



## Original Article

# An evaluation of supervised and unsupervised classification techniques for marine benthic habitat mapping using multibeam echosounder data

Jay Calvert<sup>1\*</sup>, James Asa Strong<sup>2‡</sup>, Matthew Service<sup>2</sup>, Chris McGonigle<sup>1</sup>, and Rory Quinn<sup>1</sup>

<sup>1</sup>Centre for Coastal and Marine Research, School of Environmental Sciences, University of Ulster, Cromore Road, Co Derry BT22 1SA, UK

<sup>2</sup>Agri-Food and Biosciences Institute, 18a Newforge Lane, Belfast BT9 5PX, UK

\*Corresponding author: tel: +44 28 701 24762; fax: +44 28 701 24428; e-mail: [calvert-j3@email.ulster.ac.uk](mailto:calvert-j3@email.ulster.ac.uk)

‡Present address: Institute of Estuarine and Coastal Studies, University of Hull, Cottingham Road, Hull HU6 7RX, UK

Calvert, J., Strong, J. A., Service, M., McGonigle, C., and Quinn, R. An evaluation of supervised and unsupervised classification techniques for marine benthic habitat mapping using multibeam echosounder data. – ICES Journal of Marine Science, 72: 1498–1513.

Received 19 June 2014; revised 7 November 2014; accepted 12 November 2014; advance access publication 12 December 2014.

Marine habitat mapping provides information on seabed substrata and faunal community structure to users including research scientists, conservation organizations, and policy makers. Full-coverage acoustic data are frequently used for habitat mapping in combination with video ground-truth data in either a supervised or unsupervised classification. In this investigation, video ground-truth data with a camera footprint of 1 m<sup>2</sup> were classified to level 4 of the European Nature Information System habitat classification scheme. Acoustic data with a horizontal resolution of 1 m<sup>2</sup> were collected over an area of 130 km<sup>2</sup> using a multibeam echosounder, and processed to provide bathymetry and backscatter data. Bathymetric derivatives including eastness, northness, slope, topographic roughness index, vector rugosity measure, and two measures of curvature were created. A feature selection process based on Kruskal–Wallis and *post hoc* pairwise testing was used to select environmental variables able to discriminate ground-truth classes. Subsequently, three datasets were formed: backscatter alone (BS), backscatter combined with bathymetry and derivatives (BSDER), and bathymetry and derivatives alone (DER). Two classifications were performed on each of the datasets to produce habitat maps: maximum likelihood supervised classification (MLC) and ISO Cluster unsupervised classification. Accuracy of the supervised habitat maps was assessed using total agreement, quantity disagreement, and allocation disagreement. Agreement in the unsupervised maps was assessed using the Cramer's V coefficient. Choice of input data produced large differences in the accuracy of the supervised maps, but did not have the same effect on the unsupervised maps. Accuracies were 46, 56, and 49% when calculated using the sample and 52, 65, and 51% when using an unbiased estimate of the population for the BS, BSDER, and DER maps, respectively. Cramer's V was 0.371, 0.417, and 0.366 for the BS, BSDER, and DER maps, respectively.

**Keywords:** habitat mapping, multibeam echosounder, supervised classification, towed video, unsupervised classification.

## Introduction

Over the last decade the demand for marine benthic mapping products has increased steadily, as an increasing global population places greater stress on the marine environment (Jackson *et al.*, 2001; Worm *et al.*, 2006). Outputs from these investigations are useful for research scientists in a range of fields including engineering (e.g. Wienberg and Bartholoma, 2005), archaeology (e.g. Plets *et al.*, 2011), the military (e.g. Blondel, 2000), and marine policy

makers (e.g. Howell *et al.*, 2010). Policies such as the Magnuson-Stevens Fishery Conservation and Management Reauthorization Act of 2006 in the United States and the Marine Strategy Framework Directive of 2008 in Europe show that governments accept the need to understand and protect the marine environment (Diaz *et al.*, 2004). To be successful, the full extent of these habitats must first be assessed. Acoustic and ground-truth data are commonly used for the production of benthic habitat maps

(e.g. Ierodiaconou *et al.*, 2011; Lucieer *et al.*, 2013). Although many techniques are used for habitat map production, methodologies fall into one of two broad categories, supervised and unsupervised classification (Brown *et al.*, 2011).

In unsupervised classification, acoustic data are segmented before being assigned a habitat type identified from ground-truth observations made by camera or physical sampling. This has traditionally been the most common method of conducting benthic mapping with varying levels of success reported (Kostylev *et al.*, 2001; Lathrop *et al.*, 2006; Brown and Collier, 2008). In supervised classification, ground-truth data are classified then used to constrain interpretation of the acoustic data. How classification is achieved for both approaches varies according to the precise method employed. In general, unsupervised methods are those that fall under data clustering or pattern recognition, and are assigned habitat type based on the spatial co-occurrence of classified map pixels and ground-truth data. The supervised methods on the other hand use the ground-truth data to form class signatures from the acoustic data and these are then used to identify similar regions in the acoustic dataset where no ground-truth data exist, to produce a full coverage habitat map.

Multibeam echosounders (MBES) are an increasingly common source of acoustic data for benthic habitat mapping (Brown *et al.*, 2011), owing to their ability to provide full-coverage bathymetry and backscatter data while simultaneously correcting for vessel movement and signal loss due to attenuation (Michaels, 2007). Ground-truth data are usually in the form of underwater video footage, stills or grabs, each of which has its strengths and weaknesses. Video systems are able to sample a larger portion of the seabed than still imagery or grab sampling (Galparsoro *et al.*, 2012), and can sample areas of hard substrata where grabs are ineffective (Brown *et al.*, 2011). These advantages are offset against lower taxonomic resolution, analyst subjectivity and an ability for only semi-quantitative analysis of community data. Nevertheless, video techniques continue to be a popular choice for collection of ground-truth data in benthic habitat mapping (Brown *et al.*, 2011). Once collected, video data are normally segmented into categories of habitat type based on observations of substrata and biota. Many bespoke classification systems are used for this segmentation (Diaz *et al.*, 2004), although it is increasingly accepted that there is a need for a more universal system to allow comparison of studies (Dauvin *et al.*, 2008). The European Nature Information System (EUNIS) habitat classification scheme aims to bring such uniformity to European studies (European Environment Agency, 2007; Galparsoro *et al.*, 2012).

The relationship between acoustic backscatter and substratum can be complicated by data acquisition parameters (McGonigle *et al.*, 2010a, b), effectively meaning that two surveys of the same area can yield variable results given different data collection or processing parameters, or by using different hardware. Further complication can arise from sessile organism cover (Holmes *et al.*, 2008), bioturbation (Urgeles *et al.*, 2002), gravel fraction (Goff *et al.*, 2004), random inhomogeneities in surficial sediments (Borgeld *et al.*, 1999) and when trying to distinguish coarse and mixed sediment classes (Diesing *et al.*, 2014). Despite these complications, many sources in the literature have been able to describe relationships between acoustic backscatter intensity and seabed composition (Jackson *et al.*, 1986a, b; Mitchell, 1993; Goff *et al.*, 2000, 2004; Collier and Brown, 2005; Ferrini and Flood, 2006; McGonigle and Collier, 2014).

Furthermore, relationships have been observed and reported between marine substrata and benthic community structure

(Van Hoey *et al.*, 2003; Beaman *et al.*, 2005; Dutertre *et al.*, 2013). In this way, backscatter data may have the capacity to be a useful proxy for abiotic classification of marine benthic habitats. It is possible then that these data could be used to infer habitats described at level 4 of the EUNIS classification scheme which are described almost exclusively based on their abiotic component. This level of the classification system has been used in broad-scale predictive mapping efforts (JNCC, 2014) and medium scale rule-based mapping using MBES data (Diesing *et al.*, 2009).

Morphological characteristics of the seabed including slope, seabed roughness, and orientation can be extracted from bathymetry data, and this approach has been well explored in the marine geological mapping literature (Herzfeld, 1993; Herzfeld and Higginson, 1996; Mitchell, 1996). These outputs are commonly derived from MBES bathymetry to represent variation on the seabed that governs exposure to hydrodynamic conditions and sediment accretion or erosion, and to provide a measure of morphological complexity (Ierodiaconou *et al.*, 2011). Many investigations have successfully integrated these data with acoustic backscatter data to enhance the predictive capacity of MBES products for marine mapping (Mitchell and Hughes Clarke, 1994; Rattray *et al.*, 2009; Rattray *et al.*, 2013). Bathymetric derivatives have also been successfully used by themselves in broad-scale seabed mapping investigations (Elvenes *et al.*, 2014). Where comparative studies are performed using these data, they tend to concentrate on one set of input data with multiple classifiers, rather than multiple sets of data and multiple classifiers.

Both supervised and unsupervised classification techniques have been widely used for benthic habitat mapping (Brown *et al.*, 2011). These include the easy-to-implement maximum likelihood and ISO cluster classifiers found in most GIS packages (Brown and Collier, 2008; Ierodiaconou *et al.*, 2011), and the more complicated statistical procedures which require specialist knowledge and software to implement. These include neural networks (Marsh and Brown, 2009), decision tree classifiers (Rattray *et al.*, 2009), random forest (Che Hasan *et al.*, 2014), and support vector machines (Che Hasan *et al.*, 2012). In addition to these habitat classification approaches, many studies employ a range of species distribution modelling techniques (e.g. Reiss *et al.*, 2011). These methods are more able to predict change (Lecomte *et al.*, 2013) but require significant training in terms of the statistical methods employed to set up the models. In this study, the focus is on two of the habitat classification methods to allow for a comparison of the two broad conceptual approaches to classification: supervised and unsupervised.

The primary aim of this investigation is to evaluate the outputs of supervised classifications and unsupervised classifications given different sets of input data. To achieve this aim, the objectives of the study are:

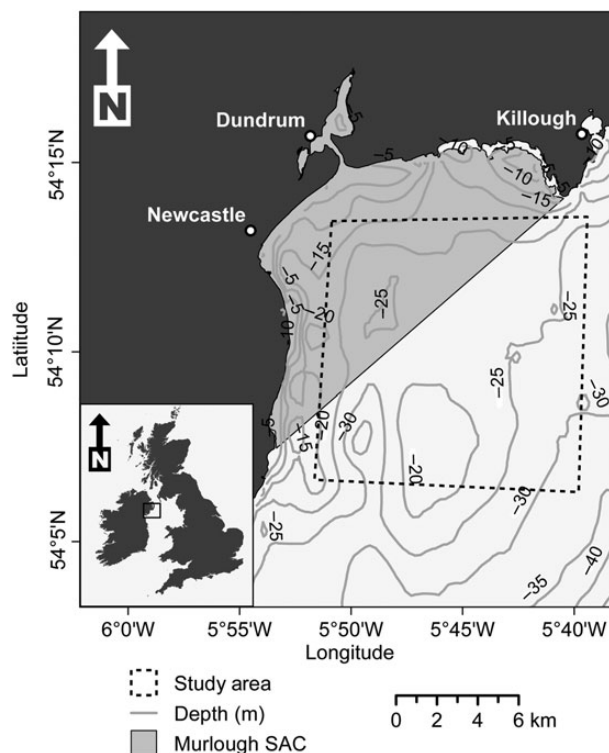
- (i) Generate benthic habitat maps using maximum likelihood supervised classification on acoustic backscatter data, backscatter in combination with bathymetry and its derivatives, and on bathymetry and its derivatives alone.
- (ii) Generate benthic habitat maps using Iterative Self-Organizing (ISO) Cluster unsupervised classification on acoustic backscatter data, backscatter in combination with bathymetry and its derivatives, and bathymetry and its derivatives alone.
- (iii) Calculate accuracy assessment and agreement statistics to facilitate comparison of the classified maps.

