

Megafaunal distribution and biodiversity in a heterogeneous landscape: the iceberg-scoured Rockall Bank, NE Atlantic

Katleen Robert^{1,*}, Daniel O. B. Jones², Veerle A. I. Huvenne²

¹School of Ocean and Earth Science, University of Southampton, Waterfront Campus, European Way, Southampton SO14 3ZH, UK

²National Oceanography Centre, European Way, Southampton SO14 3ZH, UK

*Corresponding author: kr2e11@soton.ac.uk

Marine Ecology Progress Series 501: 67–88 (2014)

Supplement 1. Effect of cryptic species on Shannon diversity index (H')

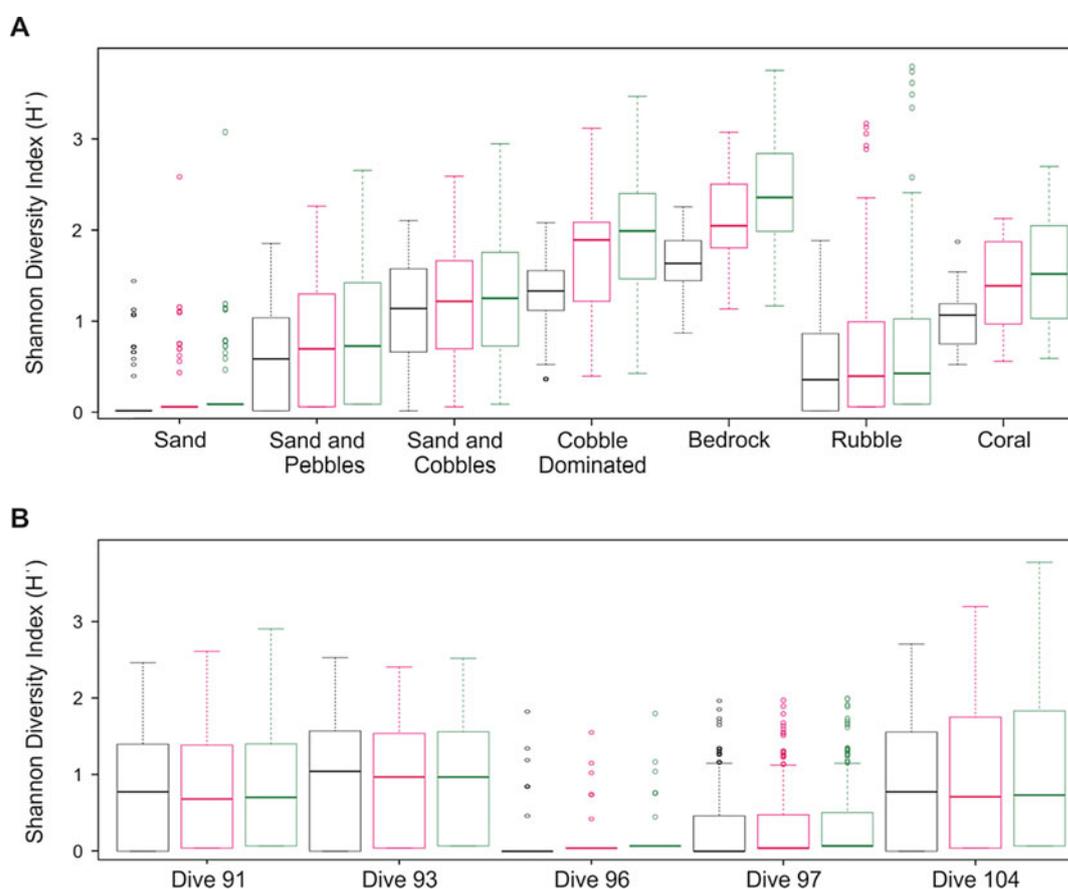


Fig. S1. Shannon biodiversity index (H') observed for (A) each substratum type and (B) each dive (black). When 2 or more specimens of a morphospecies of ascidians (111), bryozoans (135), cerianthids (8), cormatulids (65), sponges (119 or 85 for demosponges), ophiuroids (205), sabellid worms (99) or sipuncula worms (144) were observed within an image, each individual was randomly assigned, with replacement, to a potential cryptic species and biodiversity recalculated. The number of potential cryptic species (previously shown in brackets) was estimated by setting a percentage (pink, 5% and green, 20%) of the number of taxa listed in OBIS (Ocean Biogeographic Information System, www.iobis.org/) for the North Atlantic Ocean

Supplement 2. Seabed facies



Fig. S2. Images representing the different seabed facies observed along the 5 remotely operated imagery transects on Rockall Bank. Lasers dots are separated by 10 cm

