



Methods and summary of results
JNCC Contract C13-0223-0670
Definition of deep-sea infaunal assemblages for inclusion in a deep-sea section of the Marine Habitat Classification for Britain and Ireland

Document History

Revision No.	Changes	Date
A	Draft for review	05 DEC 13
B	Revised draft following final project meeting and JNCC comments	10 JAN 14
C	Datasets listed in Section 6	15 JAN 2014

1 Purpose

The purpose of this procedure is to define and document the methods and controls to be applied when assessing and defining deep-sea infaunal assemblages. A brief summary of the results obtained by applying these methods to infaunal datasets from the Rockall Trough and Faroe-Shetland Channel is also included.

2 Competence requirements

Personnel are deemed qualified to undertake tasks within this procedure on the basis of their experience and specific training. For the purposes of this SOP competence requirements are:

2.1.1 A general understanding of the deep-sea environment and the problems associated with sampling, identification of fauna and combining large datasets is useful.

2.1.2 Training and or experience in use of relevant multivariate software and undertaking such analyses, while not essential, is advisable.

2.1.3 Specific training in cluster analysis and SIMPER would be of benefit.

3 Software & data

Abundance data from separate cruises were provided in .xls format. Older versions of Excel are limited to 65,536 rows by 256 columns. Due to the large numbers of columns (stations) encountered when working with combined data sets, Excel 2010 (Windows 7) or Excel 2011 (Mac) were used to compile the data (1,048,576 rows by 16,384 columns). PRIMER v 6.1.6 (www.primers-e.com) was used for Bray-Curtis cluster analysis and graphic output. Adobe Illustrator CS3 was used to prepare the graphic output from PRIMER for printing.

4 References/associated documents

Atlantic Frontier Environmental Network (AFEN) 2000. *Environmental Surveys of the Seafloor of the UK Atlantic Margin CD-ROM*. United Kingdom Offshore Operators Association (UKOOA).

Bett, B.J. 2012. *Seafloor biotope analysis of the deep waters of the SEA4 region of Scotland's seas*. JNCC Report No. 472.

Clarke, K.R., Gorley, R.N. 2006. PRIMER v6: User Manual/Tutorial. [PRIMER-E, Plymouth](#).

Howell, K.L. 2010. A benthic classification system to aid in the implementation of marine protected area networks in the deep/high seas of the NE Atlantic. *Biological Conservation* 143, 1041 – 1056.

Piechaud, N. and Howell, K. 2013. *Definition of epifaunal assemblages for inclusion*

in a deep sea section of the Marine Habitat Classification for Britain and Ireland: Methods report. JNCC report (number TBC).

5 Procedure

The following sections describe the methods and controls applied when assessing and defining infaunal assemblages from the deep-sea datasets provided.

6 Data selection

The appropriate benthic infaunal datasets were selected. Station identification, i.e. assigning unique identifiers or codes to each station was the first task performed. For replicate samples, or already combined replicates, station identification was established at this stage. If replicates were to be combined, the cumulative areas sampled at each station were noted. At this stage, qualitative data, inappropriate data etc. were removed from the dataset. At this stage if the faunal list had separate columns for genus/species, these two columns were concatenated with a space separating the names, the data were arranged with taxa in rows, and Stations/Samples arranged in columns.

Macrofaunal datasets used in this study were collected on the following sampling cruises:

- RRS *Charles Darwin* cruise CD101C Leg 2, 14 July – 20 August 1996 (AFEN, 2000)
- RRS *Charles Darwin* cruise CD112C, 19 May – 24 June 1998 (AFEN, 2000)
- RV *Kommandor Jack* cruise for Department of Trade & Industry Strategic Environmental Assessment Area 7 survey, August-September 2005 (Data held at SAMS)

7 Data checking

In order to check the data before proceeding, the preliminary faunal list was run through the WoRMS database (<http://www.marinespecies.org/>). The column containing taxa was copied and pasted into a new worksheet and saved as either a .csv, .txt or .xls file and uploaded to WoRMS and processed through their Match Data facility. Since the data were to be analysed at Family level only, the Classification checkbox was ticked before matching the fauna. At this stage, the WoRMS matching procedure will upload the species list that has been generated, and ambiguous taxa will be highlighted for clarification by the user. Taxa whose names have been misspelled will be identified in the matched faunal list that may be downloaded from WoRMS. This is often the most laborious stage of the process, as misspelled taxon names may be very difficult to resolve, especially if the spellings deviate significantly from the accepted version. Obsolete groupings will arise at this stage, and may be investigated via the Search Taxa facility at WoRMS to ascertain current nomenclature. If this process fails to resolve the ambiguity, wider database searches may indicate which now-obsolete taxon may be resolved and renamed with current nomenclature.