## Methods

### Study area

The study area of Dundrum Bay is off the coast of Co. Down in Northern Ireland (Figure 1). The area was surveyed as part of the INIS (Ireland, Northern Ireland, and Scotland) Hydro survey, (INIS Hydro, 2012). The bedrock geology of the region is dominated by Silurian greywackes, with overlying Pleistocene deposits composed predominantly of till and sand (GSNI, 1997). Holocene sand, muddy sand, and gravel dominate the surficial sediments in the region, with a large area of cohesive mud further offshore (Atkins, 1997). Approximately 30% of the study area lies within the Murlough Special Area of Conservation (SAC). The area of sublittoral sand banks in the region is one of the qualifying features for the designation of the offshore part of the SAC, the other being the common seal *Phoca vitulina* (JNCC, 2011).

Bathymetry in the study area ranges from 15 m below chart datum in the northwest to 30 m in the southeast over a distance of 16 km, producing a very shallow gradient (Figure 1). The area experiences the largest tidal variations in Northern Ireland (Atkins, 1997), with a mean spring range of 5 m (Jackson *et al.*, 2005). This tidal variation produces weak streams ( $<0.2 \text{ ms}^{-1}$ ) that run along the shore north and east with the flood tide, and west and south with the ebb tide (Great Britain Hydrographic Department, 1985). Net sediment transport by these tidal streams is in a northerly direction, with net accretion observed in the northern section of the bay (Atkins, 1997; Cooper and Navas, 2004). Due to this net accretion, the morphology of the seabed changes significantly with time, which in turn affects incoming wave energy. In the outer bay, defined here as the portion of the bay seaward of a line



**Figure 1.** Location of the study area in Dundrum Bay off the coast of Co Down in Northern Ireland. Contours were generated from the General Bathymetric Chart of the Oceans (GEBCO) 30 arc-second dataset (vertical datum is mean sea level).

joining Newcastle and Killough (Figure 1), net wave energy moves in a westerly direction. In the inner bay, net wave energy moves in a north north-westerly direction (Cooper and Navas, 2004). Prevailing winds below gale force originate in the southwest, with the majority of winds blowing at gale force or stronger ( $>13.9 \text{ ms}^{-1}$ ) coming from the south to southeast (Atkins, 1997).

### Acoustic data

Geophysical data were acquired during June and July 2011 aboard the Agri-Food and Biosciences Institute's *R.V. Corystes*, using a hull-mounted Kongsberg EM3002 single head MBES system operating at 300 kHz. Before the survey, two Valeport Midas water-level recorders were deployed, one north and one south of the survey area. A patch test was performed before beginning survey work to calibrate the MBES system for any error in mounting, heading, or positioning. The survey was undertaken at an average speed of  $4 \text{ ms}^{-1}$ , with an average ping rate of 8 Hz. Swath-width was limited to a total angular coverage of  $120^\circ$  with an overlap of 100% maintained throughout the survey. Positioning was achieved using a Kongsberg Seapath 200 DGPS system integrated with a Kongsberg MRU 5 motion reference unit for heave, pitch, and roll corrections. Bathymetry and backscatter data were logged using Kongsberg Seafloor Information Systems (SIS) software. Sound velocity profiles were taken using a Yellow Springs International CastAway CTD at a minimum interval of every 4 h or when there was suspected stratification as indicated by the sound velocity sensor on the sonar head. Sound velocity profiles were subsequently loaded into SIS and applied to the incoming sonar data before being logged.

Post-processing of the acoustic data was carried out to IHO Order 1a standard using Caris HIPS & SIPS version 7.1. This included correcting for relative position of the sonar head and vessel draft, tidal corrections to chart datum (LAT), and visual inspection of positioning and motion reference data. A Combined Uncertainty and Bathymetric Estimator (CUBE) surface was generated for the study area, and the data were filtered to reject any soundings lying outside the 95% confidence level for the CUBE surface. Bathymetric data were exported as a 32-bit floating point Bathymetric Attributed Grid. Backscatter data were processed using the Geocoder engine (Fonseca and Calder, 2005) in Caris SIPS. Beam pattern and time varying gain corrections were applied, and the data were mosaicked and exported as a 24-bit uncompressed GeoTiff. Both the bathymetry and backscatter data were exported with a horizontal resolution of 1 m.

### Video ground-truth survey

Underwater video ground-truth data were acquired in September 2011 on-board *R.V. Corystes*. Stations were predetermined independently of the MBES data using a regularly spaced grid to objectively direct the sampling effort as there was no *a priori* knowledge of seabed composition. This resulted in a greater independence of the two datasets, and therefore a more robust statistical comparison of the final habitat maps. Video data were recorded at a total of 18 stations spaced 2.8 km apart. Equipment consisted of a drop frame fitted with a Simrad underwater camera, with an arrangement of four lasers spaced at 25 cm from each other to provide scale. Positioning of the camera was achieved using a TrackLink Ultra Short Baseline system. Each video tow was conducted for 15 min at an average speed of  $0.5 \text{ ms}^{-1}$ , and the camera was suspended an average height of 1 m from the seafloor giving an approximate field of view of  $1 \text{ m}^2$ . Video data were categorized using the EUNIS habitat classification scheme. For its broad applicability,

level 4 was chosen to classify habitats in this study. Distinction between habitat types at this level is based mainly on the abiotic component of the environment for unconsolidated substrata. The only biotic information is contained within the description and indicator species are often infaunal. For hard substrata, habitats at this level are distinguished by introducing a description of community structure. Video data were originally recorded as continuous lines of habitat codes, breaking only where a change in habitat was observed. These were subsequently split into 1 m segments for use in further analyses.

### Bathymetric derivatives

Seven bathymetric derivatives were generated in the study, all calculated using ArcGIS v10.1 (ESRI, 2012); slope, northness, eastness, topographic roughness index, vector rugosity measure, curvature, and profile curvature. These derivatives were chosen to describe the seabed in terms of exposure to wave and current energy (aspect), the likelihood of sediment accretion (slope), and seabed complexity (rugosity) (Ratray *et al.*, 2009). To overcome issues of using radial data to represent aspect (e.g.  $360^\circ = 0^\circ$ ), this layer was further split into northness (cosine of aspect) and eastness (sine of aspect), where 1 represents perfect north/east, and  $-1$  represents perfect south/west, respectively (Zar, 1999). A low-pass filter was applied to the bathymetry data before calculating each of the aspect layers to reduce very small-scale ( $<10$  m) variation in the data. After testing a number of ranges over which to apply the filter, a distance of 200 m was determined to provide an acceptable balance between filtering very small-scale variation while still representing variation over distances of tens of metres.

Four measures of seabed complexity were identified from the literature and were calculated from the bathymetric data; curvature and profile curvature (Wilson *et al.*, 2007), the Topographic Roughness Index (TRI) (Riley *et al.*, 1999), and the Vector Ruggedness Measure (VRM) (Sappington *et al.*, 2007). The TRI and VRM rasters were calculated to show roughness over a range of 25 m. After testing over several ranges, this distance was found to show the largest amount of variation with the minimal amount of surface smoothing for this resolution of data. It should be noted here that this is not an exhaustive list of available environmental descriptors, but a selection of those most commonly found in the habitat mapping literature.

Prior to classification, Kruskal–Wallis tests were performed to identify the variables most able to discriminate the ground-truth classes. If a significant result was obtained for the main test, *post hoc* pairwise tests were carried out using the routine described in Siegel and Castellan (1988) and implemented through function `kruskalmc` in R package `pgirmess` (Giradoux, 2014). Variables were kept only if the pairwise tests showed significant differences 90% or more of the time, that is, if the variable was able to discriminate at least 9 of the 10 pairs of ground-truth classes. Further redundancy was eliminated by performing principal components analysis (PCA) on the bathymetry and derivatives data.

### Habitat map production, accuracy, and agreement statistics

Classified maps were produced using unsupervised and supervised classification methods on three sets of input data: (i) backscatter data alone (BS), (ii) backscatter data in combination with bathymetry and its derivatives (BSDER), and (iii) bathymetry and its derivatives by themselves (DER). Before any classifications were carried out, all variables were scaled to have a minimum of 0 and maximum

of 1 to satisfy the assumptions of the ISO Cluster that all variables are described on similar scales (ESRI, 2012).

Unsupervised classification was performed using the ISO Cluster algorithm in ArcGIS v10.1. This technique organizes the data in the input raster into a user-defined number of groups to produce signatures which are then used to classify the data using the MLC function using the same set up parameters as for the supervised classification. The number of iterations for the clustering procedure was set to 200, as it was found that larger numbers of iterations had negligible effect on the clustering results with significant increases in computing time. Number of classes was set to correspond to the number of EUNIS classes observed in the ground-truth video footage. Dendrograms were plotted to show similarities between ISO Cluster classes created using each dataset.

For each map, a contingency table was created based on the spatial co-occurrence of the ISO Cluster classes and the ground-truth data. Contingency tables were used to perform  $\chi^2$  tests for independence. Since  $\chi^2$  is sensitive to the sample size  $n$ , with large  $n$  producing large and therefore significant  $\chi^2$ , results of the  $\chi^2$  analyses were used to compute Cramer's  $V$  statistics (Cramer, 1946). Cramer's  $V$  has a range from 0 (no relationship) to 1 (perfect relationship) and is not sensitive to sample size. The variable  $V$  was calculated for both the entire contingency table and for each individual class. Mosaic plots (Zeileis *et al.*, 2007) were generated for each contingency table to graphically represent the values in the table. Cells of the mosaic are scaled so that their size in the  $x$ -direction represents the proportion of pixels assigned to an ISO Cluster class, and the size of the cell in the  $y$ -direction represents the proportion of ground-truth data in each class. Circles represent zero values in the contingency table. Plots were shaded based on the Pearson residuals of the  $\chi^2$  test for that table, where the units of the residuals are standard deviations from the expected value. Therefore, any cell of the table that has a Pearson residual of  $\pm 2$  indicates a relationship that is statistically significant at the 95% confidence level. Cramer's  $V$  calculations and mosaic plots were carried out in the R environment using packages `lsr` (Navarro, 2014) and `vcd` (Meyer *et al.*, 2013), respectively.

Supervised classification was performed using the Maximum Likelihood Classifier (MLC) in ArcGIS v10.1. The MLC used 80% of the ground-truth data from each class as training data to create class signatures from the input raster, retaining the other 20% to validate the resulting classified maps. A stratified sampling strategy was used to ensure all observed EUNIS categories and ground-truth stations were evenly represented in the training and validation. This was achieved by randomly sampling 20% of segments from each ground-truth code within each station without replacement. The MLC was set up to classify each pixel with none being rejected due to low confidence, and all classes having equal probability of being assigned. Dendrograms were plotted to show similarities between class signatures created using each dataset.