Supplement 3. Backscatter and sediment interpretation maps

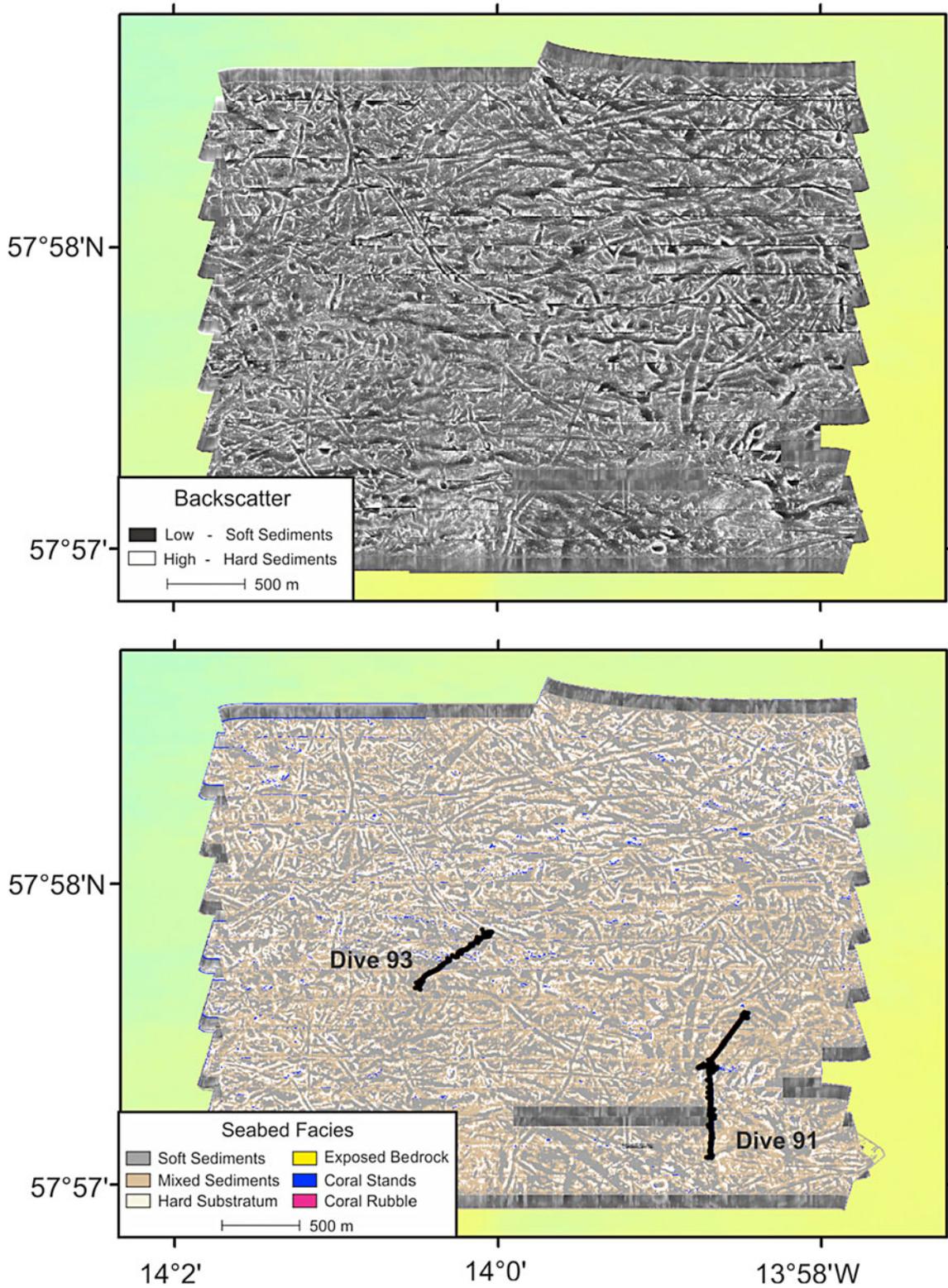


Fig. S3. Autonomous underwater vehicle mission 043 sidescan sonar backscatter map (top) and sediment interpretation map with overlaid remotely operated vehicle imagery transects (bottom)

