Centre for Environment Fisheries & Aquaculture Science

Cefas

Offshore Wind Evidence + Change Programme

North Sea Net Gain (NSNG)

Keith M. Cooper, Anna Downie, Matthew Curtis

April 2022

Cefas Document Control

Submitted to:	Isabelle Grieveson (The Crown Estate)
Date submitted:	28/04/2022
Project Manager:	Andrew Soanes (Cefas)
Report compiled by:	Keith Cooper (Cefas)
Quality control by:	Andrew Gill (Cefas)
Approved by and date:	Andrew Gill 10/01/2022, Andrew Soanes 28/04/2022
Version:	3
Recommended citation for this report:	Cooper, K. M., Downie, A.-L. and Curtis, M. (2022). North Sea Net Gain (NSNG). Cefas Project Report for The Crown Estate, 57 pp.

Version control history

Version	Author	Date	Comment
0.1	KC, AD, MC	01/12/2021	Draft
0.2	AG	10/01/2022	QC
0.3	KC, AD	11/01/2022	QC issues addressed
1.0	AS	11/01/2022	Draft report finalised
1.1	KC, AD	01/02/2022	Comments address
2.0	AS	11/02/2022	Report sent to customer
2.1	KC	25/04/2022	Address customer comments on Exec summary
3.0	AS	28/04/2022	Final report issued

Contents

1. Executive summary

The expansion of offshore wind (OW) energy generation is an important part of efforts to tackle climate change. However, to be sustainable such development must also address the twin challenge of biodiversity loss. This requirement is now recognised in both UK (Environment Act, 2021) and European (EU Biodiversity Strategy) legislation, through concepts of biodiversity Net Gain (NG) and No Net Loss (NNL) respectively, reflecting aspirations to enhance or stem the loss of biodiversity.

Whilst it is not yet clear what the long-term consequences of OW energy production will be for seabed biodiversity, strategies to achieve NG or NNL are being developed. For example, the Netherlands is pursuing a policy of Nature Inclusive Design (NID), where steps are actively taken to promote biodiversity within development areas. This is supported by the Rich North Seas (RNS) initiative (https://www.derijkenoordzee.nl/en/our-approach) that seeks to develop solutions (e.g. introduction of artificial reef structures or habitat creation) which can then be adopted by OWF developers[1].

To help gauge the success of these biodiversity strategies (NG/NNL/NID), and to support decision making (i.e. where developments should take place), there is an urgent need to improve understanding of the biodiversity of the seabed, particularly at broad spatial scales which provide important context.

This North Sea Net Gain project was funded under the Offshore Wind Evidence and Change Programme (OWEC) to seek to generate data beyond the confines of national boundaries to meet this aim. This innovative project involved an international collaboration between the UK (The Crown Estate, Cefas) and European partners (RNS, Flanders Marine Institute (VLIZ)), allowing for inclusion of data from outside the UK EEZ, and development of biodiversity layers which cross transnational boundaries. The project has been supported by a project advisory group which included members from Natural England and The Joint Nature Conservation Committee (JNCC), who provided valuable guidance throughout the project. The project considered benthic data from around the UK and across the North Sea in its scope, drawing in data from the waters of seven countries.

To date, there has been a reliance on physical based habitat maps which are often used as a proxy for biodiversity. With limited quantities of biological data, this is a logical approach. However, the data landscape is now very different and there are multiple online repositories, including the Crown Estate's Marine Data Exchange (https://www.marinedataexchange.co.uk/), where seabed biological data can be obtained. Bringing these datasets together gives us, for the first time, an opportunity to map benthic biodiversity directly.

[1] In this report, NID refers to options that can be integrated into or added to the design of offshore wind infrastructure to create suitable habitat for native species (or communities).

The OneBenthic (OB) initiative (https://sway.office.com/HM5VkWvBoZ86atYP?ref=Link) was set up to capitalise on the availability of seabed biological samples, and to allow for the adoption of big data approaches, providing new scientific insights and leading to improved sustainability. The current North Sea Net Gain project provides further development of the OB initiative, with a view to supporting the continued sustainable expansion of the OW industry by improving understanding of benthic biodiversity. This includes evolving existing OB digital infrastructure and carrying out an international data harvesting exercise to draw together standardised data which can then be harnessed in the creation of novel modelled biodiversity layers. The collaboration of international partners within the project has also supported enhanced flow of data between UK & European data repositories, widening the accessibility of open benthic data.

As the project sought to develop both the data and use of the data across the North Sea, the **key project objectives** included:

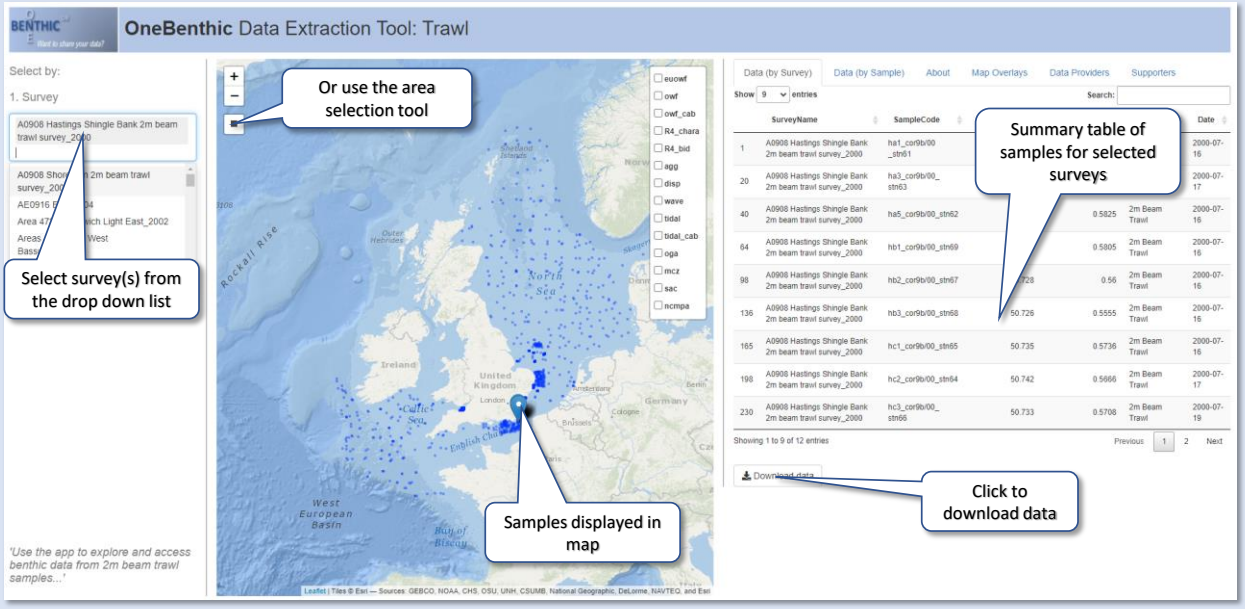
1. Development of OneBenthic infrastructure to create a dedicated database to hold trawl sample data, to sit alongside the existing grab & core sample database. This will allow data collected via trawl surveys to be drawn into OneBenthic, expanding the range of species represented to include epifauna (organisms attached to a surface) as well as infauna (organisms living within bottom sediments).
2. International data harvesting to dramatically increase the number of samples available to be drawn into analysis work, including the development of the transnational biotope and species distribution modelling layers within this project (Objectives 5 & 6). This will extend the coverage of samples available via OneBenthic to cover the Greater North Sea area as well as the UK exclusive economic zone (EEZ), facilitating the consideration of benthic biodiversity from an international perspective.
3. Creation of a web application to allow open access to the publicly available data contained within the new OneBenthic trawl sample database created as part of Objective 1.
4. Creation of a web application to provide access to modelled benthic biodiversity layers developed under this and other R&D projects. This app displays the modelled layers in the context of relevant overlay layers such as environmental designations & offshore wind farm locations, and also allows for interpretation of model metrics.
5. Modelling of benthic faunal assemblages across the Greater North Sea Area. This approach brings together tens of thousands of benthic samples and uses data describing the physical environment to model the distribution of distinct groups of species or taxa that occur together in space (i.e. community types, or faunal assemblages).
6. Species distribution modelling (SDM) carried out across the Greater North Sea Area for a variety of benthic taxa. Data pulled together through this project will be used to model the distributions of eight species around the UK and across the North Sea, increasing our understanding of biodiversity across this region.

In addition to this report, the outputs of this project include two web applications: **OneBenthic Data Extraction Tool: Trawl** and **OneBenthic Layers.** These apps allow access to the trawl data pulled

together as part of the project, and make available the modelled biodiversity layers respectively (see boxes below for further details).

