Ocean and Coastal Management 198 (2020) 105361

Contents lists available at ScienceDirect

Ocean and Coastal Management

journal homepage: http://www.elsevier.com/locate/ocecoaman

Capacity-building paper

A new machine learning approach to seabed biotope classification

Keith M. Cooper *, Jon Barry

Centre for Environment, Fisheries and Aquaculture Science, Lowestoft Laboratory, Lowestoft, Suffolk, NR33 0HT, UK

ARTICLE INFO

Keywords:
Benthos
Mapping
Macrofauna
Classification
Biotope
Clustering
Machine learning
K-means
R shiny

ABSTRACT

Effective management in the marine environment requires a thorough understanding of the distribution of natural resources, including that of the benthos, the animals living in and on the seabed. Hitherto, it has been difficult to identify broadscale patterns in the benthos as the faunal clusters identified from individual surveys are not directly comparable. As a result, much reliance has been placed on one-off broadscale spatial surveys or matching samples to a common set of biotopes. In this study, new benthic macrofaunal data from discrete surveys are matched to existing broadscale cluster groups identified using unsupervised machine learning (k-means). This objective approach allows for continual improvements in our understanding of macrofaunal distribution patterns, thereby supporting ongoing conservation and marine spatial planning efforts. Other benefits are discussed. Finally, an R shiny web application is presented, allowing users to biotope match their own data.

1. Introduction

The growth in demand for living and non-living resources has increased pressure on the marine environment (Chapman, 2017) and raised the potential for conflicts between different user groups competing for the available space. The process of marine spatial planning (Maes et al., 2005; Douvere, 2008; Ehler and Douvre, 2009; Rempis et al., 2018; Maes, 2008; Tsilimigkas and Rempis, 2017, 2018) aims to address these issues, and is thus an important component of an ecosystem-based approach to management of the ocean and seas. However, critical to the success of marine spatial planning is a proper understanding of the distribution of natural resources, including that of the seabed macrofauna (benthos), the animals living in and on the seabed. This component of the marine ecosystem plays a key role in functioning, providing food for demersal fish and contributing to nutrient cycling and carbon sequestration (Solan et al., 2014). The extent to which planned activities are likely to affect the benthos must therefore be a consideration in any marine licensing decisions.

Knowledge of the benthos comes from seabed grab surveys routinely undertaken for the purposes of characterisation, monitoring and research. In the UK alone, more than 20,000 such samples have been acquired since 2000 (Cooper and Barry, 2017). Macrofaunal data (taxon abundance) arising from these surveys have typically been analysed using hierarchical clustering techniques (Clarke and Warwick, 1994), providing local insight into the spatial distribution of macrofaunal assemblages. However, as the cluster groups arising from discrete surveys are entirely dependent on the input data, results of one survey are not directly comparable to another. This makes the challenge of identifying broadscale spatial patterns, essential for marine spatial planning and effective management (Douvere, 2008), extremely difficult. As a result, attempts have been made to identify such patterns from spatial surveys, both at regional (e.g. Cooper et al., 2007; Fielding, 2011; Tappin et al., 2011) and national/international scales (e.g. Duineveld et al., 1991; Rees et al., 1999; Kröncke et al., 2011; Barrio Froján et al., 2012; Buhl-Mortensen et al., 2015). Whilst useful, these studies are necessarily limited in sampling density due to the high costs of sample collection and processing. Another widely used approach has been to assign samples to a common set of broadscale biotopes, such as those provided by the European Nature Information System (EUNIS) marine habitat classification (Davies et al., 2004). This approach also has its limitations including subjectivity, reliance on few species and the time-consuming nature of the matching process (Galparsoro et al., 2012). Furthermore, a recent study (Cooper et al., 2019) suggests that EUNIS biotopes (levels 3 & 4) are limited in their capacity to reflect benthic assemblages. It is of course possible to analyse data from disparate surveys together, but is impractical to do so every time new data arise, and the computational complexity of hierarchical clustering algorithms limits their application in large-scale data sets (Saxena et al., 2017).

A possible solution to the challenge of identifying broadscale macrofaunal distribution patterns may be derived from an extension of the work of Cooper and Barry (2017). This study brought together 33,198 macrofaunal samples from 777 surveys around the UK and identified

* Corresponding author.
E-mail address: keith.cooper@cefas.co.uk (K.M. Cooper).

https://doi.org/10.1016/j.ocecoaman.2020.105361
Received 18 October 2019; Received in revised form 15 August 2020; Accepted 24 August 2020
Available online 18 September 2020