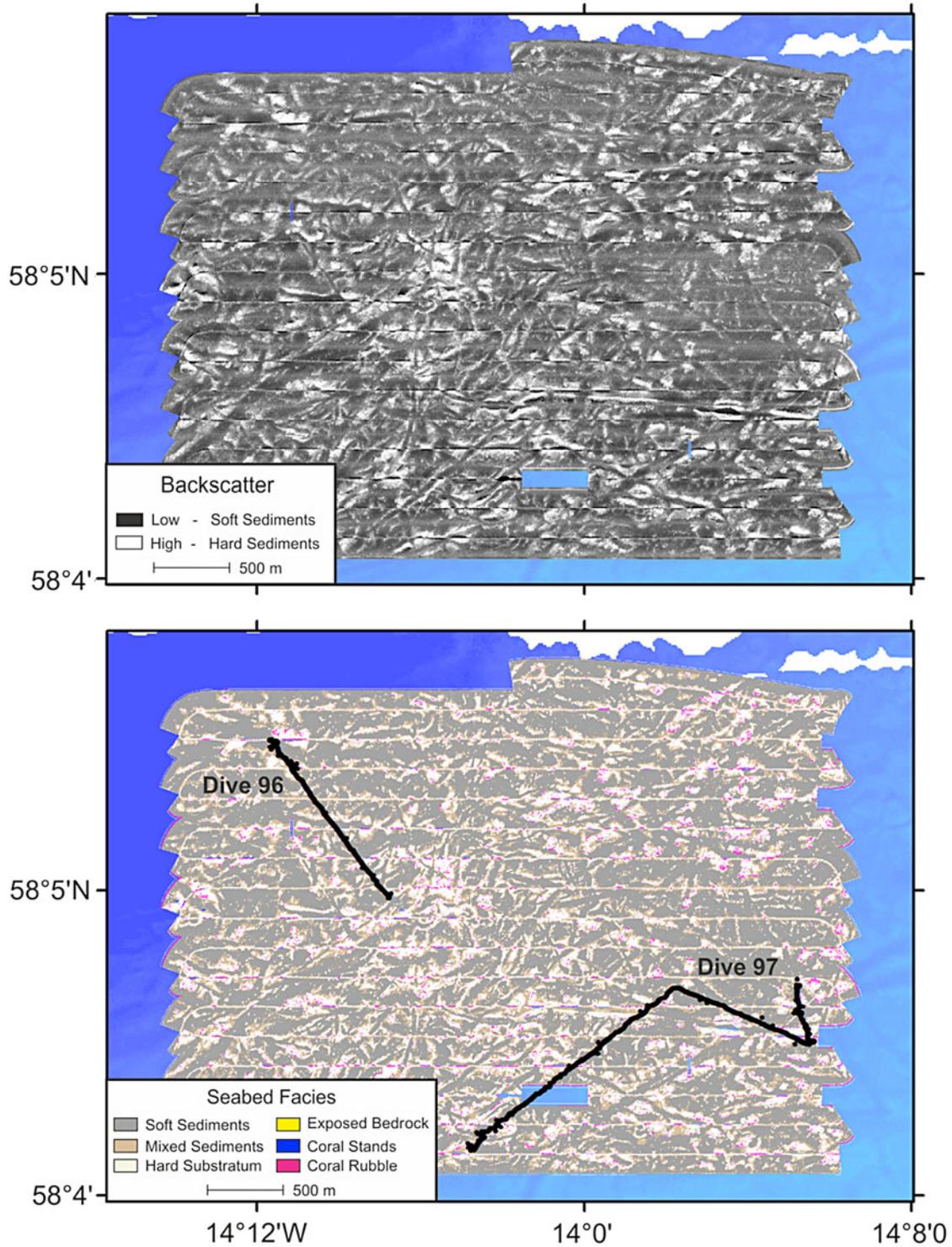


Fig. S4. Autonomous underwater vehicle mission 044 sidescan sonar backscatter map (top) and sediment interpretation map with overlaid remotely operated vehicle imagery transects (bottom)

