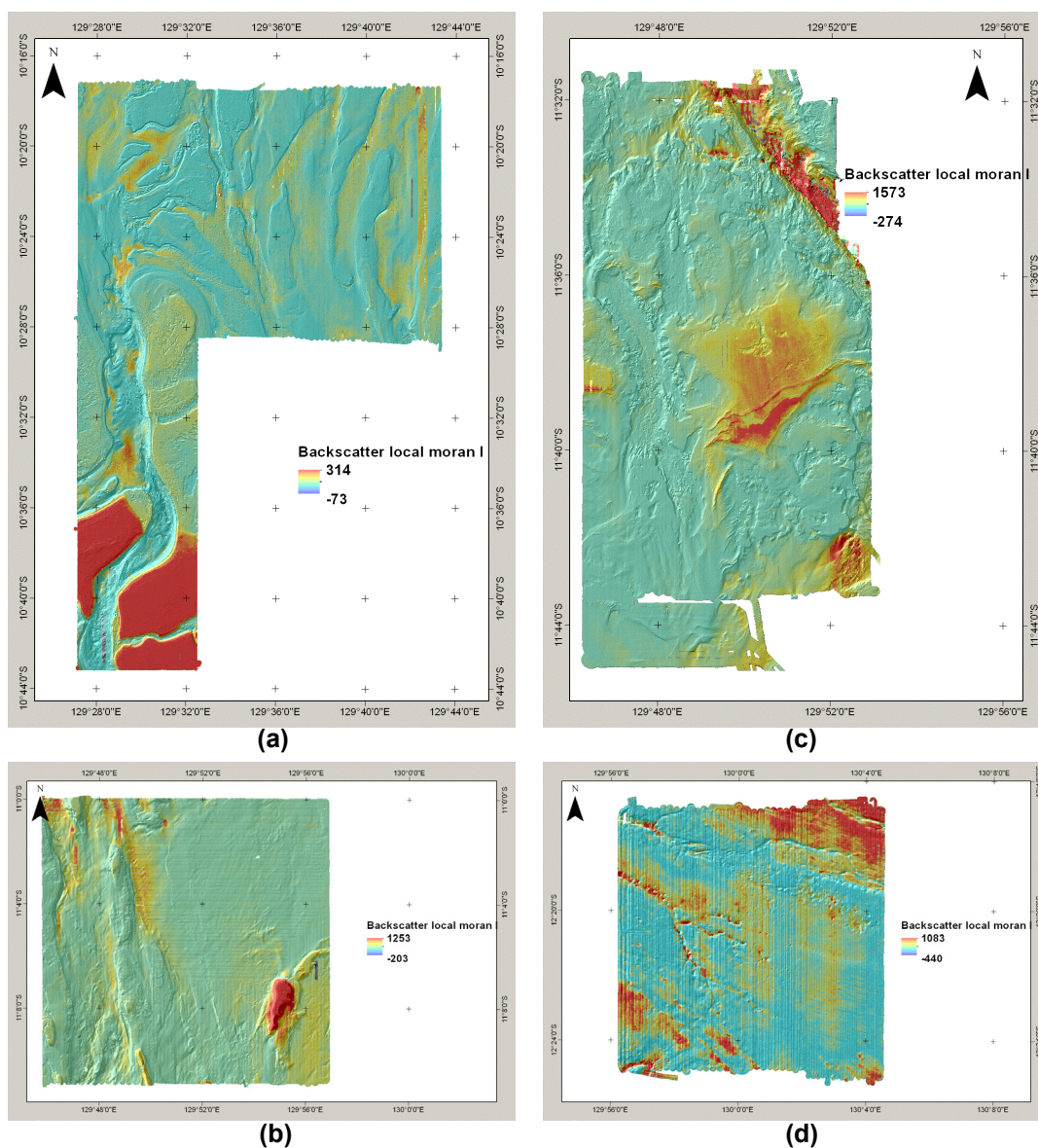


Figure 4.23: Backscatter Local Moran I maps of the four study areas, Van Diemen Rise.

4.3.2. Angular Backscatter Response (AR) p-Value

The p -value for areas of hard seabed ranged from 0 to 1 (Figures 4.24 to 4.27). A high p -value is typically associated with the “hard” seabed class as derived from video and sled tows whereas low p -value is associated with the “soft” seabed class. The threshold value for separating “hard” and “soft” seabed influences the classification accuracy when compared to video observations. Thus, for a threshold p -value ≥ 0.05 the overall classification accuracy is 74% for complete video tracks and 84% for sampling locations. If however the threshold is increased to p -value ≥ 0.2 , the overall classification accuracy increases to 82% and to 90%, respectively.

The main limitation of the angular backscatter response approach is the low spatial resolution because the analysis is limited to half of the swath even though the angular resolution is high. This becomes problematic in areas with high spatial variability or high patchiness that creates irregularities in the angular backscatter response across the swath. This explains why there are a number of locations in Areas A and D identified as “hard” seabed that are associated with low p -value. Even though the backscatter grid suffers loss of angular resolution as a result of the removal of angular dependence, it maintains higher spatial resolution. This is useful in providing a global trend overview, which in turn helps explaining the discrepancy mentioned above. The fact that the video has a smaller coverage and a layback issue gives rise to this discrepancy in a highly patchy region. Water depth also complicates this issue as the acoustic footprint increases with water depth whereas the optical footprint of the video is roughly independent of depth. If a patch is much smaller than the acoustic footprint, it might become acoustically transparent but not optically transparent.

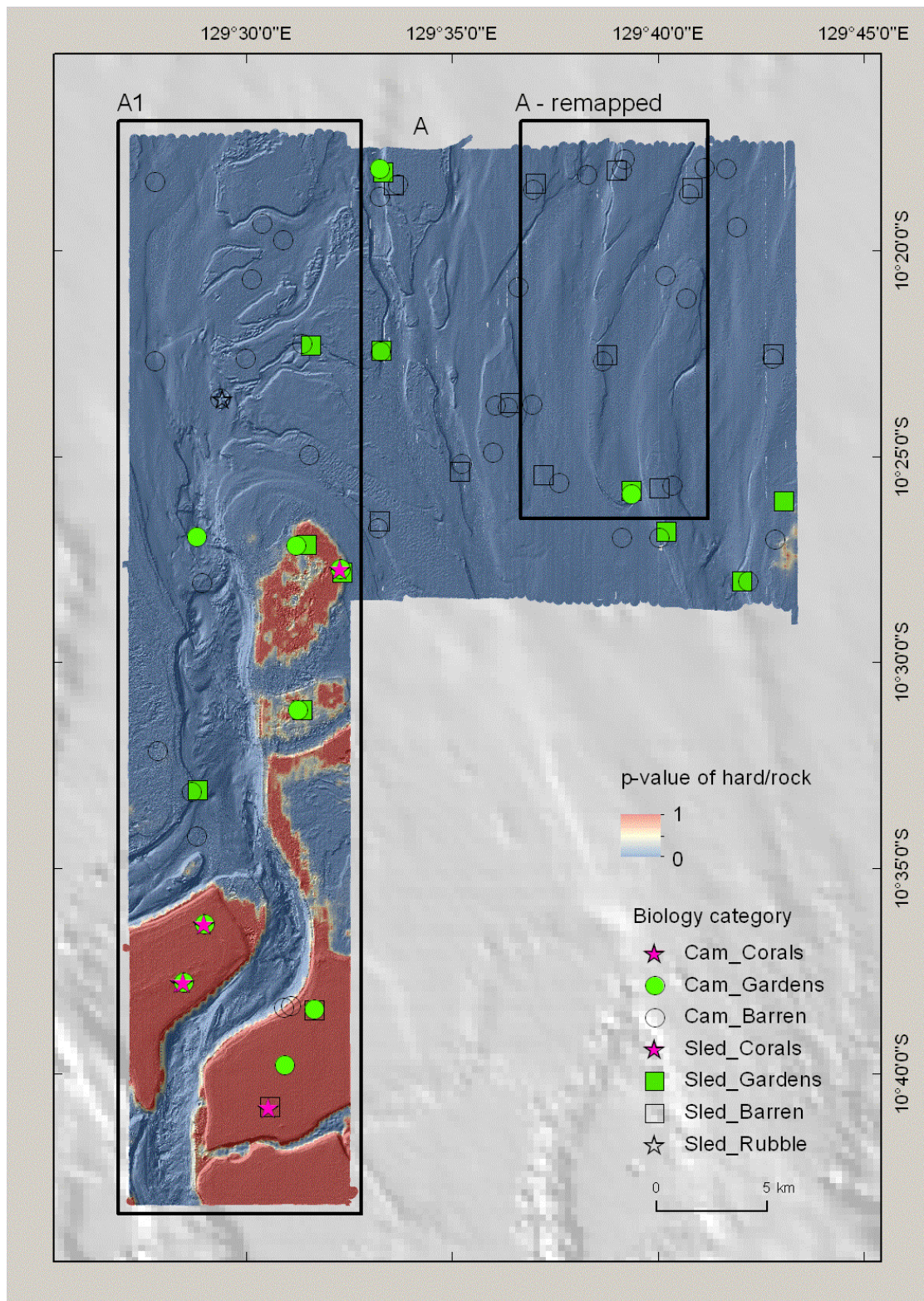


Figure 4.24: False colour p-value map of seabed hardness for Area A, with biology categories indicated. All areas inside the black boxes mapped during survey SOL5117 and all others during survey SOL4934.

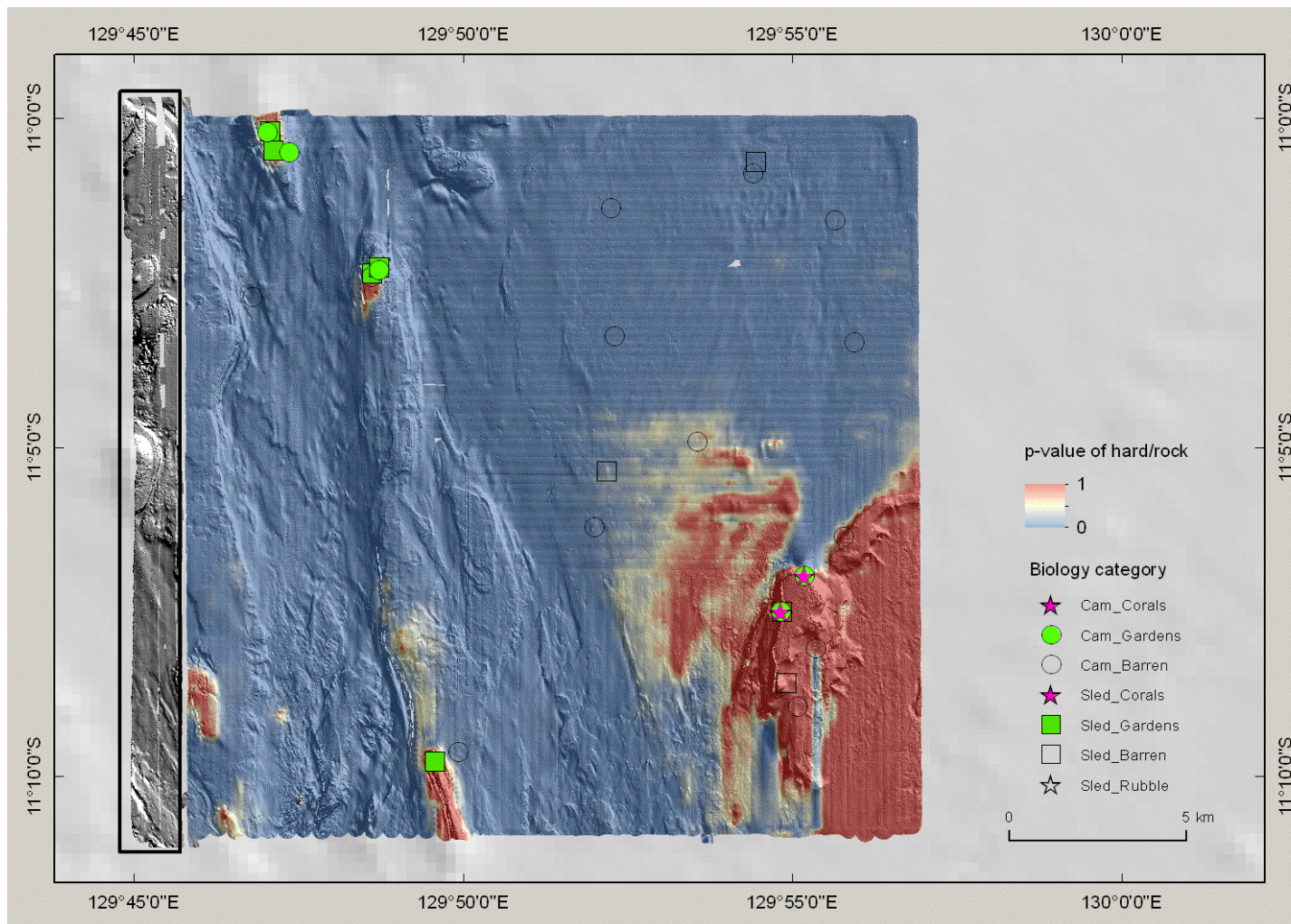


Figure 4.25: False colour p-value map of seabed hardness for Area B, with biology categories indicated.