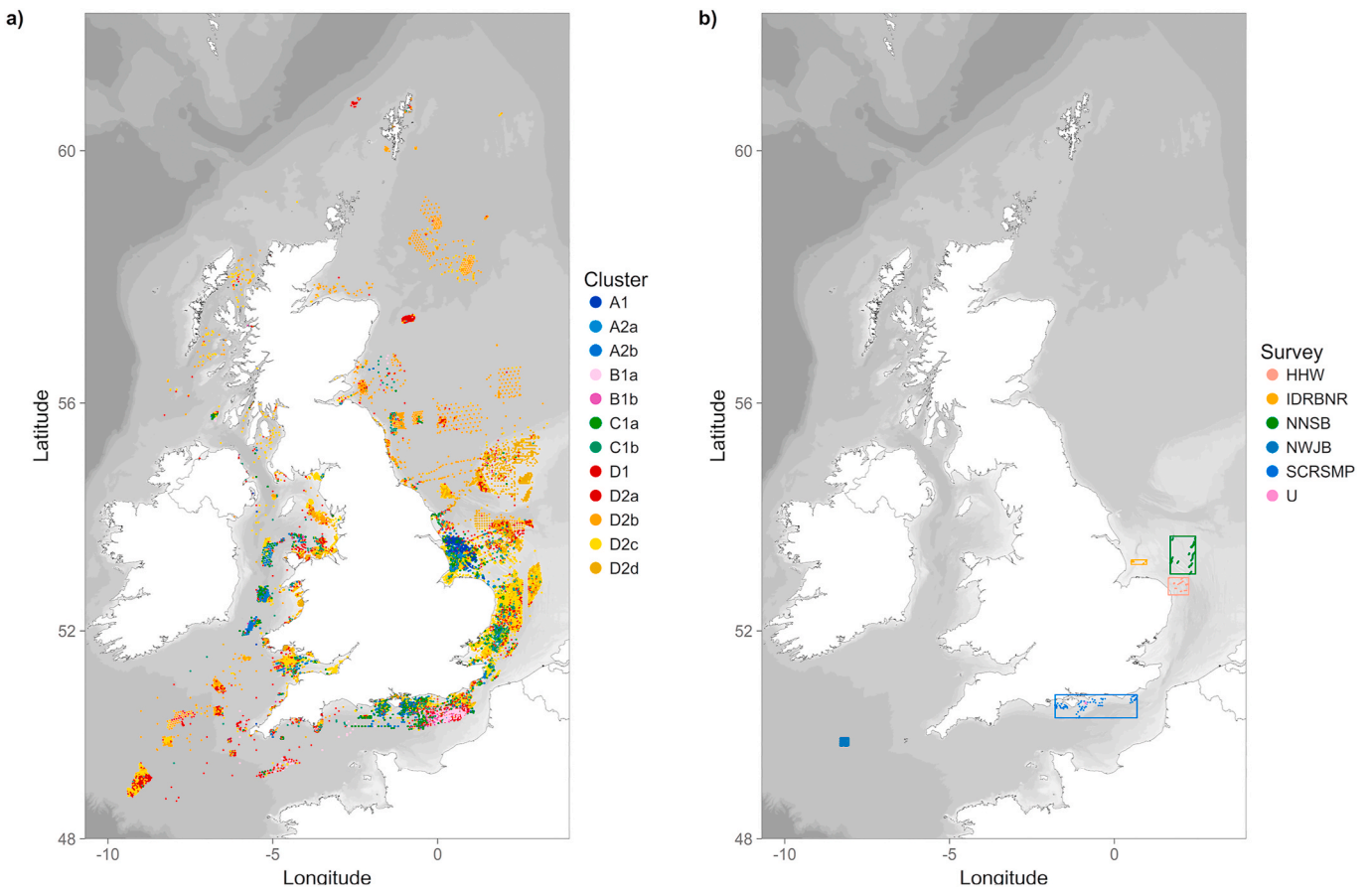

Fig. 1. a) Faunal cluster identity of baseline training samples (Figure taken from Cooper and Barry, 2017). b) Test data sample locations by survey (see Table 1 for full survey names).

assemblages using k-means, an unsupervised machine learning approach to clustering. K-means employs search algorithms to find the clustering solution that minimises the within cluster sum of squares (Hartigan and Wong, 1979). The extensive nature of this dataset means that the identified clusters are likely to be broadly representative of the entire UK shelf. Furthermore, k-means allows new sample data to be objectively matched to existing cluster groups. This potentially allows for continual improvements in our understanding of macrofaunal distribution patterns, thereby supporting ongoing conservation and marine spatial planning efforts. The purpose of this study is to: i) set out the method for matching new macrofaunal data to the existing baseline macrofaunal cluster groups identified in Cooper and Barry (2017); ii) examine the utility of the approach by comparing spatial patterns based on new data with those of the underlying baseline dataset; and iii) develop a web-based tool to allow third parties to undertake analysis (i.e. biotope classification) of their own data.

2. Materials and methods

2.1. Cluster group identification

Matching data requires both 'training' and 'test' datasets. The training dataset used in this study was the taxon (family level or higher) abundance by sample matrix used in Cooper and Barry (2017). Data from this study are stored in the 'OneBenthic' database (see https://openscience.cefas.co.uk/obdash/ for further information). This matrix incorporates data from a comparable subset of 27,432 samples across 703 taxa. All samples were collected using a 0.1 m^2 grab and processed over a 1 mm sieve. Extracted fauna were identified to the lowest possible taxonomic level and enumerated. Colonial taxa were recorded as present

Table 1
Details for case study sites (test dataset).

Survey	Year	*n*	Gear	Latitude (max, min)	Longitude (max, min)
Haisborough, Hammond and Winterton cSACSCI (HHW)	2016	165	0.1m^2 Hamon	52.96870	2.25908
				52.65830	1.63901
Inner Dowsing, Race Bank and North Ridge cSACSCI (IDRBNR)	2016	119	0.1m^2 Hamon	53.28629	0.98484
				53.20440	0.51241
North Norfolk Sandbanks and Saturn Reef cSACSCI (NNSB)	2016	156	0.1m^2 Hamon	53.71136	2.46463
				53.03373	1.69991
North West Jones Bank (NWJB)	2017	82	0.1m^2 Day/Nioz	49.98435	−8.06323
				49.83386	−8.31094
South Coast Regional Seabed Monitoring Programme (SCRSMP)	2017/18	74	0.1m^2 Hamon	50.80682	0.68890
				50.36368	−1.79703
Utopia (U)	2016	40	0.1m^2 Hamon	50.66383	−0.84551
				50.63572	−0.88876