For accuracy assessment of the MLC maps, contingency tables were produced using the 20% of video data withheld from the training set, based on their spatial co-occurrence with the classified map outputs. From these contingency tables, total agreement, quantity disagreement, and allocation disagreement were calculated (Pontius and Millones, 2011). These authors point out problems with the commonly quoted Kappa statistics for map comparisons and suggest that the two measures of disagreement should be used instead. Quantity disagreement is defined as “the amount of difference between the reference data and a comparison map that is due to the less than perfect match in the proportions of the categories”

and allocation disagreement is defined as “the amount of difference between the reference data and a comparison map that is due to the less than optimal match in the spatial allocation of the categories, given the proportions of the categories in the reference and comparison maps” (Pontius and Milliones, 2011). Furthermore, they suggest that summary statistics computed for map comparison should be based on an unbiased estimate of the population matrix given by

$$p_{ij} = \left( \frac{n_{ij}}{\sum_{j=1}^J n_{ij}} \right) \left( \frac{N_i}{\sum_{j=1}^J N_i} \right),$$

where  $p_{ij}$  is elements of the estimated population matrix,  $n_{ij}$  is elements of the sample matrix, and  $N_i$  is the total counts for the population, where  $i$  and  $j$  refer to the classified map and ground-truth categories, respectively, and  $J$  is the total number of classes (Pontius and Milliones, 2011). Summary statistics are calculated here using both the unbiased estimate of the population contingency table and the sample contingency table. Significance of difference between supervised classifications was assessed using McNemar’s tests for related samples with continuity correction as documented in Foody (2004).

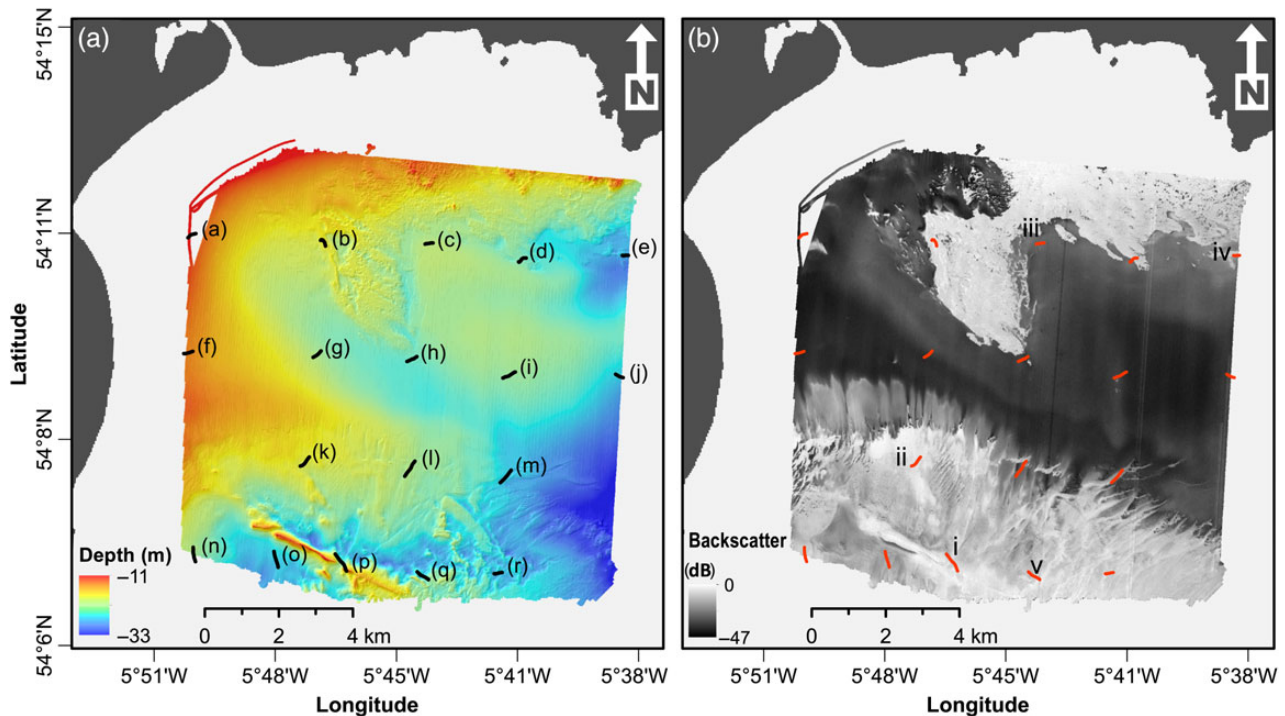
Contingency tables for the MLC maps were represented graphically by Bangdiwala agreement plots (Bangdiwala, 1988) produced using the vcd package in R (Meyer et al., 2013). Plots were drawn so that the size of the box in the  $x$ -direction represents the proportion of ground-truth data available for that class, and the size of the box in the  $y$ -direction represents the proportion of pixels classified as belonging to that category by the MLC. Each box in the plot shows agreement, omission error, and commission error for that class

where omission is a measure of between class discrimination and commission is a measure of within class discrimination.

## Results

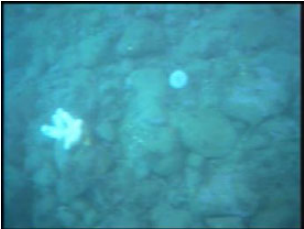
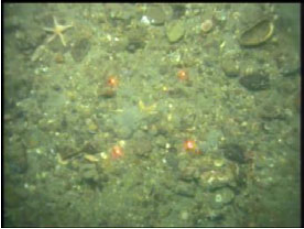
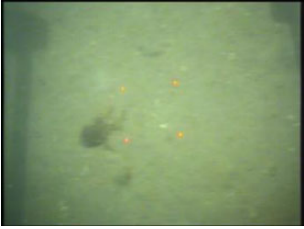
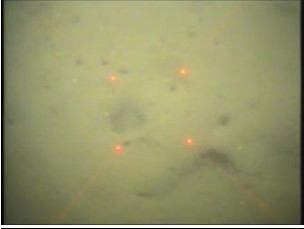

### Acoustic data

The MBES bathymetry data (Figure 2a) depict the seabed sloping from  $\sim 11$  m in the northwest of the study area, to  $\sim 33$  m in the southeast. The shallowest areas are observed in the north and west of the study site, except for an obvious shoal in the form of a linear feature to the south of the site. The feature rises from a depth of 28 m on the northern side to a minimum depth of 13.5 m before sloping to 24 m on the southern side. Based on the linearity, high-backscatter signature (Figure 2b) and orientation of the feature, it is interpreted as an igneous intrusion, probably associated with the Tertiary dyke swarm in the nearby Mourne Mountains. The region to the east of the centre of the study area displays the least variation in terms of both backscatter and bathymetry (Figure 2a and b). High-backscatter intensity is observed in the central northern region and the entire southern region of the study area. These areas demonstrate little variation in broadscale bathymetry but show the greatest fine-scale bathymetric variation in the form of morphological complexity or rugosity. This leads to the conclusion that the central northern region and entire southern region are dominated by hard substratum. Several curvilinear high-backscatter features are observed in the MBES data in the southeast of the study area, also indicating hard substratum. The central region of the study area is dominated by an area of low backscatter intensity. The sole area of low backscatter intensity in the southwest of the study area is found adjacent and to the south of the igneous intrusion, indicating sediment accretion on the steep margin of the dyke (Figure 2a and b).



**Figure 2.** Results of the acoustic survey showing (a) bathymetry (corrected to chart datum, LAT) and (b) backscatter data. Ground-truth stations are displayed on both the bathymetry and backscatter data. In (a), letters (a) to (r) correspond to (a) to (r) in Figure 3. Locations of the images in Table 2 (i–v) are displayed with the backscatter data.

**Table 1.** Example images and descriptions of EUNIS level 4 habitats (Figure 3) identified from the video ground-truth data (European Environment Agency, 2007).

Image	EUNIS Class	EUNIS habitat description	Training data (m)	Validation data (m)
(i) 	A4.21	Occurs on wave exposed, moderately strong to weakly tide-swept, circalittoral bedrock and boulders. Echinoderms, faunal and algal crusts (red encrusting algae) dominate this biotope.	95	23
(ii) 	A5.14	Tide-swept circalittoral coarse sands, gravel and shingle generally in depths of over 15–20 m.	498	124
(iii) 	A5.26	Circalittoral non-cohesive muddy sands with the silt content of the substratum typically ranging from 5 to 20%.	2384	595
(iv) 	A5.36	Sublittoral muds, occurring below moderate depths of 15–20 m, relatively stable conditions often lead to the establishment of communities of burrowing megafaunal species.	198	49
(v) 	A5.44	Mixed sediment habitats in the circalittoral zone (generally < 15–20 m) including well-mixed muddy gravelly sands or very poorly sorted mosaics of shell, cobbles and pebbles embedded in or lying upon mud, sand, or gravel.	1520	379

Lasers in the images are spaced 25 cm apart. Images (i) to (v) correspond to (i) to (v) in Figure 2b. The amount of data available for training and validation of the habitat maps is also shown.

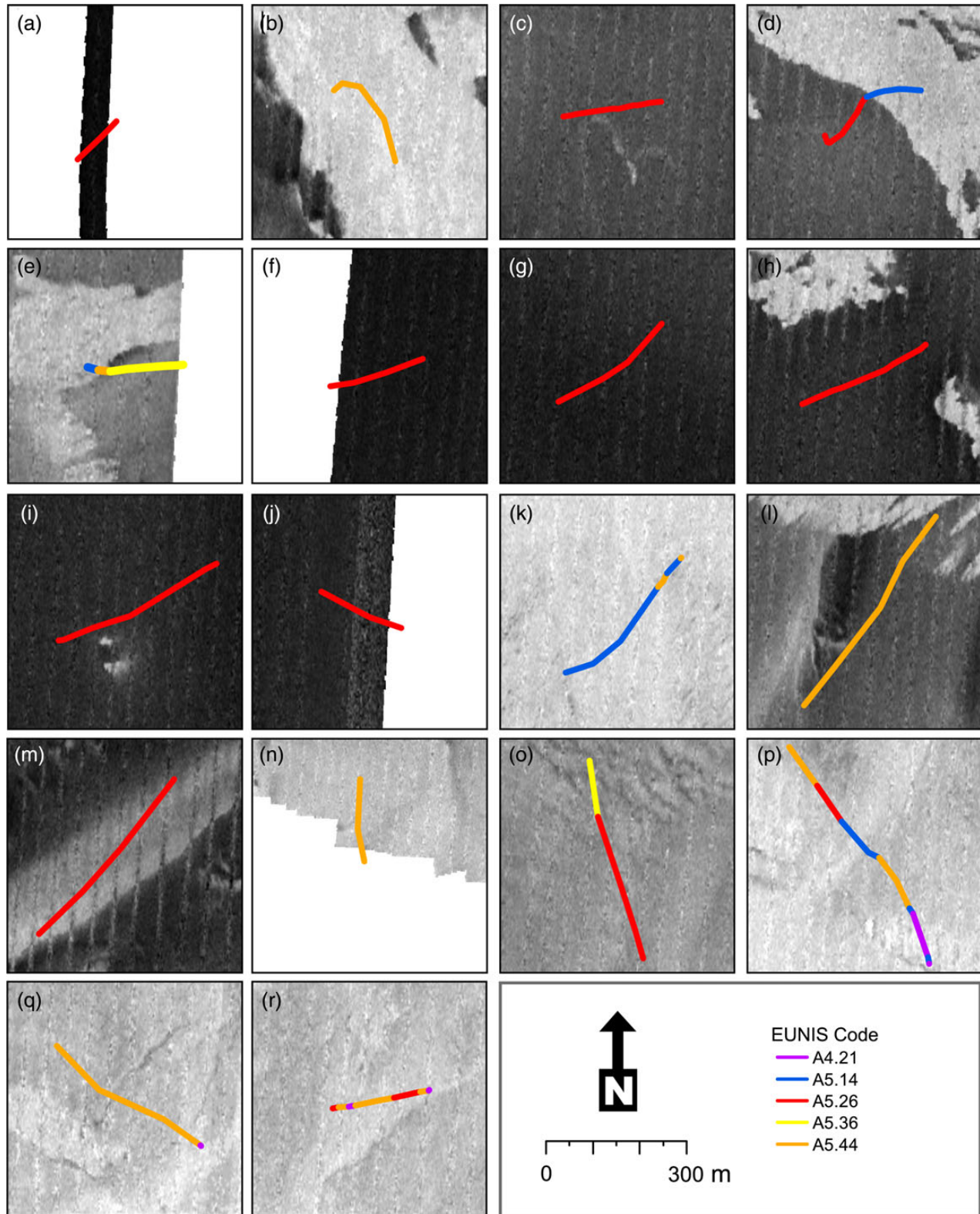
### Ground-truth data

Five classes at level 4 of the EUNIS habitat classification scheme are observed in the video footage (Table 1 and Figure 3). Classification of the footage is broken down into 5865 1 m segments. The segmentation comprises 118 m of class A4.21 (*Echinoderms and crustose communities on circalittoral rock*), 622 m of class A5.14 (*Circalittoral coarse sediments*), 2979 m of class A5.26 (*Circalittoral muddy sand*), 247 m of class A5.36 (*Circalittoral fine mud*), and 1899 m of class A5.44 (*Circalittoral mixed sediments*). As well as being the most abundant, class A5.26 is the most widespread, but dominates in the central regions where the ground-truth stations are observed to contain only this class. Class A4.21 is observed next to the dyke in the