8 Data preparation

8.1 Infaunal sample data

Data were prepared for analysis once all ambiguous or misspelled taxa were resolved; the WoRMS-approved names and classification were pasted into the faunal database where the quantitative data reside, then all data above the level of Family (eg. Order, Class, Phylum, etc.) were removed. All data were normalised to abundance per m². This required information on the area of seabed sampled by each device (e.g. megacorer, boxcorer, grab) multiplied by the number of replicates in the sample.

When all abundance data were listed at Family level, they were summed for each Family and only the summed data retained. At this stage an end-of-row check and end-of-column check was performed, by summing all data for each Family (end-of-row), and for each station (end-of-column). Any zero sums indicated missing values for the entire station or Family, so these data were checked prior to deletion of the Family or station data. As two or more datasets were to be combined for global analysis, then common taxon lists were prepared, so that all abundance data could be combined on the same spreadsheet.

The infaunal samples analysed in this study came from several different survey cruises staffed by different seagoing personnel. The station labelling system was not consistent between cruises, and some of the individual station labels were lengthy, easy to misspell and/or not indicative of geographic location or water depth (e.g. 3BA800, Site 1.5, 1AA2, SAMS-2-P#5). To reduce the possibility of confusion and to make the station labels more informative, a simple, consistent labelling system was devised based on geographic location and water depth. Stations located south of the Wyville Thomson Ridge were designated RT-E (Rockall Trough East) or RT-W (Rockall Trough West). RT-E stations were those located on or close to the Hebridean Slope along the eastern margin of the Rockall Trough. RT-W stations were located on the offshore banks (Rockall, Hatton or George Bligh Banks) on the western side of the trough. Stations north of the Wyville Thomson Ridge were designated FSC-W (Faroe-Shetland Channel West) or FSC-N (Faroe-Shetland Channel North), corresponding to the two geographic groupings of infaunal biotopes identified by Bett (2012). For each station, the geographic identifier was followed by a number indicating the water depth (in metres), giving revised labels of the form RT-E-190, RT-W-740, FSC-W-538, FSC-N-915, etc. Stations at identical depth in the same geographic area were distinguished by lower-case letters (e.g. FSC-W-123a, FSC-W-123b). Stations are listed using both original and revised labels in the final data spreadsheets:

C13-0223-0670_FINAL Biological data.xlsx

C13-0223-0670_FINAL Environmental data.xlsx

8.2 Station environmental data

Data for station bottom water temperature (mean, minimum and range) and sediment granulometry (% mud) were obtained from the CD-ROM compilation of data from the 1996 and 1998 Atlantic Margin survey cruises (AFEN, 2000). These data are listed in the Environmental Data spreadsheet (**C13-0223-0670_FINAL Environmental data.xlsx**). As they were collected simultaneously with the infaunal

sampling, these data therefore represent environmental conditions at the times and locations sampled. Both water temperature and sediment data were available for 322 stations in the Faroe-Shetland Channel. A further 14 FSC stations had recorded water temperatures but lacked sediment granulometry data. Bottom water temperature was not recorded for 53 Rockall Trough stations sampled in 1996/98, but sediment granulometry data were available. The reasons behind these data gaps are not known. At the 26 RT-W stations visited in August 2005, infaunal sampling was carried out opportunistically and no contemporaneous environmental data were available.

9 Multivariate analysis

PRIMER allows the importation of Excel data, or creation of new data files by copying and pasting from the Excel spreadsheet. If the Copy/Paste method is used, count the number of columns and rows the abundance data occupy in Excel, and enter this value in the relevant areas in the PRIMER dialog box. The Abundance Data button should be selected.