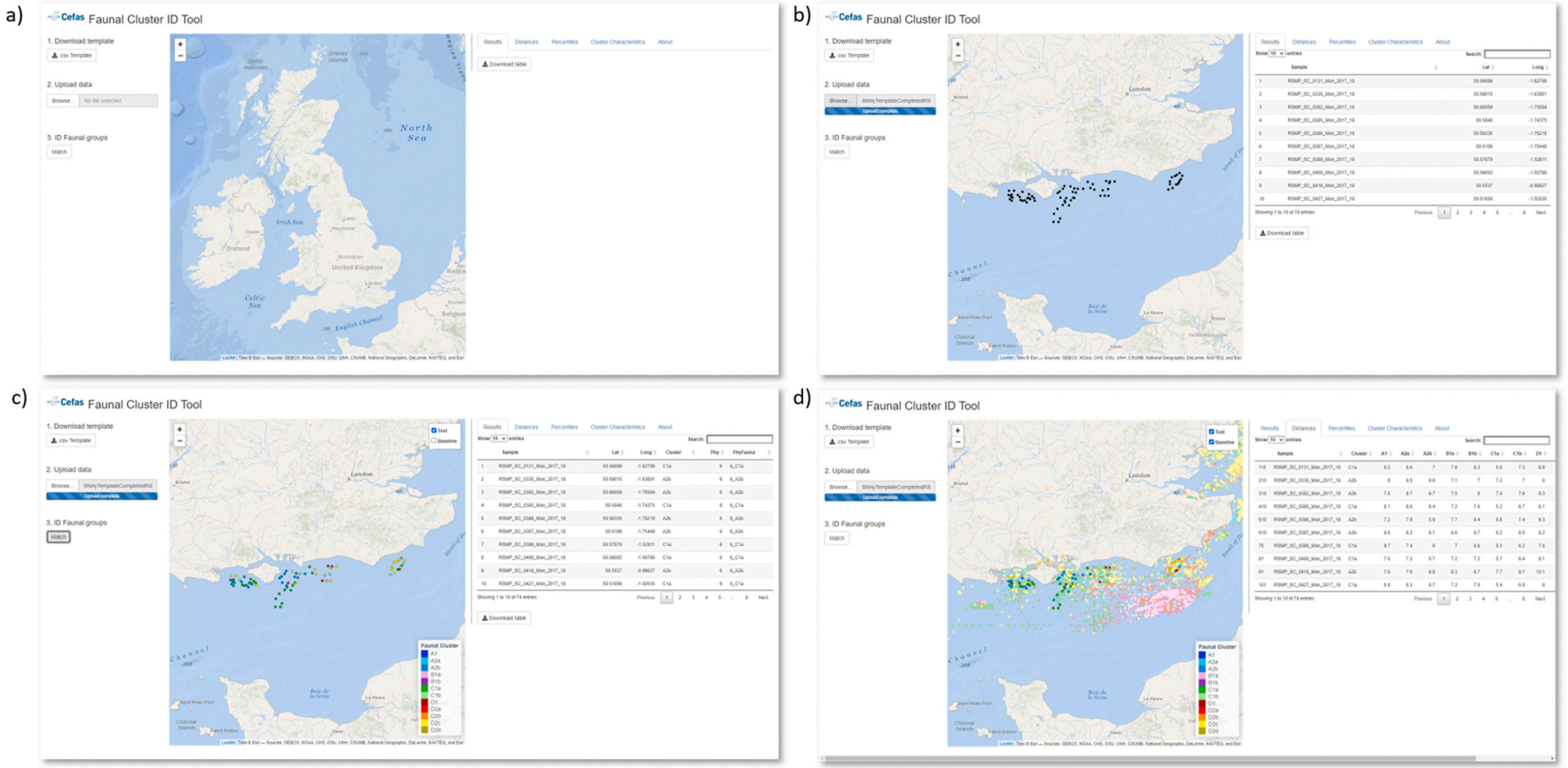

Fig. 2. Screenshots from the R shiny *Faunal Cluster ID Tool* showing: a) home screen, b) location of test samples following upload of completed template, c) cluster group identity of test samples following 'Match,' and d) cluster group identity of test and underlying baseline data (train). Distance to cluster centres, percentiles and cluster characteristics can be viewed by clicking on the relevant tabs.

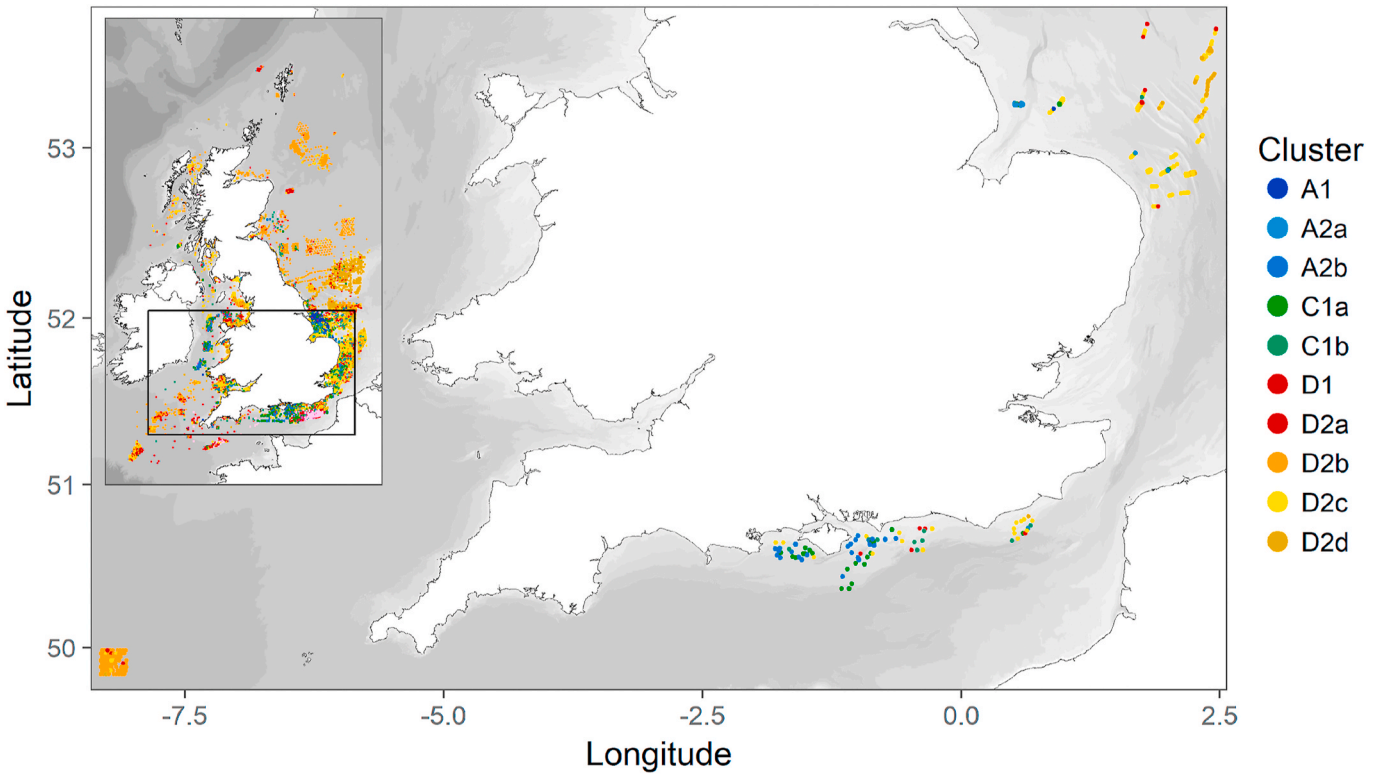

Fig. 3. Faunal cluster identity of test dataset samples showing broad agreement with the baseline spatial patterns reported in Cooper and Barry (2017) – see inset.

