Appendix 7. Method Statement – Statistical Analyses

Verification survey of intertidal rocky shore features in the Drigg Coast EMS

1.1 Programmes used for statistical analysis

Microsoft Excel 2010 was used for general data formatting and exploration. PRIMER v6 was used for the multivariate statistical analysis carried out and is one of the most common statistical packages used for analysis of biological assemblages. MiniTab12 was used to carry out an ANOVA to test for differences in species richness between Transects.

The objective of these analyses was twofold:

1. Determine if there was any difference on species richness across the site; and
2. Determine the species composition of hard substrata biotopes.

1.2 Truncation and data consolidation

Data were transferred by the surveying taxonomists from field notes to electronic files in a standard format (see Appendix 6) to automatically create factors for use in the cluster and ordination analyses, e.g. shore height, biotope allocated, physical data, etc, and enable the data to be easily manipulated into the correct format for PRIMER without losing any detail.

Once data had been transferred to the standard format, each taxonomist's data were checked by the corresponding senior taxonomist on the same survey to ensure all species names and percentage composition for each physical data category (i.e. all physical data added up to 100%) were consistent across the teams. Any data editing, e.g. species names, were made at this point following discussion with other taxonomists to ensure consistency across projects.

Final Analytical Quality Control (AQC) of the quadrat data was carried out by the project manager to ensure there were no spelling or transcription mistakes, all relevant fields had been completed and the species were in order of their species directory code. Final automated consistency checks were also made to ensure data were complete and correct.

1.3 Species richness

Species richness (number of taxa) was calculated using the 'Count' function in Excel. This allowed the number of taxa per quadrat to be determined. No other useful diversity indices could be calculated as the data were a combination of percentage coverage of encrusting, colonial or canopy-forming species, e.g. macroalgae and barnacles, and actual abundances of free-living species, e.g. *Littorina* spp. and *Nucella lapillus*, which cannot be directly compared due to the different units of measurement used. For the purposes of statistical analysis and for general description purposes, qualifiers were ignored, e.g. Cirripedia were all juveniles thus were considered as Cirripedia rather than Cirripedia and Cirripedia #juv.

The Analysis of Variance (ANOVA) test was carried out in MiniTab12 to test for differences in species richness between Transects. This test indicated whether there were any significant differences in species richness between Transects around the Drigg Coast. The null hypothesis (H_0, no difference in species richness across transects) was tested at the 5% significance level. Where a significant difference was recorded, post-hoc pair-wise tests were conducted using the Least Significant Difference (LSD) Test aka Fisher's Test to identify which of the Transects were significantly different to each other.

1.4 Transformation

As different units of measurement cannot be directly compared and the data matrix contained both simple counts and percentage coverage data, a presence/absence transformation was applied. This type of transformation gives less abundant species in the matrix equal weight to more abundant species. Whilst this approach allows the use of all species data it precludes the use of quantitative information in the analysis of biological assemblages.

To retain quantitative information in the global analysis a possible alternative would have been to adapt the actual counts and percentage cover values to a common semi-quantitative estimation scale such as the Braun-Blanquet scale[1] (r, +, 1, 2, 3, 4, 5) or SACFOR abundance scale[2]. For a more equal weight the standardised scores could have been replaced by their numerical equivalents (i.e. median or mean abundances). However, there is no easy way to assign comparable values to percentage cover. In addition, the lower levels of the percentage cover scales (r, +, or O, R) are based more on the abundance of the species rather than on its percentage cover requiring the assignment of arbitrary values[3]. Although these alternative approaches may be possible, they introduce assumptions which may obscure the interpretation of the results. Therefore, we used the simplest variant of such numerical alternatives which is presence/absence (0-1) data.

1.5 Resemblance (similarity) matrix

To enable any multivariate analysis to be carried out, an appropriate definition of resemblance between samples must be provided to signify the similarity between samples. The Jaccard index was used in the current analysis. This similarity measure eliminates matching attributes that share a 0 value as evidence of similarity and is recommended for presence/absence data. The index syntax is given by the formula:

$J = (100*a)/(a+b+c)$

where a is the number of species present in both samples; b is the number of species present in sample 1 but absent from sample 2; and c is the number of species absent in sample 1 but present in sample 2.