south of the region, and on the curvilinear high-backscatter features to the east. Classes A5.14 and A5.44 are observed in the north and south medium backscatter regions, with much of the latter seen where there is a change in backscatter intensity (Figure 2a and b). The study area lies on the margin of the Northern Irish Sea mud patch (Hill *et al.*, 1995), and as such the only areas of class A5.36 are on the seaward fringes of the study area, to the east and the south.

### Map production, accuracy, and agreement assessments

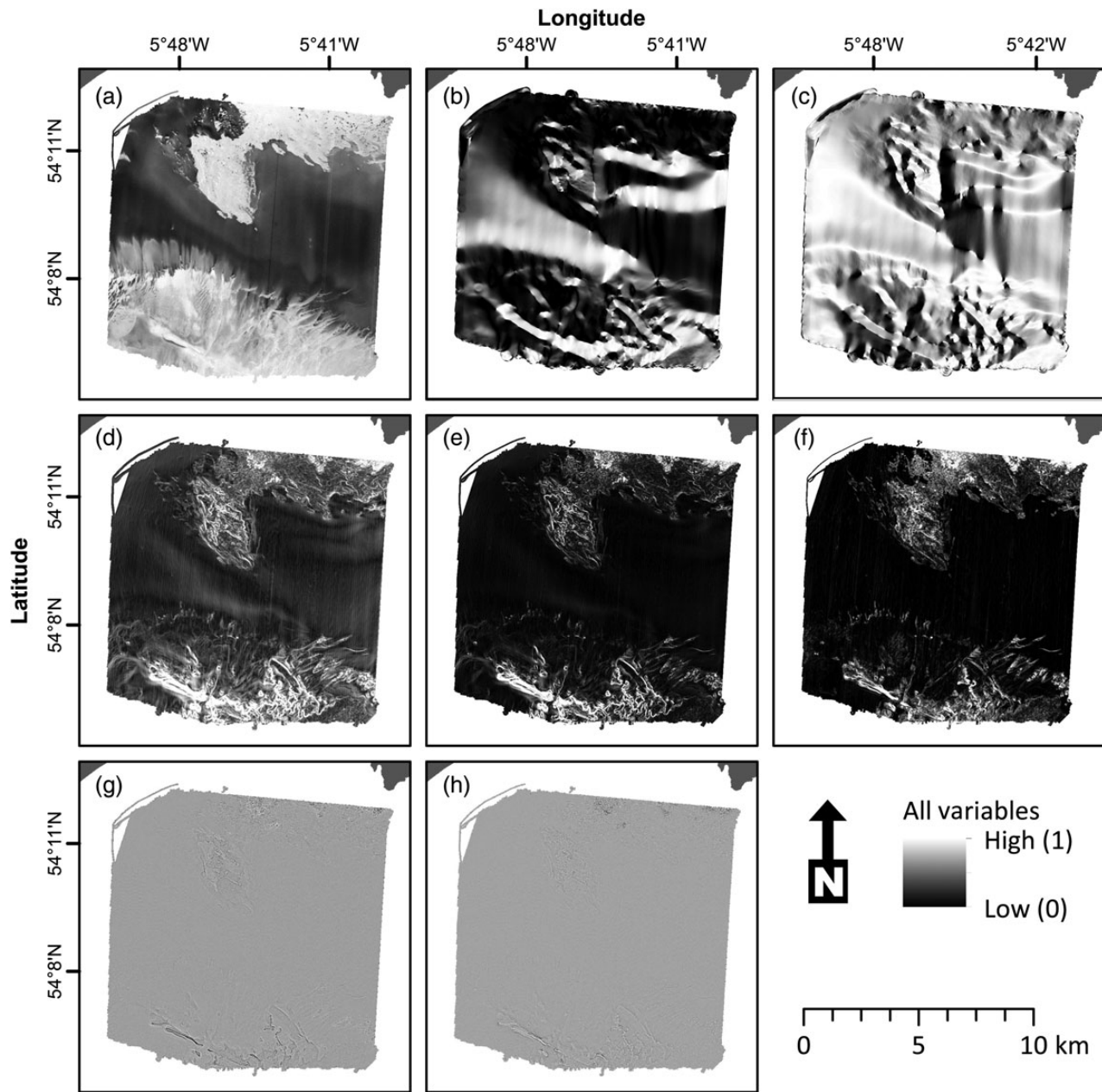
For each of the nine environmental variables created (Figure 4), Kruskal–Wallis tests show significant differences between ground-truth classes with  $p < 0.01$ . Therefore, pairwise tests were carried



**Figure 3.** Ground-truth stations showing segmentation based on EUNIS habitat codes. Letters (a) to (r) correspond to (a) to (r) in Figure 2a.

out for each of the nine variables to identify which pairs cause the differences. Bathymetry, backscatter, TRI, and slope discriminate ground-truth classes in at least 9 of the 10 pairwise tests (Table 2 and Figure 5). These four variables were subsequently subject to

PCA. This results in two components accounting for 99.99% of the variation in the original data, composed primarily of bathymetry and TRI. The output PCA data form part of the BSDER and all the DER datasets used for map production.



**Figure 4.** (a) Processed backscatter data, (b) northness calculated from the filtered bathymetry data, (c) eastness calculated from the filtered bathymetry data, (d) topographic roughness index, (e) slope, (f) vector rugosity measure, (g) profile curvature, and (h) curvature calculated from the bathymetry data in Figure 2a) and used in the Kruskal–Wallis tests and PCA. Note these are displayed after each variable had been standardized.

Maps resulting from the unsupervised classifications on BS, BSDER, and DER are shown in Figure 6a–c, respectively. It should be noted here that the ISO Cluster class names are assigned randomly by ArcGIS, so class numbers do not necessarily correspond from one map to another, and are displayed using one colour scheme for convenience only. Each of the ISO Cluster maps picks out the regions indicated as more topographically complex in the south of the region, the BS and BSDER maps pick out the topographically complex region to the north, and the main source of disagreement for all three is in the central areas.

Mosaic plots reveal a one-to-many relationship between each of the ISO Cluster outputs and the ground-truth data (Figure 7). Main sources of confusion are classes A5.26 and A5.44 for the BS and

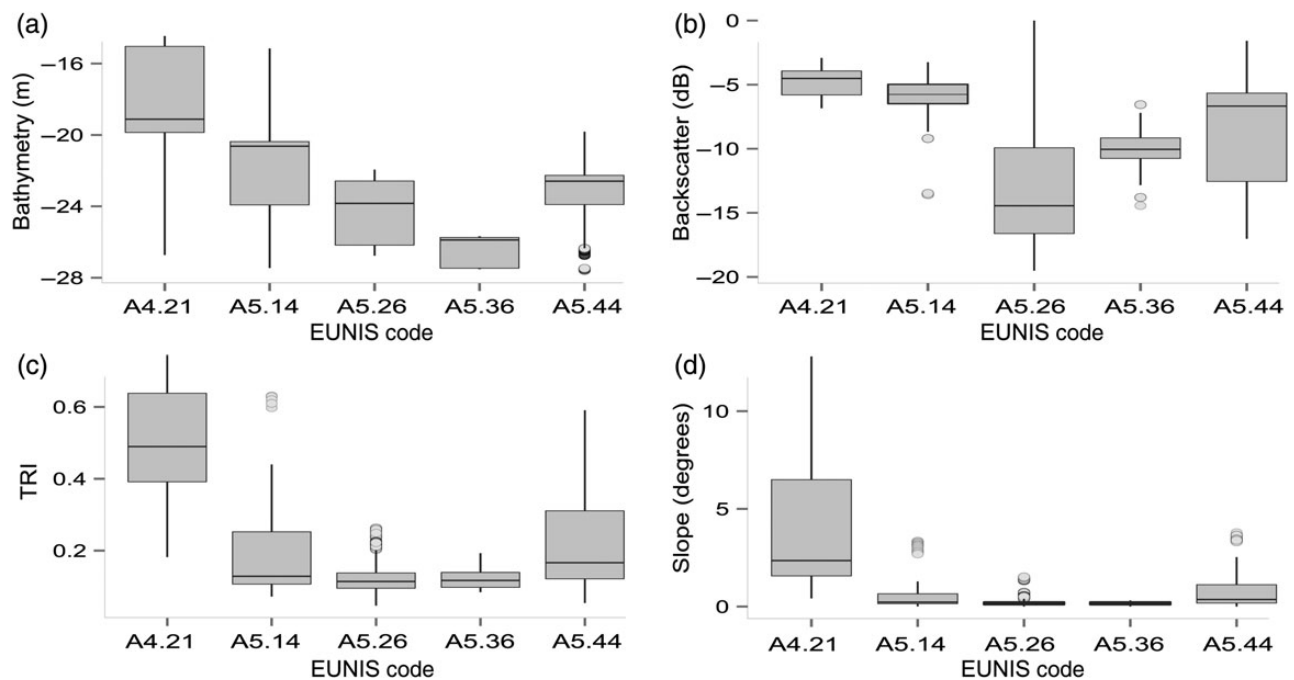
BSDER maps, and classes A5.14, A5.26, and A5.44 for the DER map. Results of the  $\chi^2$  analyses were significant with  $p < 0.001$  for all three contingency tables. It is likely however that the extreme values of  $\chi^2$  are due to large sample size. Relationships are therefore assessed based on Pearson residuals in the mosaic plots and Cramer's  $V$  statistics. Pearson residuals indicate relationships at the 95% confidence level or above between all but two cells of the contingency tables for the BS maps, and all cells for the BSDER and DER maps. Overall, Cramer's  $V$  is 0.371, 0.417, and 0.366 for the BS, BSDER, and DER maps, respectively. The lowest individual class Cramer's  $V$  is 0.232 for EUNIS class A5.26 for the BS map, while the highest is 0.843 for class A4.21 for the BS map. Across all three ISO Cluster maps, class A5.26 is the least well represented



**Table 2.** Results of the *post hoc* multiple comparison tests for discriminating ground-truth classes.

	Bathymetry	Backscatter	TRI	Slope	VRM	Northness	Eastness	Prof. curvature	Curvature
A4.21–A5.14	x	–	x	x	x	x	–	–	–
A4.21–A5.26	x	x	x	x	x	x	–	x	–
A4.21–A5.36	x	x	x	x	x	–	x	x	–
A4.21–A5.44	x	x	x	x	x	x	–	–	–
A5.14–A5.26	x	x	x	x	x	x	–	–	–
A5.14–A5.36	x	x	x	x	x	x	x	–	–
A5.14–A5.44	x	x	x	x	–	x	–	–	–
A5.26–A5.36	x	x	–	–	–	–	x	–	–
A5.26–A5.44	x	x	x	x	x	–	–	x	x
A5.36–A5.44	x	x	x	x	x	–	x	–	–

Cells containing an x represent cases where a significant difference was observed at  $p < 0.01$ , and cells containing a – represent cases where there was no significant difference.