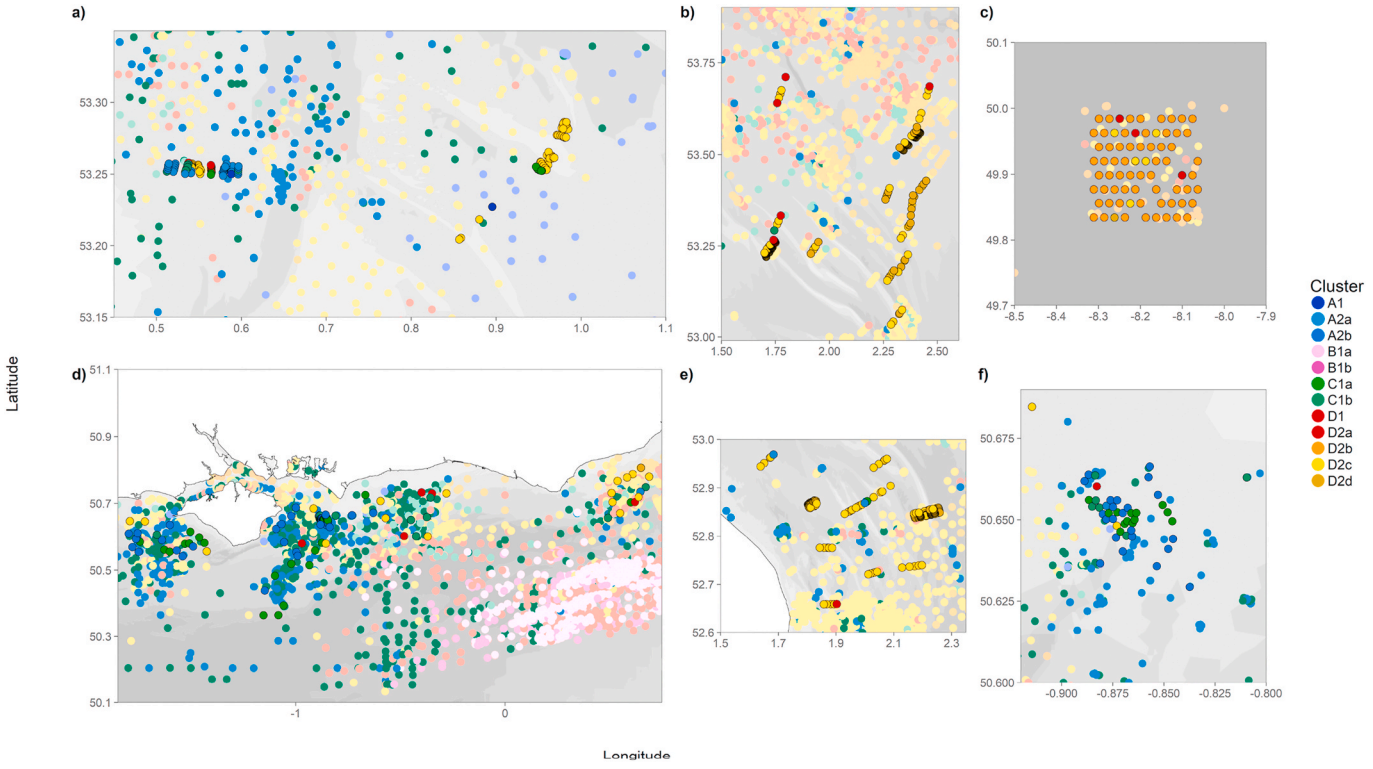

Fig. 4. Plots showing faunal cluster identity of test samples overlaid on the baseline training dataset for a) IDRBNR, b) HHW, c) NWJB, d) SCRSMP, e) NNS and f) U. For survey codes see Table 1.