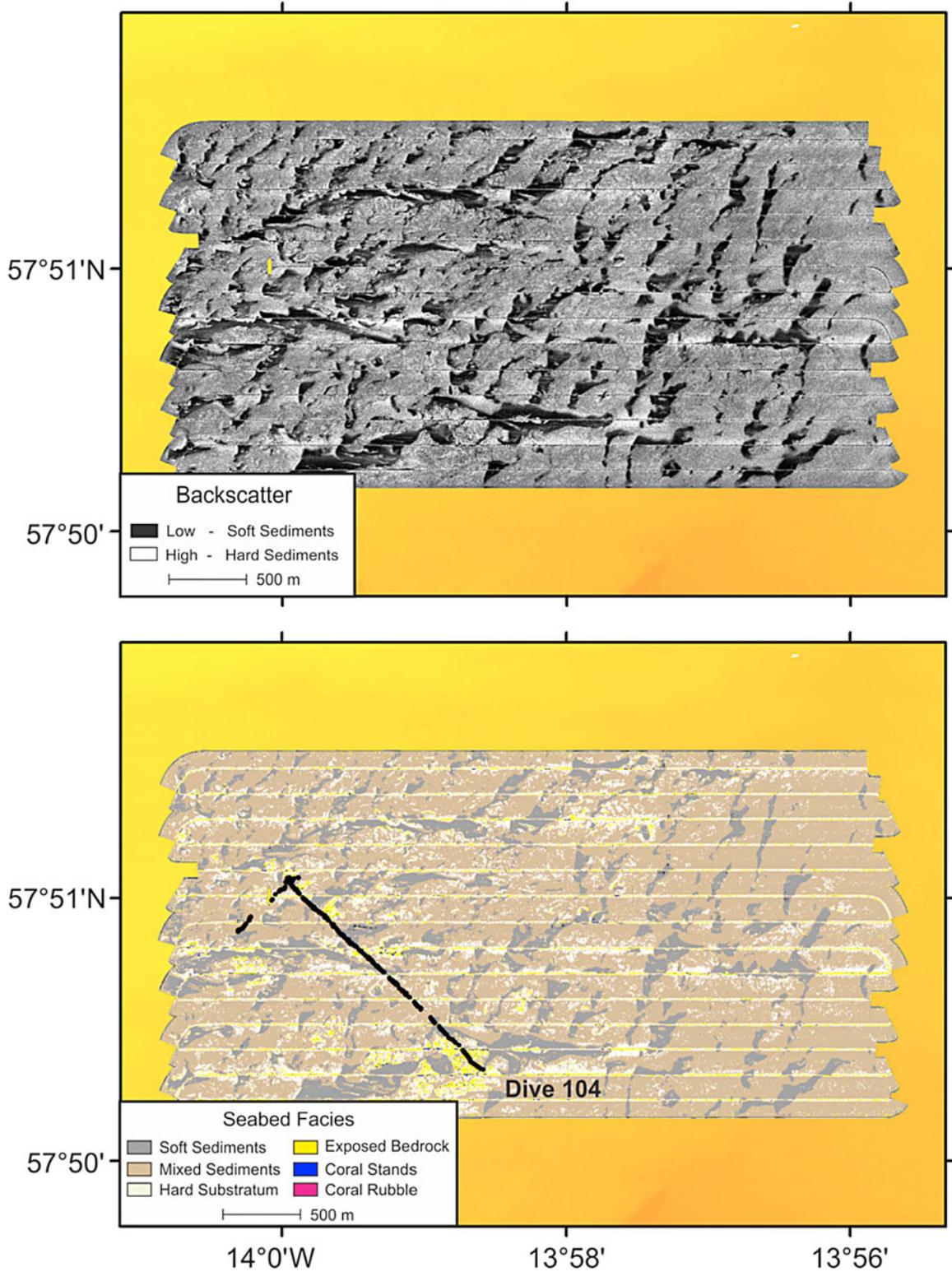


Fig. S5. Autonomous underwater vehicle mission 045 sidescan sonar backscatter map (top) and sediment interpretation map with overlaid remotely operated vehicle imagery transects (bottom)

Supplement 4. Detailed descriptions of the statistical techniques employed

Environmental variable selection

Redundancy analysis (RDA) (van den Wollenberg 1977) was chosen because an adjusted R^2 could be obtained using Ezekiel's adjustment formula (Peres-Neto et al. 2006). Because any explanatory variable included in a model increases the R^2 regardless of whether it models responses or noise, this inflation must be taken into account and the resulting adjusted R^2 used instead. This was considered particularly important for this study as a large number of environmental variables were involved (e.g. percentage cover of each substratum type, class and landscape metrics as well as bathymetric and backscatter derived variables). In RDA, species are assumed to be linearly related to environmental variables and second degree explanatory variables for substratum composition were also introduced into the model to account for potential non-linear relationships (e.g. unimodal biodiversity relationships to percentage cover of rubble).

Model significance was assessed using permutation tests with a minimal permutation number of 999 (Makarek & Legendre 2002). If found significant, forward selection by permutation tests based on p-values was applied to reduce the number of explanatory variables and obtain a more parsimonious model (Borcard et al. 2011). Owing to the large number of available explanatory variables, forward selection was first applied to each environmental dataset separately (substratum percentage cover, class metrics, landscape metrics and bathymetric variables). The selected variables were then combined, a second forward selection process was applied and variance inflation factors (VIF) were used to exclude additional explanatory variables which showed strong collinearity with others present within the model. Only explanatory variables with VIF below 5 were retained (Stine 1995). When examining the influence of the previously described environmental factors on biodiversity, multiple linear regression was used and forward selection was based on the Akaike Information Criterion (AIC), which is a measure based on the compromise between model simplicity and increased fit obtained when using additional parameters (Crawley 2005). For both the multivariate and univariate analyses, the forward selection process was carried out twice: once including all environmental variables and once including only sonar-derived variables.

