Predictive Modelling of Seabed Sediment Parameters Using Multibeam Acoustic Data: A Case Study on the Carnarvon Shelf, Western Australia

Z. Huang[1], S. Nichol[2], J. Daniell[3], J. Siwabessy[4], B. Brooke[5]

[1]Geoscience Australia, GPO Box 378, Canberra ACT 2601, Australia
Telephone: +61 2 62495876
Fax: +61 62499920
Email: Zhi.Huang@ga.gov.au

[2]Geoscience Australia, GPO Box 378, Canberra ACT 2601, Australia
Telephone: +61 2 62499346
Fax: +61 62499920
Email: Scott.Nichol@ga.gov.au

[3]Geoscience Australia, GPO Box 378, Canberra ACT 2601, Australia
Telephone: +61 2 62499691
Fax: +61 62499920
Email: James.Daniell@ga.gov.au

[4]Geoscience Australia, GPO Box 378, Canberra ACT 2601, Australia
Telephone: +61 2 62499514
Fax: +61 62499920
Email: Justy.Siwabessy@ga.gov.au

[5]Geoscience Australia, GPO Box 378, Canberra ACT 2601, Australia
Telephone: +61 2 62499434
Fax: +61 62499920
Email: Brendan.Brooke@ga.gov.au

1. Introduction

Previous studies have shown that seabed sediment parameters such as %Mud, %Sand, and %Gravel are useful surrogates for predicting the distribution of benthic species (e.g., Beaman and Harris 2007; Degraer et al. 2008). Typically, these parameters are derived from a limited number of widely distributed sediment grab samples. To improve predictions from these point data, continuous layers of these parameters are needed.

Apart from often used geostatistic techniques, predictive modelling techniques can be used for large area mapping. In particular, machine learning models offer most potential because they are able to handle both linear and non-linear relationships.

Multibeam data with high resolution coverage is now routinely collected in marine surveys. From multibeam bathymetry we can derive a range of terrain and morphometric variables that have known relationships with sediment distribution patterns. Multibeam backscatter intensity depends on both acoustic impedance contrast and the roughness of the seafloor, which are seabed habitat dependent. Various first and second order texture measures derived from backscatter data may be useful in predicting sediment. Variables that measure spatial autocorrelation are also considered to be useful.

This paper reports the results of predictive spatial modeling of two seabed sediment parameters: %Mud and %Sand for a 700 km^2 area of the Carnarvon Shelf, Western Australia. Multiple machine learning models were applied to create prediction maps and prediction uncertainty maps.

2. Materials and Methods

In August and September 2008 the CERF Marine Biodiversity Hub (http://www.marinehub.org/index.php/site/home) conducted a marine survey of three strategically selected study areas on the southern Carnarvon Shelf, Western Australia (Brooke et al. 2009). The sediment samples were collected using a standard seabed grab sampler in a water depth up to 100 metres. Summary statistics for the two sediment parameters are listed in Table 1.

Variable	Mean	STD	Min	Max
%Mud	3.46	7.03	0.00	34.82
%Sand	80.17	21.27	3.17	100.00

Table 1. Properties of %Mud and %Sand

The multibeam data were collected using a Simrad EM 3002D 300 kHz sonar system operated in a single head configuration. The high quality bathymetry and backscatter datasets were gridded at 3 metre and 5 metre resolutions, respectively. A range of secondary variables at multiple scales were derived from the two variables (Tables 2 & 3).

Variable	Description	Scales
Bathymetry	Seabed water depth	Not Applicable
Slope	Slope gradient	9 m, 15 m, 33 m, 93 m
Relief	Topographic relief	9 m, 15 m, 33 m, 93 m
Surface Area	"true" surface area, an indicator of surface rugosity	9 m, 15 m, 33 m, 93 m
TPI	Topographic Position Index (Weiss, 2001)	9 m, 15 m, 33 m, 93 m
Planar Curvature	The curvature of the surface perpendicular to the slope direction	9 m, 15 m, 33 m, 93 m
Profile Curvature	The curvature of the surface in the direction of slope	9 m, 15 m, 33 m, 93 m
Fuzzy Morphometric Features	Peakness, Pitness, Passness, Ridgeness, Channelness, and Planarness (Wood, 1996)	93 m
Local Moran I	An indicator of spatial autocorrelation	9 m, 15 m, 33 m, 93 m