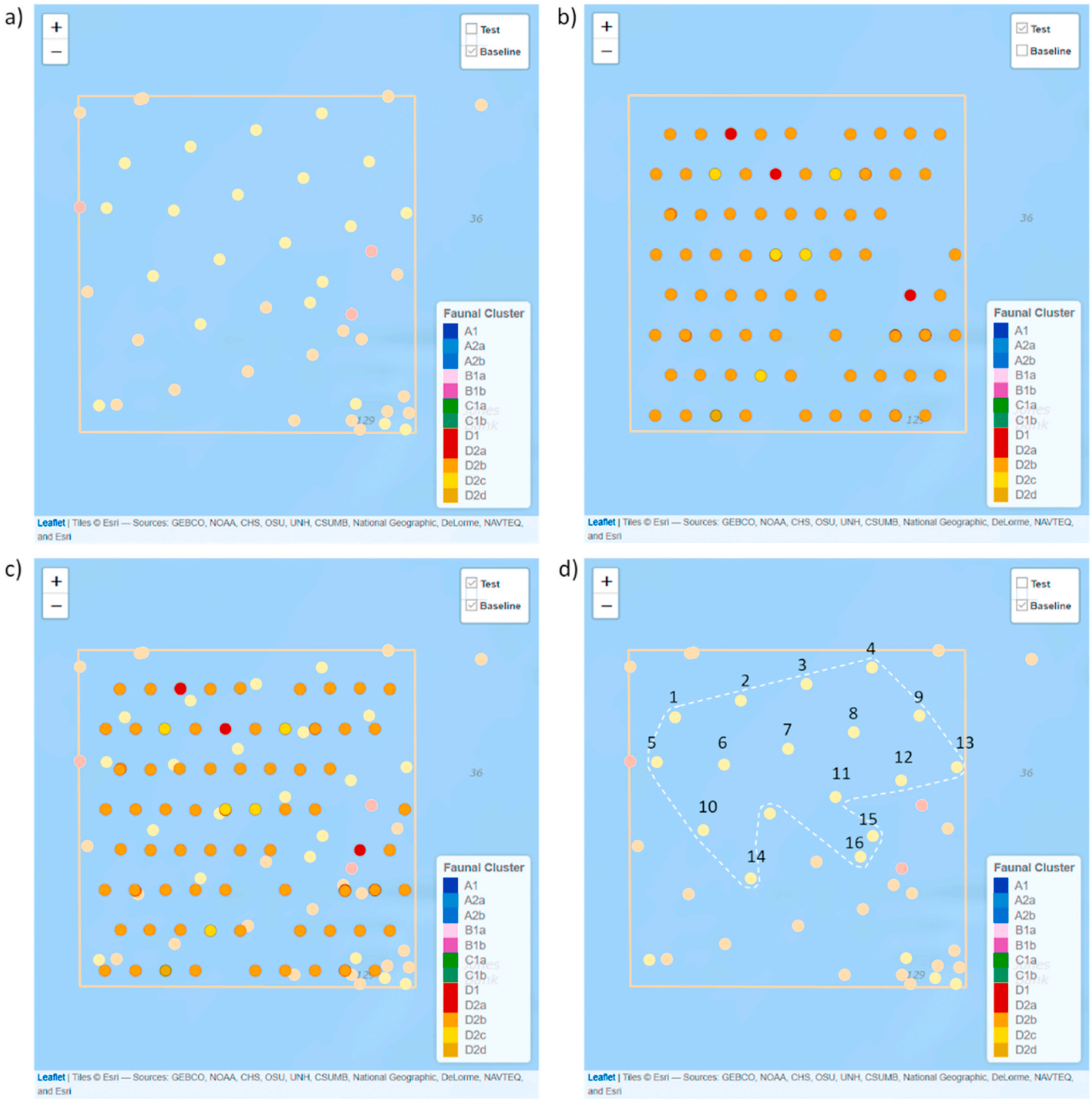

Fig. 5. Faunal cluster identity of NWJB samples during: a) baseline (2012), b) monitoring (2017) and c) 2012 and 2017 surveys combined. Part d) shows the area (dashed white line) within which faunal cluster identity has shifted from D2c (yellow) to D2b (orange) and red (D2a). Numbers are the sample codes (see Table 2 for associated 'distance to cluster centres' and percentiles). The orange box is the perimeter of the North West Jones Bank MCZ. Plots produced using the *Faunal Cluster ID Tool.* (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Cluster group, 'distance to cluster centre' and percentiles for baseline samples from NWJB within the area of apparent change (see Fig. 3). Whilst all samples belonged to cluster group D2c, underlined values indicate next nearest cluster group. Note that for most samples, the percentile value is outside the 3rd quartile.

Sample	Cluster	Distance to cluster centre				Percentile	Within 3rd quantile?
		D2a	D2b	D2c	D2d	(D2c)	
1	D2c	4.3	4.1	3.4	4.4	68	✓
2	D2c	4.3	4.7	3.7	4.6	78	×
3	D2c	5.3	5.8	4.6	5.5	95	×
4	D2c	4.4	4.1	4.0	4.7	86	×
5	D2c	3.5	3.5	3.0	3.6	54	✓
6	D2c	4.4	4.5	3.7	4.7	78	×
7	D2c	3.6	4.1	2.1	3.8	22	✓
8	D2c	4.0	4.1	3.8	4.8	80	×
9	D2c	3.7	4.1	3.5	4.3	71	✓
10	D2c	4.2	4.0	3.8	4.4	80	×
11	D2c	4.1	3.8	3.4	4.4	68	✓
12	D2c	4.5	4.8	4.1	4.8	88	×
13	D2c	5.1	5.6	5.0	5.8	98	×
14	D2c	4.8	4.3	4.2	4.7	89	×
15	D2c	4.8	5.3	4.6	5.4	95	×
16	D2c	4.1	4.3	3.3	4.4	66	✓

and given an abundance value of 1 in the dataset (see Cooper and Barry, 2017, for further details). Data were subsequently reduced to family level or higher to take account of differences in taxonomic discrimination between surveys, and to reduce any noise associated with seasonal and inter-annual changes at the level of genus and species. After applying a fourth-root transformation, the training dataset was then subject to k-means clustering, following the methodology in Cooper and Barry (2017) and using the statistical package R (R Core Team, 2017). This analysis identified 12 faunal clusters in the data (see Fig. 1a).

The test (i.e. new) dataset was the family level faunal data from six new surveys (see Table 1 and Fig. 1b). Raw macrofaunal data from these surveys were converted from species (or genus) to family level using aggregation data available from the WORMS taxa match tool (http://www.marinespecies.org/aphia.php?p=match). Family level data were then compiled into a template including all training dataset taxon names. Of the 311 families present in the test dataset, only two: Laonidae (MolluscaBase, 2019) and Mimosellidae (Bock and Gordon, 2019) were not present in the training dataset (703 families), and thus had to be removed. These two families represent only 0.3% of the total number of taxa considered in the analysis.

The faunal cluster group identity of the test samples was then determined using the *predict* function from the *flexclust* package (Leisch, 2006) in R. This function works by calculating the sum (over the transformed species counts) of the squared distances from the test sample to each of the 12 multivariate cluster centres. Mathematically, we can write the distance between the test sample (t) and the kth cluster centre (c_k) as

$$d_k = \sum_{j=1}^{s} \left(t_j - c_{kj}\right)^2 \qquad \mathrm{k} = 1, \ldots, 12 \qquad (1)$$

where t_j is the vector of transformed species counts for the test sample and c_{kj} is the vector of centres for the kth cluster. To determine the cluster membership for the test sample we choose the k for which d_k is the minimum of $d_1, d_2, \ldots, d_{12}$. Samples from both the training and test datasets were then mapped. To discriminate between datasets, the same but muted colours were used for the baseline (training) dataset.

2.2. Strength of cluster group association

To provide some insight into how representative a sample is of its assigned cluster group, its distance to the cluster centre d_k (equation (1)) was evaluated as a percentile of the distribution of distances for the members of that cluster to their centre. Specifically, if there are n_k members of the kth cluster, each with distance d_{km} (m = 1, …, n_k) to the cluster centre, we form the vector (d_k, d_{k1}, …, d_{kn_k}). We then estimate the percentile represented by d_k in this vector. We use the percentile estimator $p_t = 100^*(\text{rank } (d_k) - 0.5)/(n_k + 1)$. Thus, low values of p_t show that the test sample is close to the cluster centre relative to the original members of that cluster.