**Figure 5.** Boxplots showing characteristics of (a) bathymetry, (b) backscatter, (c) TRI, and (d) slope for each of the five EUNIS habitat types observed in the ground-truth data.

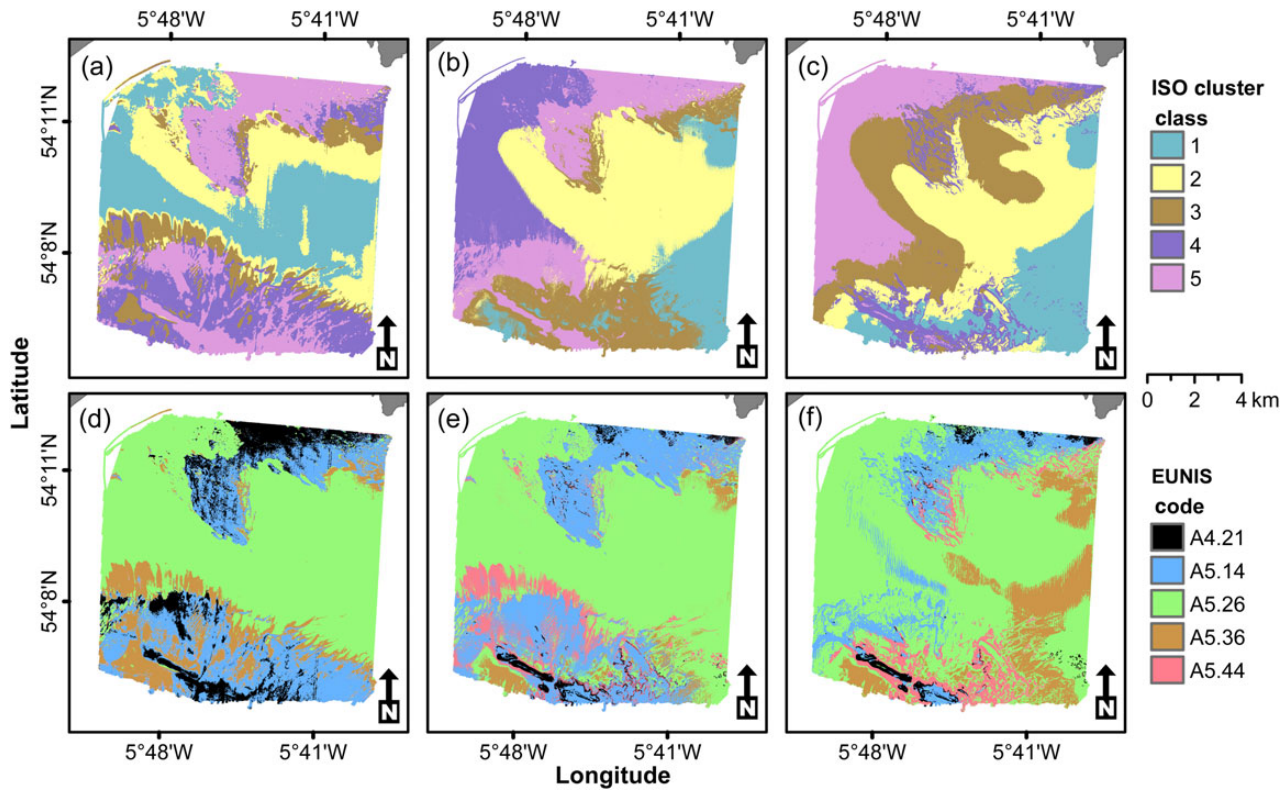
by any of the ISO Cluster classes, with Cramer's  $V$  of 0.429 or less for each map.

Dendrograms drawn for each of the ISO Cluster outputs reveal varying degrees of similarity between classes depending on input data (Figure 7). For the BS data, classes 1, 2, and 3 are completely dissimilar to classes 4 and 5. For the maps created with the BSDER and DER datasets, classes 4 and 5, respectively, are completely dissimilar to all other classes. For the BS map, all other classes are similar at  $<50\%$ . Overall, ISO Cluster classes generated using the BSDER and DER data are less similar than those generated using the BS data.

Maps resulting from the supervised classifications on BS, BSDER, and DER are shown in Figure 6d–f, respectively. The highest degree of visual similarity is observed between the maps produced using the BS and BSDER data. In all three maps, the same topologically complex regions have been identified in the north and south of the region. In contrast to the ISO Cluster maps, each of the MLC outputs classifies the central regions of the study area to a similar category, A5.26, which is by far the most common category in all three maps.

Results of the accuracy assessment for the three MLC maps (Figure 8) using the sample contingency table resulted in total agreement of 46, 56, and 49% for the MLC maps produced using BS, BSDER, and DER, respectively. For the BS map, the main source of disagreement when using the sample contingency table is quantity disagreement at 36%, indicating high intensity losses or gains among the different classes. For the BSDER and DER maps, the main source of disagreement is allocation disagreement at 29 and 28%, respectively, suggesting a large amount of confusion between classes. Bangdiwala agreement plots (Figure 8) show that the main source of agreement for each map comes from class A5.26. Most notably, the BS map fails to correctly classify any of classes A4.21 or A5.44. The map produced with the BSDER data includes the most even representation of each ground-truth class, while the DER map represents contain a less even spread than the BSDER map but a more even spread than the BS map.

Calculating the accuracy assessment statistics using an unbiased estimate of the population matrix markedly affects results. Total agreement using these data is 52, 65, and 51% for the BS, BSDER,



**Figure 6.** Classified maps using (a) ISO Cluster on BS, (b) ISO Cluster on BSDER, (c) ISO Cluster on DER, (d) MLC on BS, (e) MLC on BSDER, and (f) MLC on DER. It should be noted that while the classes in the ISO Cluster maps appear spatially correlated, they are arbitrarily assigned class names. Class names in the two MLC maps are the EUNIS habitat codes as identified from the video ground-truth data.

and DER maps, respectively. These correspond to quantity disagreement of 28, 25, and 25%, respectively, where quantity disagreement is the main source of disagreement for all three maps.

Dendrograms for the class signatures generated to classify the MLC maps show that for the BS map (Figure 9) ground-truth data classes A5.44 and A5.36 are most similar, followed by classes A4.21 and A5.14. Class A5.26 linked with the grouping of A5.44 and A5.36 at  $\sim 50\%$  similarity. For the BSDER map, the output signature file reveals the highest degree of similarity between ground-truth data classes A5.14 and A5.44. These two classes were further linked to classes A5.26 at the 55% level of similarity and A5.36 at the 30% level of similarity. For this map, class A4.21 is completely dissimilar to all other classes. The dendrogram for the DER map reveals the highest level of similarity between classes A5.44 and A5.14 at 80%, which join with class A5.26 at 60%. Class A5.36 joins with these three at 15%, and class A4.21 is completely dissimilar.

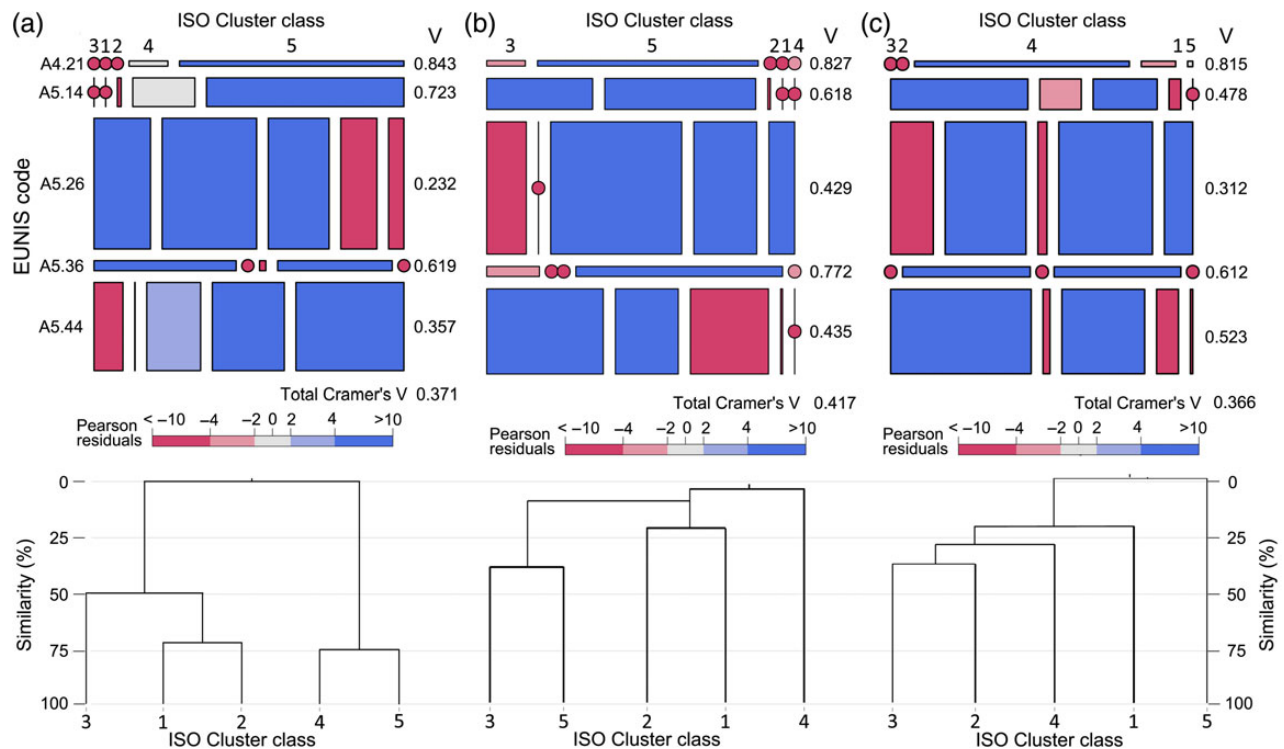
Statistics calculated from contingency tables (Table 3) for pairwise MLC map comparisons show that the BS map is 71 and 51% similar to the BSDER and DER maps, respectively, and the BSDER map is 64% similar to the DER map. The main source of agreement for these comparisons is due to the widespread nature of class A5.26 in all three maps. McNemars's tests show that all three maps are significantly different at  $p < 0.01$  (Table 3).