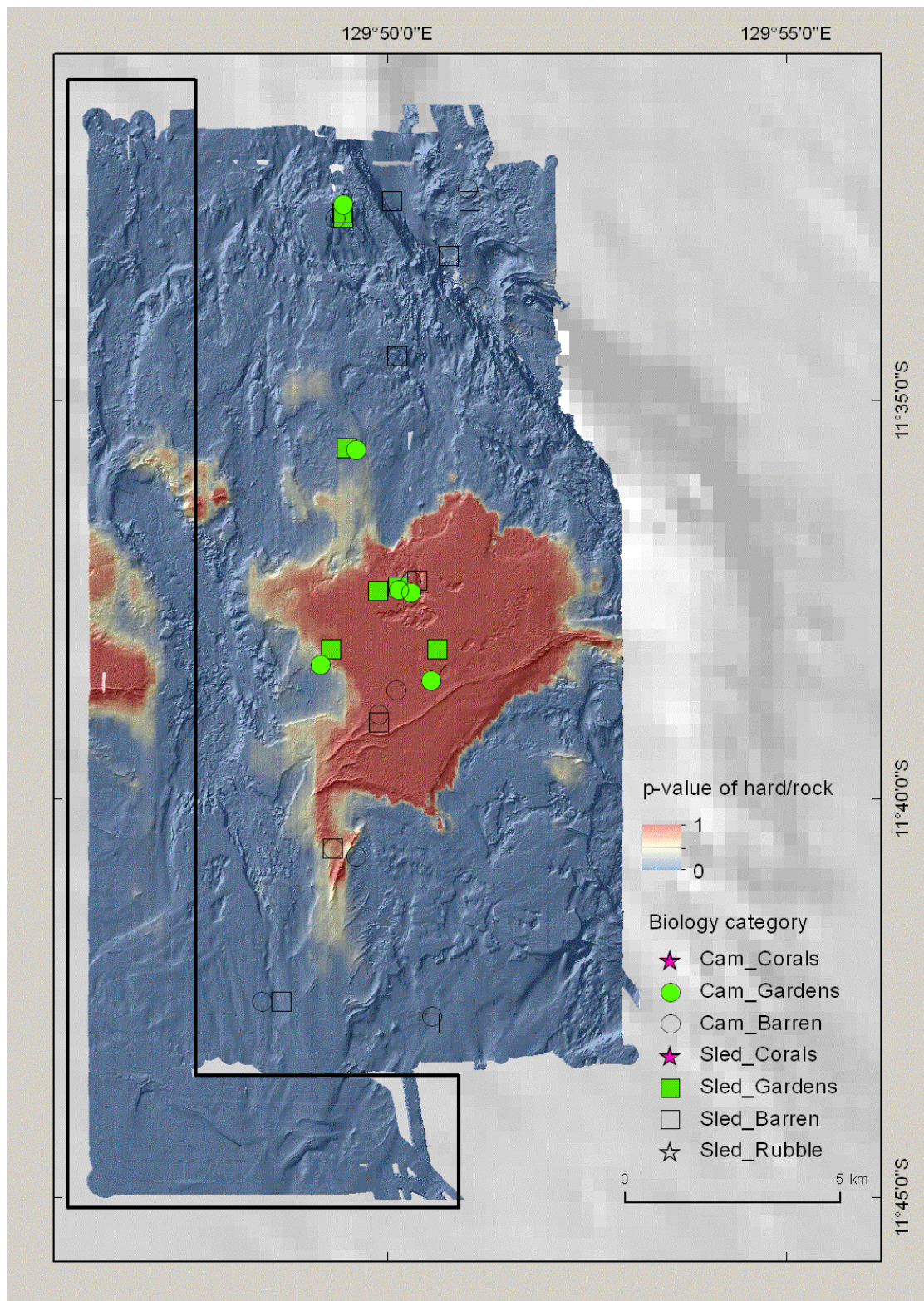


Figure 4.26: False colour *p*-value map of seabed hardness for Area C, with biology categories indicated.

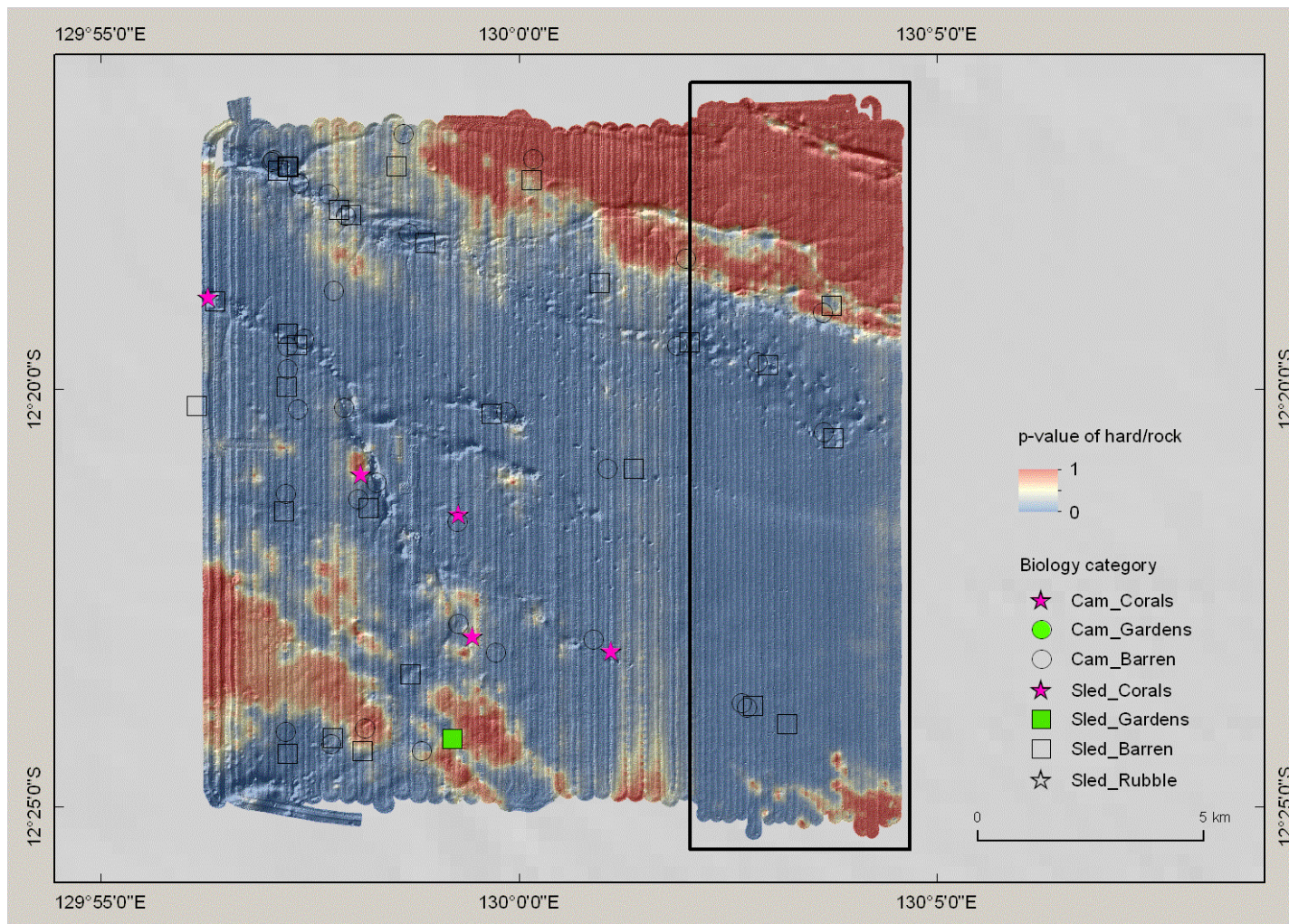


Figure 4.27: False colour p-value map of seabed hardness for Area D, with biology categories indicated.

4.3.3. Hardness Prediction

The predictive variables (Appendix D) are extracted from the grid values (Figures 4.2 to 4.9, 4.14 to 4.23 & 4.24 to 4.27) at each sampling station (shown as white dots in Figures 4.2 to 4.9). There are clear similarities between the assigned “hard” and “soft” video classes and the classes based on the angular response derived p-values. However, for reasons mentioned in Section 4.3.2, some samples had extreme discrepancies between the video class and the p-value and were consequently reassessed by looking at neighbouring trends in backscatter and bathymetry grids. For example, if the original video was classified as “hard” but the backscatter level was low and the bathymetric feature appeared smoother, the video class was then reassigned as “soft” and vice versa. Of these samples, 13 were consequently reassigned to different classes.

The correlations between hardness and some predictors were very strong (Table 4.2 & Appendix E). Based on these predictors (Table 4.2), a final predictive model was developed using Random Forest. Of these twelve predictors, the *p*-value of hard seabed, Backscatter (BS), Bathymetry and BS local Moran were the four most important predictors (Figure 4.28). In addition, three out of these four were backscatter-derived parameters. This suggests that the backscatter strength played an important role in hardness prediction and supports Huang *et al.* (2011) in their work on using backscatter-derived variables for the prediction of %sand and %gravel.

This model produced a perfect fit to the data, that is, the fitted values fell into the predicted hard category for all 38 hard samples and soft for all 99 soft samples. Consequent predictive accuracy is also very high. On the basis of 10-fold cross-validation, the predictive accuracy is 0.7937 in terms of kappa value and 92.15% in terms of a correct classification rate (Table 4.3).

According to Fielding & Bell (1997), the agreement between the predicted values and the tested values is excellent if kappa is more than 0.75. The correct classification rate for hardness predictions exceeded 92%. Hence the accuracy of the model developed for predicting the seabed hardness in this study is exceptionally high. In addition, the findings further confirm the superior performance of the Random Forest method observed previously by Li *et al.* (2010; 2011a, b). Therefore, we conclude that:

- the sample size used in this study was adequate for modelling the hardness;
- seabed hardness was appropriately classified based on ground-truthed video data;
- appropriate predictors have been identified and incorporated;
- appropriate modelling methods were employed;
- a robust predictive method has been developed;
- the hardness of the seabed is predictable; and
- the hardness can be predicted with a high accuracy.

Applying the most important predictors to the aforementioned model resulted in prediction maps of seabed hardness (Figures 4.29 to 4.32). Obvious trends were observed between the hardness prediction (Figures 4.29 to 4.32) and the backscatter (cf Figures 4.6 to 4.9) as well as the bathymetry (cf Figures 4.2 to 4.5). The hard seabed areas found primarily in shallow water are predominantly associated with the highest backscatter values, namely geomorphic features such as banks and deep slope. Valleys found in relatively deeper water are mostly associated with the lowest backscatter values and fall within the predicted soft-bottomed category.

The prediction map and video data in selected areas of Area A were compared ([Figure 4.33](#)). The agreement between the two datasets is reasonably good. The layback issue with video data collected in 2009 was observed in [Figure 4.33](#). This becomes more pronounced in highly patchy regions ([Figures 4.33c & 4.33d](#)).

Table 4.2. Spearman correlation among the hardness and the predictors.

	Easting	Northing	p -value of hard	Bathy	BS	Bathy local Moran	Planar curvature	Profile curvature	Relief	Slope	Surface	TPI	Homogeneity	BS local Moran	Variance
Easting	1	-0.85	0.37	-0.51	0.43	0.06	0.05	-0.11	-0.35	-0.24	-0.38	-0.1	0.3	-0.05	-0.33
Northing	-0.85	1	-0.53	0.7	-0.65	-0.22	-0.06	0.09	0.19	0.13	0.19	0.07	-0.41	-0.07	0.4
p -value of hard	0.37	-0.53	1	-0.78	0.8	0.45	-0.13	-0.05	-0.14	-0.11	-0.14	0.07	0.42	0.32	-0.37
Bathy	-0.51	0.7	-0.78	1	-0.91	-0.18	-0.01	0.12	0.22	0.19	0.19	0.08	-0.59	-0.18	0.55
BS	0.43	-0.65	0.8	-0.91	1	0.24	-0.04	-0.03	-0.11	-0.09	-0.07	0.03	0.56	0.1	-0.52
Bathy local Moran	0.06	-0.22	0.45	-0.18	0.24	1	-0.15	0.04	0.16	0.12	0.14	0.15	-0.02	0.57	0.06
Planar curvature	0.05	-0.06	-0.13	-0.01	-0.04	-0.15	1	-0.19	0.01	-0.04	-0.04	-0.51	-0.03	0.01	0.01
Profile curvature	-0.11	0.09	-0.05	0.12	-0.03	0.04	-0.19	1	0.13	0.1	0.18	0.7	-0.05	-0.09	0.04
Relief	-0.35	0.19	-0.14	0.22	-0.11	0.16	0.01	0.13	1	0.9	0.93	0.14	-0.32	0.19	0.37
Slope	-0.24	0.13	-0.11	0.19	-0.09	0.12	-0.04	0.1	0.9	1	0.84	0.18	-0.24	0.09	0.29
Surface	-0.38	0.19	-0.14	0.19	-0.07	0.14	-0.04	0.18	0.93	0.84	1	0.16	-0.33	0.16	0.38
TPI	-0.1	0.07	0.07	0.08	0.03	0.15	-0.51	0.7	0.14	0.18	0.16	1	0.01	-0.02	-0.02
Homogeneity	0.3	-0.41	0.42	-0.59	0.56	-0.02	-0.03	-0.05	-0.32	-0.24	-0.33	0.01	1	-0.06	-0.92
BS local Moran	-0.05	-0.07	0.32	-0.18	0.1	0.57	0.01	-0.09	0.19	0.09	0.16	-0.02	-0.06	1	0.14
Variance	-0.33	0.4	-0.37	0.55	-0.52	0.06	0.01	0.04	0.37	0.29	0.38	-0.02	-0.92	0.14	1
Hardness	-0.1	-0.07	0.62	-0.47	0.6	0.3	-0.13	-0.05	0.18	0.14	0.21	0.08	0.2	0.31	-0.13

Table 4.3. Confusion matrix between predictions and observed values based on the average of 100 times of 10-fold cross validation.

		Observed			
Prediction		Soft	Hard	Total	User's Acc.
	Soft	96.77	8.52	105.29	0.9191
	Hard	2.23	29.48	31.71	0.9297
	Total	99	38	137	
	Producer's Acc.	0.9775	0.7758		
Overall Accuracy					0.9215

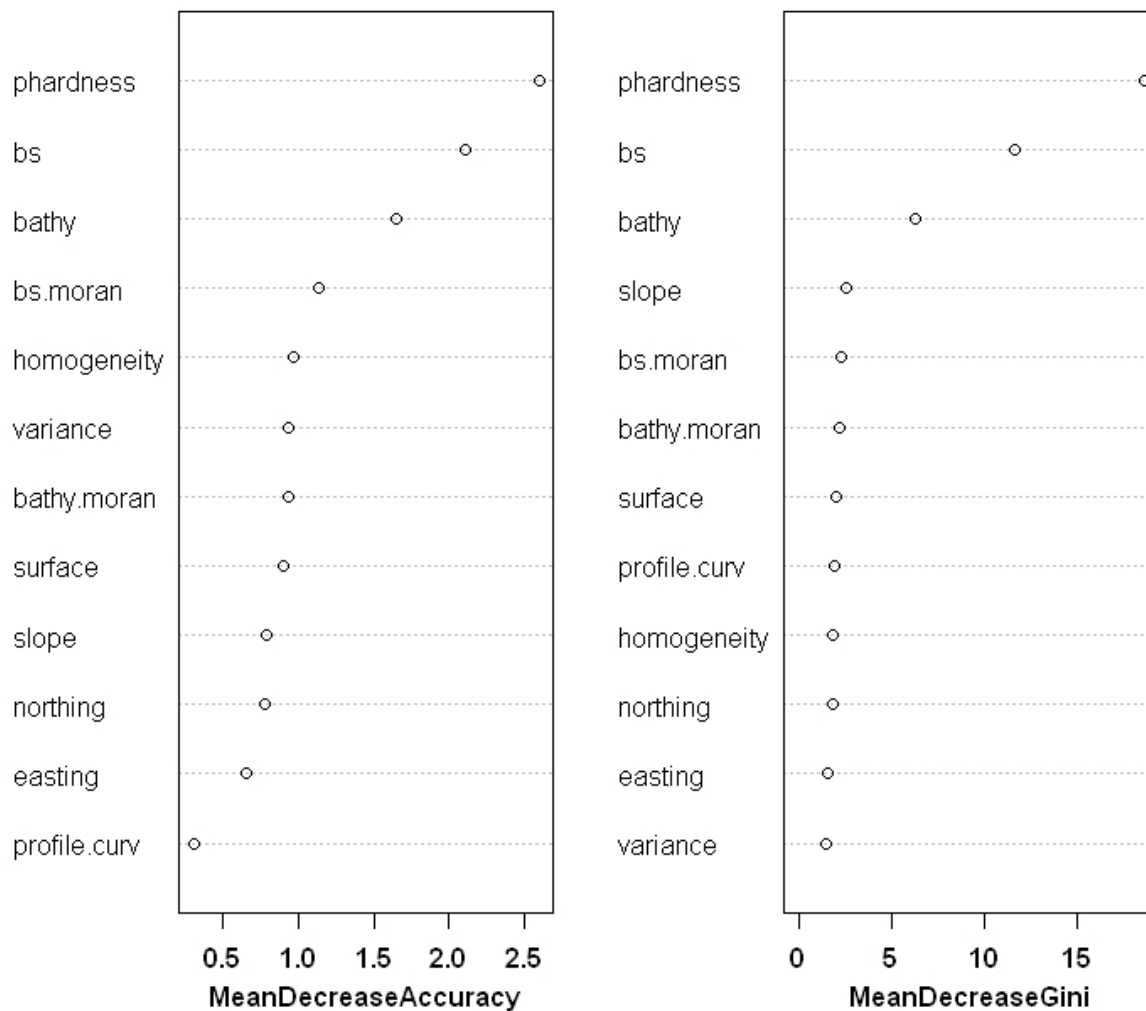


Figure 4.28: The importance of the predictors for modelling the hardness using the Random Forest method.