Table 2. Variables derived from Bathymetry

Variable	Description	Scales
Backscatter	Seabed backscatter intensity	Not Applicable
Local Moran I	An indicator of spatial autocorrelation	15 m, 35 m,

		95 m
Homogeneity	GLCM Homogeneity (Haralick et al. 1973); Four directions (North, East, North-East, and South-East)	15 m, 35 m, 95 m
Variance	GLCM Variance (Haralick et al. 1973); One direction (North-East)	15 m, 35 m, 95 m

Table 3. Variables derived from Backscatter

The three predictive models used to simulate the non-linear sediment-environment relationships are Boosted Decision Tree (BDT) (Friedman 1999), Support Vector Machine (SVM) (Cortes and Vapnik 1995), and General Regression Neural Network (GRNN) (Specht 1990). The models' performance was evaluated against a separated test set (78 out of 259 samples) using three statistics. They include R^2 (proportion of variance explained by model), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). To investigate the sensitivity of model performance to the number of explanatory variables, the secondary variables were added to the models one at a time until the best performance was reached.

3. Results and Discussion

The models' performance for predicting %Mud is good with over 70% of variance explained and low RMSE and MAE values (Table 4). The models did not perform as well for predicting %Sand, with 50% of the variance explained (Table 4). However, the RMSE and MAE values for %Sand are less than one STD, which indicates a satisfactory performance.

%Mud	**R^2**	**RMSE**	**MAE**	**%Sand**	**R^2**	**RMSE**	**MAE**
BDT	0.70	4.04	2.50	BDT	0.49	16.45	11.72
SVM	0.77	3.56	2.00	SVM	0.44	16.87	11.24
GRNN	0.72	3.91	1.95	GRNN	0.48	16.60	11.28

Table 4. Models' Statistical Performance

When training models for %Mud, we discovered that BDT and SVM need only one secondary variable: Local Moran I for Bathymetry to obtain the best performance. GRNN used secondary variables of Homogeneity, Local Moran I for Backscatter, Variance, and Planar Curvature to reach the best performance. For %Sand, the best performing model, BDT, did not use Slope, Surface Area, TPI and Homogeneity in East direction. The best performing SVM model used the secondary variables of Homogeneity in South-East direction, Slope, Variance, Local Moran I for Bathymetry, Surface Area, Homogeneity in North direction, and Relief. Three secondary variables including Local Moran I for Bathymetry, Local Moran I for Backscatter, and Homogeneity in South-East direction were used to obtain the best performing GRNN. The findings indicate the importance of spatial autocorrelation in mapping seabed sediment parameters.

As an example, Figures 1 and 2 display the prediction maps for Point Cloates, which show similar spatial patterns among the three models for both sediment parameters. Percentage mud generally increases with water depth (Figure 1A-C). Figure 1D indicates higher uncertainty in areas with high %Mud. The average error for %Mud prediction is 1.5% with less than 4% of the survey area having a standard deviation error greater than 7%. The opposite spatial distribution patterns were observed for the predictions of %Sand (Figure 2A-C). The average error for %Sand prediction is 4.6% with only 1% of the area having a standard deviation error greater than 21% (Figure 2D). The general patterns for the two sediment parameters are consistent with our knowledge of the survey areas, based on physical samples and underwater video.