## Discussion

The primary aim of this investigation was to evaluate outputs from unsupervised and supervised approaches to benthic habitat mapping, by performing ISO Cluster unsupervised classification and maximum likelihood supervised classification (MLC) on

three sets of input data. Video ground-truth data classified to level 4 of the European Nature Information System habitat classification scheme (European Environment Agency, 2007) revealed five seabed classes in the study area, so the MLC produced maps containing five habitat types. The same number of classes was produced during the ISO Cluster classifications, and statistics were derived to evaluate the outputs of the two methods. Direct statistical comparison of the two approaches was not attempted due to differences in assumptions in the statistical analyses. The  $\chi^2$  analyses assume independence in the categorical data being tested, so it would not be appropriate to use  $\chi^2$  for the MLC outputs since one dataset has been used to predict the other. Similarly, agreement and disagreement analysis assumes that the operator knows which categories are paired, so for the ISO Cluster class names being assigned randomly in ArcGIS, this is impractical. While a statistical comparison was not attempted, it is possible to evaluate qualitatively the outputs from the two methods.

Accuracy of the MLC outputs was assessed using both a sample contingency table and an unbiased estimate of the population matrix following recommendations in Pontius and Millones (2011). The procedure of estimating the population matrix is cited by these authors as vital for producing unbiased summary statistics for accuracy assessment; however, no instances of this step are found in the benthic habitat mapping literature. In the scientific literature, accuracy assessments are commonly reported in purely numeric terms (e.g. Diesing *et al.*, 2014; Stephens and Diesing, 2014), with total agreement or Cohen's Kappa being the most widely quoted statistics. Visual analysis of agreement plots generated for the MLC outputs in this study clearly show that, for the BS and DER maps, almost all the agreement is due to the spread of class A5.26, and very



**Figure 7.** Mosaic plots of contingency tables and dendrograms produced for the ISO Cluster maps using (a) BS, (b) BSDER, and (c) DER. Cells of the mosaic are proportional to the counts for that cell. Shading is based on the Pearson residuals of the  $\chi^2$  test, where units are standard deviations from the expected value assuming independence. Cramer's V statistics are displayed to indicate agreement between individual ground-truth categories and ISO Cluster classes as well as overall for each classification.

little is due to agreement in any other class. The exclusion of either the backscatter or bathymetric derivatives from the input data for the MLC therefore seems to have a negative effect on the predictive capacity of the data. Nevertheless, accuracy of the MLC maps in this study is comparable with that seen in other studies using the same classifier (Ierodiakonou *et al.*, 2011), although these authors show that the MLC performs below the levels of other classifiers.

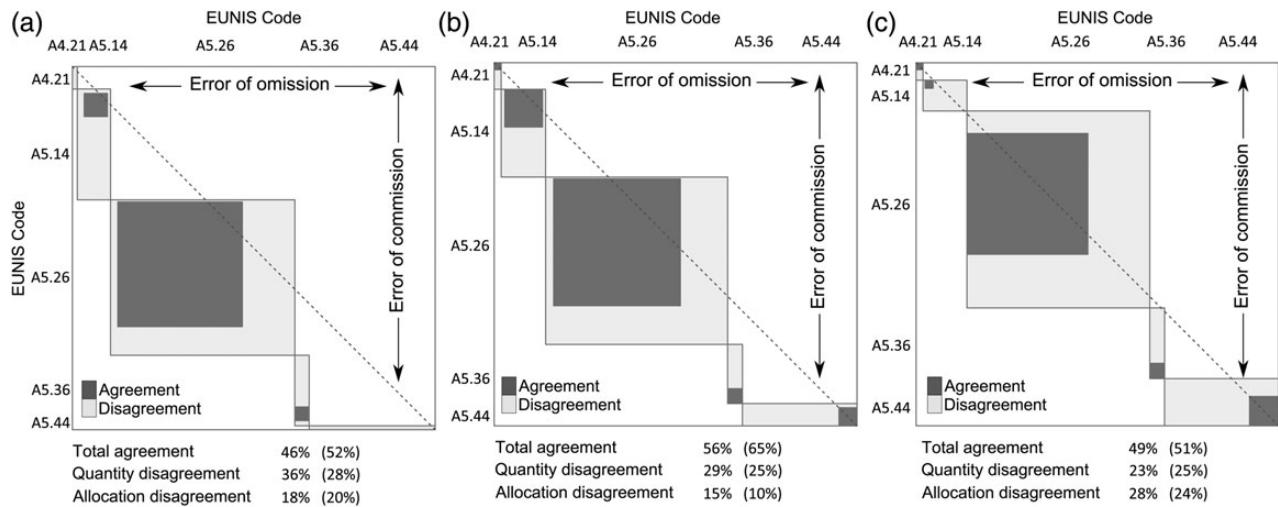
The ISO Cluster unsupervised classification used in this study is not truly unsupervised. To attempt a comparison between the classified acoustic and ground-truth data, the ISO Cluster was constrained to create five classes out of the two sets of input data. This partitioning into a user-defined number of classes assumes prior knowledge of the data, when in fact, in comparison to the acoustic data, very little of the seabed has been qualified by ground-truth data. The dendrogram produced from the ISO Cluster analysis on the backscatter data (Figure 7a) shows that classes 1, 2, and 3 are ~50% similar, while at 75% classes 4 and 5 are even more similar, suggesting that only two or three distinct clusters can be observed in the backscatter data. Jones and Brewer (2012) report using a similar method with the main difference being that they merged classes which displayed high levels of similarity. This however would not have allowed for a comparison with the ground-truth data, so was not considered in this study.

In this study, a grid of ground-truth sampling locations was predefined independently of the acoustic data, rather than using the latter to direct the sampling effort, to objectively direct the sampling of ground-truth data. Comparison of the BS, BSDER, and DER maps would have been biased if the ground-truth sampling had been based on the backscatter data or any other assumed *a priori* knowledge of the seabed. As pointed out above, a maximum of

three (but more likely two) ground types would have been sampled (Figure 7) if this were the case. Based on the EUNIS level 4 habitat codes though, a minimum of five ground cover types exist in the region. Kostylev (2012) states that there is an urgent need for a change in how data for marine habitat mapping are sampled and interpreted, and suggests that acoustic backscatter alone should not form the sole basis for inferring benthic habitat nor directing sampling effort.

Nevertheless, backscatter data have been shown to be a good proxy for substrate type (Goff *et al.*, 2000, 2004; Collier and Brown, 2005; McGonigle and Collier, 2014). Since the habitat codes at level 4 of the EUNIS habitat classification system describe the seabed almost exclusively based on their abiotic component, maps were produced using three different datasets to investigate the effects of their inclusion; backscatter alone, backscatter in combination with bathymetry and derivatives, and bathymetry and derivatives by themselves. In this study, the use of the BSDER data is shown to achieve more accurate results for the MLC classifications than using BS.

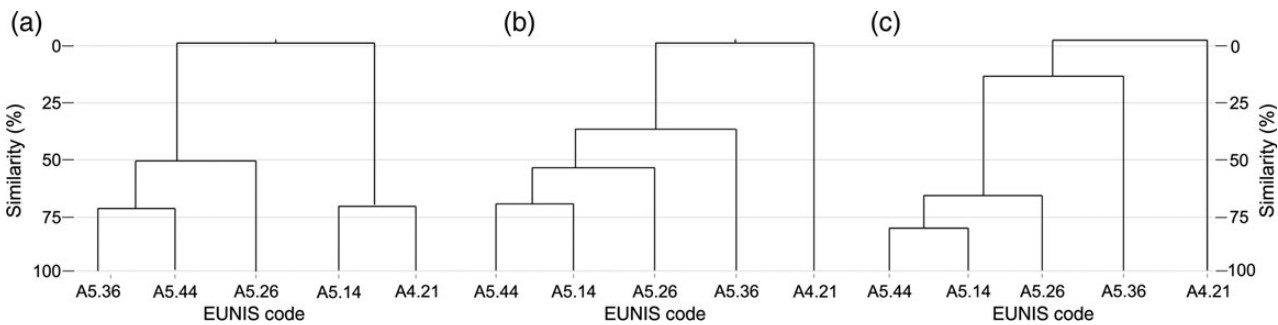
For the ISO Cluster unsupervised classifications, Cramer's V results show that the exclusion of backscatter data does not negatively impact the classifier as it did with the MLC maps. To the contrary, inclusion of bathymetric derivatives resulted in better relationships between the ISO Cluster classes and EUNIS classes A5.26, A5.36, and A5.44 for the BSDER map, and classes A5.26 and A5.44 for the DER map. Class separation as observed in the dendrograms for the ISO Cluster classifications (Figure 7) is also much better for the maps with bathymetric variables included, which is not the case for the MLC dendrograms (Figure 9). These results suggest that in areas of hard substrata



**Figure 8.** Bangdiwala agreement plots produced using the contingency tables produced for the MLC maps using (a) BS, (b) BSDER, and (c) DER. The total size of each box is proportional to the counts for that cell of the table, the dark box represents perfect agreement, and the lighter box is error. Error in the x-direction is error of omission and error in the y-direction is error of commission. Results for total agreement, quantity disagreement, and allocation disagreement are also displayed. Values in parentheses are based on an unbiased estimate of the population matrix as described in Pontius and Millonnes (2011). Values outside the parentheses are based on the sample matrix.

**Table 3.** Pairwise map comparisons for the three maps produced using the MLC and results of McNemar’s related samples tests for significance.

Dataset	Agreement	Dis. quantity	Dis. allocation	McNemar’s $\chi^2$	p
BS and BSDER	71	18	11	16.2	< 0.01
BS and DER	51	15	34	48.7	< 0.01
BSDER and DER	64	15	21	18.2	< 0.01



**Figure 9.** Dendrograms produced from the signature files for the maximum likelihood classification on (a) BS, (b) BSDER, and (c) DER.

with high morphological complexity, unsupervised classifications using bathymetric derivatives may be as good for deriving meaningful segmentations as backscatter data are.

Jackson *et al.* (2005) reported that the beaches in the littoral zone of Dundrum Bay are composed of a thin surficial veneer of fine sediment on top of a harder substratum, with cobbles frequently exposed in the area. While wave energy is likely to be much reduced in the middle of the Bay, and so sediment retention should be higher, it is not unreasonable to assume that there would be areas of surficial veneer offshore. Lucieer *et al.* (2013) found that their classification system was least accurate when trying to classify a seabed characterized by a sediment veneer on a harder underlying substratum with ground-truth data identified

from video imagery. A similar situation is reported in Kostylev *et al.* (2001), and is a potential source of some of the disagreement seen in this study. Ground-truth station *m* (Figure 3), for example, appears acoustically different from station *g* or *h*; however, video footage reveals that, on the surface at least, they are each composed of class A5.26. This is interpreted as volume scattering either from muddy sand overlying a coarse substratum or a coarser fraction mixed through the sand which was unobservable in the video footage. Thus, describing the seabed in purely acoustic terms ignores the possibility that two different acoustic signatures are being assigned one ground-truth class. On the other hand, describing habitat based solely on video data ignores the possibility that hard substrata can underlie surficial sediment (e.g. Callaway *et al.*, 2009).

The EUNIS habitat classification scheme has successfully been applied in low-resolution, broad-scale mapping projects (Diesing *et al.*, 2009; JNCC, 2014). For the higher resolution MBES data used in this study, it was found that some of the EUNIS habitat codes did not provide sufficient allowances for discriminating areas of seabed. For example, class A5.44 (Circalittoral mixed sediments) covers well-mixed muddy gravel, poorly sorted shell mosaics, and cobbles and pebbles lying on mud or sand. Class A5.14 (Circalittoral coarse sediments) covers a similarly large range of substrata, including coarse sand, gravel, and shingle. Within each of these classes there are many sediment types that will be classified in the same way based on visual classification of the ground-truth data, but will produce acoustic signatures that differ immensely. In the BS map created using the MLC, the greatest amount of confusion was observed between the circalittoral mud class and the mixed sediment class. This is potentially due to the mixed sediment class including pebbles and cobbles lying on mud, although as the amount of pebble/cobble increases, this confusion should diminish as the larger fraction begins to dominate the backscatter signal. For the BSDER and DER maps created using the MLC, the greatest amount of confusion was observed between the mixed sediment class and the coarse sediment class. Again, this is to be expected since the mixed sediment class includes mosaics of shell, pebble, or cobble on coarse sand and gravel.