The morphospecies by images collected matrix was $\log_{10}(x + 1)$ transformed before the analysis to reduce the influence of abundant species (Clarke & Warwick 2001). Variation partitioning was used to examine how much variation could be explained by environmental variables obtained from the imagery or the sonar maps when controlling for the effect of the other (Borcard & Legendre 2002, Peres-Neto et al. 2006). Variation partitioning is a technique which, using multiple RDAs, allows for the fraction of variation explained by a set of explanatory variables (e.g. imagery derived environmental descriptors) to be separated from the fraction of variation explained by another set of explanatory variables (e.g. sonar derived environmental descriptors). As some of the variation will be explained by both sets of explanatory variables, the total amount of variation explained will be less than the sum of the 2 fractions examined (Borcard et al. 2011).

Species assemblages

K-means partitioning was used on the morphospecies matrix to separate images into clusters representing different species assemblages. This method aims to separate observations into a predefined number of clusters based on sample proximity when data points are positioned in a space where each species represents an axis. Data points are assigned a cluster in such a manner as to minimize the distance between individual samples and the center of their assigned cluster (Hartigan & Wong 1979). Points are swapped between clusters, and cluster centers are updated accordingly until a local optimal solution is achieved. As this is an iterative method and a

predefined number of clusters is required, the analysis was successively conducted using 2 to 10 clusters and 100 random starting configurations. The number of clusters was chosen using the ‘simple structure index’, as defined in Dimitriadou et al. (2002), and which takes into account elements describing intra and inter cluster structure. Analysis of similarities (ANOSIM) was also carried out in R to assess significant differences between the identified clusters (based on 999 permutations and Euclidian distance). The representative species for each cluster were determined using the ‘species indicator values’ (Dufrêne & Legendre 1997).

Model evaluation

In order to evaluate the goodness-of-fit of the model, assess its ability to predict new data points and, as a single dataset was available to evaluate the model, a ‘holdout partition’ approach was taken (Kohavi 1995, Manel et al. 1999, Verfaillie et al. 2009). In a ‘holdout partition’ approach, a subset of the data, in this case 300 randomly selected images (with proportional representation by transect), is removed from the original dataset and the model (RDA) parameters are recomputed. The removed data points are then reclassified based on the recomputed model, and the percentage of data points reassigned to the same cluster is calculated. The initial sample is then returned to the dataset and a subsequent sample is removed. The process is repeated multiple times to estimate variability in the results obtained. For this study, the environmental variables associated with each removed image were used to predict updated positions in ordination space (using RDA axis 1 and 2) based on the RDA model recomputed when the 300 randomly selected images were removed. The Euclidean distance to the centroid of each K-means cluster was then computed and the removed images were assigned the class of the closest cluster (Guisan & Zimmermann 2000, Anderson & Willis 2003, Wang et al. 2004). The percentage of removed images reassigned to the same cluster was calculated and the process was repeated 100 times.

Spatial dependency assessment

To determine whether the spatial structure present in species distribution was the result of spatial structuring in the environmental factors (induced spatial dependence), we carried out a Principal Coordinates of Neighbour Matrices (PCNM) approach (Borcard & Legendre 2002, Dray et al. 2006). PCNMs are a type of distance-based eigenvector maps, and are computed using a truncated Euclidean distance matrix based on sample location (distance threshold of 110 m) followed by a Principal Coordinate Analysis. Only eigenvectors with associated positive eigenvalues and a significant Moran’s index of spatial autocorrelation (Moran 1950, Sokal & Oden 1978, Dray et al. 2006) (an index used to test that the pattern represented is spatially structure and not random) are retained. As these eigenvectors are representations of spatial structure based on distances between sampling sites, RDA is used to select the eigenvectors which significantly represent the spatial structure also present in the biological data. To determine which environmental variables explained the spatial structure detected, the resulting significant canonical axes were regressed against the environmental dataset. Gaussian variograms were also fitted to the significant canonical axes. Variograms describe the relationship between the variance in values observed at points with regards to their separation in space and can be used to estimate the range of spatial dependency (Bellier et al. 2007, Loots et al. 2011). Analyses were carried out using the ‘vegan’, ‘labdsv’ and ‘gstat’ package in R.