2.3. Tool development

To allow third parties to match their own data to the existing cluster groups, the R code was also implemented in an R shiny application and deployed on the Cefas Open Science website (https://openscience.cefas.co.uk/). The application (*Faunal Cluster ID Tool)* comprises of a user interface (UI) and server function. The UI is split into 3 sections (Fig. 2a). On the left of the screen users can download an empty template for test sample data (Step 1). This template includes columns for: Sample code, Latitude (decimal degrees, WGS84), Longitude (decimal degrees, WGS84), and all the taxon families present in the training set. Once completed, the template can then be uploaded to the application (Step 2). Following upload, the map will display the test sample locations, with sample codes and coordinates appearing in the table (Fig. 2b). In step 3, the user clicks the 'Match' button to identify the faunal cluster identity of the test samples. At this point the map and table will update to show cluster group identity (Fig. 2c). The faunal cluster group of baseline stations can be viewed by checking the 'Baseline' button on the map. The 'distance to cluster centres' and the associated percentiles for the assigned group are also available by clicking on the appropriate tab (see Fig. 2d). The same information can be viewed for baseline samples by clicking on samples in the map, with information appearing in a pop-up. The 'Cluster Characteristics' tab provides a table showing the characterising taxa and mean taxon richness and total abundance associated with each cluster group.

3. Results

3.1. Cluster group identification

The spatial distribution of faunal clusters in the test dataset shows close agreement with that of the baseline training dataset (see Fig. 3 and Fig. 4). At NWJB, however, the faunal cluster identity of test samples showed some deviation from the underlying baseline data, with an apparent shift from the yellow cluster D2c (baseline training samples, 2012) to the orange cluster D2b (test samples, 2017) and the red cluster D2a (test samples, 2017) across much of the survey area (Figs. 4c and 5). Examination of the 'distance to cluster centre' and percentile values for baseline samples within the affected area showed that the majority (63%) had values outside the 3rd quartile (i.e. > 75%) (see Fig. 5 and Table 2). This may help to explain some of the apparent shift to other cluster groups seen in samples from the 2017 survey.

4. Discussion

In this study, new benthic macrofaunal data from discrete surveys were successfully matched, using a shiny R application (https://openscience.cefas.co.uk/faunal_cluster_id/), to the existing baseline faunal cluster groups identified in Cooper and Barry (2017). Spatial patterns in the new data were largely consistent with the underlying baseline (training) dataset, providing confidence in the approach. At the one survey site where new data showed some obvious deviation from the baseline patterns (i.e. NWJB), the apparent discrepancy in assemblage composition (D2c to D2b and D2a) involved three very closely related cluster groups (see Cooper and Barry, 2017), as opposed to major changes in assemblage composition. When assessing apparent changes, it is important to recognise that cluster groups are a simplistic

representation of a complex picture and belie what is a continuum in faunal composition. As such, samples will vary in how typical they are of their assigned group. In the case of NWJB, examination of the 'distance to cluster centres' for the baseline samples suggested that many had been close to the boundary between different groups. As such, the observed discrepancies with test data may not be particularly unexpected. This capacity to interrogate the percentile values of underlying data, to help understand the potential reasons behind any lack of fit of test data, is an important feature of the tool.

The ability to simply and objectively match new sample data to the existing baseline groups means that it will be possible, as new data become available, to construct a very detailed map of faunal cluster distribution, without the need to resort to one-off broadscale surveys or subjective biotopes. Such a map will be important for effective marine spatial planning (Maes et al., 2005; Douvere, 2008; Ehler and Douvre, 2009; Rempis et al., 2018; Maes, 2008; Tsilimigkas and Rempis, 2017, 2018) and conservation efforts, helping to inform decisions about the siting of developments.

The approach described in this study and enacted in the *Faunal Cluster ID Tool* (https://openscience.cefas.co.uk/faunal_cluster_id/) will have other important applications. For instance, it will allow offshore developers to identity the nature of the fauna in their area of development, and to place these findings into the wider context. This is particularly important for the marine aggregate dredging sector where developers are required, under the Regional Seabed Monitoring Programme (Cooper, 2013; Barrio Froján et al., 2016), to characterise the macrofauna at sampling stations within their area of potential effect, and to subsequently monitor the sediment composition to make sure it remains suitable for supporting a return of the original assemblage type after dredging (Cooper and Barry, 2017).

In addition to mapping, the approach may also be useful for monitoring the condition or health of known faunal communities. Such assessments are typically made using a variety of approaches including, *inter alia*, basic univariate faunal metrics and their derivatives (e.g. species richness, abundance, Shannon diversity), multivariate approaches (see Clarke and Warwick, 1994) and more complex multimetric approaches (Borja et al., 2011). Knowing whether the assemblage type has changed is, however, a fundamental consideration that could contribute to assessments of the status of the macrofauna.

Unsurprisingly, developer-led seabed grab sampling efforts tend to focus on licensed areas and their surroundings. Understanding of the bigger picture (i.e. UK scale) is however an issue for government, and here assessments typically rely on snapshots, modelled data and ongoing, but spatially limited monitoring programmes (Defra, 2005; UK Marine Monitoring Assessment Strategy Community, 2010). It has long been recognised that industry and government monitoring efforts could be brought together to help better understand the bigger picture (Barrio Froján et al., 2016), but barriers including access to data (ABPmer, 2015), and appropriate data analysis methodologies (Defra, 2005) have limited progress. In recent years, however, data have become much more freely available through various online portals, including the Crown Estate's Marine Data Exchange (MDE), and ongoing efforts to collate and standardise these data (e.g. Cooper and Barry, 2017; https://openscience.cefas.co.uk/obdash/). The next challenge is how to best use these data, and the application presented in the present study (https://openscience.cefas.co.uk/faunal_cluster_id/) has a role to play, allowing macrofaunal data from any source (industry or government) to be analysed in a standardised way, improving our knowledge on the distribution and status of benthic macrofaunal communities. Furthermore, the approach can be used to validate predictions regarding the sensitivity of faunal communities to various anthropogenic pressures (see Hiscock and Tyler-Walters, 2006).