Some of this inter-class confusion could be eliminated by collecting a larger amount of ground-truth data. Collecting larger volumes of these data will however significantly increase survey cost (Holmes *et al.*, 2008; Clements *et al.*, 2010). The result of this situation is a comparatively small amount of ground-truth data that represents the seabed at a much higher resolution than the full-coverage acoustic data. The sampling unit of the video data in this study is the same as the resolution of the acoustic data, at  $\sim 1 \text{ m}^2$  field of view per frame of video footage. While it has been shown that the fine-scale variability of the seabed is what drives backscatter strength in unconsolidated substrata (Jackson *et al.*, 1986a), the small-scale variation observed in video data is often not evident in the acoustic data (Diesing *et al.*, 2009). This potentially accounts for some of the inaccuracies in the current study. Che Hasan *et al.* (2014) state that accuracy assessment of habitat maps needs to be considered not only in terms of the statistics, but also with consideration of the scheme used to classify the ground-truth data. Put simply, inaccuracies of the output habitat maps may be due to a comparative lack of ground-truth data or the inadequacies of the habitat classification scheme, rather than survey or experimental design.

Scale and resolution will also affect bathymetric derivatives. Recent work has shown that changing the scale over which bathymetric derivatives are calculated, as well as calculating the same derivatives using different resolutions of input data will cause large differences in the output layers (Wilson *et al.*, 2007). For example, calculating slope or rugosity over large scales will have a smoothing effect on the surface, to the point that all the variation has essentially been averaged out. For example, for the data in the current study artificially coarsening the MBES data to 10 m resolution changes the maximum slope in the area from  $67^\circ$  to  $16^\circ$ . Varying scales of these bathymetric derivatives will therefore affect the success of habitat modelling studies (Lundblad *et al.*, 2006). Recent work has shown promise in terms of varying scales of analysis for habitat suitability modelling (Giusti *et al.*, 2014).

The BSDER and DER datasets used here included a number of bathymetric derivatives, but were not exhaustive. As a result, there could be others that explain more variation than these, or explain

variation in a way more suited to describing benthic habitat distribution. Furthermore, the ISO cluster assumes that data are described on similar scales, so all variables were transformed prior to classification. This has the effect of giving each variable equal weight which may not be the case in nature. For example, while there are tidal streams observed in the study area, they were found to be imperceptible in the middle of the Bay (Great Britain Hydrographic Department, 1985). The two aspect variables were included in initial tests due to their expected influence on exposure to the current regime in the bay; however, they do not account for the spatial variation seen in the strength of the currents within the study area. In addition to this spatial variation, temporal variation is observed in benthic habitat structure. Even over short-time scales, distinct differences can be seen in seabed habitats that need to be understood before mapping methods can be accurately assessed (Anderson *et al.*, 2008).

A further potential source of inaccuracy with the methodology used in this paper is that the maximum likelihood classifier assumes a Gaussian distribution when assigning classes to pixels (Ierodiaconou *et al.*, 2011). Habitat distributions, however, are likely to be multimodal in nature. In this investigation, modal frequencies were observed in classes A5.26 and A5.44. Moreover, except features like rocky outcrops and sorted bedforms (Murray and Thiel, 2004), habitats are not likely to display the types of discrete boundary being modelled in maps of the type created in this study. Rather, the boundaries between habitats would exist as fuzzy boundaries (Lucieer and Lucieer, 2009) displaying features that resemble more closely an ecotone rather than an ecotone (Attrill and Rundle, 2002).

Diesing *et al.* (2014) state that there is a need for more comparative marine mapping studies to fully inform best practice for this type of investigation. Stephens and Diesing (2014) produced one such study for the assessment of a range of supervised classifiers. Based on the results of the ISO Cluster classifications produced in this investigation, it is also recommended that the use of unsupervised methods is explored to appreciate the full range of methods available to habitat mappers.

## Conclusion

This paper describes methods for classifying benthic habitats using unsupervised and supervised classifications on three different sets of input data. For habitat maps produced using a supervised classification on backscatter, backscatter with bathymetry and derivatives, and bathymetry and derivatives alone, large differences were observed in output maps, and significant increases in accuracy were observed when using the backscatter and bathymetry combination. For habitat maps produced using an unsupervised classification on backscatter, backscatter with bathymetry and derivatives, and bathymetry and derivatives alone, little effect on the overall performance of the classifier is observed. Accuracy of the supervised outputs varies based on how the accuracy statistics are generated; from the sampled values or from an estimate of the population values. A class-wise assessment of the unsupervised outputs in this study reveals that the unsupervised method is optimal for distinguishing coarse substrata and poorer at distinguishing soft substrata when using bathymetric derivatives alone. The study shows that there are issues associated with using the EUNIS habitat classification scheme for classifying video ground-truth data, and that the use of this system could be a source of error in trying to match these data with the ISO Cluster classes. Many areas for improvement have been identified with the EUNIS system and work is ongoing to

improve it (Galparsoro *et al.*, 2012), so many of the issues highlighted here may not be valid in future. Most importantly though, this paper demonstrates the clear need to conduct more research into how best to approach benthic habitat mapping in the future. It is recommended that more investigations are carried out to assess variability due to the types of input data, the resolution of the input data, and the classifier used to generate the map.

## Acknowledgements

Many thanks to the captain and crew of *RV Corystes* for logistical support and guidance in conducting the acoustic and ground-truth surveys. Thanks to the surveyors from the UKHO and Marine Institute for advice and assistance. The acoustic data used in this project were made available from the INTERREG IVa funded INIS Hydro project.

## References

- Anderson J. T., Van Holliday D., Kloser R., Reid D. G., and Simard Y. 2008. Acoustic seabed classification: Current practice and future directions. *ICES Journal of Marine Science*, 65: 1104–1011.
- Atkins W. S. 1997. Geology and physical environment. *In* Coasts and Seas of the United Kingdom. Region 17, Northern Ireland, pp. 19–40. Ed. by Barne, Robson, Kaznowska, Doody, Davidson, and Buck. Joint Nature Conservation Committee, Peterborough.
- Attrill M. J., and Rundle S. D. 2002. Ecotone or Ecocline: Ecological boundaries in estuaries. *Estuarine, Coastal and Shelf Science*, 55: 929–936.
- Bangdiwala S. I. 1988. The Agreement Chart. Department of Biostatistics, University of North Carolina at Chapel Hill, Institute of Statistics Mimeo Series No. 1859.
- Beaman R. J., Daniell J. J., and Harris P. T. 2005. Geology-benthos relationships on a temperate rocky bank, eastern bass strait, Australia. *Marine and Freshwater Research*, 36: 943–958.
- Blondel P. 2000. Automatic mine detection by textural analysis of COTS sidescan sonar imagery. *International Journal of Remote Sensing*, 21: 3115–3128.
- Borgeld J. C., Hughes Clarke J. E., Goff J. A., Mayer L. A., and Curtis J. A. 1999. Acoustic backscatter of the 1995 flood deposit on the Eel shelf. *Marine Geology*, 154: 197–210.
- Brown C. J., and Collier J. S. 2008. Mapping benthic habitat in regions of gradational substrata: An automated approach utilising geophysical, geological, and biological relationships. *Estuarine, Coastal and Shelf Science*, 78: 203–214.
- Brown C. J., Smith S. J., Lawton P., and 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.
- Callaway A., Smyth J., Brown C. J., Quinn R., Service M., and Long D. 2009. The impact of scour processes on a smothered reef system in the Irish Sea. *Estuarine, Coastal and Shelf Science*, 84: 409–418.
- Che Hasan R., Ierodiaconou D., Laurenson L., and Schimel A. 2014. Integrating multibeam backscatter angular response, mosaic and bathymetry data for benthic habitat mapping. *PLoS ONE*, 9: e97339.
- Che Hasan R., Ierodiaconou D., and Monk J. 2012. Evaluation of four supervised learning methods for benthic habitat mapping using backscatter from multi-beam sonar. *Remote Sensing*, 4: 2427–3443.
- Clements A., Strong J. A., Flanagan C., and Service M. 2010. Objective stratification and sampling-effort allocation of ground-truthing in benthic-mapping surveys. *ICES Journal of Marine Science*, 67: 628–637.
- Collier J. S., and Brown C. J. 2005. Correlation of sidescan backscatter with grain size distribution of surficial seabed sediments. *Marine Geology*, 214: 431–449.
- Cooper J. A. G., and Navas F. 2004. Natural bathymetric change as a control on century-scale shoreline behavior. *Geology*, 32: 513–516.
- Cramer H. 1946. *Mathematical methods of statistics*. Princeton University Press, Princeton. 282 pp.
- Dauvin J. C., Bellan G., and Bellan-Santini D. 2008. The need for clear and comparable terminology in benthic ecology. Part I. Ecological concepts. *Aquatic Conservation: Marine and Fresh Water Ecosystems*, 18: 432–445.
- Diaz R. J., Solan M., and Valente R. M. 2004. A review of approaches for classifying benthic habitats and evaluating habitat quality. *Journal of Environmental Management*, 73: 165–181.
- Diesing M., Coggan R., and Vansten K. 2009. Widespread rocky reef occurrence in the central English Channel and the implications for predictive habitat mapping. *Estuarine, Coastal and Shelf Science*, 83: 647–658.
- Diesing M., Green S. L., Stephens D., Lark R. M., Stewart H. A., and Dove D. 2014. Mapping seabed sediments: Comparison of manual, geostatistical, object-based image analysis and machine learning approaches. *Continental Shelf Research*, 84: 107–119.
- Dutertre M., Hamon D., Chevalier C., and Ehrhold A. 2013. The use of the relationships between environmental factors and benthic macrofaunal distribution in the establishment of a baseline for coastal management. *ICES Journal of Marine Science*, 70: 294–308.
- Elvenes S., Dolan M. F. J., Buhl-Mortensen P., and Bellec V. K. 2014. An evaluation of compiled single-beam bathymetry data as a basis for regional sediment and biotope mapping. *ICES Journal of Marine Science*, 71: 867–881.
- ESRI. 2012. *ArcGIS Desktop: Release 10.1*. Environmental Systems Research Institute, Redlands, CA.
- European environment agency. 2007. EUNIS habitat classification 200711. <http://eunis.eea.europa.eu/habitats.jsp> (last accessed 07 October 2012).
- Ferrini V. L., and Flood R. D. 2006. The effects of fine scale surface roughness and grain size on 300 kHz multibeam backscatter intensity in sandy marine sedimentary environments. *Marine Geology*, 228: 158–172.
- Fonseca L., and Calder B. 2005. Geocoder: An efficient backscatter map constructor. *Proceedings of the U.S. Hydrographic Conference 2005*, San Diego.
- Foody G. M. 2004. Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering and Remote Sensing*, 70: 627–633.
- Galparsoro I., Connor D. W., Borja A., Aish A., Amorim P., Bajjouk T., Chambers C., *et al.* 2012. Using EUNIS habitat classification for benthic mapping in European seas: Present concerns and future needs. *Marine Pollution Bulletin*, 64: 2630–2638.
- Girardoux P. 2014. *Pgirmess: Data analysis in ecology*. R package version 1.5.9. <http://cran.r-project.org/web/packages/pgirmess/index.html>. (last accessed 24/09/2014).
- Giusti M., Innocenti C., and Canese S. 2014. Predicting suitable habitat for the gold coral *Savalia savaglia* (Bertoloni, 1819) (Cnidaria, Zoantharia) in the South Tyrrhenian Sea. *Continental Shelf Research*, 81: 19–28.
- Goff J. A., Kraft B. J., Mayer L. A., Schock S. G., Sommerfield C. K., Olson H. C., Gulick S. P. S., *et al.* 2004. Seabed characterization on the New Jersey middle and outer shelf: Correlatability and spatial variability of seafloor sediment properties. *Marine Geology*, 209: 147–172.
- Goff J. A., Olson H. C., and 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.
- Great Britain Hydrographic Department. 1985. *Irish coast pilot: Offshore and coastal waters round Ireland including routes to the Irish Sea from Atlantic Ocean landfalls*. Hydrographer of the Navy, Taunton. 504 pp.
- GSNI. 1997. *Geological map of Northern Ireland*. Geological Survey of Northern Ireland, Belfast.