LITERATURE CITED

- Anderson MJ, Willis TJ (2003) Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. *Ecology* 84:511–525
- Bellier E, Monestiez P, Durbec JP, Candau JN (2007) Identifying spatial relationships at multiple scales: principal coordinates of neighbour matrices (PCNM) and geostatistical approaches. *Ecography* 30:385–399
- Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecol Modell* 153:51–68
- Borcard D, Gillet F, Legendre P (2011) *Numerical ecology with R*. Springer, New York, NY
- Clarke KR, Warwick RM (2001) *Change in marine communities: an approach to statistical analysis and interpretation*, 2nd edn. PRIMER-E, Plymouth
- Crawley MJ (2005) *Statistics: an introduction using R*. John Wiley & Sons, Chichester
- Dimitriadou E, Dolničar S, Weingessel A (2002) An examination of indexes for determining the number of clusters in binary data sets. *Psychometrika* 67:137–159
- Dray S, Legendre P, Peres-Neto PR (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecol Modell* 196:483–493
- Dufrêne M, Legendre P (1997) Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecol Monogr* 67:345–366
- Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecol Modell* 135:147–186
- Hartigan JA, Wong MA (1979) Algorithm AS 136: a K-means clustering algorithm. *J R Stat Soc Ser C Appl Stat* 28:100–108
- Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proc 14th Int Joint Conf on Artificial Intelligence*, Montreal 2:1137–1143
- Loots C, Vaz S, Planque B, Koubbi P (2011) Understanding what controls the spawning distribution of North Sea whiting (*Merlangius merlangus*) using a multi-model approach. *Fish Oceanogr* 20:18–31
- Makarenkov V, Legendre P (2002) Nonlinear redundancy analysis and canonical correspondence analysis based on polynomial regression. *Ecology* 83:1146–1161
- Manel S, Dias JM, Ormerod SJ (1999) Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecol Modell* 120:337–347
- Moran PAP (1950) Notes on continuous stochastic phenomena. *Biometrika* 37:17–23
- Peres-Neto PR, Legendre P, Dray S, Borcard D (2006) Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology* 87:2614–2625
- Sokal RR, Oden NL (1978) Spatial autocorrelation in biology: 1. Methodology. *Biol J Linn Soc* 10:199–228
- Stine RA (1995) Graphical interpretation of variance inflation factors. *Am Stat* 49:53–56
- van den Wollenberg A (1977) Redundancy analysis an alternative for canonical correlation analysis. *Psychometrika* 42:207–219
- Verfaillie E, Degraer S, Schelfaut K, Willems W, Van Lancker V (2009) A protocol for classifying ecologically relevant marine zones, a statistical approach. *Estuar Coast Shelf Sci* 83:175–185
- Wang L, Sousa WP, Gong P (2004) Integration of object-based and pixel-based classification for mapping mangroves with IKONOS imagery. *Int J Remote Sens* 25:5655–5668

Supplement 5. Morphospecies list

Table S1. List and count of the morphospecies observed along the 5 remotely operated imagery transects on Rockall Bank. Sample images were deposited in the SERPENT media archive (<http://archive.serpentproject.com/>)

Index	Name\description	Count	Index	Name\description	Count
1	<i>Parastichopus tremulus</i>	185	42	Porifera Cup sp. 2	1
2	<i>Munida sarsi</i>	2373	43	<i>Caryophyllia</i> sp. 2	6
3	<i>Cidaris cidaris</i>	59	44	Cephalopoda sp. 1	2
4	<i>Asterias rubens</i>	28	45	Hydrozoa sp. 2	5
5	<i>Henricia</i> sp. 1	117	46	Bryozoa sp. 2	10
6	<i>Reteporella</i> sp. 1	942	47	Holothuroidea sp. 1	7
7	<i>Lophelia pertusa</i>	39	48	Hydrozoa sp. 3	2
8	Sabellidae sp. 1	272	49	Holothuroidea sp. 2	2
9	Actinaria sp. 1	227	50	Porifera Orange Lobose sp. 2	1
10	<i>Echinus</i> sp. 1 (possibly <i>E. acutus</i>)	19	51	Anthozoa sp. 2	1
11	Portunidea sp. 1	24	52	Pennatulacea sp. 1	7
12	Caridea sp. 1	80	53	Porifera Cup sp. 3	1
13	Sipuncula sp. 1	233	54	<i>Stylaster</i> sp. 1	2
14	Porifera Cup sp. 1	44	55	Paguridae sp. 1	1
15	<i>Hippasteria</i> sp. 1	19	56	<i>Ceramaster</i> sp. 1	1
16	<i>Echinus</i> sp. 2	10	57	Desmosongiae Pink Encrusting sp. 1	1
17	Ascidiacea sp. 1	96	58	<i>Bolocera tuediae</i>	1
18	Cyclostomatida sp. 1	175	59	Pycnogonidae sp. 1	1
19	Desmosongiae Orange Encrusting sp. 1	138	60	Asteroidea sp. 2	1
20	Bryozoa sp. 1	152	61	Cormatulida sp. 1	5
21	Porifera Brown Lamellate sp. 1	69	62	<i>Kophobelemnion stelliferum</i>	5
22	Desmosongiae White Encrusting sp. 1	445	63	Buccinidae sp. 1	3
23	Ceriantharia sp. 1	13	64	Custacea sp. 1	2
24	<i>Caryophyllia</i> sp. 1	92	65	<i>Nephrops norvegicus</i>	1
25	Porifera Yellow Columnar sp. 1	28	66	<i>Caryophyllia smithii</i>	1
26	Ophiuroidea sp. 1	263	67	Asteroidea sp. 3	1
27	Desmosongiae Yellow Encrusting sp. 1	705	68	Unknown 1	10
28	<i>Porania pulvillus</i>	10	69	Asteroidea sp. 4	3
29	<i>Hymedesmia</i> sp. 1	38	70	Porifera White Lamellate sp. 1	6
30	Porifera Orange Branching sp. 1	19	71	Ophiuroidea sp. 2	6
31	Hydrozoa sp. 1	35	72	Porifera Yellow Lobose sp. 1	10
32	Porifera Tan Columnar sp. 1	48	73	Cnidaria sp. 1	1
33	Porifera Spherical sp. 1	36	74	Brachyura sp. 1	1
34	<i>Echnius</i> sp. 3 (possibly <i>E. elegans</i>)	11	75	Cnidaria sp. 2	1
35	Desmosongiae Red Encrusting sp. 1	60	76	Porifera Spherical sp. 2	1
36	<i>Actinauge richardi</i>	6	77	Porifera sp. 1	17
37	<i>Phelliactis</i> sp. 1	4	78	Asteroidea sp. 5	1
38	Asteroidea sp. 1	4	79	Unknown 2	2
39	Porifera Yellow Branching sp. 1	7	80	Alcyonacea sp. 1	2
40	Porifera Orange Lobose sp. 1	5	81	<i>Paromola cuvieri</i>	3
41	Porifera Orange Branching sp. 2	2			
Total			7267		