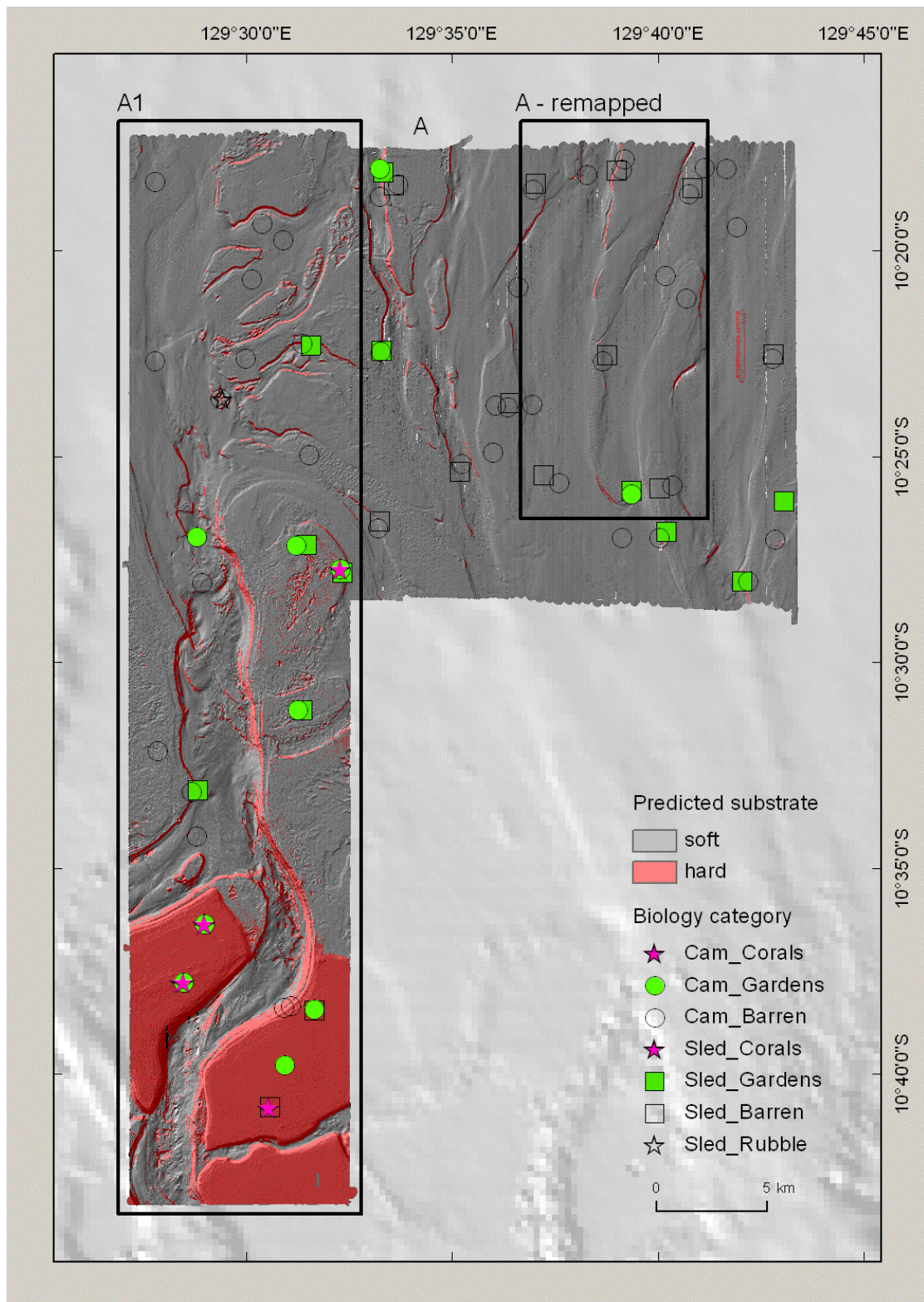


Figure 4.29: Seabed hardness prediction map of Area A.

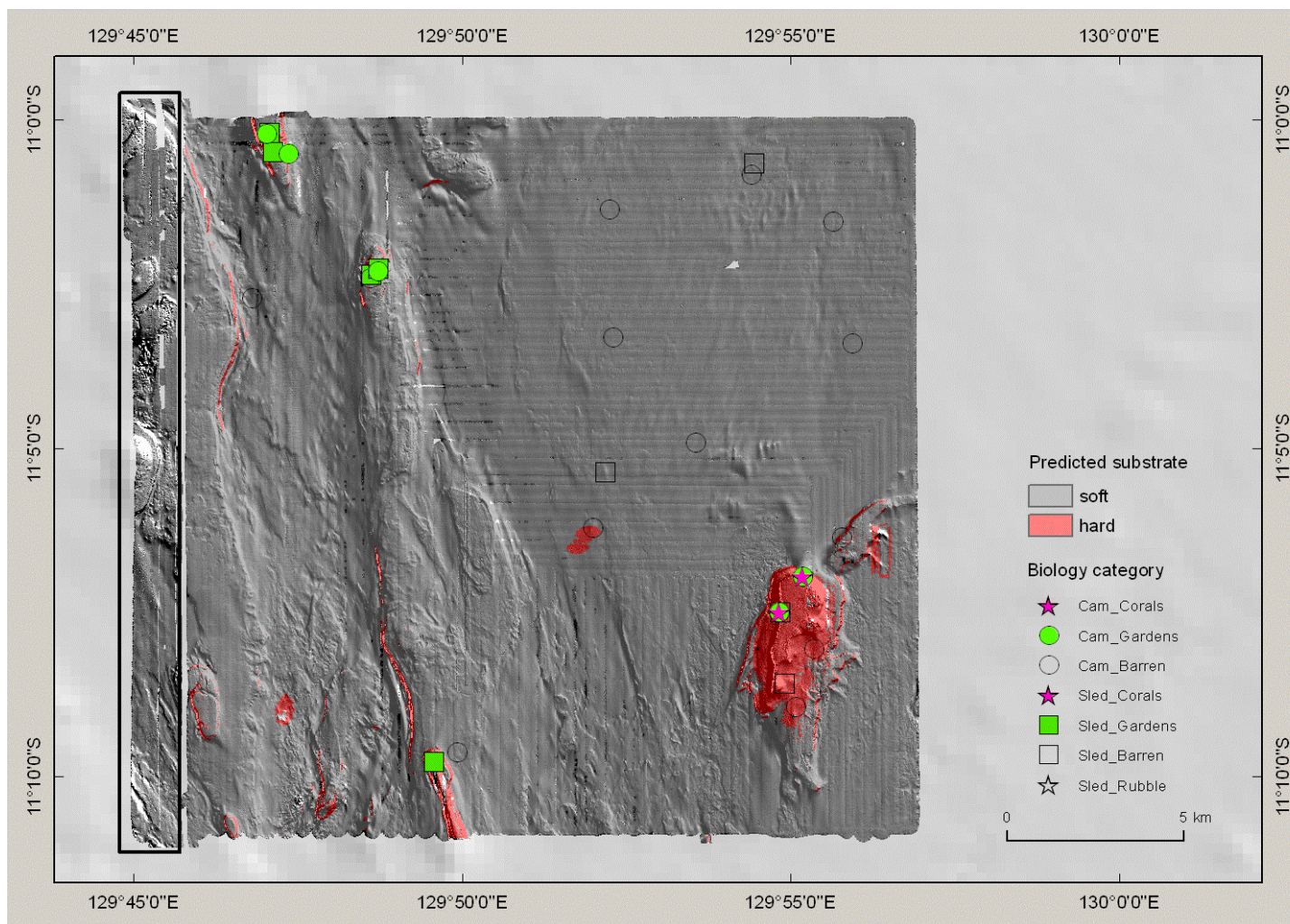


Figure 4.30: Seabed hardness prediction map of Area B.

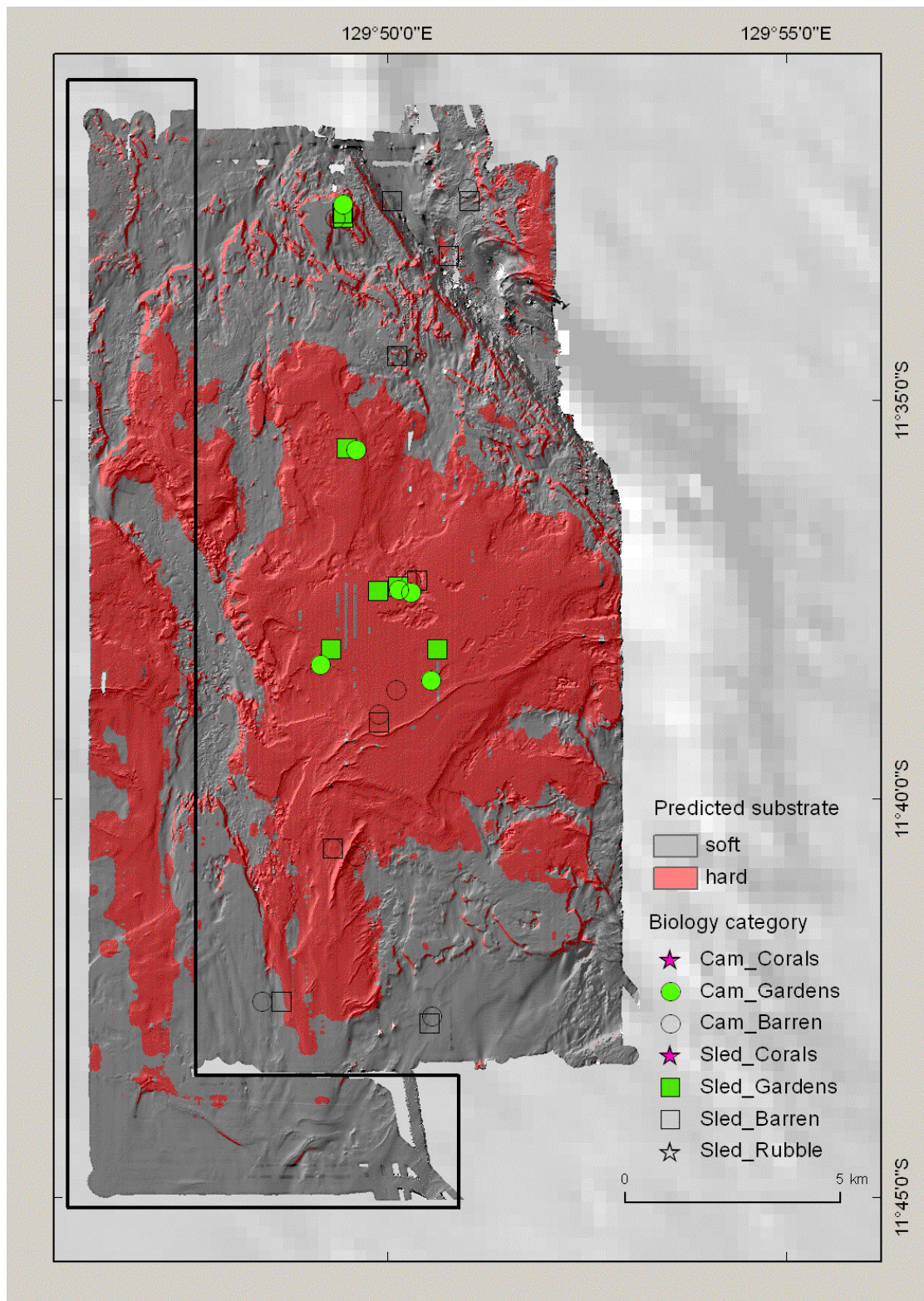


Figure 4.31: Seabed hardness prediction map of Area C.

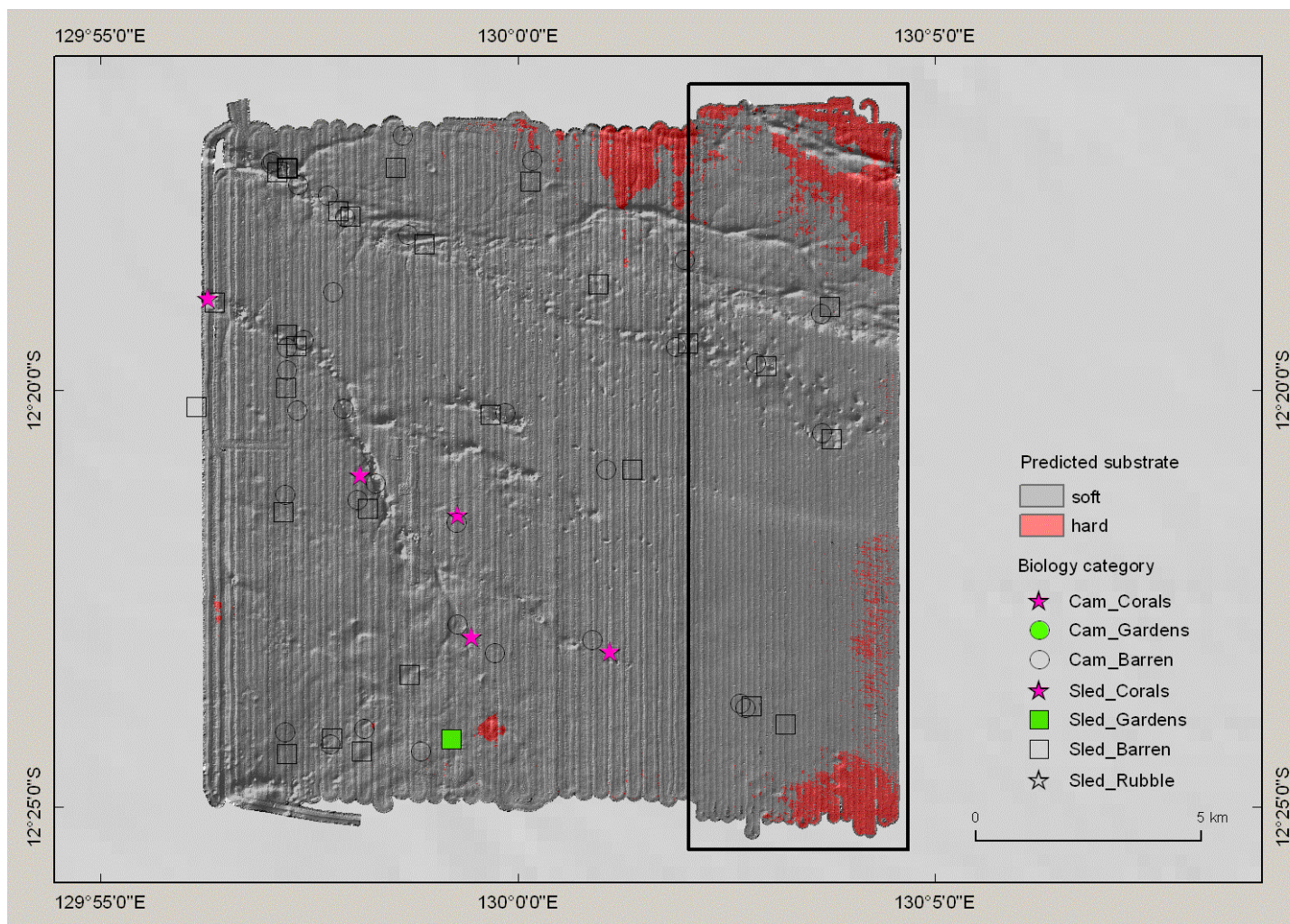


Figure 4.32: Seabed hardness prediction map of Area D.