OneBenthic Data Extraction Tool: Trawl

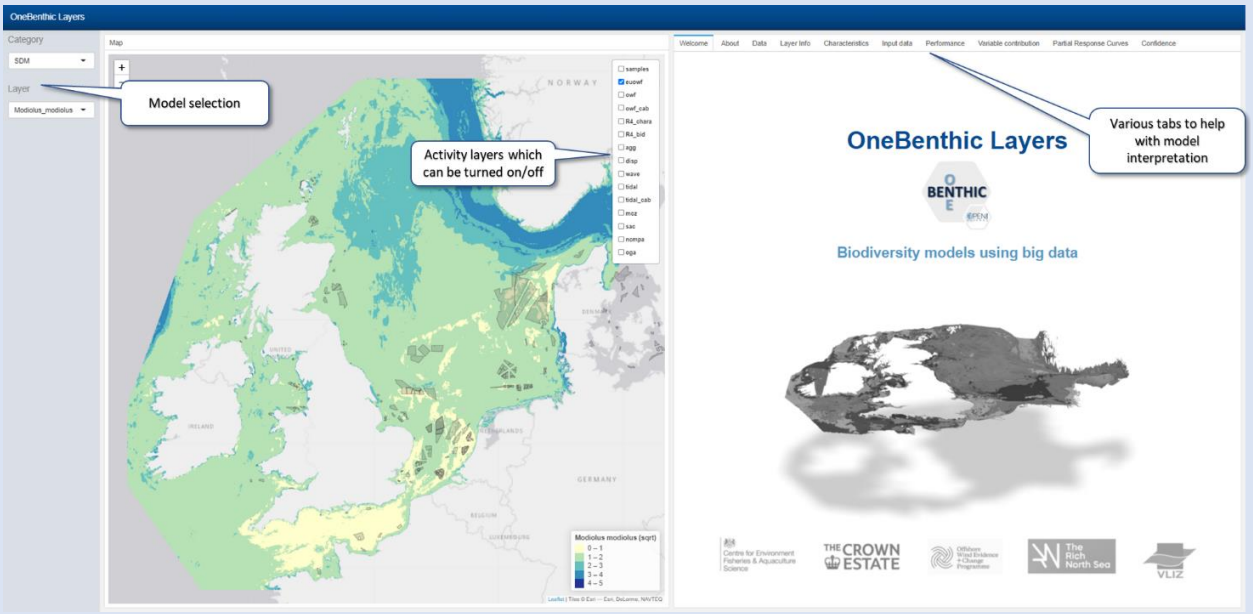
https://rconnect.cefas.co.uk/onebenthic_dataextractiontrawl/



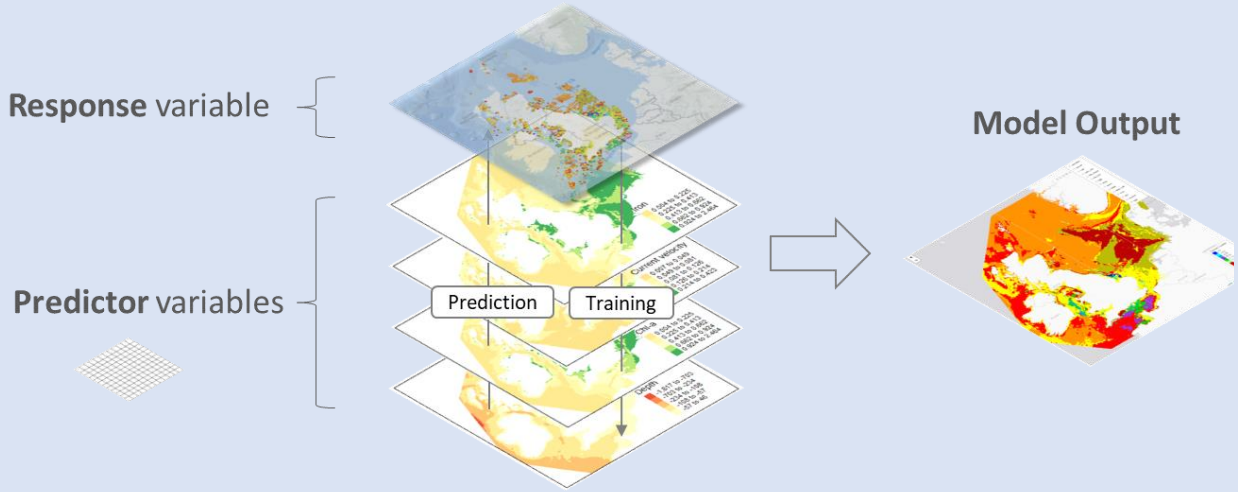
This application allows users to explore and download data from >1500 2m beam trawl samples collated under this project and stored in the new OneBenthic Trawl sample database (OBT). Such data can support developers in efforts to characterise the seabed within their area of interest. Moreover, the standardised dataset can now be used to generate new scientific insights, thereby supporting sustainable development in the marine environment.

OneBenthic Layers Tool

https://rconnect.cefas.co.uk/onebenthic_layers/



The OneBenthic Layers tool provides a platform for sharing modelled biodiversity layers based on the OneBenthic dataset. In this project we developed nine such models, one for seabed macrofaunal assemblages (i.e. community types), and eight for individual species. All models were based on the enhanced dataset produced in this project, including >5k samples from the EurOBIS repository (https://www.eurobis.org/) and covering areas of the North Sea outside the UK EEZ. We employed a Species Distribution Modelling (SDM) approach based on machine learning (Random Forest) in which point sample data (assemblage class or species abundance) are overlaid on a stack of raster (gridded data) predictor variable layers. These predictors are physical environmental variables with relevance for seabed animal communities. Values of each predictor variable are extracted at each of the point sample locations. The relationship between the response variable (the thing we want to model) and the physical predictors is then used to make predictions in places where there are no point sample data. In this way, point sample data are used to build a 100% coverage map across the study area. A range of model outputs, including an associated confidence map, are presented to help with model interpretation.



Big data approaches are likely to play a significant role in the sustainable development of the North Sea. This study has supported this endeavour through development of big data infrastructure, expansion of an existing dataset, and generation of new science outputs. Furthermore, the enhanced dataset is already being used in other projects to develop additional biodiversity layers which will further enhance understanding of the seabed. These layers will be added to the OneBenthic Layers tool in due course. The approach taken here could provide a useful template for other regions around the world where offshore wind energy generation is planned.

A number of recommendations are made:

Project Recommendations

- Infrastructure and processes should be maintained to facilitate continued flow of data into OneBenthic from sources including industry and the appropriate Data Archiving Centre (DASSH), as well as enabling the flow of data internationally to EurOBIS via DASSH. The functionality of OneBenthic should be built upon to enhance its analytical capabilities, enabling big data approaches which have huge potential to improve sustainability by providing new scientific insights.
- The data layers produced as part of this project complement existing approaches for characterising offshore wind development areas by providing an enhanced understanding of benthic biodiversity. “Nature knows no boundaries”, and within environments such as the North Sea, biodiversity should be considered at an ecologically relevant scale using datasets that don’t stop at national borders. We recommend future consideration of international data set development and use to drive this agenda forward.
- A wealth of data collected via industry and other programmes is now available that hasn’t yet been incorporated into the current benthic mapping systems used in UK (or EU) decision making. This project demonstrates the potential to harness these data to provide new insights into benthic biodiversity. We recommend that further work is required to consider how these data and new layers could be incorporated into the decision-making process. In the UK, the OWEC POSEIDON and Defra’s marine Natural Capital and Ecosystem Assessment (mNCEA) projects offer opportunities to address this.
- It is recommended that consideration should be made of the Rich North Sea’s approach to Nature Inclusive Design and how this could be adopted in UK waters to contribute to biodiversity enhancement. The layers produced as part of this project can be used to help inform spatial decisions on where Nature Inclusive Design initiatives might be most appropriate.
- Outputs from this project can help identify areas of seabed that are less well characterised (e.g. Celtic Sea) and hence where future sampling can benefit understanding. This is a key goal of the OWEC POSEIDON project (see https://www.thecrownestate.co.uk/en-gb/media-and-insights/news/the-crown-estate-invests-over-12million-in-new-research-to-help-protect-the-uk-marine-environment/).

2. Introduction

To mitigate the worst effects of climate change, governments around the world are increasingly adopting policies of Net Zero carbon (see Energy & Climate Intelligence Unit, 2021), whereby emissions are reduced or offset. As one of the largest contributors to greenhouse gases, energy production (Herzog, 2005) has attracted considerable attention, with increasing moves away from the use of fossil fuels towards renewables, including wind. As a result of pressure on land use, and lower onshore wind speeds (Akhtar et al., 2021), many coastal states are now pursuing policies of offshore wind energy production. For example, in Europe there are plans to increase capacity from ~17 GW in 2019 to 450 GW by 2050 (Akhtar et al., 2021). Around 47% of this development is expected to take place in the North Sea (WindEurope, 2019).