- Herzfeld U. C. 1993. A method for seafloor classification using directional variograms, demonstrated for data from the western flank of the Mid-Atlantic Ridge. *Mathematical Geology*, 25: 901–924.
- Herzfeld U. C., and Higginson C. A. 1996. Automated geostatistical seafloor classification—Principles, parameters, feature vectors, and discrimination criteria. *Computers and Geosciences*, 22: 35–41, 43–52.
- Hill A. E., Brown J., and Fernand L. 1995. The Western Irish Sea gyre: A retention system for Norway Lobster (*Nephrops norvegicus*). *Oceanologica Acta*, 19: 357–368.
- Holmes K. W., Van Niel K. P., Radford B., Kendrick G. A., and Grove S. L. 2008. Modelling distribution of marine benthos from hydroacoustics and underwater video. *Continental Shelf Research*, 28: 1800–1810.
- Howell K. L., Davies J. S., and Narayanaswamy B. E. 2010. Identifying deep-sea megafaunal epibenthic assemblages for use in habitat mapping and marine protected area network design. *Journal of the Marine Biological Association of the United Kingdom*, 90: 33–68.
- Ierodiconou D., Monk J., Rattray A., Laurenson L., and Versace V. L. 2011. Comparison of automated classification techniques for predicting benthic biological communities using hydroacoustics and video observations. *Continental Shelf Research*, 31: 528–538.
- INIS Hydro. 2012. [www.inishydro.eu](http://www.inishydro.eu) (last accessed 07 October 2012).
- Jackson D. R., Baird A. M., Crisp J. J., and Thomson P. A. G. 1986a. High-frequency bottom backscatter measurements in shallow water. *Journal of the Acoustical Society of America*, 80: 1188–1199.
- Jackson D. R., Winebrenner D. P., and Ishimaru A. 1986b. Application of the composite roughness model to high-frequency bottom backscattering. *Journal of the Acoustical Society of America*, 79: 1410–1422.
- Jackson J. B. C., Kirby M. X., Berger W. H., Bjorndal K. A., Botsford L. W., Bourque B. J., Bradbury R. H., et al. 2001. Historical overfishing and the recent collapse of coastal ecosystems. *Science*, 293: 629–638.
- Jackson D. W. T., Cooper J. A. G., and Del Rio L. 2005. Geological control of beach morphodynamic state. *Marine Geology*, 216: 297–314.
- JNCC. 2011. Natura 2000 Data Form, Murlough. <http://jncc.defra.gov.uk/protectedsites/sacselection/n2kforms/UK0016612.pdf> (last accessed 21 January 2013).
- JNCC. 2014. UKSeaMap 2010 – Predictive mapping of seabed habitats. <http://jncc.defra.gov.uk/page-2117> (last accessed 01 May 2014).
- Jones D. O. B., and Brewer M. E. 2012. Response of megabenthic assemblages to different scales of habitat heterogeneity on the Mauritanian slope. *Deep Sea Research Part I: Oceanographic Research Papers*, 67: 98–110.
- Kostylev V. A., Todd B. J., Fader G. B. J., Courtney R. C., Cameron G. D. M., and Pickrill R. A. 2001. Benthic habitat mapping on the Scotian Shelf based on multibeam bathymetry, surficial geology and sea floor photographs. *Marine Ecology Progress Series*, 219: 121–137.
- Kostylev V. E. 2012. Benthic habitat mapping from seabed acoustic surveys: Do implicit assumptions hold? *In Sediments, Morphology and Sedimentary Processes on Continental Shelves: Advances in Technologies, Research, and Applications*. Ed. by Li, Sherwood, and Hill. John Wiley & Sons Ltd, Chichester, UK.
- Lathrop R. G., Cole M., Senyk N., and Butman B. 2006. Seafloor habitat mapping of the New York Bight incorporating sidescan sonar data. *Estuarine, Coastal and Shelf Science*, 68: 221–230.
- Lecomte J. B., Benoît H. P., Etienne M. P., Bel L., and Parent E. 2013. Modeling the habitat associations and spatial distribution of benthic macroinvertebrates: A hierarchical Bayesian model for zero-inflated biomass data. *Ecological Modelling*, 265: 74–84.
- Lucieer V., Hill N. A., Barret N. S., and Nicholl S. 2013. Do marine substrates 'look' and 'sound' the same? Supervised classification of multibeam acoustic data using autonomous underwater vehicle images. *Estuarine, Coastal and Shelf Science*, 117: 94–106.
- Lucieer V., and Lucieer A. 2009. Fuzzy clustering for seafloor classification. *Marine Geology*, 264: 230–241.
- Lundblad E. R., Wright D. J., Miller J., Larkin E. M., Rinehart R., Naar D. F., Donahue B. T., et al. 2006. A Benthic Terrain classification scheme for American Samoa. *Marine Geodesy*, 29: 89–111.
- Marsh I., and Brown C. 2009. Neural network classification of multi-beam backscatter and bathymetry data from Stanton Bank (Area IV). *Applied Acoustics*, 70: 1269–1276.
- McGonigle C., Brown C. J., and Quinn R. 2010a. Operational parameters, data density and benthic ecology: Considerations for image-based classification of multibeam backscatter. *Marine Geodesy*, 33: 16–38.
- McGonigle C., Brown C. J., and Quinn R. 2010b. Insonification orientation and its relevance for image-based classification of multibeam backscatter. *ICES Journal of Marine Science*, 67: 1010–1023.
- McGonigle C., and Collier J. 2014. Interlinking backscatter, grain size and benthic community structure. *Coastal and Shelf Science*, Estuarine.
- Meyer D., Zeileis A., and Hornik K. 2013. vcd: Visualizing Categorical Data. R package version 1.3–1. <http://cran.r-project.org/web/packages/vcd/index.html>. (last accessed 15/05/2014).
- Michaels W. L. 2007. Acoustic classification of marine physical and biological landscapes. ICES cooperative research report No. 286.
- Mitchell N. C. 1993. A model for attenuation of backscatter due to sediment accumulations and its application to determine sediment thickness with GLORIA sidescan sonar. *Journal of Geophysical Research*, 98: 22477–22493.
- Mitchell N. 1996. Processing and analysis of Simrad multibeam sonar data. *Marine Geophysical Researches*, 18: 729–739.
- Mitchell N. C., and Hughes Clarke J. E. 1994. Classification of seafloor geology using multibeam sonar data from the Scotian Shelf. *Marine Geology*, 121: 143–160.
- Murray A. B., and Thieler E. R. 2004. A new hypothesis and exploratory model for the formation of large-scale inner-shelf sediment sorting and “rippled scour depressions”. *Continental Shelf Research*, 24: 295–315.
- Navarro D. J. 2014. Learning statistics with R: A tutorial for psychology students and other beginners. (Version 0.4) University of Adelaide, Adelaide, Australia.
- Plets R., Quinn R., Forsythe W., Westley K., Bell T., Benetti S., McGrath F., et al. 2011. Using Multibeam Echo-Sounder Data to Identify Shipwreck Sites: Archaeological assessment of the Joint Irish Bathymetric Survey data. *International Journal of Nautical Archaeology*, 40: 87–98.
- Pontius R. G., and Millones M. 2011. Death to kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, 32: 4407–4429.
- Rattray A., Ierodiconou D., Laurenson L., Burq S., and Reston M. 2009. Hydro-acoustic remote sensing of benthic biological communities on the shallow South East Australian continental shelf. *Estuarine, Coastal and Shelf Science*, 84: 237–245.
- Rattray A., Ierodiconou D., Monk J., Versace V. L., and Laurenson L. J. B. 2013. Detecting patterns of change in benthic habitats by acoustic remote sensing. *Marine Ecology Progress Series*, 477: 1–13.
- Reiss H., Cunze S., König K., Neumann H., and Kröncke I. 2011. Species distribution modelling of marine benthos: A North Sea case study. *Marine Ecology Progress Series*, 442: 71–86.
- Riley S. J., DeGloria S. D., and Elliot R. 1999. A terrain ruggedness index that quantifies topographic heterogeneity. *Intermountain Journal of Sciences*, 5: 23–27.
- Sappington J. M., Longshore K. M., and Thompson D. B. 2007. Quantifying landscape ruggedness for animal habitat analysis: A case study using Bighorn Sheep in the Mojave Desert. *Journal of Wildlife Management*, 71: 1419–1426.
- Siegel S., and Castellan N. J. 1988. *Non Parametric Statistics for the Behavioural Sciences*. McGraw-Hill, Inc., New York. 312 pp.
- Stephens D., and Diesing M. 2014. A comparison of supervised classification methods for the prediction of substrate type using multibeam acoustic and legacy grain-size data. *PLoS ONE*, 9: e93950.
- Urgeles R., Locat J., Schmitt T., and Hughes Clarke J. E. 2002. The July 1996 flood deposit in the Saguenay Fjord, Quebec, Canada:

- Implications for sources of spatial and temporal backscatter variations. *Marine Geology*, 184: 41–60.
- Van Hoey G., Degraer S., and Vincx M. 2003. Macrobenthic community structure of soft bottom sediments at the Belgian Continental Shelf. *Estuarine, Coastal and Shelf Science*, 59: 599–613.
- Wienberg C., and Bartholomä A. 2005. Acoustic seabed classification in a coastal environment (outer Weser Estuary, German Bight) – A new approach to monitor dredging and dredge spoil disposal. *Continental Shelf Research*, 25: 1143–1156.
- Wilson M. F. J., O'Connell B., Brown C., Guinan J. C., and Grehan A. J. 2007. Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the Continental slope. *Marine Geodesy*, 30: 1–35.
- Worm B., Barbier E. B., Beaumont N., Duffy J. E., Folke C., Halpern B. S., Jackson J. B. C., *et al.* 2006. Impacts of biodiversity loss on ocean ecosystem services. *Science*, 314: 787–790.
- Zar J. H. 1999. *Biostatistical Analysis*. Prentice Hall, New Jersey. 663 pp.
- Zeileis A., Meyer D., and Hornik K. 2007. Residual-based shadings for visualizing (conditional) independence. *Journal of Computational and Graphical Statistics*, 16: 507–525.

*Handling editor: Steven Degraer*