Despite the benefits, it is important to recognise and consider the limitations of the approach. Firstly, the large training data set used in this study is an amalgamation of macrobenthic abundance data from 27,432 grab samples taken from around the UK over a 48-year period. Clearly, there are assumptions made when clustering using these data. One is that community characteristics don't change within this 48-year time frame, and the temporal analysis reported in Cooper and Barry (2017), although limited to a visual assessment of changes in patterns of faunal cluster distribution, supports this. The other inherent issue is the assumption that surveys identified species correctly and consistently. Whilst considerable effort has gone into standardising the data and checking for errors, possible issues cannot be ruled out entirely. We continually update data accuracy and check on assumptions within our big data set. In addition, test data can only include taxa present in the training dataset. This should not present a serious problem, given that the training dataset is very large, has broadscale coverage and that data are analysed at family level. The likelihood of a test dataset comprising additional taxa is, resultingly, very limited, particularly in already heavily sampled areas. Once significant quantities of new data have been acquired, it may be appropriate to undertake a re-clustering of the expanded dataset.

Further work is now needed to ensure that the insight provided by new surveys is effectively captured and turned into new information (see Wilding et al., 2017), providing benefits for industry and government and leading to greater efficiency and improved sustainability (Etzion and Aragon-Correa, 2015; Barrio Froján et al., 2016). Given positive changes in attitude towards data sharing (ABPmer, 2015; McMeel et al., 2017), this should be possible, but a formal mechanism for capturing this new information will need to be identified. Furthermore, it should be possible to develop a similar tool for epifaunal data (the benthic component living on the seabed and generally sampled using trawls or video approaches), although this would require an initial collation of such data.

Author contributions

This study was conceived and undertaken by K.M.C. J.B. provided statistical advice, contributed to the manuscript and some code within the Faunal Cluster ID tool. Neither author has any conflict of interest to declare.

5. Data accessibility

The training dataset used in this study is available from the Cefas Data Hub (https://doi.org/10.14466/cefasdatahub.34). The test data, R script and supporting files are also available from the Cefas Data Hub (https://doi.org/10.14466/CefasDataHub.72).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The baseline grab sample data came from Cooper and Barry (2017), and we gratefully acknowledge the contribution of individual data providers. Test sample data were provided by Defra (surveys: HHW, IDRBNR, NNSB, NWJB, U) and the South Coast Dredging Association (survey: SCRSMP). We are grateful to colleagues for their insightful comments on previous drafts of the manuscript, to Paul McIlwaine (Cefas) for providing much of the test data, and to Roi Martinez (Cefas) for assistance with the development of the app. This study was funded by the British Marine Aggregates Producers Association (BMAPA) and the Crown Estate under project C7912 (Marine Aggregate Regional Seabed Monitoring Programme: further development, support and delivery – Part 2), and by Defra under project C5786 (Marine Conservation Zone programme).

References

ABPmer, 2015. A Review of Access to Industry Environmental Data. A Report Produced by. ABP Marine Environmental Review Ltd for Productive Seas Evidence Group. November 2015.

Barrio Froján, C.R.S., Bolam, S.G., Eggleton, J.D., Mason, C., 2012. Large-scale faunal characteristics of marine benthic sedimentary habitats around the UK. J. Sea Res. 69, 53–65.

Barrio Froján, C.R.S., Cooper, K.M., Bolam, S.G., 2016. Towards an integrated approach to marine benthic monitoring. Mar. Pollut. Bull. 104, 20–28.

Bock, P., Gordon, D., 2019. World list of bryozoa. Mimosellidae hincks, 1877, Accessed through: World Register of Marine Species at: http://www.marinespecies.org/aphia.php?p=taxdetails&id=110793 on 2019-06-28.

Borja, Á., Barbone, E., Basset, A., Borgersen, G., Brkljacic, M., Elliott, M., Garmendia, J. M., Marques, J.C., Mazik, K., Muxika, I., Neto, J.M., Norling, K., Rodríguez, J.G., Rosati, I., Rygg, B., Teixeira, H., Trayanova, A., 2011. Response of single benthic metrics and multi-metric methods to anthropogenic pressure gradients, in five distinct European coastal and transitional ecosystems. Mar. Pollut. Bull. 62, 499–513. https://doi:10.1016/j.marpolbul.2010.12.009.

Buhl-Mortensen, L., Buhl-Mortensen, P., Dolan, M.F.J., Holte, B., 2015. The MAREANO programme – a full coverage mapping of the Norwegian off-shore benthic environment and fauna. Mar. Biol. Res. 11 (1), 4–17. https://doi.org/10.1080/17451000.2014.952312.

Chapman, P.M., 2017. Assessing and managing stressors in a changing marine environment. Mar. Pollut. Bull. 124 (2), 587–590. https://doi.org/10.1016/j.marpolbul.2016.10.039.

Clarke, K.R., Warwick, R.M., 1994. Change in Marine Communities: an Approach to Statistical Analysis and Interpretation. Natural Environment Research Council, Plymouth Marine Laboratory, Plymouth, p. 144.

Cooper, K.M., October 2013. Marine Aggregate Dredging: A New Regional Approach to Environmental Monitoring. Thesis submitted for degree of PhD by Publication.

Cooper, K.M., Barry, J., 2017. A big data approach to macrofaunal baseline assessment, monitoring and sustainable exploitation of the seabed. Sci. Rep. 7, 12431. https://doi.org/10.1038/s41598-017-11377-9.

Cooper, K., Boyd, S., Aldridge, J., Rees, H., 2007. Cumulative impacts of aggregate extraction on seabed macro-invertebrate communities in an area off the east coast of the United Kingdom. J. Sea Res. 57, 288–302.

Cooper, K.M., Bolam, S.G., Downie, A.-L., Barry, J., 2019. Biological- based habitat classification approaches promote cost- efficient monitoring: an example using seabed assemblages. J. Appl. Ecol. 56, 1085–1098. https://doi.org/10.1111/1365-2664.13381.

Davies, C.E., Moss, D., Hill, M.O., 2004. EUNIS habitat classification revised 2004. Report to the European topic centre on nature protection and biodiversity. European Environment Agency. http://www.eea.europa.eu/themes/biodiversity/eunis/eunis-habitat-classification#tab-documents.