If plans for offshore windfarm (OWF) development are to be as sustainable as possible, then they must be cognizant of the twin challenge of biodiversity loss (Barnosky et al., 2015), given its link to ecosystem functioning (Gamfeldt et al., 2015), goods and services provision and ultimately human prosperity (Naeem et al., 2016). This is now recognised in UK law, with the Environment Act (2021) requiring most developments to deliver at least a 10% Biodiversity Net Gain (BNG) as a result of their activities. The UK National Infrastructure Commission (UKNIC) seeks to go further, advocating for Environmental Net Gain (ENG), an approach to development that leaves both biodiversity and the environment in a measurably better state than prior to the development, as measured by biodiversity measures, ecosystem services and environmental metrics (UKNIC, 2021). Whilst the details of what NG will mean for developers in the UK remains to be determined, in particular for marine development (see ABPmer, 2019), initiatives in other countries bordering the North Sea are of interest.

The EU's Biodiversity Strategy (European Commission, 2020) is heralded as a comprehensive, ambitious, and long-term plan to protect nature and reverse the degradation of ecosystems. Central to the strategy is the concept of No Net Loss (NNL) of biodiversity. Steps to realise this ambition are being taken in The Netherlands through a policy of Nature Inclusive Design (NID), whereby developers are required to *'take measures to increase the suitable habitat for species naturally occurring in the North Sea'* (Hermans et al., 2020). The introduction of this policy is supported by the Rich North Seas (RNS) initiative (https://www.derijkenoordzee.nl/en/our-approach) that seeks to develop solutions which can be adopted by OWF developers, including the introduction of reef structures (Seaman and Lindberg, 2009) to promote colonisation by naturally occurring reef forming species (e.g. European oyster – *Ostrea edulis* , horse mussel – *Modiolus modiolus*, tube worms – *Sabellaria spinulosa*). These species are often associated with biodiversity hotspots (see van der Reijden et al., 2021). OWFs may also provide benefits for benthic biodiversity through reductions in fishing pressure, either as a result of exclusion or avoidance by boats, facilitating natural recovery of the seabed (Coates, et al., 2016). Thus the roll out of OWFs across the North Sea may present opportunities for biodiversity enhancement or so-called North Sea Net Gain (NSNG).

To help support the expansion of offshore wind (OW), and to assess whether there is evidence of NSNG, there is an urgent need for high resolution maps depicting benthic biodiversity, and for development of approaches to assess temporal change. This is important given the placing of turbines (or their anchoring equipment, in the case of floating devices), hard substrate scour and cable protection on the seabed. These maps could go on to support licensing decisions and provide a benthic faunal baseline against which changes resulting from the development (positive or negative) can be assessed. They can also support decisions around NID by raising awareness of the distribution (known and potential) of species of interest (e.g. *Sabellaria spinulosa*, *Modiolos modiolus*, *Ostrea edulis*). The challenge for NID will be to ensure that any interventions either conserve or enhance the functioning of the local ecosystem.

Data concerning benthic biodiversity are often available for individual development sites, as a result of characterisation surveys undertaken in support of an Environmental Impact Assessment (EIA). What's generally lacking, however, is high resolution contextual information showing how the site fits into the broader spatial picture of benthic changes across the seabed, both nationally and internationally. At present, much reliance is placed on physical based habitat classifications schemes such as the European nature information system (EUNIS) which is used as a proxy for benthic biodiversity (https://www.emodnet-seabedhabitats.eu/about/euseamap-broad-scale-maps/). Whilst it is true that the variables used in the EUNIS marine system are biologically relevant (e.g. water depth, sediment composition, light penetration), recent evidence suggests that EUNIS classes, at the mapped levels 3 and 4, do not discriminate well between benthic faunal assemblages, particularly for the coarse and mixed sediment groups (Cooper et al., 2019). In addition, biodiversity is multifaceted (see Cochrane et al., 2016), and doesn't lend itself to presentation in a single map. Simplifying the environment in this way, whilst conceptually appealing, is likely to mask complexity.

Big data[2] provides an opportunity to build our understanding of benthic habitats through creation of broadscale high resolution biodiversity maps. In recent years the quantity of benthic information in the public domain has risen sharply, and there are now multiple repositories where these data can be accessed (e.g. European ocean biodiversity information system (EurOBIS, https://www.eurobis.org/); The Crown Estate's Marine Data Exchange (MDE, https://www.marinedataexchange.co.uk/); The archive for marine species and habitats data (DASSH, https://www.dassh.ac.uk/); UKbenthos database, https://oguk.org.uk/product/ukbenthos-database-5-14/). In addition, many public and private sector organisations make their data available through websites or upon request.

The OneBenthic (OB) big data initiative (https://sway.office.com/HM5VkWvBoZ86atYP?ref=Link) was set up to allow for the implementation of big data and associated analytical approaches. Data are harvested from different sources, standardised according to the World Register of Marine Species (https://www.marinespecies.org/) and then used to generate new science (e.g. Cooper et al., 2019; Cooper and Barry, 2020; Thompson et al., 2020), taking advantage of contemporary data

[2] For an explanation of big data see https://en.wikipedia.org/wiki/Big_data

science tools for storage, analysis and sharing of information (see Figure 1). As highlighted by Runting et al. (2020), such techniques have huge potential for improving sustainability by providing new insights.

The present study sought to enhance the existing OB infrastructure to further support big data science, and the continued expansion of offshore wind in the North Sea (and beyond), including assessments and future discussions/decisions relating to NG. This was achieved through:

- development of a dedicated trawl sample database (Objective 1), augmenting the OB initiative with an additional type of benthic data
- data harvesting to increase the number of samples available within OB for big data analysis, focusing on the Greater North Sea beyond the UK EEZ (Objective 2)
- development of a web application to allow access to the publicly available data contained within the new trawl sample database (Objective 3)
- development of a web application to allow access to modelled biodiversity layers developed under this and other R&D projects (Objective 4)

These web tools complement others available from the Cefas Open Science site: https://openscience.cefas.co.uk/. Furthermore, the enhanced dataset was used to:

- model benthic faunal assemblages (biotopes) across the Greater North Sea area (Objective 5), expanding on previous work by Cooper et al. (2019) for the UK shelf, and
- model species distributions (SDM) for a variety of benthic taxa (Objective 6).

Outputs from this work are intended to complement existing approaches to assessing benthic biodiversity change and to assist developers, regulators and their advisors in future decision making, thus helping to accelerate the roll out of offshore wind to help meet both Net Zero and NG/NNL objectives.