Labels for Samples (stations) and Variables (families) may be entered at this stage, by copying and pasting from the Excel abundance spreadsheet (N.B. if station labels are at the top of columns in Excel, they will have to be Copy/Paste Special transposed first in Excel before copying and pasting into PRIMER).

If there are any factors (e.g. depth, temperature etc.) to be entered into the analysis, they may be entered at this stage from the Edit/Factor menu.

Prior to the similarity matrix being generated, data transformation must be considered from square root, fourth root, $\log(x+1)$. As the abundance data contain many zero counts, $\log(x+1)$ may be appropriate, and was used in this study. After the data have been transformed, the similarity matrix may be generated by the resemblance analysis. Tick the desired matrix, e.g. Group Average, and Plot dendrogram. To view the results as a dendrogram of the cluster, the cluster analysis is selected from the Analyse menu.

At this stage prior to the initial faunal assemblage determination, the factors may be overlaid on the cluster analysis. With the cluster diagram as the active window, a right click will initiate a dialog. Select Special. Next select the More button, the Data labels & symbols tab. Plot the labels, and plot the Symbols by factor (choosing your selected factor from the pull-down menu of your entered factors. You may edit the symbols and colours of the selected factors at this stage.

If you wish to assess the contribution of each Family in the dataset to the factors you have chosen, you may test these by selecting the transformed data level in the left-hand PRIMER pane and choosing SIMPER from the Analyse menu. If you are only analysing one factor at a time, select One way, and the desired factor from the pull-down menu. You may repeat this procedure for all desired factors. The results generated from each SIMPER analysis may be saved as RTF files for ease of viewing with a word processing application or simple text editor.

10 Infaunal assemblage definition

10.1 Initial identification of station clusters

Infaunal assemblages were identified from the dendrogram showing Bray-Curtis similarities of stations based on Family-level abundance data [$\log(x + 1)$ -transformed]. Two versions of this dendrogram, with stations assigned to Folk scale sediment categories or to bottom water temperature range (where data were available) are provided as PDF files:

C13-0223-0670_dendrogram_Folk sediment categories.pdf

C13-0223-0670_dendrogram_Temperature range.pdf

Station clusters (i.e. representing putative infaunal biotopes) were distinguished by close examination of the dendrogram. In a highly complex dendrogram such as this, identification of clusters that are considered meaningful is a subjective procedure governed by criteria that must be set by the observer. The PRIMER analytical software will group the stations according to the data and the type of analysis selected, but will not identify which if any of the groupings are ecologically significant or useful in biotope definition. In this analysis, putative biotopes were required to achieve a 50% (or higher) level of similarity (on the dendrogram Y-axis) and to include a minimum of 10 individual stations. It was decided that a 50% similarity cut-off provided a manageable number of clusters while the 10-station minimum was intended to increase the likelihood that clusters represent distinguishable, repeated ecological categories rather than stochastic associations of stations driven by local patchiness of the seabed environment. However, it is important to recognise the element of subjectivity in this process and to note that using different criteria, a different set of clusters/biotopes could have been distinguished. Using the criteria outlined here, 11 station clusters (putative biotopes) labelled A-K, were identified, accounting for 340 of the total 415 stations (i.e. ~82% of stations) included in the analysis.

10.2 Cluster definition by characterising and distinguishing taxa

Once station clusters have been identified according to the chosen criteria, the relative importance of the various taxa (in this case, infaunal families) in differentiating between them can be investigated using the SIMPER routine in PRIMER. SIMPER calculates an average similarity between stations in a cluster followed by a list of families which shows the decreasing percentage contribution of each family to the overall similarity of stations within a cluster. SIMPER also indicates the relative contribution of each family to the average dissimilarity between station clusters.

To undertake the analysis all the SIMPER results need to be copied into Excel.

- i. In Excel, label each worksheet with your chosen groupings i.e. Cluster A, Cluster B etc.
- ii. Copy the entire list showing the average similarity and the contribution of each family to this.