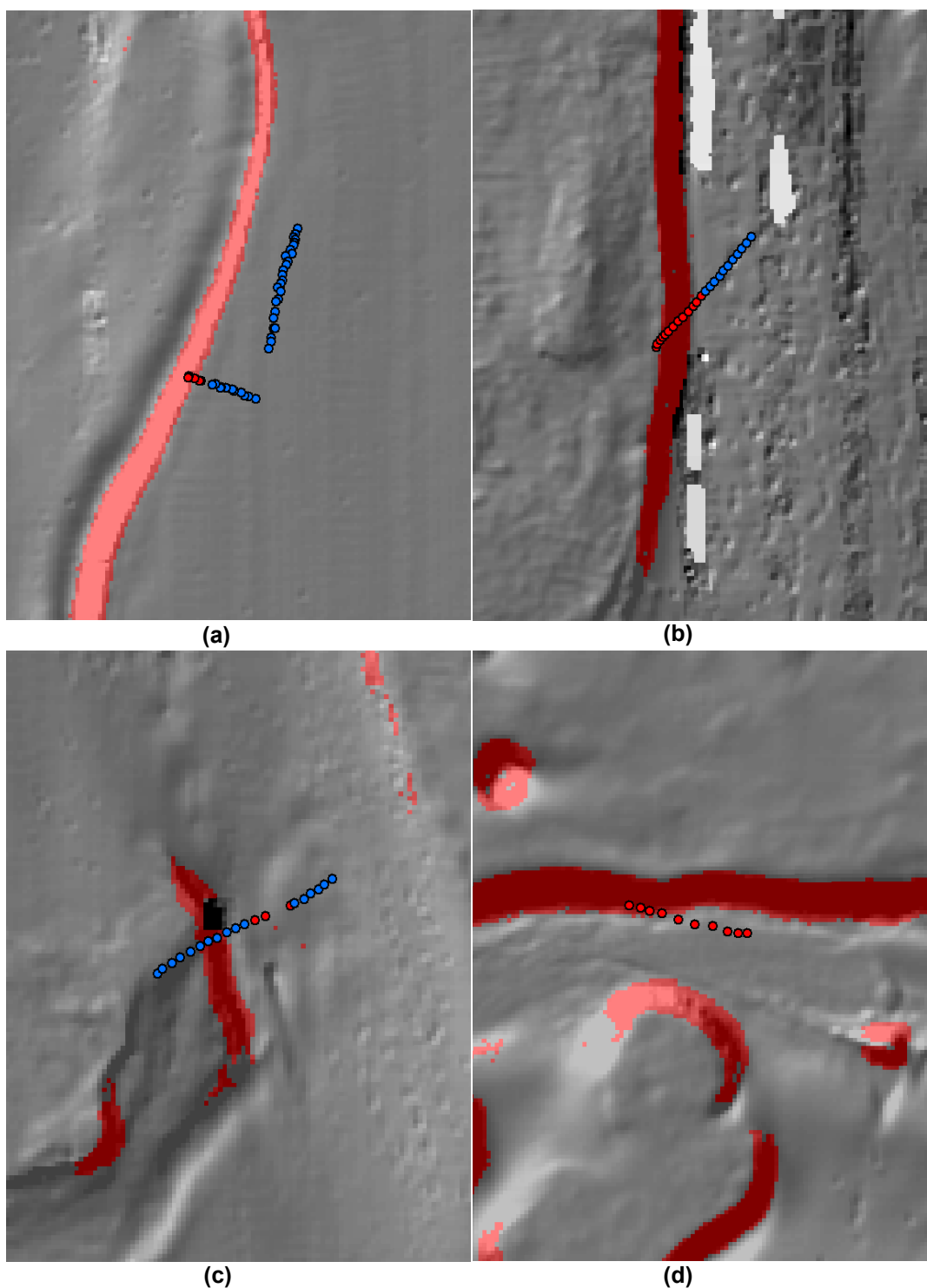


Figure 4.33: Representative examples of comparison between video observations and prediction map with (a) & (b) showing good alignment with video track; and (c) & (d) showing misalignment with video due to the layback issue.

5. Summary

5.1. Approach Overview

Two approaches have been presented to enable acoustic seabed segmentation from multibeam bathymetry and backscatter and their derivatives into two major seabed types; “hard” and “soft”. Both approaches are robust and have been successfully applied to data for an area of complex geomorphology on the carbonate banks of the Van Diemen Rise, Timor Sea.

The first approach was classification-based involving two stages: an initial classification using a Self Organising Map; and a second stage classification utilising hierarchical clustering based on the Random Forest method (Vesanto & Alhoniemi, 2000). To our knowledge, this study is the first to map seabed substrate in this way and using full angular backscatter response curves. The success of the approach is primarily due to the fact that the angular backscatter response is a direct intrinsic property of the seabed. While the idea stems from the approach developed by Hamilton & Parnum (2011), the method presented in this study has the advantage of providing a visualisation of cluster relationships (through the denrogram and Self Organising Map) at both stages of processing.

The two-stage classification method is also relatively straight-forward to execute. Thus, once the angular backscatter response data for each half swath have been prepared, the seabed segmentation using this approach is simple and can be applied immediately as it requires neither selection of training sets, interpolation, feature extraction, nor modelling. The limitation of this approach however, is in the generation of the angular backscatter response which assumes homogeneity across half the swath and within a sliding window along the swath track. In reality, this is not always the case. This also results in a low spatial resolution inhibiting a construction of a fine resolution seabed habitat map.

The second, prediction-based approach using a Random Forest model was applied to bathymetry, backscatter data and their derivatives, including the angular backscatter response-derived product (i.e. the p -value of hard bottom). This study is among a small number of studies that have directly used multibeam bathymetry and its derivatives (Leecaster 2003; Verfaillie *et al.*, 2006, 2009; Li *et al.*, 2010), in combination with backscatter and its derivatives (Huang *et al.*, 2011) to predict and map continuous seabed substrate. Additionally, this study is the first to utilise angular backscatter response-derived products as a predictor. The model developed has been used to successfully predict the two major seabed types at a spatial resolution that matches that of the bathymetry grid. Unlike the first approach, this approach enables the generation of a continuous map of the major seabed types. However, this approach requires selection of proper training sets of known seabed types. In this respect, the information provided by the ground-truthed video footage is essential. Video data also helps establish a reference for the angular backscatter response curves later applied to the Kolmogorov-Smirnov goodness of fit to generate the p -value of hard seabed.

5.2. Comparison of Methods

Results from the two segmentation methods are compared here for a representative part of study area A, using a confusion matrix against the video-derived class at the nearest sampling station. To do this, it was necessary to reduce the number of classes derived from unsupervised classification to two

classes to match the number of video classes. Using the dendrogram (Figure 3.2), the number of classes was reduced at a level where classes B and C are merged. The confusion matrix suggests that the overall classification accuracy is 78.26% for the classification-based approach (Table 5.1). For the Random Forest method, predicted classes were extracted from the continuous prediction layer to match the same locations on the unsupervised classification. The overall classification accuracy is 86.96% (Table 5.2). The higher overall classification accuracy of the prediction-based method is not unexpected given that the video-derived class adopted as the observed variable dictate the predictive model. This emphasises the important role *a-priori* knowledge can play. This, on the other hand, is not required in the unsupervised classification-based approach.

Table 5.1. Confusion matrix between classification map and video class at sampling stations within Area A.

Groundtruth (video)					
Classification		Soft	Hard	Total	User's Acc.
	Soft	13	5	18	0.7222
	Hard	0	5	5	1.0000
	Total	13	10	23	
	Producer's Acc.	1.0000	0.5000		
Overall Accuracy					0.7826

Table 5.2. Confusion matrix between prediction map and video class at sampling stations within Area A.

Groundtruth (video)					
Prediction		Soft	Hard	Total	User's Acc.
	Soft	13	3	16	0.8125
	Hard	0	7	7	1.0000
	Total	13	10	23	
	Producer's Acc.	1.0000	0.7000		
Overall Accuracy					0.8696

The *p*-value for hard seabed has been shown to be the most important predictor in the prediction approach. This is because it is derived from the full angular backscatter response which is a direct inherent property of the seabed. This is supported by the success of the classification-based approach utilising the full angular backscatter response alone. To be included in the prediction approach, however, it requires an interpolation to generate a continuous map of the *p*-value of hard seabed. One of the criticisms of this is that it might be dependent on the technique used for interpolation. Nonetheless, a good agreement has been observed between the interpolated *p*-value of hard seabed and the backscatter mosaic and the bathymetry.

5.3. Limitations

To verify and validate the acoustic classification and prediction results, we relied on video and information from sediment grabs. The limitation of these approaches is that they provide little to no information about the sub-surface. Certain conditions and situations allow penetration of the acoustic

energy below the surface. It is therefore difficult to rely on the information provided from the video to validate hard grounds underneath a thin veneer of soft sediments. In addition, the information used for validation basically relied on a video-derived classification which can be somewhat subjective if only one observer was involved in the classification process.

Angular backscatter response curves have become popular in various fields for seabed classification and characterisation as they preserve seabed properties in greater angular resolution than backscatter mosaics can offer. They are however lacking in spatial resolution due to the nature of their construction because they are generated by taking the average of the stack of backscatter intensity as a function of incidence angles within a sliding window. In addition, this is based on the assumption that the seabed is homogeneous across half the swath which is not always the case.

5.4. Summary Comments and Recommendations

The use of angular backscatter response curves for seabed mapping is important because the backscatter strength from different seabed types is presented at different incidence angles and provides more information than single normalised backscatter values. The success of the classification-based approach in utilising the full angular backscatter response alone is to be expected due to the nature of the angular backscatter response that inherits the seabed property. This is supported by the prediction approach as it ranks the p -value of hard seabed, estimated from the angular backscatter response, as the highest among the most important predictors for generating seabed substrate maps. The angular backscatter response alone seems to be sufficient to distinguish hard from soft terrain. This is perhaps due to the small number of classes of interest (i.e. “hard” and “soft”) but is nonetheless a key outcome for future seabed mapping.

The prediction-based approach also indicates that bathymetry is one of the most important predictors of substrate type. The two adopted approaches, the classification and prediction approaches, in association with appropriate data quality control would nevertheless provide broad scale insight of seabed substrate type; distinguishing hard from soft terrain. In addition they offer fast and inexpensive assessment in comparison to intensive seabed sampling. Significantly, these methods could potentially be extended to areas where only multibeam acoustic data are available.

The classification-based approach has shown its strength to visualise the process of split and merge. It enables an open number of classes and hence allows rationalising of the relationship with other data, including video for ground-truthing. The visualisation capability of this approach has allowed us to determine that a more than two class solution would potentially provide better results.

The choice between the two methods comes down only to the availability and the reliability of the *a-priori* knowledge which in fact determines whether or not the path of the prediction approach becomes possible. In the absence of the *a-priori* knowledge, a combination of the two approaches is advocated and may potentially be a better choice. In this scenario, the first approach is used to establish an initial grouping which then becomes an observed variable dataset used in the second approach to establish a predictive model.

Since the Random Forest approach produced continuous predictions with improved predictive accuracy, it is recommended if continuous predictions of the hardness terrain are required and the information for relevant predictors is available.

References

- Anderson, T.J., Nichol, S., Radke, L., Heap, A.D., Battershill, C., Hughes, M., Siwabessy, P.J., Barrie, V., Alvarez de Glasby, B., Tran, M., Daniell, J. & Shipboard Party, 2011. *Seabed Environments of the Eastern Joseph Bonaparte Gulf, Northern Australia: GA0325/Sol5117 - Post-Survey Report*. Geoscience Australia, Record 2011/08, 59pp.
- Anderson, T.J., Cochrane, G.R., Roberts, D.A., Chezar, H., & Hatcher, G., 2008. "A rapid method to characterize seabed habitats and associated macro-organisms." *In*: Todd, B.J., & Greene, H.G. (Eds.), *Mapping the Seafloor for Habitat Characterization*. Geological Association of Canada, Special Paper 47, 71-79.
- APL, 1994. *APL-UW High-Frequency Ocean Environment Acoustic Models Handbook*. APL_UW TR9407.
- Brown, C.J., Smith, S.J., Lawton, P., & Anderson, J.T., 2011. "Benthic habitat mapping: A review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques." *Estuarine, Coastal, and Shelf Science* 92, 502-520.
- Cochrane, G.R., & Lafferty, K.D., 2002. "Use of acoustic classification of sidescan sonar data for mapping benthic habitat in the Northern Channel Islands, California." *Continental Shelf Research* 22, 683-690.
- Cohen J., 1960. "A coefficient of agreement for nominal scales." *Educational and Psychological Measurement* 20, 37-46.
- Daniell, J., Jorgensen, D.C., Anderson, T., Borissova, I., Burq, S., Heap, A.D., Hughes, M., Mantle, D., Nelson, G., Nichol, S., Nicholson, C., Payne, D., Przeslawski, R., Radke, L., Siwabessy, J., Smith, C., & Shipboard Party, 2010. *Frontier Basins of the West Australian Continental Margin: Post-survey Report of Marine Reconnaissance and Geological Sampling Survey GA2476*. Geoscience Australia, Record 2009/38, 229pp.
- Daniell, J., Fellows, M., & Fountain, L., 2007. *Methodologies for seafloor substrate characterisation using available physical datasets on the northwest margin of Australia*. Geoscience Australia, Interim Report, 143pp.
- De Falco, D., Tonielli, R., Di Martino, G., Innangi, S., Simeone, S., & Parnum, I.M., 2010. "Relationships between multibeam backscatter, sediment grainsize and Posidonia oceanica seagrass distribution." *Continental Shelf Research* 30, 1941-1950.
- de Moustier, C.P., & Alexandrou, D., 1991. "Angular dependence of 12-kHz seafloor acoustic backscatter" *The Journal of the Acoustical Society of America* 90 (1), 522-531.
- de Moustier, C.P., & Matsumoto, H., 1993. "Seafloor acoustic remote sensing with multibeam echosounders and bathymetric sidescan sonar systems." *Marine Geophysical Research* 15, 27-42.
- Dunn, D.C., & Halpin, P.N., 2009. "Rugosity-based regional modeling of hard-bottom habitat." *Marine Ecology Progress Series* 377, 1-11.
- Erdey-Heydorn, M.D., 2008. "An ArcGIS Seabed Characterization Toolbox Developed for Investigating Benthic Habitats," *Marine Geodesy* 31, 318 - 358.
- Ferrini, V.L., & Flood, R.D., 2006. "The effects of fine-scale surface roughness and grainsize on 300kHz multibeam backscatter intensity in sandy marine sedimentary environments." *Marine Geology* 228, 153-172.
- Fielding, A.H., & Bell, J.F., 1997. "A review of methods for the assessment of prediction errors in conservation presence/absence models." *Environmental Conservation* 24, 38-49.
- Fonseca, L., & Mayer, L., 2007. "Remote estimation of surficial seafloor properties through the application Angular Range Analysis to multibeam sonar data." *Marine Geophysical Researches* 28, 119-126.