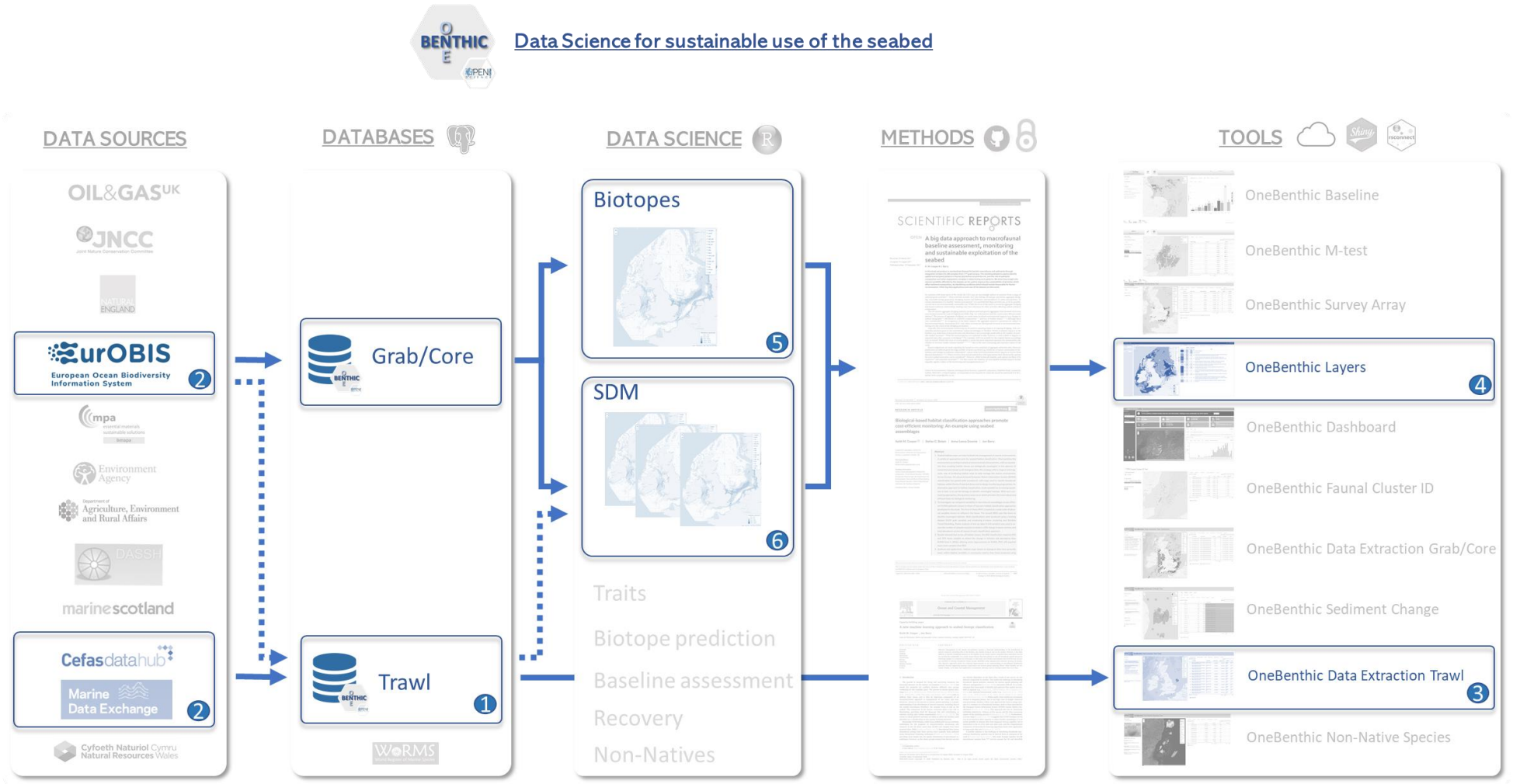


Figure 1. OneBenthic workflow highlighting (in blue) steps undertaken in the current project (numbers relate to project objectives). Dashed lines indicate planned work which could not be undertaken within the scope and timeframe of the project.

3. OneBenthic Infrastructure Enhancement

This project included three objectives to enhance the OB infrastructure, thereby supporting the delivery of big data science. These objectives relate to the development of a new trawl sample database (Objective 1), data harvesting to increase the number of samples available for analysis (Objective 2), and development of web applications to allow for interaction with data and data products (Objective 3).

3.1. Trawl Sample Database (Objective 1)

Benthic surveys, particularly those concerning seabed characterisation, often include the collection of 2m beam trawl samples. This device targets larger epibenthic taxa which are not sampled effectively using a grab or core device. Hitherto, these data have not been included in the OB initiative. The purpose of this objective was to create a dedicated OB trawl sample database (OBT), thereby expanding the scope of the OB initiative and its capacity to support the roll out of offshore wind.

OBT was designed using the online database design tool Vertabelo (https://vertabelo.com/), starting with the existing database model for the OB Grab and Core sample database (OBGC), and adapting it. This involved the removal of the sediment data schema, and addition of new required fields pertinent to beam trawl data. The completed database model is shown in Appendix 1.

3.2. Data Harvesting (Objective 2)

Data harvesting sought to increase the number of samples within OBT and OBGC for use in big data work, including the development of the transnational biotope and SDM layers in the present study. The main area of search was defined by the Greater North Sea ICES Ecoregion (https://www.marineregions.org/gazetteer.php?p=details&id=36317), but also covered the full extent of the raster layers used in modelling (see section 3.1.3).

For OBT, efforts were focused on data held by Cefas, with some additional data acquired from the MDE. For OBGC, the focus was on acquiring data held in EurOBIS, the European node of the International Ocean Biodiversity Information System (OBIS) data repository (https://www.eurobis.org/). Harvesting of EurOBIS data was undertaken by the Flanders Marine Institute (VLIZ), with funding from the RNS programme. EurOBIS data were supplied in Darwin Core (DwC) format (https://dwc.tdwg.org/), an agreed standard, involving up to three tables, for sharing of biodiversity information. DwC tables were uploaded to a temporary database, and R (R Core Team, 2020) script used to output information in the form required for ingestion into OBGC. Some differences in how DwC fields were used between surveys meant that tweaking of the R script was sometimes necessary. Where abundance and biomass were reported in per m^2 it was necessary to change values to per 0.1 m^2, thereby ensuring consistency with existing data held in OBGC. These changes were recorded in OBGC and stored metadata include a URL link back to the original EurOBIS record.

Data harvesting resulted in the addition of 7,456 samples to OneBenthic, raising the total number to 45,935, a 19% increase (December, 2021). These new samples include 1,528 2m beam trawls added to OBT (see Figure 2a) and 5,928 grab and core samples added to OBGC (Figure 2b). For further details of these data see Appendix 2 for OBT and Appendix 3 for OBCG.

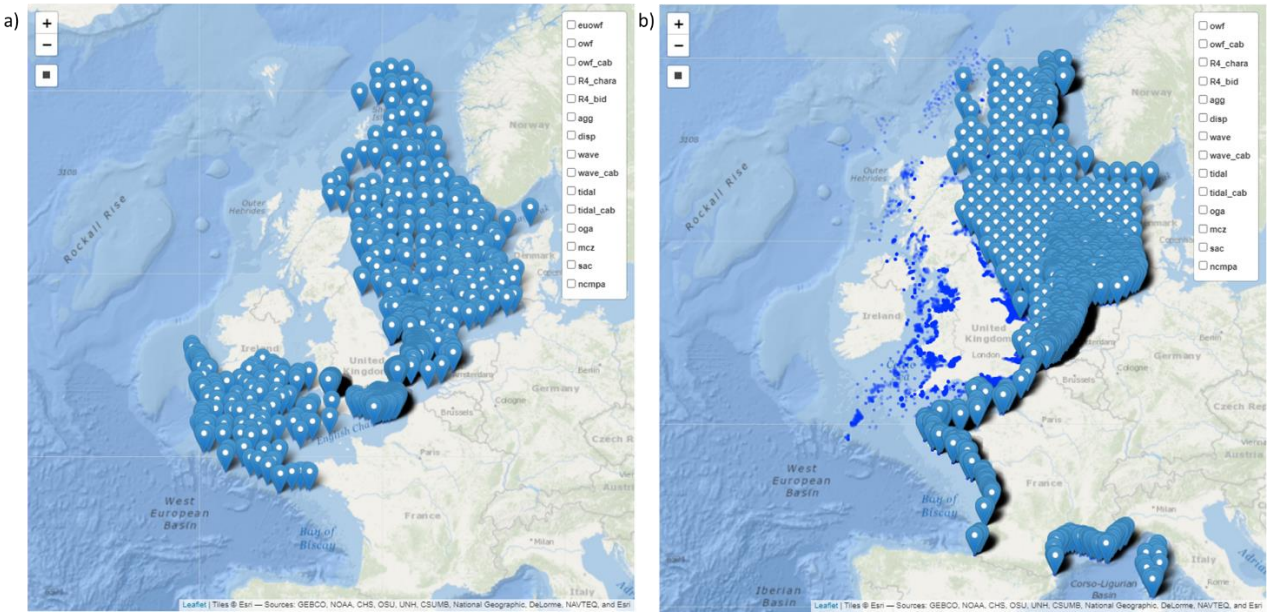


Figure 2. Location of a) 2m beam trawls samples added to OBT, and b) grab/core samples added to OBGC. Blue dots show existing sample locations.

With limited project resources, data harvesting necessarily targeted datasets with good spatial coverage. However, both EurOBIS and the MDE include multiple other datasets which could, in future, usefully be incorporated in OBT and OBGC.

The expanded OBGC dataset can be explored in more detail using the OneBenthic Dashboard (https://rconnect.cefas.co.uk/onebenthic_dashboard/) and a YouTube fly-through (https://www.youtube.com/watch?v=mPZubod_z14).

3.3. App Development

A series of existing web applications allow users to interact with data and science outputs from the OB initiative (see Table 1). These apps are available from the Cefas Open Science site (https://openscience.cefas.co.uk/).

Table 1. Existing OB apps available from the Cefas Open Science site.

App	URL
OneBenthic Baseline Tool	https://openscience.cefas.co.uk/ob_baseline/
OneBenthic M-test Tool	https://openscience.cefas.co.uk/ob_mtest/
OneBenthic Non-Native Species Tool	https://rconnect.cefas.co.uk/onebenthic_nonnativespecies/
OneBenthic Faunal Cluster ID Tool	https://rconnect.cefas.co.uk/onebenthic_faunalclusterid/
OneBenthic Sediment Change Tool	https://rconnect.cefas.co.uk/onebenthic_sedimentchange/
OneBenthic Survey Array Tool	https://rconnect.cefas.co.uk/onebenthic_surveyarray/
OneBenthic Dashboard	https://rconnect.cefas.co.uk/onebenthic_dashboard/
OneBenthic Data Extraction Tool	https://rconnect.cefas.co.uk/onebenthic_dataextractiongrabcore/

Two further apps were developed in the current project to allow users: to i) access data held in the OBT database (**OneBenthic Data Extraction Tool: Trawl**), and ii) access benthic biodiversity models developed under this and other R&D projects (**OneBenthic Layers Tool**). Each app is described in more detail below.