- iii. Copy each section showing the dissimilarity between each pair of clusters – making sure only to copy the relevant ones for the worksheet you have.
- iv. Ensure that the average abundance column of the cluster you are looking at appears prior to the column showing another cluster’s average abundance (this is really important when analysing the next stage).
- v. Sort each pair of clusters based on the family list (i.e. do it alphabetically). This needs to be done for every pair.
- vi. Sum the “Average Abundance” for every cluster that you are comparing with your main cluster.
- vii. Do this for Average Dissimilarity, Average Dissimilarity/SD, Contribution % and Cumulative %.
- viii. Average all of these columns by the total number of clusters you are comparing with your main cluster i.e. if you have 12 clusters in total you need to divide the total values for average dissimilarity etc. by 11 (n - 1).
- ix. Copy and “Paste Special” these values into new columns.
- x. Rank the values based on “Contribution %”
- xi. Copy the top 5 families and paste into a new set of columns
- xii. Rank the families based initially on Contribution % and then by Average Abundance of your main cluster.

The Excel file **C13-0223-0670_SIMPER biotope comparisons.xlsx** contains tables showing the SIMPER output and application of the procedure described above for each of the 11 identified clusters.

When considering defined station clusters as biotopes, the Bett (2012) trinomial labelling system was applied. This labels each putative biotope using the names of the two highest-ranking characterising families (contributing the to the similarity of stations within the biotope), followed by the name of the highest-ranking distinguishing family (which contributes most to the dissimilarity between that biotope and all others). These three family names were then followed by a short summary of the geographic region, substratum type and approximate water depth in which the biotope was located. Biotope labels therefore take the following form:

Biotope A: Sabellidae-Spionidae-Poecilochaetidae in Atlantic mud and sandy mud (900-2050 m)

When using this labelling system it is very important to note that the characterising families are not necessarily present at particularly high abundance within the defined biotope, and the biotope should not be thought of as being “dominated” by these families.

The PRIMER routine SIMPROF was not used to test for statistically significant differences between the clusters. The routine was not used by Bett (2012), whose work provided the model for this study, and is suggested to be inadvisable for complex dendrograms with a very large number of nodes, such as the example generated here (Clarke & Gorley, 2006).

11 Assignment of biotopes to broader community categories

Broad community categories are assigned based on overall faunal composition, geographic location and the general physical environment in which a community is found. See the “Broad community description” tab in the Excel spreadsheet **C13-0223-0670_biotope description templates.xlsx** for summary descriptions of the biotopes defined in this study.

For this study, the Broad Community name for each biotope is “Mixed infauna dominated by polychaetes”, a descriptor noting that Polychaeta is generally by far the most abundant higher taxon (as is the case in most deep-sea macrofaunal communities). This is followed by a more detailed biotope descriptor using Bett’s (2012) trinomial labelling system. Depth limits are rounded to the nearest 50 m.

The “Associated physical environment” column refers to the water mass in which each biotope occurs. “Atlantic waters” refers to the temperate water environment of the Rockall Trough, and also to geographically located on the West Shetland Shelf, north of the Wyville Thomson Ridge, but at relatively shallow depths where they are bathed in Atlantic water flowing north along the Hebridean margin. “Arctic waters” refers to the cold water mass of the deeper Faroe-Shetland Channel, north of the Wyville Thomson Ridge. The combined term “Atlanto-Arctic waters” is used for biotopes located in the hydrodynamically-complex depth zone of the Faroe-Shetland Channel where these two water masses impinge upon each other. Biotopes in this category may be exposed to widely fluctuating temperatures. The “Associated physical environment” also includes a general description of the substratum type (e.g. sand, muddy sand, mud).

The biological community in each case is summarised as “Polychaetes, bivalves, peracarid crustaceans and various minor groups”. A more distinctive (and concise) descriptor for each community is not practicable, as these biotopes are distinguished by relatively small variations in relative abundance of a large number of families, and are not conveniently “dominated” by any single family. More detailed information on community structure of each biotope is presented in the “Detailed community description” tab of the Excel spreadsheet **C13-0223-0670_biotope description templates.xlsx**.

Entries in the “Parent habitats” column refer to the depth zones defined for the temperate NE Atlantic by Howell (2010) and for the Faroe-Shetland Channel by Bett (2012). If JNCC’s preferred deep-sea zonation scheme changes as a result of further study, these zonal references may need to be revised.

The “Child biological assemblages” column notes that these putative macrofaunal biotopes are not “dominated” by any single family or group of families (in any way that lends itself to a concise description). See the “Distinguishing family abundance” tab of the Excel spreadsheet **C13-0223-0670_biotope description templates.xlsx** for further details.