- Fonseca, L., Brown, C., Calder, B., Mayer, L., & Rzhano, Y., 2009. "Angular range analysis of acoustic themes from Stanton Banks Ireland: A link between visual interpretation and multibeam echosounder angular signatures." *Applied Acoustics* 70 (10), 1289-1304.
- Gavrilov, A.N., Duncan, A.J., McCauley, R.D., Parnum, I.M., Penrose, J.D., Siwabessy, P.J.W., Woods, A.J., & Tseng, Y-T., 2005a. "Characterization of the Seafloor in Australia's Coastal Zone using acoustic techniques." *Proceedings of the International Conference "Underwater Acoustic Measurements: Technologies & Results"*, Crete, Greece, 1075-1080.
- Gavrilov, A.N., Siwabessy, P.J.W., & Parnum, I.M., 2005b. *Multibeam echo sounder backscatter analysis*, Centre for Marine Science and Technology, Perth, Australia, CA3.03.
- Gavrilov, A.N., & Parnum, I.M., 2010. "Fluctuations of seafloor backscatter data from multibeam sonar systems." *IEEE Journal of Oceanic Engineering* 35 (2), 209-219.
- Goff, J.A., Olson, H.C., & Duncan, C.S., 2000. "Correlation of side-scan backscatter intensity with grain-size distribution of shelf sediments, New Jersey margin." *Geo-Marine Letters*, 20, 43-49.
- Gratwicke, B., & Speight, M.R., 2005. "The relationship between fish richness, abundance and habitat complexity in a range of shallow tropical marine habitats." *Journal of Fish Biology* 66, 650-667.
- Hammerstad, E., 2000. *Backscattering and seabed image reflectivity*. EM Technical Note Kongsberg Horton Norway 6pp.
- Hamilton, L.J., & Parnum, I.M., 2011. "Acoustic seabed segmentation from direct statistical clustering of entire multibeam sonar backscatter curves." *Continental Shelf Research* 31, 138-148.
- Hasan, R.C., Ierodiaconou, D., & Laurenson, L., 2012. "Combining angular response classification and backscatter imagery segmentation for benthic biological habitat mapping." *Estuarine, Coastal and Shelf Science* 97, 1-9.
- Haralick, R.M., Shanmugan, K., & Dinstein, I., 1973, "Textural features for image classification." *IEEE Transactions on Systems, Man, and Cybernetics* 3, 610-621.
- Heap, A.D., Przeslawski, R., Radke, L., Trafford, J., Battershill, C., & Shipboard Party, 2010. *Seabed Environments of the Eastern Joseph Bonaparte Gulf, Northern Australia. Sol4934 - Post-survey Report*. Geoscience Australia, Record 2010/09, 78pp.
- Heap, A.D., & Harris, P.T., 2008. "Geomorphology of the Australian margin and adjacent seafloor." *Australian Journal of Earth Sciences*, 55, 555-585.
- Hensley, R.H., 1996. "Preliminary survey of benthos from the Nephrops norvegicus mud grounds in the north-western Irish Sea." *Estuarine, Coastal and Shelf Science* 42, 457-465.
- Holmes, K.W., Van Neil, K.P., Radford, B., Kendrick, G.A., & Grove, S.L., 2008. "Modelling distribution of marine benthos from hydroacoustics and underwater video," *Continental Shelf Research* 28, 1800-1810.
- Huang, Z., Nichol, S., Siwabessy, P.J.W., Daniell, J., & Brooke, B.P., 2011. "Predictive Modelling of Seabed Sediment Parameters Using Multibeam Acoustic Data: A Case Study on the Carnarvon Shelf, Western Australia." *International Journal of Geographical Information*, doi:10.1080/13658816.2011.590139.
- Hughes-Clarke, J.E., 1994. "Toward remote seafloor classification using the angular response of acoustic backscattering: a case study from multiple overlapping GLORIA data." *IEEE Journal of Oceanic Engineering* 19(1), 112-127.
- Hughes-Clarke, J.E., Danforth, B.W., & Valentine, P., 1997. "Aerial seabed classification using backscatter angular response at 95kHz." *Shallow Water, NATO SACLANTCEN, Conference Proceedings Series CP*, vol. 45, 243-250.
- Iampietro, P.J., Kvitek, P.G., & Morris, E., 2005. "Recent advances in automate genus-specific marine habitat mapping enabled by high-resolution multibeam bathymetry." *Marine Technology Society Journal* 39, 83-93
- Jackson, D.R., Briggs, K.B., Williams, K.L., & Richardson, M.D., 1996. "Tests of models for high-frequency seafloor backscatter." *IEEE Journal of Oceanic Engineering* 21 (4), 458-470.

- Jackson, D.R., Winebrenner, D.P., & Ishimaru, A., 1986. "Application of the composite roughness model to high-frequency bottom backscatter," *Journal of the Acoustical Society of America* 79, 1410-1422.
- Kaufman, L., & Rousseeuw, P.J., 1990. *Finding Groups in Data: An introduction to Cluster Analysis*. John Wiley, New York.
- Kloser, R.J., Bax, N., Ryan, J.T., Williams, A., & Barker, B.A., 2001. "Remote sensing of seabed types in the Australian South East Fishery; development and application of normal incident acoustic techniques and associated 'ground truthing'." *Marine and Freshwater Research*, 52, 475-489.
- Kloser, R.J., Penrose, J.D., & Butler, A.J., 2010. "Multi-beam backscatter measurements used to infer seabed habitats." *Continental Shelf Research* 30, 1772-1782.
- Kohonen, T., 1995. *Self-organizing maps*. Springer Series in Information Sciences vol. 30. Springer. Berlin.
- Kohonen, T., Hynninen, J., Kangas, J., & Laaksonen, J., 1996. SOM_PAK: *The self-organising-map program package*. Helsinki University of Technology, Laboratory of Computer and Information Science. Technical Report A31, Helsinki, Finland, 25pp.
(http://www.cis.hut.fi/research/som_lvq_pak.shtml).
- Lamarche, G., Lurton, X., Verdier, A-L., & Augustin, J-M., 2010. "Quantitative characterisation of seafloor substrate and bedforms using advanced processing of multibeam backscatter-Application to Cook Strait, New Zealand." *Continental Shelf Research* 31 (S-1), 93-S109.
- Lanier, A., Romsos, C., & Goldfinger, C., 2007. "Seafloor Habitat Mapping on the Oregon Continental Margin: A Spatially Nested GIS Approach to Mapping Scale, Mapping Methods, and Accuracy Quantification," *Marine Geodesy* 30, 51-76.
- Leecaster, M., 2003, "Spatial analysis of grain size in Santa Monica Bay." *Marine Environmental Research* 56, 67-78.
- Li, J., Heap, A.D., Potter, A., & Daniell, J.J., 2011a. *Predicting Seabed Mud Content across the Australian Margin II: Performance of Machine Learning Methods and Their Combination with Ordinary Kriging and Inverse Distance Squared*. Geoscience Australia, Record 2011/07, 69pp.
- Li J., Heap A.D., Potter A., & Daniell J., 2011b. "Application of machine learning methods to spatial interpolation of environmental variables." *Environmental Modelling & Software* 26, 1647-1659.
- Li, J., & Hilbert, D.W., 2008. "LIVES: a new habitat modelling technique for predicting the distributions of species' occurrence using presence-only data based on limiting factor theory." *Biodiversity Conservation* 17, 3079-3095.
- Li, J., Potter, A., Huang, Z., Daniell, J.J., & Heap, A.D., 2010. *Predicting Seabed Mud Content across the Australian Margin: Comparison of Statistical and Mathematical Techniques Using a Simulation Experiment*. Geoscience Australia, Record 2010/11, 146pp.
- Liaw, A., & Wiener, M., 2002. "Classification and regression by Radom Forest." *R News* 2, 18-22.
- Lucieer, V.L., 2008. "Object-oriented classification of sidescan sonar data for mapping benthic marine habitats." *International Journal of Remote Sensing* 29, 905-921.
- Lundblad, E.R., Wright, D.J., Miller, J., Larkin, E.M., Rinehart, R., Naar, D.F., Donahue, B. T., Anderson, S. M., & Battista, T., 2006. "A Benthic Terrain Classification Scheme for American Samoa." *Marine Geodesy* 29, 89-111.
- Lyons, A.P., Anderson, A.L., & Dwan, F.S., 1994. "Acoustic scattering from the seafloor: modeling and data comparison." *The Journal of the Acoustical Society of America* 95 (5), 2441-2451.
- Mackie, A.S.Y., Oliver, P.G., & Rees, E.I.S., 1995. "Benthic biodiversity in the southern Irish Sea." *Studies in marine biodiversity and systematics from the National Museum of Wales. BIOMAR Reports*, no. 1, 263pp.
- Marsh, I., & Brown, C., 2009. "Neural Network classification of multibeam backscatter and bathymetry data from Stanton Bank (Area IV)." *Applied Acoustics* 70 (10), 1269-1276.
- McArthur, M.A., Brooke, B.P., Przeslawski, R., Ryan, D.A. Lucieer, V.L., Nichol, S., McCallum, A.W., Mellin, C., Cresswell, I.D., & Radke, L.C. 2010. "On the use of abiotic surrogates to describe marine benthic biodiversity." *Estuarine Coastal and Shelf Science* 88, 21-32.

- Mortensen, P.B., & Buhl-Mortensen, L., 1992. "Distribution of deep-water gorgonian corals in relation to benthic habitat features in the Northeast Channel (Atlantic Canada)." *Marine Biology* 144, 1223-1238.
- Newell, R.C., Seiderer, L.J., & Robinson, J.E., 2001. "Animal/sediment relationships in coastal deposits of the eastern English Channel." *Journal of the Marine Biological Association of the UK* 81 (01), 1-9.
- Nichol, S.L., Anderson, T.J., McArthur, M., Barrett, N., Heap, A.D., Siwabessy, P.J.W., & Brooke, B., 2009. *Southeast Tasmania Temperate Reef Survey, Post Survey Report*. Geoscience Australia, Record 2009/43, 73pp.
- Nitsche, F.O., Bell, R., Carbotte, S.M., Ryan, W.B.E., & Flood, R. D., 2004. "Process-related classification of acoustic data from the Hudson River Estuary." *Marine Geology* 209, 131-145.
- Novarini, J.C., & Caruthers, J.W., 1998. "A simplified approach to backscattering from a rough seafloor with sediment inhomogeneities." *IEEE Journal of Oceanic Engineering* 23 (3), 157-166.
- Parnum, I.M., 2007. *Benthic habitat mapping using multibeam sonar systems*. Ph.D. Thesis, Curtin University of Technology, Western Australia, 213pp. (http://espace.library.curtin.edu.au:80/R?func=dbin_jump_full&object_id=18584&local_base=gen01-era02).
- Parnum, I.M., & Gavrilov, A.N., 2011. "High-frequency multibeam echo-sounder measurements of seafloor backscatter in shallow water: Part 1 - Data acquisition and processing." *International Journal of the Society for Underwater Technology* 30 (1), 3-12.
- Parnum, I., Gavrilov, A., Siwabessy, J., Duncan, A., 2006. "Analysis of High-frequency Multibeam Backscatter Statistics from Difference Seafloor Habitats." ECUA 8th European Conference on Underwater Acoustics, Portugal, 775-782.
- Parnum, I.M., Siwabessy, P.J.W. & Gavrilov, A.N., 2004. "Identification of Seafloor Habitats in Coastal Shelf Waters Using a Multibeam Echosounder." *Acoustics 2004*, Proceedings of the Conference of the Australian Acoustical Society, Gold Coast, Australia, 181-186.
- Przeslawski, R., Daniell, J., Anderson, T., Barrie, J.V., Heap, A., Hughes, M., Li, J., Potter, A., Radke, R., Siwabessy, J., Tran, M., Whiteway, T., & Nichol, S., 2011. *Seabed Habitats and Hazards of the Joseph Bonaparte Gulf and Timor Sea, Northern Australia*. Geoscience Australia, Record 2011/40, 69 pp.
- R Development Core Team, 2011. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna.
- Richardson, M.D., & Young, D.K., 1980. "Geoacoustic models and bioturbation." *Marine Geology* 38, 205-218.
- Sammon, J.W., 1969. "A nonlinear mapping for data structure analysis." *IEEE Transactions for Computing* 18, 401-409.
- Simpson, K.W., Fagnani, J.P., Bode, R.W., DeNicola, D.M. & Abele, L.E., 1986. "Organism-substrate relationships in the main channel of the Lower Hudson River." *Journal of the North American Benthological Society* 5 (1), 41-57.
- Siwabessy, P.J.W., Gavrilov, A.N., Duncan, A.J., & Parnum, I.M., 2006a. "Analysis of statistics of backscatter strength from different seafloor habitats." *Acoustics 2006*, Proceedings of the Conference of the Australasian Acoustical Societies, Christchurch, New Zealand, 507-514.
- Siwabessy, P.J.W., Parnum, I.M., Gavrilov, A., & McCauley, R.D., 2006b. *Overview of coastal water habitat mapping research for Coastal CRC*. Coastal CRC Technical Report No. 86, 27pp.
- Stein, D.L., Tissot, B.N., Hixon, M.A., & Barss, W.H., 1992. "Fish-habitat associations on a deep reef at the edge of the Oregon continental shelf." *Fishery Bulletin* 90, 540-551.
- Sutherland, T.F., Galloway, J., Loschiavo, R., Levings, C.D., & Hare, R., 2007. Calibration techniques and sampling resolution requirements for groundtruthing multibeam acoustic backscatter (EM3000) and QTC VIEW classification technology. *Estuarine Coastal and Shelf Science* 75, 447-458.
- Talukdar, K.K., Tyce, R.C., & Clay, C.S., 1995. "Interpretation of Sea Beam backscatter data collected at the Laurentian fan off Nova Scotia using acoustic backscatter theory." *The Journal of the Acoustical Society of America* 97 (3), 1545-1558.