3.3.1. OneBenthic Data Extraction Tool: Trawl (Objective 3)

This R Shiny application allows users (e.g. developers, regulators, advisors and researchers) to access publicly available benthic trawl sample data from the OBT database, potentially reducing the need for collection of new samples. The app can be accessed either via the Cefas Open Science Site (https://openscience.cefas.co.uk/), or directly (https://rconnect.cefas.co.uk/onebenthic_dataextractiontrawl/).

The app screen is spilt into three parts (Figure 3). On the left the user makes selections, in the middle there is an interactive map, and on the right there are various tabs with results and further information. Data can be extracted in two ways. Firstly, by selecting one or more survey names from a drop-down list. Sample locations for the selected surveys are shown in the map. The map also allows for different polygons to be overlaid, including European offshore windfarm areas supplied by EMODnet (https://www.emodnet-humanactivities.eu/). Further details for the map overlays can be found in the 'Data' tab. Basic metadata for selected samples are displayed in a table on the 'Data (by survey)' tab. Data can be downloaded, in csv format, using the 'Download data' button. Alternatively, samples of interest can be selected using the draw rectangle tool in the map. Results from this search are displayed in the 'Data (by sample)' tab. Similarly, data can be downloaded in csv format using the 'Download data' button. Downloaded data include multiple metadata fields in addition to the species abundance data. All data are exported as valid names/aphia ids according to the World Register of Marine Species (https://www.marinespecies.org/). Extracted data are in long format. To create a data matrix the user simply needs to pivot the relevant columns (e.g. Samplecode, ValidName, Abund). App code is stored in the gitHub online repository https://github.com/keithmcooper/OneBenthicDataExtractionToolTrawl.

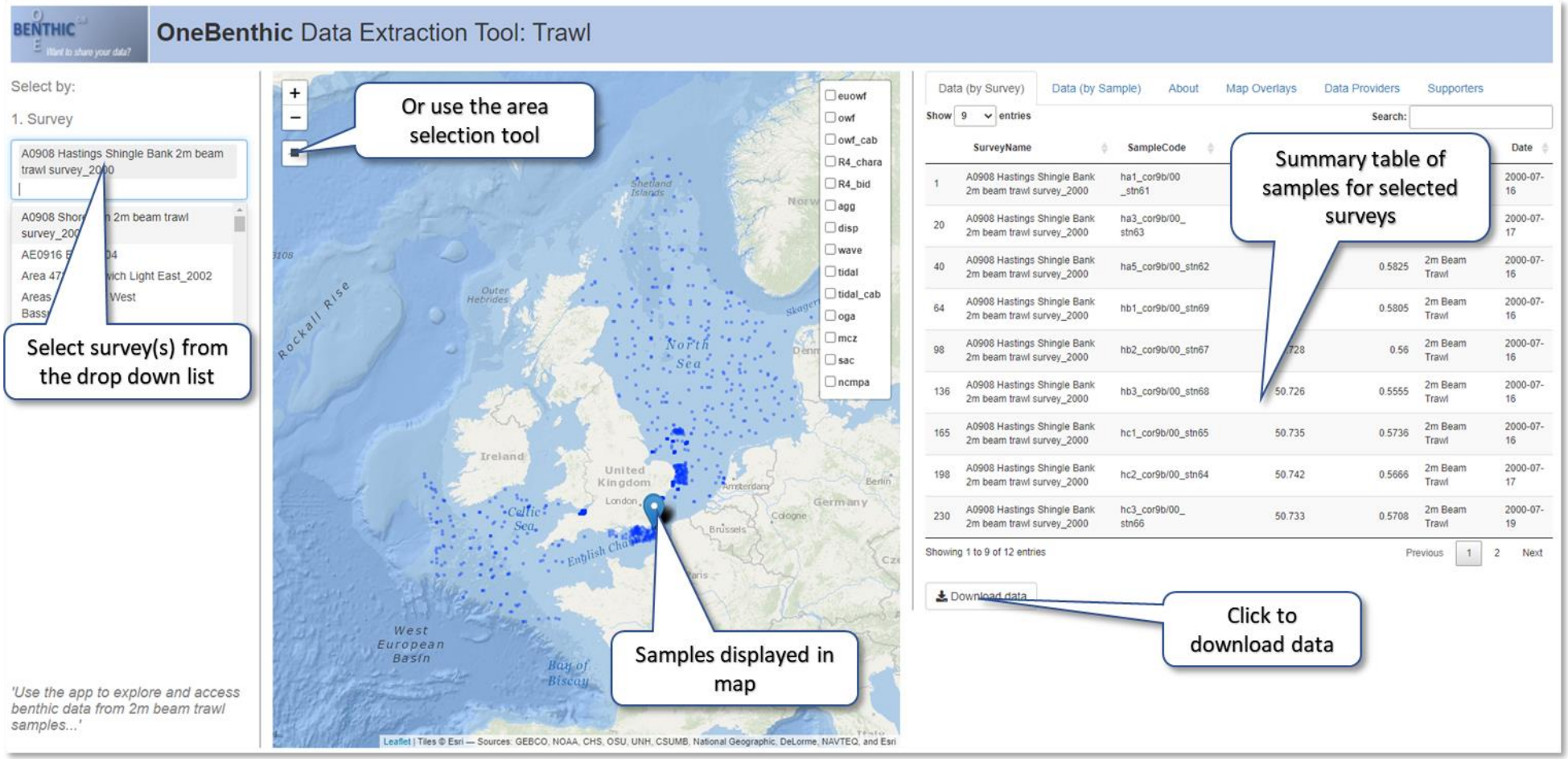


Figure 3. Screenshot of the OneBenthic Data Extraction Tool: Trawl.

3.3.2. OneBenthic Layers Tool (Objective 4)

This R Shiny application, created using the 'flexdashboard' package, allows users (e.g. developers, regulators, advisors and researchers) to access modelled biodiversity layers developed under this (i.e. Biotopes – see Section 4.1 and Species Distribution Models – see Section 4.2) and other R&D projects. The app can be accessed either via the Cefas Open Science Site (https://openscience.cefas.co.uk/) , or directly (https://rconnect.cefas.co.uk/onebenthic_layers/).

The app screen is split into 3 parts (Figure 4). On the left the user makes selections, in the middle there is an interactive map, and on right there are various tabs with model outputs and further information. Using the selection boxes on the left of the screen, users must first select a category of modelled layer, choosing from the following options: Structure or SDM (further options will be added in due course). Having made a choice, the user then selects a layer to display. The modelled data will then appear in the map. Various map overlays can also be displayed, including European offshore windfarm areas supplied by EMODnet (https://www.emodnet-humanactivities.eu/). Further details for the map overlays can be found in the 'Data' tab.

On the right of the screen are several tabs which show various model outputs. The 'Layer Info' tab provides a description of the modelled layer, a reference (if appropriate), a link to download data (if applicable) and details of funders or partners involved in the production of the layer. The 'Characteristics' tab applies only to categorical layers and details the nature or characteristics of individual modelled classes. The 'Input data' tab shows a map of distribution records and a histogram (numeric models) or bar chart (categorical models) for the data used in the model. The 'Performance' tab provides various metrics of model performance, together with a confusion matrix (class models) or scatterplot (continuous models). The 'Variable Contribution' tab shows how much information the predictor variable contributes to the model, in the presence of all other variables. The 'Partial Response Curves' tab show the effect each predictor variable has on the response variable. The curves are produced by varying each predictor variable over its range in turn and predicting with the model, whilst all other variables are held constant at their mean (or in the case of a factor variable, most common) value. Finally, the 'Confidence' tab contains a map showing spatial confidence in the model, based on multiple model runs. For categorical variables (e.g. faunal cluster group), the map shows confidence, with high values (darker shades) indicating areas of high confidence. For numerical variables (e.g. species abundance), the map shows the coefficient of variation (CV). In this case, high values (darker shades) indicate higher variability and thus lower confidence in the model. App code is stored in the GitHub online repository https://github.com/keithmcooper/OneBenthicOneBenthicBiodiversityLayersApp.