END OF PROCEDURE

12 Summary of project results

12.1 Station clustering pattern

11 station clusters met the criteria for recognition as putative biotopes, i.e. 10 or more stations clustered at 50% or higher similarity level. The 11 clusters accounted for 340 out of the total 415 stations (i.e. ~82% of stations). The 11 putative biotopes, labelled A-K, are shown in the following dendrograms:

C13-0223-0670_dendrogram_Folk sediment categories.pdf
C13-0223-0670_dendrogram_Temperature range.pdf

The 11 clusters included most stations from the Faroe-Shetland Channel (FSC) and Hebridean Slope, eastern Rockall Trough (RT-E). Stations from the western Rockall Trough (RT-W) mostly fell outside these clusters and were not included in the putative biotopes. The inability to resolve the RT-W stations is considered to reflect the low density of sampling and the very heterogeneous seabed topography in this area (i.e. the offshore banks and seamounts), in comparison with the larger number of stations from the more continuous topography of the Hebridean Slope.

The 415-station dendrogram showed a complex branching pattern. Two putative RT-E biotopes (A and B) were grouped closely together, with a series of five FSC biotopes (C-G) coming from the same node. The remaining four FSC biotopes (H-K) joined the combined RT-E/FSC branch at a lower level of similarity. This pattern indicates that at the taxonomic level of Family, infaunal communities did not show a complete separation between RT-E and FSC stations, although the RT-E branch was well-defined.

The putative FSC biotopes (C-K) consisted predominantly (with some outliers) of stations from either the northern (FSC-N) or western (FSC-W) regions identified by Bett (2012). The FSC biotopes fell into two groups broadly defined by water depth (with associated differences in substratum type and water temperature). Biotope cluster C-G covered a total depth range of 600-2300 m, with cluster H-K consisting of shallower stations in the 100-500 m depth range.

12.2 Summary information for putative infaunal biotopes (with Bett trinomial labels)

A. Sabellidae-Spionidae-Poecilochaetidae in Atlantic mud and sandy mud (900-2050 m)

24 RT-E stations, 4 RT-W stations (28 total)

Temp. range 3 – 6° C

B. Ampeliscidae-Paraonidae-Poecilochaetidae in Atlantic sand and muddy sand (400-1000 m)

14 RT-E stations

Temp. range 7 – 9° C

- C. Capitellidae-Oweniidae-Myriotrochidae in Arctic mud and sandy mud (1300-2300 m)**
 34 FSC-N stations, 8 FSC-W stations (42 total)
 Temp. range 0 – 4° C
- D. Oweniidae-Amphinomidae-Thyasiridae in Arctic mud and sandy mud (1050-1400 m)**
 18 FSC-W stations
 Temp. range 0 – 4° C
- E. Cirratulidae-Spionidae-Sphyrapodidae in Arctic mud and sandy mud (600-900 m)**
 12 FSC-N stations
 9 stations in temp. range 0 – 4° C, 1 station in 4 – 7° C, 2 stations in 7 – 9° C
- F. Cirratulidae-Maldanidae-Maldanidae in Arctic sand and sandy mud (600-1200 m)**
 52 FSC-W stations, 1 FSC-N station (53 total)
 48 stations in temp. range 0 – 4° C, 1 station in 4 – 7° C, 4 stations in 7 – 9° C
- G. Oweniidae-Cirratulidae-Glyceridae in Arctic mud and sandy mud (750-1250 m)**
 37 FSC-N stations, 3 FSC-W stations (40 total)
 Temp. range 0 – 4° C
- H. Spionidae-Syllidae-Syllidae in Atlantic sand and muddy sand (100-250 m)**
 37 FSC-W stations
 18 stations in temp. range 0 – 4° C, 19 stations in 4 – 7° C
- I. Spionidae-Terebellidae-Enchytraeidae in Atlanto-Arctic sand and muddy sand (200-600 m)**
 56 FSC-W stations, 8 FSC-N stations, 2 RT-E stations (66 total)
 3 stations in temp. range 4 – 7° C, 28 stations in 7 – 9° C, 33 stations in > 9° C,
 2 stations with no temp. data
- J. Spionidae-Syllidae-Maldanidae in Atlanto-Arctic sand and muddy sand (200-500 m)**
 11 FSC-W stations, 1 FSC-N stations, 4 RT-E stations (16 total)
 2 stations in temp. range 4 – 7° C, 7 stations in 7 – 9° C, 3 stations in > 9° C, 4
 stations with no temp. data.
- K. Spionidae-Paraonidae-Golfingiidae in Atlantic sand and muddy sand (150-250 m)**
 13 FSC-N stations, 1 FSC-W station (14 total)
 Temp. range 4 – 7° C