1.6 Hierarchical Cluster Analysis

Cluster analysis was used to visualise the groupings of samples based on their faunal composition. Agglomerative, hierarchical clustering was carried out on the Jaccard's resemblance (similarity) matrix. The method groups the samples into small groups first (i.e. those with the highest levels of similarity based on faunal composition). These first groups are subsequently grouped together into larger groups, based on group averages, lowering the level of similarity until all of the samples are in a single cluster at the lowest level of similarity between samples. A dendrogram is then used to show the results of this clustering and indicates the level of similarity between each group of samples.

[1] Westhoff. V. & van der Maarel, E. 1978. The Braun-Blanquet approach. In: Whittaker, R.H. Classification of Plant Communities. The Netherlands:p. 289-312

[2] Connor, D.W., & Hiscock, K. 1996. Data collection methods (with Appendices 5 – 10). In: Marine Nature Conservation Review: rationale and methods, ed. by K. Hiscock, 51-65, 126-158. Peterborough, Joint Nature Conservation Committee. (Coasts and Seas of the United Kingdom. MNCR series.)

[3] Lepš J., Šmilauer P. 2003. Multivariate analysis of ecological data using CANOCO. Cambridge University Press, Cambridge.

The similarity profile test (SIMPROF) was also implemented as part of the hierarchical clustering to identify how many distinct groups existed based on the null hypothesis (H_0) that the resultant sample clusters do not share a significant group structure. This test does not consider samples to be divided into groups prior to analysis and considers each sample independently. This test was carried out during the hierarchical cluster analysis using group average and the default SIMPROF setting in PRIMER for permutations (Mean: 1000, Simulations: 999) and significance level (5%).

1.7 Multidimensional Scaling (MDS) Ordination

The multidimensional scaling (MDS) ordination technique uses a similar principle to hierarchical cluster analysis but places the samples into a multidimensional space so that the similarity between all samples can be visually assessed according to their distance in space.

Those samples most similar to each other are placed closer together and those most dissimilar placed further apart and this is then presented as a 2D or 3D plot. It is important to remember that these plots represent a multidimensional configuration. Where more samples are included, the accuracy of the plot generally decreases (indicated by increasing stress values). The stress value stated on the ordination plot corresponds to the accuracy of result: stress <0.2 shows a good ordination, stress 0.2-0.3 shows a useful configuration but with a low level of accuracy thus should not be considered in detail, stress >0.3 indicates samples have been placed in an arbitrary fashion and should not necessarily be regarded as similar to one another. The default options provided in PRIMER were chosen for this analysis: 25 restarts (50 restarts were also attempted but the resultant stress level and general pattern of ordination were almost identical), Kruskal fit scheme 1, minimum stress 0.01, and configuration plot selected.

1.8 ANOSIM

The Analysis of Similarities (ANOSIM) was used to test for differences in the species abundance and composition between Transects and also between Shore Height as individual one-way tests. ANOSIM is a resemblance-based permutation test used to evaluate differences in assemblage composition between pre-defined groups of samples (*a priori* design factors). The ANOSIM test is analogous to the parametric ANOVA and tests the null hypothesis (H_0) of the lack of a distinct multimetric structure across the pre-defined groups. The global test produces a summary statistic termed 'R' which is an indication of the degree of separation of the pre-defined groups, and a p-value to assess the difference between the groups. R = 0 indicates a lack of structure and completely overlapping groups. Values approaching 1 indicate that there is a strong separation between groups. Both the R and p-value are required for a valid interpretation of the results. Finally, the ANOSIM routine provides pair-wise R values to evaluate the degree of separation between groups.

Natural England is here to secure a healthy natural environment for people to enjoy, where wildlife is protected and England's traditional landscapes are safeguarded for future generations.

Natural England publications are available as accessible pdfs from www.gov.uk/natural-england.

Should an alternative format of this publication be required, please contact our enquiries line for more information: 0300 060 3900 or email enquiries@naturalengland.org.uk.

ISBN 978-1-78354-993-1

Catalogue code: NECR408

NATURAL ENGLAND