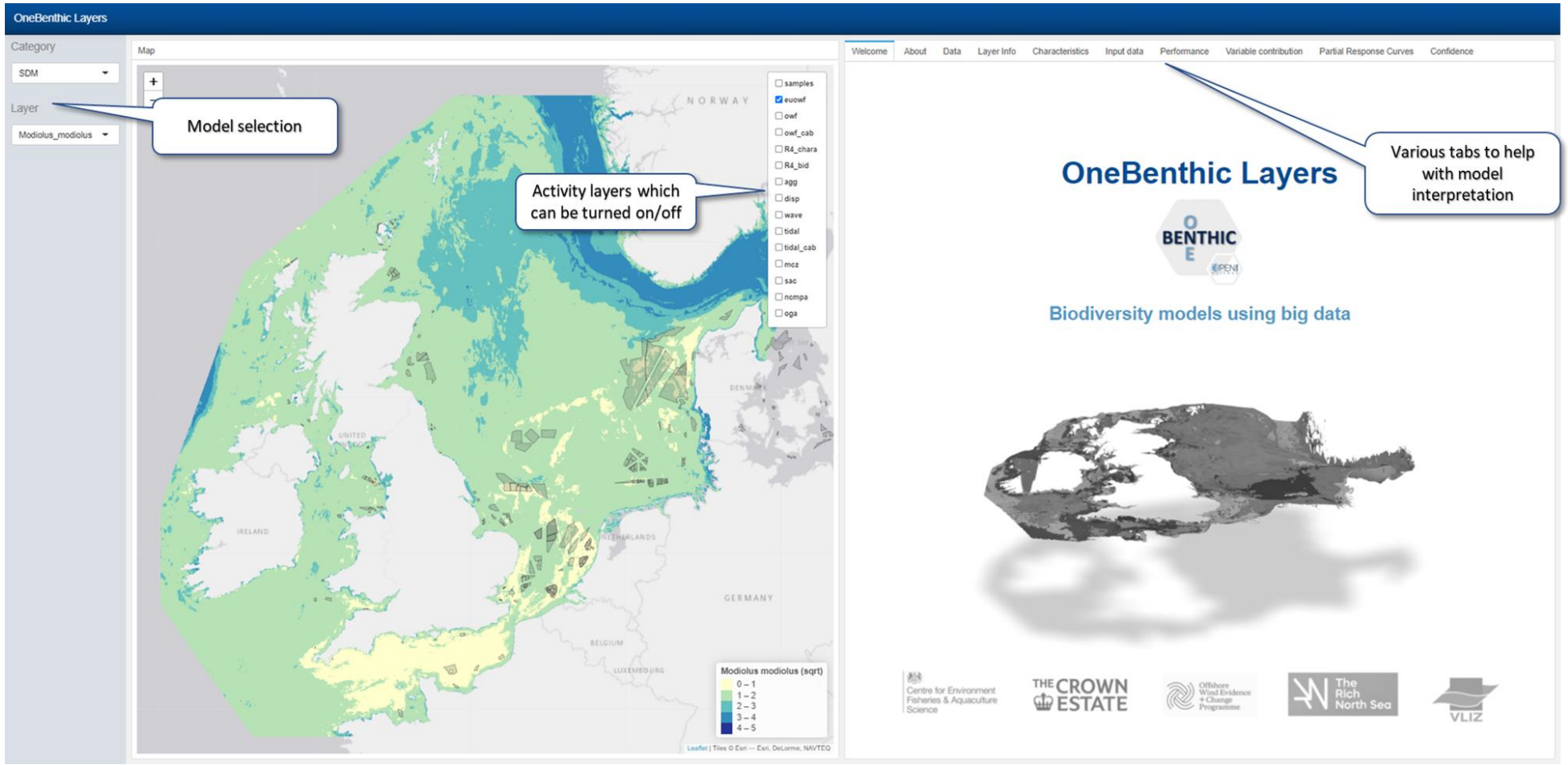


Figure 4. Screenshots from the OneBenthic Layers Tool.

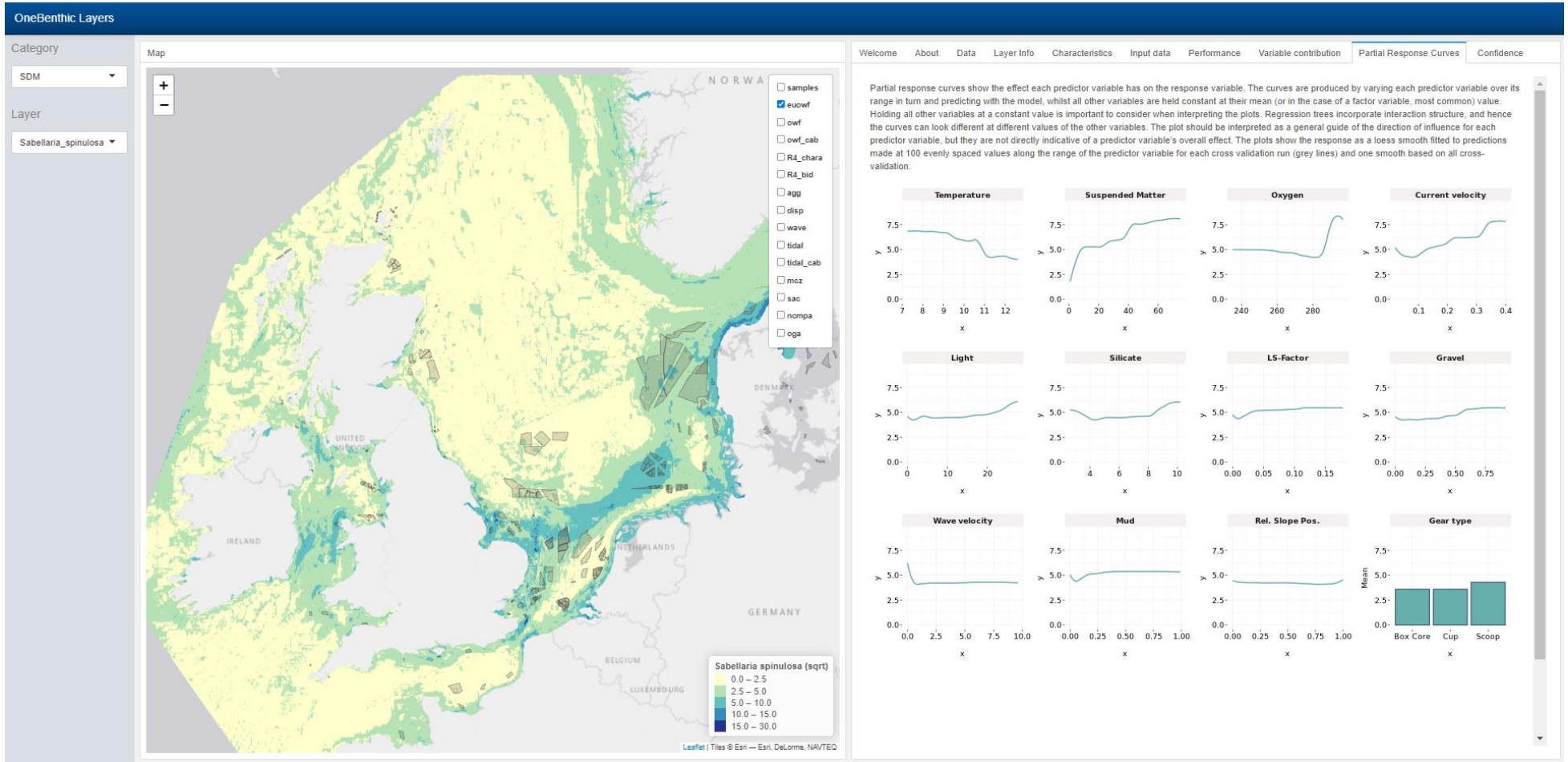


Figure 4. Cont'd.

4. Layer Development

This project supported development of a variety of benthic biodiversity layers, using data from OBGC. These layers included biotopes, and eight species distribution models.

4.1. Biotope Map (Objective 5)

The purpose of this objective was to produce a macrofaunal assemblage[3] (biotope) layer based on the methodology in Cooper et al. (2019) and covering the Greater North Sea. This necessitated use of the expanded dataset from OBGC, and a reclustering of the data followed by Random Forest modelling (Breiman, 2001).

4.1.1. Dataset

From the expanded OBGC dataset of 44,407 samples, we selected a subset of 31,845 for which data were considered comparable (i.e. sample acquired using a 0.1 m^2 grab or core, processed using a 1 mm sieve and not taken from a known impacted site). Colonial taxa were included and given a value of one.

To take account of potential differences in taxonomic resolution between surveys, macrofaunal data were aggregated to family level using the taxonomic hierarchy provided by the World Register of Marine Species (https://www.marinespecies.org/). This reduced the number of taxa from 3,659 to 750.

To address spatial autocorrelation in the data, and in keeping with the previous approach, samples closer than 50 m were removed from the dataset, reducing the overall number to 18,348. A fourth-root transformation was then applied to the data to down weight the influence of highly abundant taxa.

4.1.2. Clustering

Prior to clustering, an elbow plot (Thorndike, 1953) was used to identify an appropriate number of cluster groups. Clustering was performed in R using a k-means approach (R function *kmeans*) with the MacQueen algorithm (MacQueen, 1967). To establish the relationship (i.e. similarity/dissimilarity) between the different faunal cluster groups, we computed the absolute distances between each of the cluster centres across all variables (R function *dist*). The resulting dissimilarity matrix was then used to generate a dendrogram based on group average hierarchical clustering (R function *hclust*). Cluster groups were matched to those previously identified in Cooper and Barry (2017) and assigned a corresponding label. Group characteristics were identified by

[3] a group of species or taxa (>1mm in size) that occur together in space.

calculating the mean richness and abundance for each group. Richness was calculated using the 'vegan' package in R (Oksanen et, 2016). Characterising taxa for each group were identified from the cluster centres (ten highest z-scores).