12.3 Community composition

The 11 putative biotopes are typically not “dominated” by any single family, but rather are differentiated by variation in the relative abundance of a subset of common families which tend to occur across the range of sampling stations. Further details are provided in the “Distinguishing family abundance” tab of the Excel spreadsheet **C13-0223-0670_biotope description templates.xlsx**. The standard deviations for family abundance are often as large as, or larger than, the mean values, indicating that there is considerable variation between stations within the same putative biotope.

Some broad family-level mean abundance patterns can be observed:

- Ampeliscidae (Amphipoda) common in Biotopes A, B and I, much less common everywhere else.
- Ampharetidae (Polychaeta) particularly common in Biotope K.
- Amphinomidae (Polychaeta) particularly common in Biotopes A, D, E, F, G.
- Cirratulidae (Polychaeta) and Golfingiidae (Sipuncula) much more common in Biotope cluster CDEFG than anywhere else. Maldanidae (Polychaeta) are also most common in this cluster, with the exception of Biotope C.
- Oweniidae (Polychaeta) abundant in all the FSC Biotopes (C-K) but much less common in the RT-E Biotopes A and B.
- Poecilochaetidae (Polychaeta) abundant in RT-E Biotopes A and B, but rare or absent in all of the FSC biotopes.
- Sabellidae (Polychaeta) particularly abundant in Biotope A, Scalibregmatidae (Polychaeta) and Sphyrapodidae (Tanaidacea) in Biotope E.
- Spionidae (Polychaeta) common everywhere except for Biotopes C and D.
- Syllidae (Polychaeta) particularly common in Biotopes H and I.
- Terebellidae (Polychaeta) particularly common in Biotopes E and I.
- Thyasiridae (Bivalvia) particularly common in Biotopes E and F.
- Yoldiidae (Bivalvia) particularly common in Biotope E.

12.4 Placement of biotopes in provisional deep-sea classification

The Excel spreadsheet **C13-0223-0670_deep sea classification.xlsx** shows the placement of the 11 putative biotopes in the provisional deep-sea classification structure. Where the depth ranges of putative biotopes extend across bathymetric zones in the provisional classification (e.g. Biotopes A, E), the biotope has been entered at both levels in the structure.

12.5 Depth zonation of biotopes

Figure 12.5.1 shows the depth limits of the 11 putative infaunal biotopes in relation to the bathymetric zonation schemes of Howell (2010) for the temperate NE Atlantic, and Bett (2012) for the Faroe-Shetland Channel. Figure 12.5.2 shows them in relation to the modified zonation scheme for both regions proposed by Piechaud & Howell

(2013) in the JNCC epifaunal biotope definition project which accompanies the present study.

Depth (m)	NE Atlantic depth zone (Howell, 2010)	RT-E biotope (this study)	FSC depth zone (Bett, 2012)	FSC-W biotope (this study)	FSC-N biotope (this study)	
100			Shelf Break	H	K	
150				I		J
200						
250	Upper Slope	B	Upper Slope	F	E	
300						
400						
500						
600						
700	Upper Bathyal	A	Lower Slope	D	G	
800						
900	Mid Bathyal	A	Deep Channel		C	
1000						
1050						
1100						
1200						
1250						
1300						
1400						
1500						
1600						
1700	Lower Bathyal		Norwegian Basin			
1800						
1900						
2000						
2100						
2200	Abyssal					
2300						
2400						
2500						
2600						
2700						
2800						
2900						
3000						

Fig. 12.5.1 Putative infaunal biotopes identified in this study (A-K) in relation to the bathymetric zonation schemes proposed by Howell (2010) for the temperate NE Atlantic and Bett (2012) for the Faroe-Shetland Channel (FSC).