Supplement 6. Choice of area for calculation of class and landscape metrics

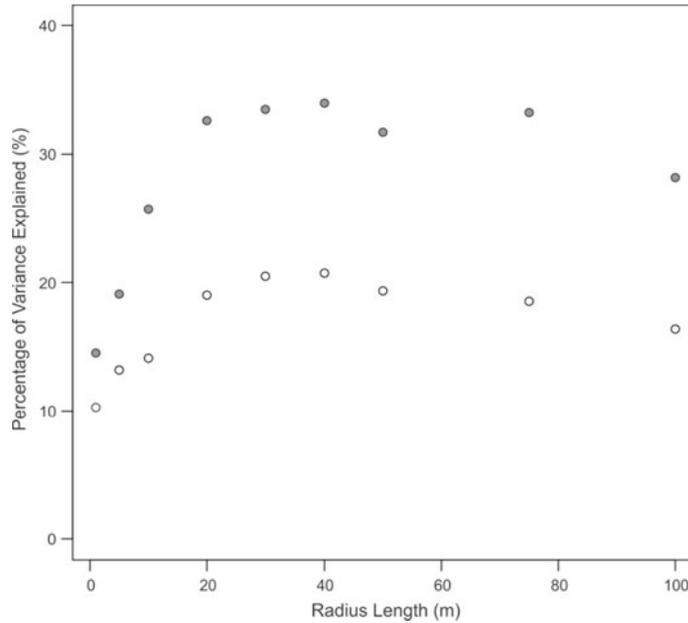


Fig. S6. Percentage of variance explained following forward selection of class and landscape metrics calculated using circular areas of varying radius length; biodiversity (H') (grey) and morphospecies composition (white)

Table S2. Comparison of the percentage of variation explained when class and landscape metrics calculated from 2 areas of different sizes are considered in combination. Italicized values show results for species composition and grey values show biodiversity. Values for chosen settings appear in bold

		Percentage of variation explained (%)								Biodiversity (H')
		Radius length (m)								
Radius Length (m)		1	5	10	20	30	40	50	75	100
1			20.0	26.8	34.2	35.2	36.0	34.1	36.9	33.5
5	13.2			26.9	33.6	34.6	35.6	34.3	37.5	34.7
10	15.5	16.3			35.4	36.3	35.3	35.5	35.5	36.5
20	20.5	21.4	21.6			37.1	37.7	38.1	41.2	39.3
30	21.6	22.6	22.8	23.8			37.0	36.8	40.7	38.3
40	22.4	23.6	23.8	25.0	23.9			36.2	40.0	37.7
50	21.6	22.9	23.8	25.2	24.2	23.3			36.8	35.0
75	22.0	23.9	24.2	26.0	26.3	25.0	23.7			35.7
100	20.6	22.0	23.1	25.7	26.0	25.1	24.0	23.2		

Species Composition

Supplement 7. Environmental variables and spatial structure as identified using PCNMs