- Toohey, B.D., Kendrick, G.A., & Harvey, E.S., 2007. "Disturbance and reef topography maintain high local diversity in *Ecklonia radiata* kelp forests." *Oikos* 116, 1618-1630.
- Utsch, A., & Simon, H.P., 1990. "Kohonen's self organizing feature maps for exploratory data analysis." *Proceedings INNC90, International Neural Network Conference*, 305-308.
- Verfaillie, E., Du Four, I., Van Meirvenne, M., & Van Lancker, V., 2009. "Geostatistical modeling of sedimentological parameters using multi-scale terrain variables: application along the Belgian Part of the North Sea." *International Journal of Geographical Information Science* 23, 135-150.
- Verfaillie, E., Van Lancker, V. & Van Meirvenne, M., 2006. "Multivariate geostatistics for the predictive modelling of the surficial sand distribution in shelf seas." *Continental Shelf Research* 26, 2454-2468.
- Vesanto, J., & Alhoniemi, E., 2000. "Clustering of the Self-Organising Map." *IEEE Transactions on Neural Networks* 11 (3), 586-600.
- Warwick, R.M., & Davies, J.R., 1977. "The distribution of sublittoral macrofauna communities in the Bristol Channel in relation to the substrate." *Estuarine, Coastal and Marine Science* 5, 267-288.
- Wentworth, C.K., 1992. "A scale of grade and class terms for clastic sediments." *The Journal of Geology* 30, 377-392.
- West, B.G., Conolly, J.R., Blevin, J.E., Miyazaki, S., & Vuckovic, V., 1992. *Petroleum Prospectivity of the East Malita Graben area, Bonaparte Basin*. Bureau of Mineral Resources, Geology and Geophysics, Australia, Record 1992/24.
- Williams, I.M. & Leach, J.H.J., 1999. "The relationship between depth, substrate and ecology: a drop video study from the southeastern Australian coast." *Oceanologica Acta* 22 (6), 651-662.
- Williams, K.L., Jackson, D.R., Thorsos, E.I., Tang, D., & Briggs, K.B., 2002. "Acoustic backscattering experiments in a well characterized sand sediment: data/model comparisons using sediment fluid and Biot models." *IEEE Journal of Oceanic Engineering* 27 (3), 376-387.
- Wilson, M.F.J., O'Connell, B., Brown, C., Guinan, J.C., & Grehan, A.J., 2007. "Multiscale Terrain Analysis of Multibeam Bathymetry Data for Habitat Mapping on the Continental Slope." *Marine Geodesy* 30, 3-35.
- Zieger, S., Stieglitz, T., & Kininmonth, S., 2009. "Mapping reef features from multibeam sonar data using multiscale morphometric analysis." *Marine Geology* 264, 209-217

Appendices

Appendix A. Comparison of Primary Bottom Classification with Hierarchical Classification (Average Linkage with 12 classes).

Note classes are presented in the same order that they occur in the classification dendrogram.

class		Pri-mud%	Pri-sand%	Pri-rubble%	Pri-rock%	Video observations
A	9	0.0	9.8	7.9	82.3	1221
	10	0.0	16.6	0.0	83.4	4590
B	2	0.4	50.1	1.0	48.5	3049
	3	0.0	66.7	2.4	30.9	2935
	4	0.0	43.5	0.4	56.1	7190
	6	0.0	3.5	13.3	83.2	315
	7	0.0	0.0	11.5	88.5	591
C	1	8.0	56.3	0.1	35.5	10078
	5	14.2	66.1	0.0	19.8	4309
	12	56.8	43.1	0.0	0.2	18009
	8	28.8	69.2	0.0	2.0	5848
	11	59.5	40.2	0.0	0.3	5568
SUMMARY	A	0.0	13.2	3.9	82.8	5811
	B	0.1	32.8	5.7	61.4	14080
	C	33.4	55.0	0.0	11.6	43812

Appendix B. Comparison of Relief Classification with Hierarchical Classification (Average Linkage with 12 classes).

Note classes are presented in the same order that they occur in the classification dendrogram.

CLASS		flat%	low%	mod%	Video observations
A	9	58.2	39.6	2.2	1276
	10	23.1	76.9	0.0	4590
B	2	72.7	26.5	0.8	3053
	3	75.4	17.7	6.8	2960
	4	37.2	40.1	22.7	7199
	6	64.3	28.2	7.5	333
	7	74.3	20.7	5.0	623
C	1	82.9	16.9	0.2	10078
	5	91.1	8.8	0.1	4309
	12	99.9	0.1	0.0	18009
	8	98.5	1.5	0.0	5848
	11	99.7	0.3	0.0	5568
SUMMARY	A	40.6	58.3	1.1	5866
	B	64.8	26.6	8.6	14168
	C	94.4	5.5	0.1	43812

Appendix C. Comparison of Biological Classification with Hierarchical Classification (Average Linkage with 12 classes).

Note classes are presented in the same order that they occur in the classification dendrogram.

CLASS		Bioturbated %	mixed patches %	mixed gardens (Hard Corals) %	mixed gardens %	Video Observations
A	9	0.0	72.3	0.0	27.7	1276
	10	0.0	0.0	47.5	52.5	4590
B	2	6.7	93.3	0.0	0.0	3053
	3	0.4	98.0	10.7	0.9	2960
	4	0.0	7.9	43.9	48.3	7199
	6	0.0	91.9	3.0	5.1	333
	7	0.0	100.0	0.0	0.0	623
C	1	31.8	68.2	0.0	0.0	10078
	5	45.1	54.9	0.0	0.0	4309
	12	99.5	0.5	0.0	0.0	18009
	8	91.9	8.1	0.0	0.0	5848
	11	98.3	1.7	0.0	0.0	5568
SUMMARY	A	0.0	36.2	23.7	40.1	5866
	B	1.4	78.2	11.5	10.9	14168
	C	73.3	26.7	0.0	0.0	43812

Appendix D. Summary statistics describing sediment samples and predictor parameters at sediment sample locations.

*Reassigned to different class.

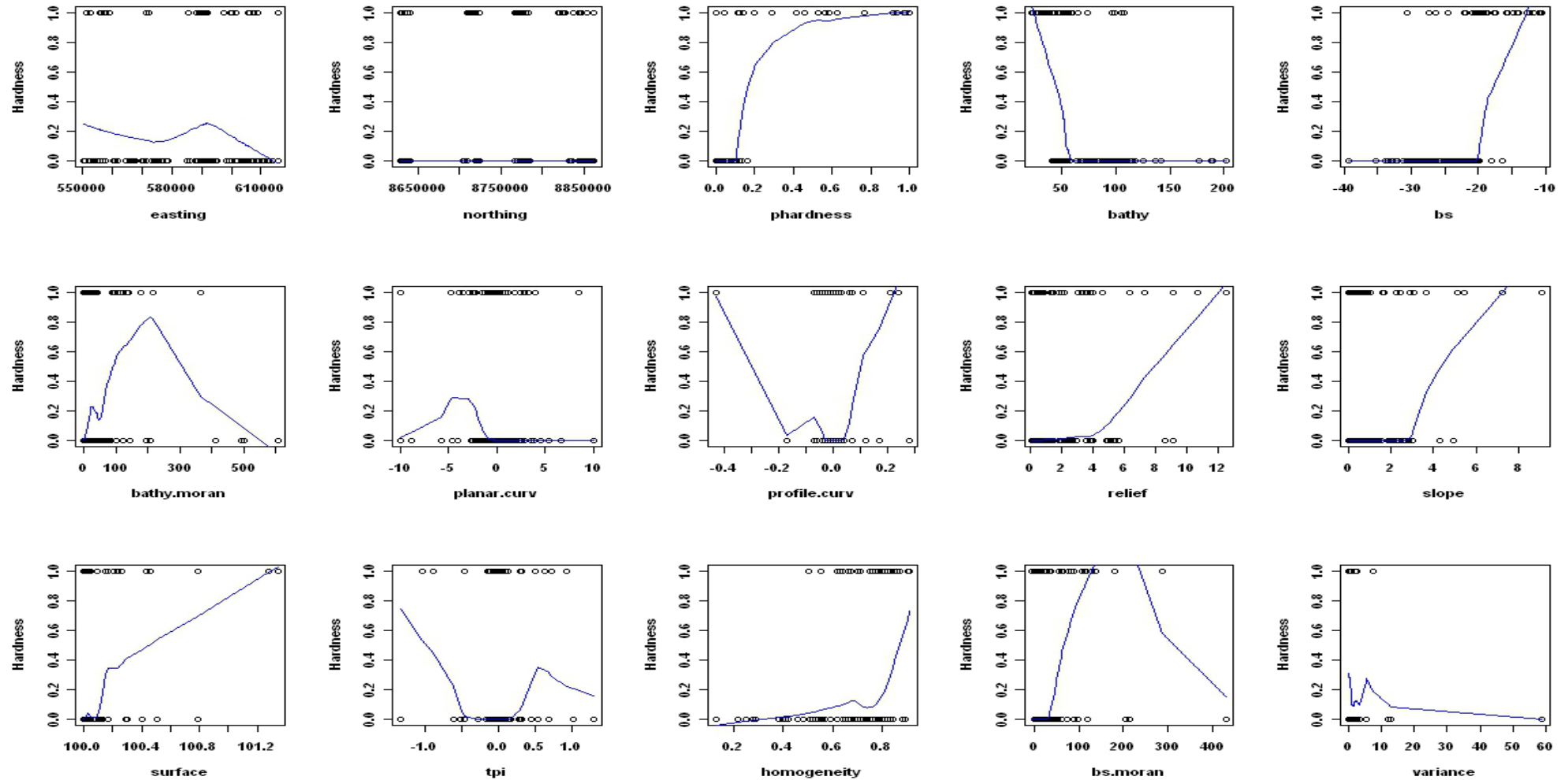
Sample ID	Video class	Area	Easting [m]	Northing [m]	Mud [%]	Sand [%]	Gravel [%]	Mean grain size [μ]	p-value of hard	Bathy [m]	BS [dB]	Bathy local Moran	Planar curvature	Profile curvature	Relief	Slope	Surface	TPI	Homo-geneity	BS local Moran	Variance
SOL4934/001GR056	hard	C	591803.53	8712946.40	0.16	53.98	45.86	678.72	1	22.12	-10.52	29.85	3.95	-0.02	1.53	0.39	100.02	-0.16	0.85	83.96	0.12
SOL4934/002GR001	hard	C	590054.20	8718688.78	0.38	33.97	65.65	565.32	0.119	32.83	-15.63	13.59	-1.54	-0.02	0.73	0.53	100.01	-0.04	0.90	12.71	0.07
SOL4934/002GR052	hard	C	590454.81	8718470.87	12.25	71.69	16.06	310.67	0.042	38.65	-15.78	7.38	0.34	0.01	1.85	0.85	100.02	-0.03	0.87	9.42	0.16
SOL4934/003GR002	soft	C	591296.64	8720738.75	19.92	53.40	26.68	145.53	0	48.17	-16.52	1.24	0.85	-0.02	1.48	1.24	100.03	-0.28	0.71	1.05	0.52
SOL4934/004GR003	soft	C	592835.80	8723871.55	0.90	81.44	17.66	619.63	0	92.35	-23.38	38.68	-0.20	-0.02	8.60	3.01	100.11	0.09	0.70	24.30	0.60
SOL4934/031GR049	hard	C	589306.24	8713387.21	3.29	44.34	52.37	478.03	0.200	35.85	-15.66	10.27	-0.56	0.06	3.84	2.88	100.15	0.29	0.77	10.34	0.33
SOL4934/032GR054	hard	C	589865.80	8724060.49	18.26	53.47	28.27	171.31	0	65.01	-20.80	2.62	-0.08	0.11	10.69	7.20	100.79	0.72	0.62	6.35	0.77
SOL4934/034GR060	hard	C	590845.00	8715069.18	0.71	50.77	48.52	494.86	1	25.81	-11.80	23.68	-2.20	0.00	0.67	0.26	100.00	0.09	0.91	59.67	0.09
SOL4934/041GR077	hard	C	591057.53	8715049.76	0.29	35.08	64.63	641.26	1	25.16	-12.33	24.59	-0.77	-0.01	0.62	0.25	100.01	-0.11	0.86	54.20	0.18
SOL4934/003GR076A1	soft	C	591169.08	8720371.97	22.87	64.15	12.98	116.50	0	56.92	-19.93	0.11	1.99	-0.02	4.02	2.35	100.10	-0.17	0.84	2.45	0.25
SOL4934/005GR005A1	soft	A	576759.52	8842744.41	24.28	64.44	11.28	86.31	0	79.93	-22.77	2.37	10.00	0.04	2.61	0.21	100.08	-0.52	0.71	8.17	0.65
SOL4934/006GR008A1	soft	A	578198.49	8853327.63	6.17	92.31	1.52	197.69	0	88.54	-29.25	0.00	0.26	0.00	1.00	0.69	100.01	-0.02	0.51	5.78	1.42
SOL4934/007GR009A1	soft	A	574293.39	8859988.29	19.78	77.44	2.79	103.15	0	101.17	-29.23	5.84	-1.50	0.00	0.31	0.18	100.00	0.06	0.57	6.72	1.14
SOL4934/008GR012A1	soft	A	570561.60	8852868.47	11.77	86.97	1.26	134.42	0	94.18	-29.13	1.22	-0.09	0.00	2.27	2.08	100.07	-0.02	0.71	4.44	0.34
SOL4934/009GR014A1	soft	A	566231.78	8850495.25	14.74	80.67	4.58	156.30	0	92.54	-27.67	0.63	0.57	0.00	0.98	0.79	100.01	-0.03	0.79	0.79	0.46
SOL4934/010GR016A1	hard	A	572116.77	8846853.40	29.44	69.87	0.69	87.06	0	106.99	-30.62	12.40	-0.98	0.00	2.25	1.63	100.05	-0.01	0.62	7.83	0.77
SOL4934/011GR018A1	soft	A	560653.92	8845137.18	6.88	89.18	3.94	273.49	0	94.42	-27.67	1.27	-0.74	0.07	3.61	2.73	100.13	0.28	0.62	1.16	1.01
SOL4934/012GR020A1	soft	A	561002.50	8853540.01	24.53	75.20	0.27	99.06	0	106.99	-33.89	12.13	-4.09	0.04	1.47	0.20	100.02	0.44	0.42	35.18	2.42
SOL4934/013GR022A1	soft	A	561641.19	8860626.50	7.11	89.95	2.94	354.10	0	90.13	-25.61	0.11	1.58	0.02	0.63	0.23	100.00	0.07	0.76	0.10	0.24
SOL4934/014GR024A1	soft	A	567496.69	8860421.11	11.77	77.55	10.68	251.68	0	89.13	-25.84	0.02	-0.17	0.03	0.93	0.80	100.01	0.10	0.81	0.27	0.14
SOL4934/018GR026A1	soft	A	575621.16	8850194.12	10.55	84.79	4.66	161.70	0	99.02	-27.46	4.07	1.79	0.00	0.25	0.18	100.00	-0.02	0.59	1.26	0.56
SOL4934/019GR028A1	soft	A	575666.22	8855818.93	17.28	81.70	1.02	92.37	0	108.30	-30.56	14.18	-1.03	-0.01	0.33	0.18	100.00	0.00	0.82	11.36	0.25
SOL4934/021GR030A1	soft*	B	600160.30	8768380.19	20.42	65.91	13.66	115.12	1	37.34	-17.70	217.95	-0.44	0.01	2.05	1.68	100.04	0.04	0.90	105.75	0.12
SOL4934/022GR032A1	hard*	B	586230.90	8783907.23	16.54	55.04	28.42	136.81	0	79.22	-26.39	0.06	0.23	0.03	5.70	4.94	100.40	1.03	0.89	27.97	0.12
SOL4934/023GR034A1	soft	B	588974.50	8779968.72	11.86	54.12	34.02	78.93	0.001	74.37	-22.95	3.97	-4.64	0.17	3.92	1.92	100.51	1.30	0.72	0.03	0.24
SOL4934/024GR036A1	soft	B	584957.21	8779360.15	17.98	75.31	6.71	95.38	0	103.40	-25.78	69.17	-0.67	-0.01	1.13	0.85	100.01	0.04	0.56	24.62	0.81
SOL4934/025GR038A1	soft	B	590517.00	8766241.15	7.22	89.48	3.30	249.20	0.063	69.52	-23.06	12.30	0.05	0.01	1.08	0.94	100.01	-0.04	0.76	0.67	0.38
SOL4934/026GR040A1	soft	B	594887.62	8773523.86	16.77	76.22	7.00	108.53	0.094	78.63	-22.77	0.11	0.57	0.00	0.54	0.38	100.00	0.02	0.74	1.86	0.27
SOL4934/027GR042A1	soft	B	598219.61	8776654.57	21.42	67.54	11.04	74.33	0.034	74.79	-21.33	2.81	0.70	0.00	0.32	0.21	100.00	0.03	0.70	13.96	0.36
SOL4934/028GR044A1	hard*	B	587972.66	8770229.15	15.64	48.95	35.42	45.74	0	108.62	-26.75	102.69	-8.77	0.03	0.40	0.06	100.00	0.10	0.56	28.54	1.33
SOL4934/029GR046A1	soft	C	588171.85	8705528.86	10.84	67.32	21.84	141.56	0	53.70	-21.12	0.05	-0.16	0.02	1.94	1.55	100.04	0.06	0.84	6.89	0.09
SOL4934/030GR048A1	soft	C	590107.52	8708911.65	0.40	93.84	5.75	758.76	0.021	46.88	-18.13	1.91	-1.31	0.12	2.72	1.07	100.12	0.54	0.57	0.20	0.55
SOL4934/033GR059A1	hard	C	590620.33	8711686.99	0.18	38.08	61.75	793.64	1	22.51	-10.89	29.70	-0.64	0.21	4.64	3.10	100.26	0.63	0.55	65.01	7.68
SOL4934/035GR063A1	soft	C	592647.21	8724260.27	0.30	27.88	71.83	351.87	0	84.96	-20.66	24.92	-0.30	0.03	3.87	2.58	100.13	0.10	0.82	4.97	0.26