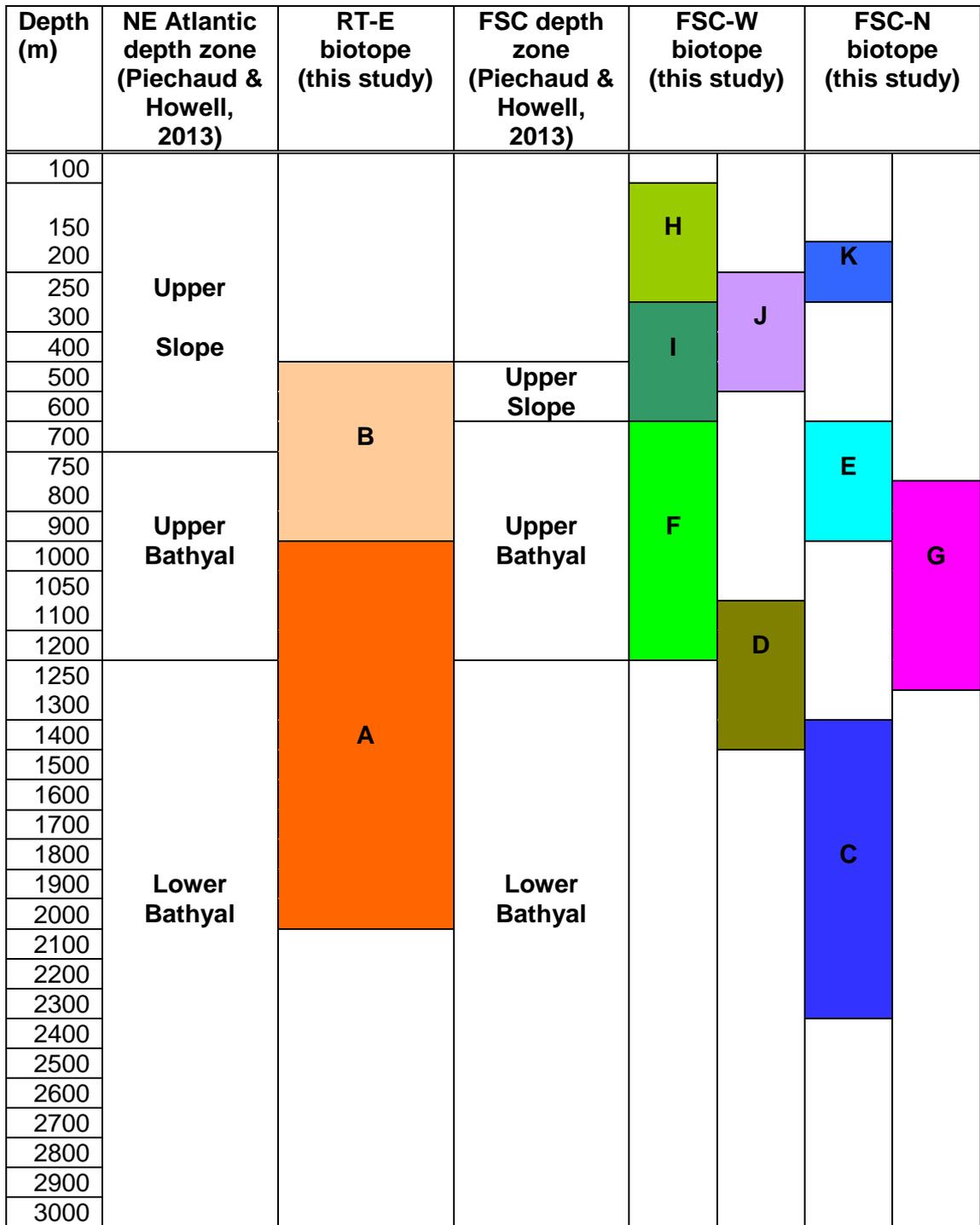


Fig. 12.5.2 Putative infaunal biotopes identified in this study (A-K) in relation to the bathymetric zonation scheme proposed by Piechaud & Howell (2013) for the temperate NE Atlantic and the Faroe-Shetland Channel (FSC).