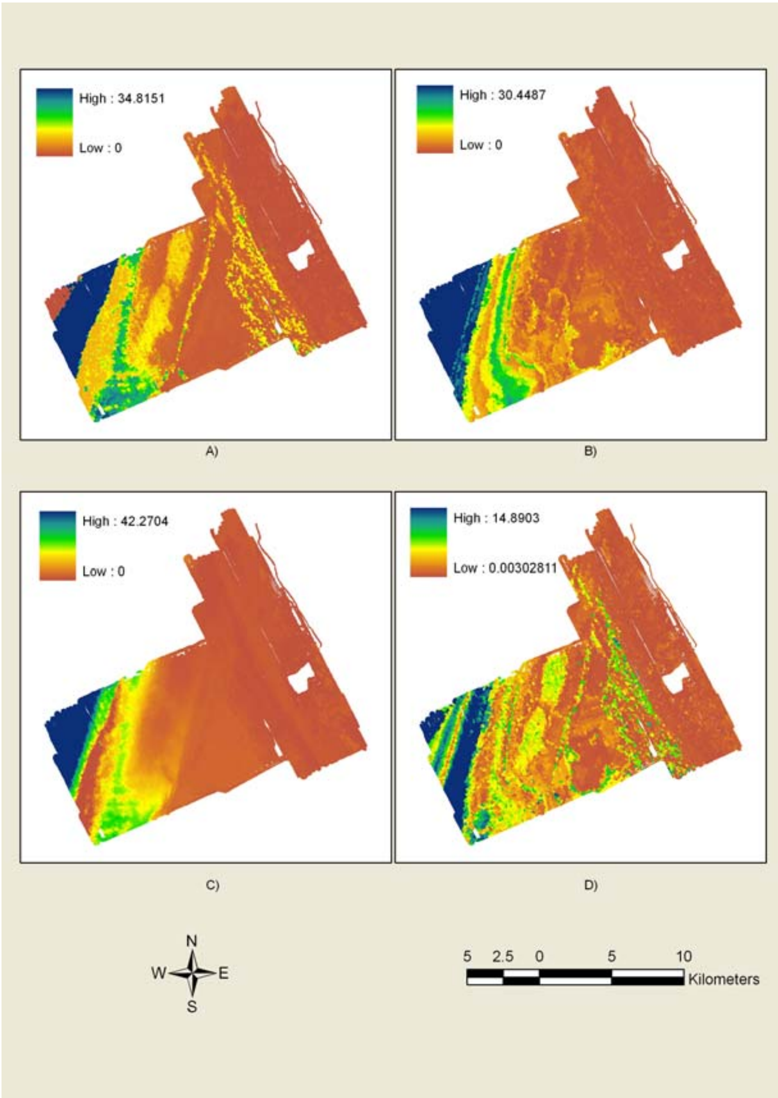

Figure 1: Predicted %Mud distribution for the Point Cloates area; A) GRNN model, B) BDT model, C) SVM model, D) Standard Deviation Error