4.1.3. Environmental Variables

A variety of raster predictor layers with relevance to benthic communities were sourced for use in modelling. These layers came from Bio-ORACLE (https://www.bio-oracle.org/; Tyberghein et al., 2012; Assis et al., 2017), and Mitchell et al. (2019). With the exception of water depth, which was sourced via the R library 'sdmpredictors', all Bio-ORACLE layers (Temperature, Salinity, Current velocity, Nitrate, Phosphate, Silicate, Dissolved molecular oxygen, Iron, Chlorophyll, Phytoplankton, Primary productivity, Light at bottom) were obtained using the Download manager with the following options:

Period: Present
Depth of layers: Benthic layers
Format of files(s): Tiff Raster file (.tif)
Bio-ORACLE version: 2.2
Layers to download: All (Mean)

Available layers from Mitchell et al. (2019) included data products (% Mud, % Sand and % Gravel; https://doi.org/10.14466/CefasDataHub.63), and associated predictor variables (Suspend inorganic particulate matter, Peak wave orbital velocity; https://doi.org/10.14466/CefasDataHub.62). Two additional environmental layers representing seafloor topography were derived from the Bio-ORACLE bathymetry layer (water depth) using SAGA GIS tools for QGIS (v. 3.2; Conrad et al., 2015): a variable combining topographic slope length and steepness (gradient over the length, the LS-factor), and the relative location along the entire length of a discrete slope ranging from 0-1 from the bottom to the top of the slope (Relative Slope Position, Böhner & Selige, 2006). The LS-Factor is used to predict erosion potential in the terrestrial environment (Desmet & Govers, 1996) and can analogously be applied in the marine context to reflect the potential stability of sediment deposits and hence the likelihood of exposed hard substrata. The Relative Slope Position can again be interpreted to represent different current conditions nearer the bottom or top of the slope. Bio-ORACLE rasters were cropped and resampled so that the spatial extent and pixel resolution matched that from Mitchell et al. (2019). Figure 5 shows thumbnails of all the candidate environmental variables over the study area.

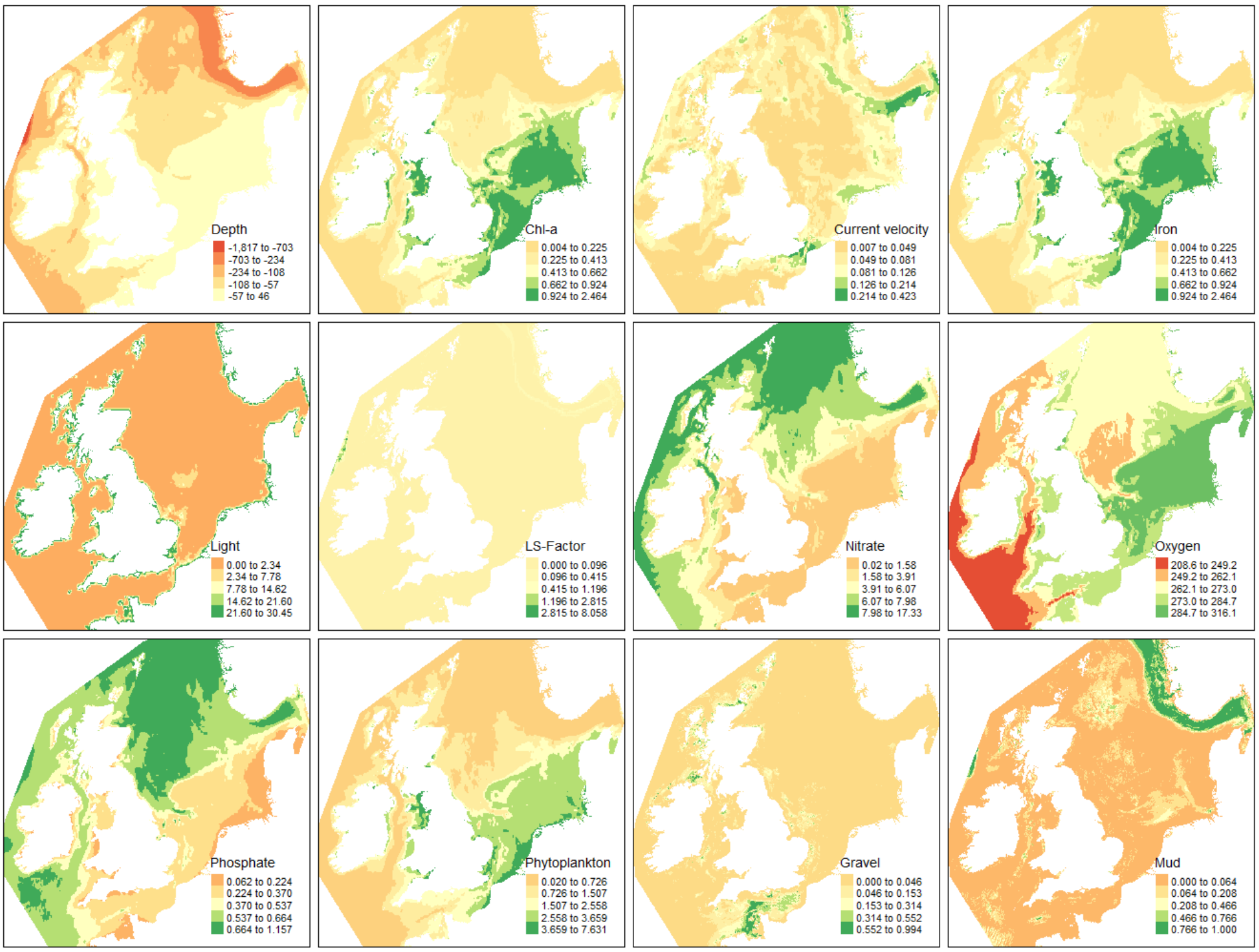


Figure 5. Candidate environmental predictor variable rasters.

4.1.4. Modelling Methodology

Full coverage maps for macrofaunal clusters were produced using Random Forest, an ensemble modelling method where a large number of decision trees (typically 500-1000) are built using random subsets of the samples and predictor variables in the input data (Breiman, 2001; Cutler et al., 2007). Classification trees are used for response variables consisting of discrete factor classes, such as the assemblage cluster groups here and spatial predictions can be made either as class-specific probabilities, derived from the proportion of component trees predicting the class, or as the class with a majority vote. Random Forest was selected because of its suitability for predicting factor-type response variables and its ability to account for the multiple interactions and nonlinear relationships between the response and predictor variables (Rodriguez-Galiano, Ghimire, Rogan, Chica-Olmo, & Rigol-Sanchez, 2012; Strobl, Malley, & Tutz, 2009). In addition, our own research on this dataset showed that it performed better than multinomial models for prediction (results not shown).

The models were built in R (v.4.0.2, R Development Core Team, 2020), using the 'randomForest' implementation of Random Forests in the *randomForest* package (Liaw and Wiener, 2002). The models were run on the default settings, using 1000 trees.

Preliminary single models using all environmental variables were run first to select best variables and remove variables with high covariance. Variables were dropped from the models based on redundancy (high correlation with a more important variable), or poorly defined relationship with the response variable (based on the response curves – see below).

Co-variance between environmental variables was investigated using values extracted from the predictor rasters at the 18,348 grab sample locations taken forward for analysis. Correlation analysis (Figure 6) and hierarchical clustering of euclidean distance plotted as a dendrogram (Figure 7) were used as two different ways to represent covariance and allow the main predictor variables to be identified. Although Random Forest models are not sensitive to covariance effects, models with fewer predictor variables are simpler and easier to interpret. Furthermore, the way variable importance statistics for the models are calculated also makes them more accurate in models with fewer variables than in models that include highly correlated predictor variables, which are interchangeable in the component trees, and hence can mask the importance of other variables.