Sample ID	Video class	Area	Easting [m]	Northing [m]	Mud [%]	Sand [%]	Gravel [%]	Mean grain size [μ]	ρ -value of hard	Bathy [m]	BS [dB]	Bathy local Moran	Planar curvature	Profile curvature	Relief	Slope	Surface	TPI	Homo-geneity	BS local Moran	Variance
SOL4934/036GR065A1	soft	C	592807.03	8722263.70	29.19	63.32	7.50	98.97	0.162	188.44	-33.52	494.27	-1.15	-0.04	5.15	2.27	100.08	0.10	0.29	213.56	12.82
SOL4934/037GR067A1	soft	C	592258.26	8723647.66	43.02	49.04	7.95	50.13	0	141.58	-25.53	207.64	-0.24	0.03	5.46	4.28	100.30	0.28	0.69	52.48	0.82
SOL4934/038GR070A1	soft	C	591649.69	8705196.89	7.18	79.64	13.18	264.06	0	67.13	-22.54	4.13	-0.45	-0.06	2.56	1.88	100.06	-0.11	0.83	17.07	0.19
SOL4934/039GR072A1	soft*	C	591402.36	8723696.63	0.10	98.76	1.14	459.93	n/a	176.66	n/a	410.74	-0.48	0.00	1.90	1.43	100.03	0.13	n/a	n/a	n/a
SOL4934/040GR074A1	soft	C	594617.38	8719814.28	14.04	49.17	36.78	163.67	0	189.57	-39.40	501.03	2.85	0.28	9.15	2.26	100.79	0.69	0.13	429.94	58.69
SOL4934/042GR080A1	soft*	D	606063.08	8641404.10	18.26	70.87	10.87	145.94	0.459	48.12	-19.86	16.76	2.90	-0.01	0.33	0.13	100.00	-0.05	0.81	3.12	0.16
SOL4934/043GR082A1	soft	D	603495.89	8641258.46	13.79	83.38	2.83	246.22	0.001	55.12	-23.29	195.15	-2.01	0.00	0.58	0.36	100.00	0.08	0.82	97.30	0.33
SOL4934/045GR086A1	soft	D	603620.85	8636627.36	15.05	68.85	16.10	193.72	0.009	49.37	-20.29	34.18	-0.52	-0.01	0.36	0.27	100.00	-0.01	0.78	-0.44	0.19
SOL4934/046GR088A1	soft	D	603557.88	8634208.85	13.26	78.57	8.16	236.35	0.009	48.35	-20.24	19.07	-0.61	0.00	0.26	0.09	100.00	0.02	0.79	-0.28	0.29
SOL4934/047GR090A1	soft	D	606376.93	8630252.58	22.84	61.23	15.93	144.36	0.009	46.33	-19.90	2.27	-2.64	-0.01	0.71	0.39	100.00	0.01	0.82	0.34	0.18
SOL4934/048GR092A1	soft*	D	605454.11	8628508.51	21.18	66.66	12.16	135.90	0.524	45.83	-18.75	0.61	0.84	0.01	0.28	0.20	100.00	0.03	0.82	28.27	0.26
SOL4934/049GR094A1	soft*	D	608798.54	8641454.89	15.80	66.19	18.01	112.71	0.934	47.61	-18.57	10.93	-3.84	0.01	0.25	0.11	100.00	0.05	0.75	28.12	0.44
SOL4934/050GR096A1	soft	D	610358.37	8638917.47	17.35	77.45	5.20	115.59	0.033	45.42	-21.14	0.06	-2.13	0.01	0.56	0.29	100.00	-0.03	0.80	8.05	0.31
SOL4934/051GR098A1	soft	D	608311.21	8635856.05	15.73	78.46	5.82	186.66	0.006	47.08	-21.82	6.79	0.78	-0.02	1.83	1.20	100.02	-0.06	0.79	21.68	0.29
SOL4934/052GR100A1	soft*	D	611242.62	8634669.35	16.93	78.82	4.25	176.39	0.066	42.54	-21.28	15.11	1.65	0.00	0.28	0.14	100.00	0.00	0.79	3.82	0.30
SOL4934/053GR102A1	soft*	D	605615.25	8634559.79	63.48	35.42	1.11	19.26	0.138	53.55	-18.62	138.32	-2.54	0.03	1.40	0.81	100.04	0.09	0.50	-4.94	2.51
SOL4934/054GR104A1	soft	D	607449.00	8633520.11	23.58	68.76	7.66	95.02	0.010	46.36	-21.72	2.49	-0.15	0.00	0.27	0.11	100.00	-0.02	0.73	4.76	0.48
SOL4934/055GR106A1	soft*	D	607601.40	8630926.03	17.69	69.87	12.44	173.93	0.195	45.86	-19.23	0.78	2.35	-0.01	0.58	0.39	100.00	-0.08	0.76	12.78	0.50
SOL4934/056GR108A1	soft	D	610607.01	8630511.05	20.72	76.81	2.48	147.03	0.005	43.22	-20.60	8.24	0.27	-0.01	0.42	0.31	100.00	-0.07	0.78	2.67	0.17
SOL4934/057GR110A1	soft	D	614136.80	8629045.21	16.74	77.28	5.99	135.26	0.001	40.67	-21.65	42.77	-1.76	0.00	0.17	0.07	100.00	-0.01	0.84	18.67	0.16
SOL4934/058GR112A1	soft	D	609504.12	8635291.97	18.22	78.31	3.47	118.32	0.004	44.30	-22.65	1.83	1.27	-0.01	0.27	0.17	100.00	-0.04	0.79	43.43	0.29
SOL4934/059GR114A1	soft	D	604312.28	8629007.01	20.56	70.01	9.43	175.63	0.094	47.73	-20.50	12.10	-0.26	0.01	0.43	0.33	100.00	0.05	0.78	0.46	0.20
SOL4934/060GR116A1	soft	D	610950.45	8632472.87	18.77	77.61	3.62	110.29	0.104	43.31	-20.57	8.20	-0.10	0.01	0.43	0.30	100.00	0.10	0.79	0.97	0.24
SOL4934/063GR118A1	soft*	C	589859.17	8708968.77	0.13	59.71	40.16	756.80	0.416	28.26	-12.70	19.88	0.31	0.07	7.37	5.10	100.43	0.12	0.62	37.75	1.15
SOL4934/044GR084A1	soft	D	603787.83	8637409.79	17.88	79.29	2.83	183.71	0.001	51.40	-22.97	76.53	1.25	0.00	0.22	0.09	100.00	-0.02	0.74	71.93	0.19
SOL5117/003GR006	soft	A	567341.32	8850832.57	18.98	78.52	2.50	102.42	0	104.65	-26.90	9.51	-0.89	-0.01	1.08	0.78	100.01	-0.06	0.25	0.23	n/a
SOL5117/022GR044	soft	A	571571.82	8844801.17	24.79	74.65	0.56	98.29	0	96.63	-29.74	2.44	-2.50	0.00	0.83	0.58	100.01	0.07	0.69	7.70	0.57
SOL5117/023GR046	soft	A	573146.73	8844772.45	20.52	66.60	12.87	126.68	0	84.65	-24.71	0.48	10.00	0.01	2.04	0.43	100.04	-0.05	0.53	2.24	2.02
SOL5117/024GR048	soft	A	578537.25	8844833.60	14.28	79.80	5.92	212.09	0.054	73.23	-25.70	8.12	-0.69	0.00	0.61	0.44	100.00	0.01	0.77	0.81	0.34
SOL5117/025GR050	soft	A	573101.44	8847166.47	7.35	85.06	7.59	267.50	0	75.27	-24.10	6.05	2.79	-0.01	0.32	0.19	100.00	-0.05	0.77	3.86	0.32
SOL5117/026GR052	soft	A	567814.43	8847577.50	16.67	81.75	1.58	145.15	0	97.85	-28.79	3.23	2.46	0.00	0.26	0.16	100.00	-0.02	0.64	3.65	0.90
SOL5117/031GR061	soft	A	573056.62	8856576.38	17.81	81.82	0.37	134.07	0	101.45	-33.50	6.18	0.33	-0.04	1.56	1.12	100.03	-0.14	0.54	26.13	1.70
SOL5117/027GR054	soft	A	571451.59	8861425.01	16.90	81.85	1.25	194.48	0	94.28	-28.97	1.26	0.17	0.00	0.34	0.28	100.00	-0.02	0.82	3.60	0.16
SOL5117/028GR056	hard	A	571252.35	8861450.85	16.44	76.38	7.18	145.76	0	95.64	-27.45	2.11	-0.61	-0.43	9.15	5.46	101.34	-1.05	0.66	0.46	0.77
SOL5117/029GR058	soft	A	574744.42	8861484.49	24.55	75.11	0.34	114.78	0	104.04	-32.16	8.76	-0.55	0.00	1.33	0.99	100.02	0.04	0.52	23.29	1.29
SOL5117/030GR060	soft	A	575561.04	8861358.90	35.60	63.73	0.67	103.19	0	104.53	-33.11	9.34	0.07	0.00	0.64	0.44	100.00	0.00	0.28	15.87	5.60
SOL5117/032GR064	soft	A	560704.08	8861363.43	12.57	71.39	16.04	199.89	0	124.97	-27.97	47.77	-0.28	0.03	2.64	1.93	100.07	0.04	0.38	0.96	5.39
SOL5117/033GR066	soft	A	576738.08	8858813.15	27.12	68.74	4.15	133.80	0	103.06	-27.73	7.74	0.49	-0.01	0.60	0.46	100.00	-0.04	0.66	1.04	0.85