Defra, 2005. Charting Progress, An Integrated Assessment of the State of UK Seas. In: Department for Environment, Food and Rural Affairs (Defra), p. 119.

Douvere, F., 2008. The importance of marine spatial planning in advancing ecosystem-based sea use management. Mar. Pol. 32, 762–771. https://doi.org/10.1016/j.marpol.2008.03.021.

Duineveld, G.C.A., Künitzer, A., Niermann, U., De Wilde, P.A.W.J., Gray, J.S., 1991. The macrobenthos of the North sea. Neth. J. Sea Res. 28 (1/2), 53–65.

Ehler, C., Douvre, F., 2009. Marine spatial planning: a step-by-step approach toward ecosystem-based management. Intergovernmental oceanographic commission manual and guides No. 53. ICAM Dossier No 6, 99pp.

Etzion, D., Aragon-Correa, J.A., 2015. Big Data, Management, and Sustainability: Strategic Opportunities Ahead, vol. 29. Organization & Environment, pp. 147–155.

Fielding, H., 2011. The humber regional environmental characterisation. Marine Aggregate Levy Sustainability Fund 345pp. OR/10/054).

Galparsoro, I., Connor, D.W., Borja, A., Aish, A., Amorim, P., Bajjouk, T., et al., 2012. Using EUNIS habitat classification for benthic mapping in European seas: present concerns and future needs. Mar. Pollut. Bull. 64, 2630–2638. https://doi.org/10.1016/j. marpolbul.2012.10.010.

Hartigan, J.A., Wong, M.A., 1979. A K-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics) 28 (1), 100–108.

Hiscock, K., Tyler-Walters, H., 2006. Assessing the sensitivity of seabed species and biotopes – the marine life information network (MarLIN). Hydrobiologia 555, 309–320. https://doi.org/10.1007/s10750-005-1127-z.

Kröncke, I., Reiss, H., Eggleton, J.D., Aldridge, J., Bergman, M.J.N., Cochrane, S., Craeymeersch, J.A., Degraer, S., Desroy, N., Dewarumez, J.-M., Duineveld, G., Essink, K., Hillewaert, H., Lavaleye, M., Moll, A., Nehring, S., Newell, R., Oug, E., Pohlmann, T., Rachor, E., Robertson, M., Rumohr, H., Schratzberger, M., Smith, R., Vanden Berghe, E., van Dalfsen, J., van Hoey, G., Vincx, M., Willems, W., Rees, H.L., 2011. Changes in North Sea macrofauna communities and species distribution between 1986 and 2000. Estuar. Coast Shelf Sci. 94, 1–15.

Leisch, Friedrich, 2006. A toolbox for K-Centroids cluster analysis. Comput. Stat. Data Anal. 51 (2), 526–544.

Maes, F., 2008. The international legal framework for marine spatial planning. Mar. Pol. 32, 797–810.

Maes, F., Schrijvers, J., Van Lancker, V., Verfaillie, E., Degraer, S., Derous, S., De Wachter, B., Volckaert, A., Vanhulle, A., Vandenabeele, P., Cliquet, A., Douvere, F., Lambrecht, J., Makgill, R., 2005. Towards a Spatial Structure Plan for Sustainable Management of the sea. Research in the Framework of the BELSPO Mixed Actions - SPSD II, June 2005, p. 539.

McMeel, O., Pirlet, H., Calewaert, J.B., 2017. Use and Sharing of Marine Observations and Data by Industry Good Practice guide. Report for Horizon 2020 COLUMBUS Proj. 17.

MolluscaBase, 2019. MolluscaBase. Laonidae pruvot-fol, 1954. http://www.marinespecies.org/aphia.php?p=taxdetails&id=848158 on 2019-06-28. Accessed through: World Register of Marine Species at:

R Core Team, 2017. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rees, H.L., Pendle, M.A., Waldock, R., Limpenny, D.S., Boyd, S.E., 1999. A comparison of benthic biodiversity in the North sea, English channel, and celtic seas. ICES (Int. Counc. Explor. Sea) J. Mar. Sci. 56, 228–246.

Rempis, N., Alexandrakis, G., Tsilimigkas, G., Kampanis, N., 2018. Coastal use synergies and conflicts evaluation in the framework of spatial, development and sectoral policies. Ocean Coast Manag. 166, 40–51.

Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O.P., Tiwari, A., Er, M.J., Ding, W., Lin, C.-T., 2017. A review of clustering techniques and developments. Neurocomputing 267, 664–681.

Solan, M., Cardinale, B.J., Downing, A.L., Engelhardt, K.A., Ruesink, J.L., Srivastava, D. S., 2014. Extinction and ecosystem function in the marine benthos. Science 306, 1177–1180.

Tappin, D.R., Pearce, B., Fitch, S., Dove, D., Gearey, B., Hill, J.M., Chambers, C., Bates, R., Pinnion, J., Diaz Doce, D., Green, M., Gallyot, J., Georgiou, L., Brutto, D., Marzialetti, S., Hopla, E., Ramsay, E., Fielding, H., 2011. The humber regional environmental characterisation. Marine Aggregate Levy Sustainability Fund 345pp. OR/10/054).

Tsilimigkas, G., Rempis, N., 2017. Maritime spatial planning and spatial planning: synergy issues and incompatibilities. Evidence from Crete island, Greece. Ocean Coast Manag. 139, 33–41.

Tsilimigkas, G., Rempis, N., 2018. Marine uses, synergies and conflicts. Evidence from Crete Island, Greece. J. Coast Conserv. 22, 235–245.

UK Marine Monitoring Assessment Strategy Community, 2010. Charting Progress 2: An Assessment of the State of UK Seas. In: Department for Environment Food and Rural Affairs.

Wilding, T.A., Gill, A.B., Boon, A., Sheehan, E., Dauvin, J.-C., Pezy, J.-P., O'Beirn, F., Janas, U., Rostin, L., De Mesel, I., 2017. Turning off the DRIP ('Data-rich, information-poor') – rationalising monitoring with a focus on marine renewable energy developments and the benthos. Renew. Sustain. Energy Rev. 74, 848–859.