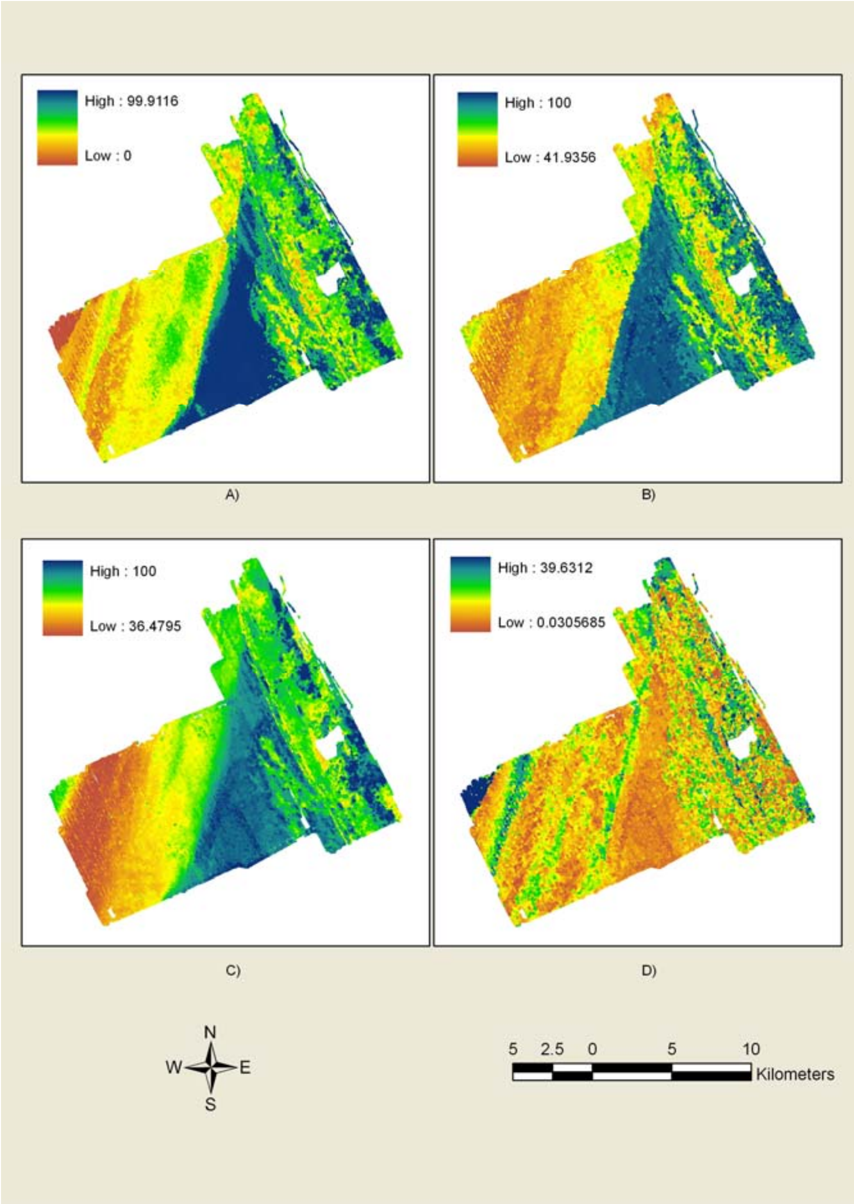

Figure 2: Predicted %Sand distribution for the Point Cloates area; A) GRNN model, B) BDT model, C) SVM model, D) Standard Deviation Error

4. Conclusion

Robust models of the spatial distribution of physical parameters are essential for testing their utility as surrogates of patterns of seabed biodiversity. The sediment prediction maps will be incorporated into the analysis of co-variance of physical and biological data for this area. The results will provide a test of the degree to which these parameters are able to explain observed biodiversity patterns in this area.

5. Acknowledgements

This work has been funded through the Commonwealth Environment Research Facilities (CERF) programme, an Australian Government initiative supporting world class, public good research.

6. References

Beaman RJ and Harris PT, 2007, Geophysical variables as predictors of megabenthos assesmblages from the northern Great Barrier Reef, Australia. in Todd BJ and Greene HG (eds), *Mapping the seafloor for habitat characterization: geological association of Canada, Special Paper*.

Brooke B, Nichol S, Hughes S, McArthur M, Anderson T, Przeslawski R, Siwabessy J, Heyward A, Battershill C, Colquhoun J and Doherty P, 2009, *Carnarvon Shelf Survey Post-Survey Report*. Geoscience Australia, Record 2009/02, 90pp.

Cortes C and Vapnik V, 1995, Support-Vector Networks. *Machine Learning*, 20: 273-297.

Degraer S, Verfaillie E, Adriaens W, Van Lancker V and Vincx M, 2008, Habitat suitability as a mapping tool for macrobenthic community: an example from the Belgian part of the North Sea. *Continental Shelf Research*, 28(3):369-379.

Friedman JH, 1999b, Stochastic Gradient Boosteding. *Technical report*, Dept. of Statistics, Stanford University.

Haralick RM, Shanmugan K and Dinstein I, 1973, Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3:610-621.

Specht DF, 1990, Probabilistic Neural Networks. *Neural Networks*, 3:109-118.

Weiss AD, 2001, Topographic Position and Landforms Analysis. in *ESRI International User Conference,* San Diego, CA.

Wood J, 1996, The geomorphological characterization of digital elevation models. Unpublished PhD thesis, Department of Geography, University of Leicester.