The first 13 canonical axes of the redundancy analysis of Principal Coordinates of Neighbour Matrices (PCNMs) against the morphospecies count matrix were significant. When the first 2 axes were regressed against the selected set of environmental variables, those related to presence, percentage cover or proportion of landscape occupied by coral, cobbles or exposed bedrock had the strongest influence ($p < 0.001$). In the case of biodiversity, the significant PCNMs showed a strong negative response to percentage cover of sand ($p < 0.001$) and a positive association to presence of bedrock ($p < 0.001$) and percentage cover of coral present ($p < 0.01$) (Fig. S7)

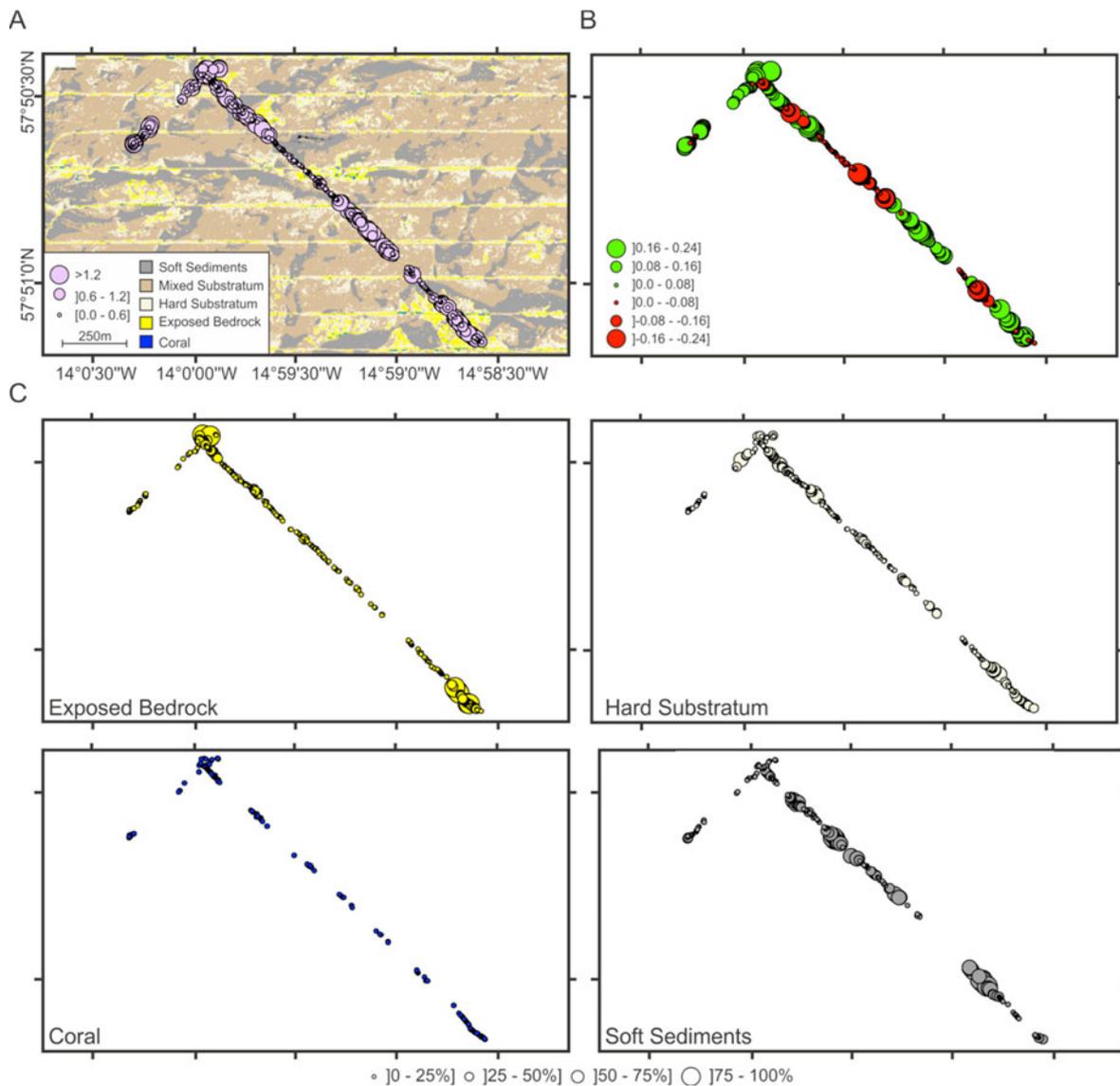


Fig. S7. (A) Distribution of Shannon diversity index (H') for each image along transect 104 overlaying a section of the sediment interpretation map from mission 45. (B) Fitted site scores obtained using the Principal Coordinates of Neighbour Matrices (PCNMs) approach. These represent the strength of the spatial structure identified as present in H' . (C) Proportions of landscape represented by different seabed facies within a 30 m radius of each image and showing similar spatial structuring as depicted by the fitted side scores