Sample ID	Video class	Area	Easting [m]	Northing [m]	Mud [%]	Sand [%]	Gravel [%]	Mean grain size [μ]	ρ -value of hard	Bathy [m]	BS [dB]	Bathy local Moran	Planar curvature	Profile curvature	Relief	Slope	Surface	TPI	Homo-geneity	BS local Moran	Variance
SOL5117/034GR068	soft	A	574079.15	8854998.56	27.28	66.98	5.74	162.84	0	87.47	-26.58	0.03	-10.00	0.00	0.20	0.03	100.00	0.02	0.73	0.00	0.27
SOL5117/035GR070	hard	A	559003.70	8843362.05	6.61	82.39	10.99	436.82	0.565	54.85	-20.50	40.04	-0.19	0.00	0.95	0.32	100.01	0.13	0.82	24.55	0.22
SOL5117/036GR071	soft	A	564312.72	8848060.74	15.09	84.46	0.45	110.99	0	117.43	-30.03	30.10	0.37	0.03	2.30	1.54	100.05	0.06	0.54	8.90	1.87
SOL5117/037GR074	soft	A	565808.48	8848935.88	15.88	80.14	3.98	161.88	0	92.43	-28.54	0.60	-2.14	-0.01	0.80	0.53	100.01	0.02	0.72	3.12	0.39
SOL5117/038GR075	soft	A	566672.57	8856028.55	40.24	59.71	0.05	96.63	0	105.93	-31.19	11.09	-0.05	0.00	0.87	0.30	100.00	-0.10	0.41	20.75	3.48
SOL5117/039GR077	soft	A	560294.98	8860335.82	28.25	67.23	4.52	125.67	0	115.19	-27.87	25.62	0.19	0.01	0.65	0.47	100.00	0.07	0.55	0.79	1.11
SOL5117/040GR079	soft	A	569914.71	8861127.35	15.21	82.94	1.85	105.43	0	113.51	-33.45	22.58	0.19	-0.01	0.71	0.36	100.00	-0.03	0.22	9.07	n/a
SOL5117/041GR087	soft	A	556533.59	8858155.01	41.88	57.97	0.16	91.99	0	103.35	-35.21	8.10	1.47	0.01	1.10	0.38	100.01	-0.12	0.46	44.72	1.49
SOL5117/042GR086	soft	A	555155.71	8856440.06	28.68	71.10	0.22	80.88	0	99.27	-35.23	4.26	-2.22	-0.03	1.29	0.31	100.00	0.00	0.52	50.27	2.45
SOL5117/043GR084	soft	A	557713.05	8848414.02	13.05	84.73	2.22	155.13	0	90.13	-30.21	0.14	-0.66	-0.07	2.26	1.36	100.06	-0.47	0.57	10.30	1.29
SOL5117/044GR082	hard	A	557185.67	8844559.28	1.03	87.31	11.67	380.37	0.911	58.03	-21.82	32.91	-4.72	0.07	1.53	0.83	100.04	0.31	0.80	16.70	0.35
SOL5117/049GR098	soft	A	552825.88	8844881.34	17.31	80.87	1.82	146.60	0	136.97	-33.15	84.23	1.23	0.02	2.00	1.22	100.02	-0.04	0.55	26.53	1.91
SOL5117/048GR096	soft	A	553762.44	8851046.41	12.06	80.48	7.46	290.87	0	90.95	-25.79	0.22	-0.69	0.07	4.94	2.61	100.29	0.30	0.65	0.52	1.74
SOL5117/047GR094	soft	A	555044.85	8852792.52	13.71	80.08	6.21	206.61	0	85.36	-27.91	0.31	3.76	-0.02	0.40	0.18	100.00	-0.09	0.78	1.68	0.27
SOL5117/046GR092	hard	A	557552.92	8853507.18	6.05	53.87	40.08	288.48	0	105.16	-24.52	9.63	-2.96	0.24	3.57	0.91	100.23	0.93	0.70	3.34	0.49
SOL5117/045GR090	soft*	A	557125.14	8837098.02	0.91	84.95	14.14	449.39	0.116	55.19	-22.05	39.07	-0.31	-0.07	3.30	2.43	100.17	-0.05	0.66	13.18	0.53
SOL5117/050GR100	hard	A	555920.63	8819234.87	1.02	59.36	39.62	511.53	1	33.73	-12.81	106.00	1.16	-0.03	0.62	0.29	100.00	-0.07	0.83	129.12	0.27
SOL5117/051GR102	hard	A	557733.71	8823621.56	0.60	90.91	8.49	553.16	1	37.23	-15.74	93.04	-10.00	0.00	0.24	0.05	100.00	0.06	0.81	89.89	0.37
SOL5117/052GR103	hard	A	551321.98	8825072.61	0.64	57.73	41.63	702.69	1	27.12	-10.71	133.49	-3.42	0.01	0.73	0.24	100.01	0.09	0.85	179.79	0.26
SOL5117/053GR105	soft	A	552199.03	8833521.14	16.52	70.14	13.34	165.48	0	67.67	-24.30	15.40	-1.07	0.02	1.53	1.02	100.04	0.31	0.82	4.27	0.24
SOL5117/054GR107	soft	A	552668.58	8831510.24	38.25	61.49	0.26	67.76	0	97.15	-32.85	2.91	0.76	-0.17	4.98	0.98	100.07	-0.61	0.48	32.69	1.81
SOL5117/055GR110	hard	A	552647.09	8827510.90	0.49	55.70	43.80	423.72	1	31.89	-14.85	112.51	8.41	-0.04	0.83	0.23	100.02	-0.47	0.70	112.16	0.38
SOL5117/056GR111	hard	A	556334.12	8821265.41	1.50	58.93	39.57	469.14	1	33.70	-12.77	106.25	-1.26	0.00	0.16	0.01	100.00	0.01	0.82	138.85	0.18
SOL5117/057GR113	hard	A	556391.79	8823715.60	2.44	51.72	45.84	364.87	0.769	59.03	-19.81	30.98	-0.20	0.11	12.53	9.08	101.28	0.50	0.75	34.56	0.38
SOL5117/059GR115	soft	A	550734.43	8835267.05	19.34	69.19	11.47	324.40	0.005	62.95	-23.36	23.05	1.22	0.04	1.90	1.24	100.02	0.14	0.75	8.63	0.32
SOL5117/060GR117	hard	A	552534.00	8842944.59	26.72	67.27	6.01	97.54	0	98.42	-26.38	3.66	0.63	-0.06	3.10	2.27	100.09	-0.14	0.67	-0.10	0.61
SOL5117/061GR119	soft	A	550757.14	8852959.36	14.11	72.11	13.78	147.26	0	112.85	-29.58	21.28	-1.38	0.02	1.40	0.99	100.02	0.17	0.66	6.01	0.53
SOL5117/062GR121	soft	A	555654.31	8858784.46	25.54	73.40	1.06	123.85	0	97.96	-32.81	3.32	10.00	0.02	1.20	0.15	100.03	-0.06	0.67	28.25	0.58
SOL5117/063GR123	soft	A	550427.61	8860532.74	27.99	68.74	3.27	145.40	0	102.75	-30.29	7.34	0.48	0.02	2.44	2.03	100.07	0.12	0.72	11.87	0.89
SOL5117/064GR126	hard	B	585693.52	8783549.25	20.41	60.28	19.30	127.72	0.626	51.53	-20.38	96.30	-0.26	0.02	0.71	0.29	100.00	0.09	0.78	24.88	0.17
SOL5117/065GR127	hard	B	588447.96	8779564.31	22.89	59.40	17.71	64.82	0.584	52.36	-20.41	90.53	-1.30	-0.02	1.38	0.70	100.02	0.03	0.64	27.83	0.59
SOL5117/066GR129	soft	B	595287.73	8781499.09	19.93	73.11	6.95	99.29	0	77.24	-24.18	0.67	3.55	-0.01	0.39	0.12	100.00	-0.02	0.73	1.87	0.45
SOL5117/067GR131	soft	B	599270.34	8782377.47	28.55	60.14	11.31	62.91	0	78.67	-24.02	0.10	-0.09	0.00	0.53	0.37	100.00	-0.02	0.84	1.71	0.13
SOL5117/068GR133	soft	B	601406.50	8781052.75	23.36	68.90	7.74	86.68	0	77.85	-24.07	0.36	0.78	0.00	0.67	0.45	100.00	-0.02	0.74	1.99	0.44
SOL5117/069GR135	soft	B	602069.72	8777928.86	24.92	62.52	12.56	58.55	0.008	74.94	-24.20	2.70	-10.00	0.02	0.39	0.06	100.00	0.15	0.71	0.32	0.46
SOL5117/070GR137	soft	B	595182.90	8777767.20	15.63	77.96	6.41	158.85	0.001	76.10	-23.40	1.48	-1.41	0.01	0.25	0.12	100.00	0.02	0.76	-0.01	0.25
SOL5117/071GR139	soft*	B	597440.92	8774877.60	20.56	73.01	6.43	183.24	0.291	73.76	-22.06	4.20	-0.70	0.02	0.55	0.37	100.00	0.13	0.80	7.16	0.36
SOL5117/073GR142	hard	B	599771.14	8770026.90	3.91	78.24	17.85	518.77	1	25.04	-13.95	363.32	-0.21	-0.02	4.04	3.67	100.21	0.04	0.63	286.09	2.43

Sample ID	Video class	Area	Easting [m]	Northing [m]	Mud [%]	Sand [%]	Gravel [%]	Mean grain size [μ]	ρ -value of hard	Bathy [m]	BS [dB]	Bathy local Moran	Planar curvature	Profile curvature	Relief	Slope	Surface	TPI	Homo-geneity	BS local Moran	Variance
SOL5117/074GR143	hard	B	600285.80	8767293.85	7.70	79.10	13.20	262.20	1	41.48	-17.46	178.49	-1.54	0.01	1.49	1.03	100.03	0.31	0.75	116.05	0.26
SOL5117/075GR145	soft*	B	601746.96	8772173.87	32.33	53.98	13.70	67.20	0.972	48.14	-18.90	120.64	0.18	0.00	0.73	0.50	100.00	0.00	0.75	55.13	0.22
SOL5117/002GR004	hard	C	591095.05	8712981.67	0.19	14.37	85.45	n/a	1	23.48	-11.00	27.56	-2.41	0.00	0.49	0.25	100.00	0.03	0.86	75.20	0.08
SOL5117/081GR147	hard	C	591559.11	8715448.61	2.72	52.07	45.20	160.56	0.973	30.32	-14.23	16.40	1.85	-0.05	6.36	5.10	100.46	-0.90	0.71	26.63	0.34
SOL5117/082GR148	hard	C	589956.93	8723518.36	15.98	44.04	39.97	97.42	0	48.65	-19.76	1.11	3.25	-0.01	2.23	0.33	100.04	-0.02	0.63	0.87	0.54
SOL5117/083GR149	soft	C	592776.41	8721867.89	7.43	77.20	15.38	446.12	0.003	202.04	-32.08	606.16	-0.35	-0.07	5.17	2.82	100.17	-1.35	0.25	205.11	11.96
SOL5117/001GR001	soft	D	605244.01	8638649.89	14.15	80.61	5.24	172.94	0.132	48.30	-20.17	19.00	-0.50	0.02	0.48	0.20	100.00	-0.04	0.88	0.26	0.09
SOL5117/004GR008	soft	D	603680.19	8628567.99	23.07	68.45	8.47	133.35	0.003	47.44	-20.64	9.76	0.99	0.00	0.58	0.13	100.00	-0.05	0.81	1.25	0.16
SOL5117/005GR010	soft	D	604750.14	8628638.25	14.90	79.43	5.67	172.51	0.075	47.10	-20.31	6.80	-0.13	-0.01	0.44	0.29	100.00	-0.01	0.83	-0.14	0.14
SOL5117/007GR014	soft	D	605366.89	8633834.07	15.87	78.76	5.37	193.69	0.012	48.02	-20.84	15.34	5.33	0.01	0.20	0.03	100.00	0.00	0.73	2.56	0.50
SOL5117/008GR016	soft	D	605156.15	8635696.18	16.95	79.39	3.66	211.79	0.005	49.67	-22.94	39.46	6.67	-0.02	0.74	0.19	100.00	-0.05	0.72	44.79	0.37
SOL5117/009GR018	soft	D	603689.21	8635976.82	16.05	72.32	11.63	160.47	0.004	49.17	-21.48	30.90	0.17	0.00	0.60	0.50	100.00	-0.01	0.84	6.88	0.14
SOL5117/006GR012	soft	D	607158.99	8628592.95	15.52	74.29	10.19	147.94	0.118	45.68	-20.16	0.34	1.09	-0.01	0.57	0.43	100.00	-0.02	0.81	0.66	0.19
SOL5117/010GR020	soft	D	606502.90	8639730.28	29.04	69.53	1.43	83.92	0.007	51.49	-24.08	79.09	0.13	0.01	0.66	0.44	100.00	-0.03	0.62	119.40	0.80
SOL5117/011GR022	soft	D	604915.53	8640449.37	12.58	70.49	16.93	210.05	0.098	51.56	-21.52	80.84	2.29	0.00	0.58	0.35	100.00	-0.05	0.80	16.72	0.20
SOL5117/012GR024	soft	D	603752.90	8641357.08	4.51	87.14	8.35	320.85	0.046	53.04	-22.51	124.74	2.56	0.00	2.38	1.31	100.04	-0.17	0.70	40.24	0.59
SOL5117/013GR026	soft	D	602115.32	8638327.00	11.69	74.96	13.36	153.47	0.004	53.63	-23.29	143.20	0.14	0.00	3.00	2.31	100.10	-0.11	0.64	88.94	1.25
SOL5117/014GR028	soft	D	603591.87	8637653.80	15.28	79.54	5.18	173.41	0.001	51.34	-23.08	73.25	-0.67	0.01	0.81	0.51	100.01	0.13	0.74	61.68	0.30
SOL5117/015GR030	soft	D	612456.52	8637315.74	16.82	80.84	2.34	160.84	0.019	43.52	-21.80	6.58	0.85	0.02	0.57	0.31	100.00	0.10	0.84	17.86	0.14
SOL5117/016GR032	soft	D	614116.42	8636875.75	14.58	81.96	3.45	194.48	0.001	42.11	-22.53	20.02	4.58	0.00	0.40	0.20	100.00	-0.05	0.79	35.09	0.28
SOL5117/017GR034	soft	D	615551.42	8635395.80	16.67	73.97	9.36	210.93	0.003	43.49	-21.11	6.20	-5.75	-0.01	1.46	0.69	100.02	-0.02	0.57	11.58	1.38
SOL5117/018GR036	soft*	D	615548.22	8638156.39	27.80	60.97	11.22	124.54	0.287	45.32	-20.30	0.01	0.37	-0.01	0.69	0.49	100.00	-0.02	0.82	0.67	0.19
SOL5117/019GR038	soft	D	614081.68	8629402.67	17.22	76.51	6.27	199.24	0.001	40.69	-21.15	42.71	-0.84	0.00	0.45	0.20	100.00	0.02	0.81	11.84	0.11
SOL5117/020GR040	soft	D	605118.35	8640296.02	13.10	79.59	7.32	244.52	0.105	50.49	-20.62	54.16	-1.28	0.01	1.44	1.05	100.02	0.12	0.83	-0.23	0.20
SOL5117/021GR042	soft	A	565770.72	8850959.42	9.22	87.52	3.25	168.25	0	91.42	-29.52	0.34	0.16	0.00	0.39	0.29	100.00	0.01	0.73	5.38	0.26
SOL5117/076VC018	soft	B	600211.47	8775051.66	n/a	n/a	n/a	n/a	0.013	74.27	-22.01	3.46	-0.43	0.01	0.48	0.37	100.00	0.06	0.80	4.51	0.12
SOL5117/077VC019	soft	B	600154.05	8780188.51	n/a	n/a	n/a	n/a	0.014	77.20	-22.48	0.66	-0.38	0.00	0.19	0.08	100.00	-0.08	0.57	4.60	1.10
SOL5117/078VC020	soft	B	599256.78	8782590.94	n/a	n/a	n/a	n/a	0	79.15	-25.12	0.02	-0.77	0.01	0.38	0.21	100.00	0.03	0.40	8.88	2.29
SOL5117/079VC021	soft	B	595175.97	8782516.79	n/a	n/a	n/a	n/a	0	78.00	-24.47	0.32	-0.65	0.01	0.42	0.21	100.00	0.09	0.73	1.67	0.23
SOL5117/080VC022	soft	B	593768.73	8776396.74	n/a	n/a	n/a	n/a	0	81.98	-24.60	0.74	0.09	-0.01	0.39	0.13	100.00	-0.11	0.70	2.52	0.70

Appendix E. The relationship of the hardness of seabed (i.e. the converted variable from soft and hard to 0 and 1) and relevant predictors.



Appendix F. The relationship among the predictors and hardness of seabed (i.e. the converted variable from soft and hard to 0 and 1).