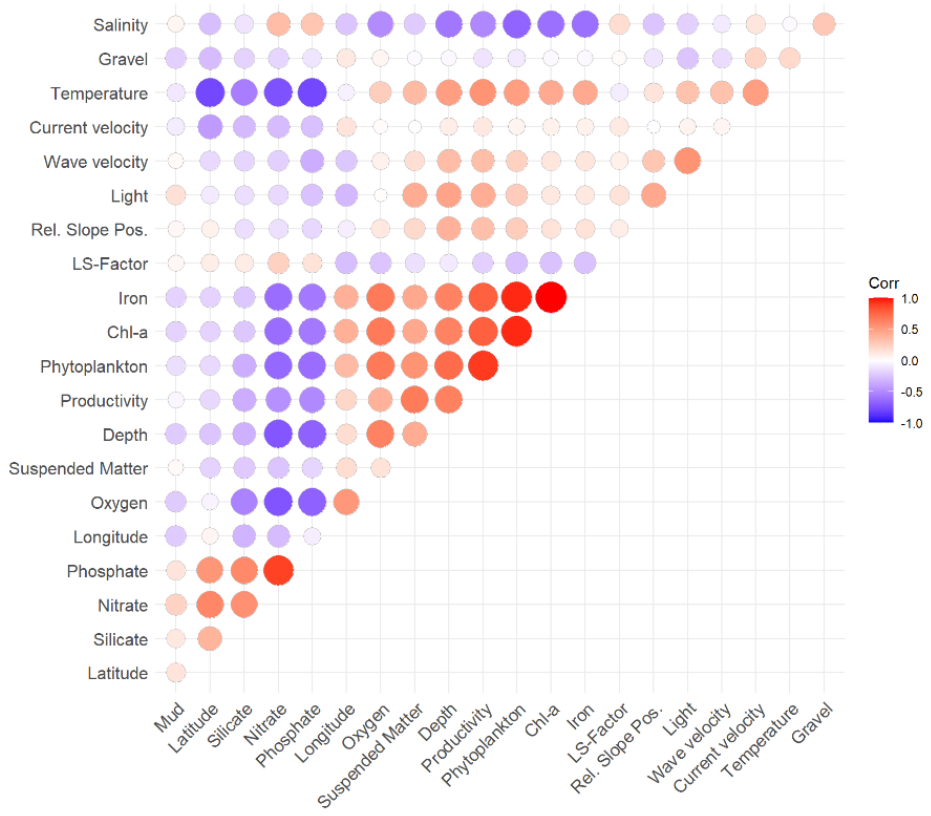


Figure 6. Correlations between environmental predictor variables. The extent of the correlation (0 to ±1) is indicated by symbol size, with colours indicating whether the correlation is positive (red) or negative (blue).

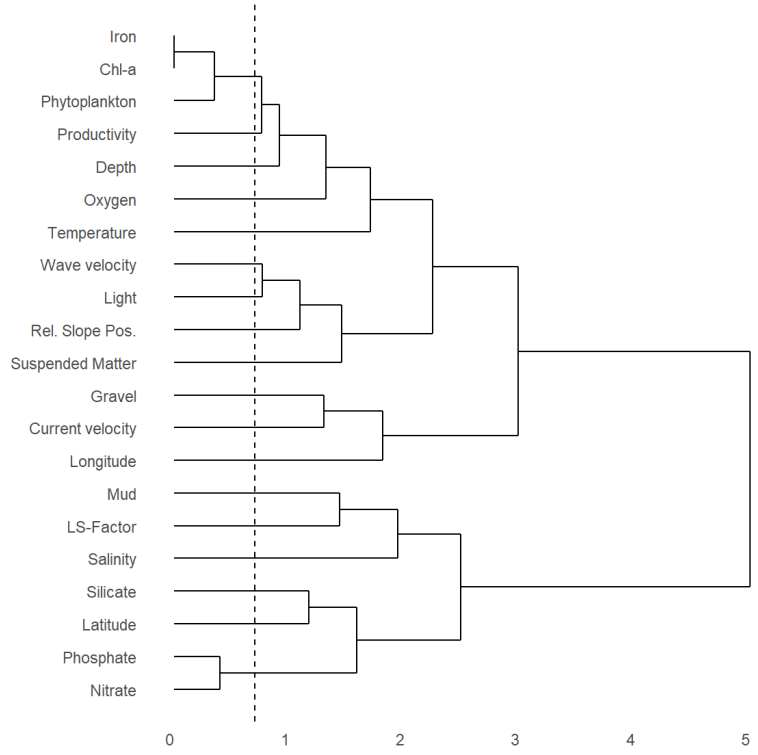


Figure 7. Dendrogram showing the relationship (i.e. similarity) between environmental variables based on euclidean distance. Closely linked variables are likely to be highly correlated.

Based on these results, a number of variables were dropped including nitrate and phosphate (covariates for silicate), iron, chlorophyll and primary productivity (covariates for phytoplankton), % sand (covariate for % gravel), latitude (covariate for temperature) (See Figure 6 and 7). The final random forest models were built with the subset of predictor variables listed in Table 2.

Table 2. Final predictor variables used for modelling.

Variable	Detail	Units	Source
Temperature	Temperature	ºC	Bio-ORACLE (download)
Salinity	Salinity	PSS	Bio-ORACLE (download)
Current velocity	Current velocity	m-1	Bio-ORACLE (download)
Silicate	Silicate	mol.m-3	Bio-ORACLE (download)
Oxygen	Dissolved molecular oxygen	mol.m-3	Bio-ORACLE (download)
Phytoplankton	Phytoplankton	umol.m-3	Bio-ORACLE (download)
Productivity	Primary productivity	g.m-3.day-1	Bio-ORACLE (download)
Phosphate	Phosphate	mol.m-3	
Iron	Iron	umol.m-3	
Light	Light at bottom	-	Bio-ORACLE (download)
Depth	Water depth	m	Bio-ORACLE (via sdmpredictors R package)
Mud	% Mud	%	Mitchell et al. (2019)
Gravel	% Gravel	%	Mitchell et al. (2019)
Suspended Matter	Suspend inorganic particulate matter	g.m-3	Mitchell et al. (2019)
Wave velocity	Peak wave orbital velocity	m.s	Mitchell et al. (2019)
LS-Factor	Topographic slope length and steepness	ratio	Based on Depth
Rel. Slope Pos.	Relative slope position	ratio	Based on Depth

Cross-validation via repeated sub-sampling was done to evaluate the robustness of the model estimate and predictions to data sub-setting and to extract additional information from the model outputs to produce maps of confidence in the predicted distribution, following the approach described in Mitchell et al. (2018). The cross-validation was done on 10 split sample data sets with 75% used to train and 25% to test models, randomly sampled within the levels of the response variable to maintain the class balance. The final model output was plotted as the cluster class with the majority vote of all 10 model runs. Three confidence map layers were also produced consisting of: (1) the frequency of the most common class, (2) the average probability of the most common class and (3) combined confidence computed by multiplying the previous two.

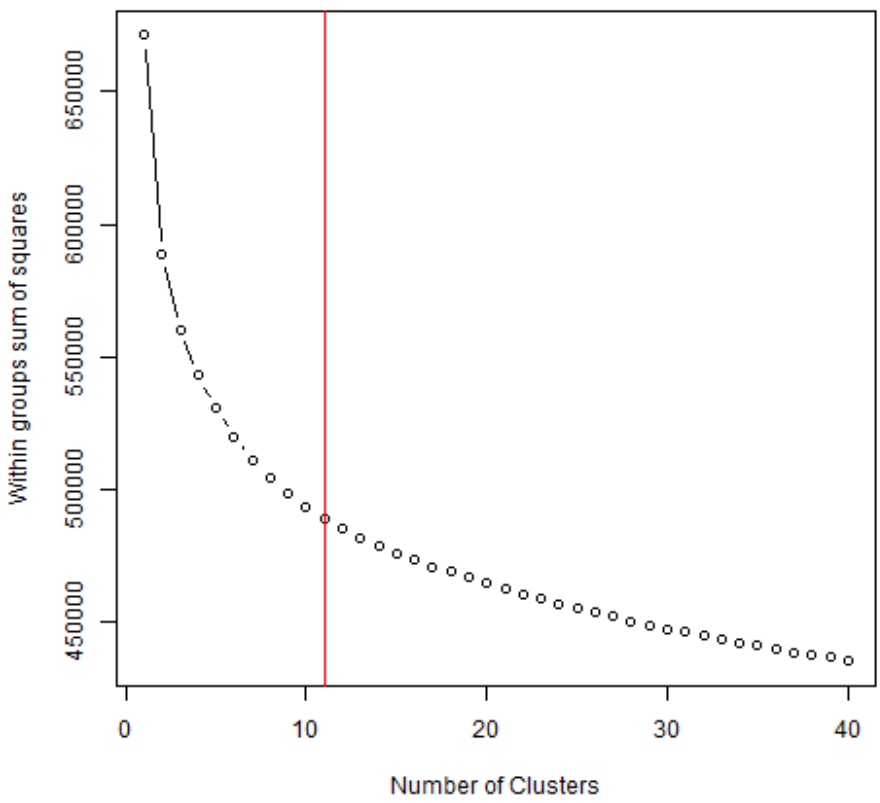
Model performance was assessed using multiple commonly used accuracy statistics calculated from a confusion matrix. Sensitivity, Specificity and Balanced Accuracy were calculated both for individual classes and for the model overall. Sensitivity, also referred to as the True Positive Rate corresponds to the proportion of observed members of a class correctly predicted as such. Conversely, Specificity or True Negative Rate is the proportion of non-members of a class correctly predicted. These can

be used to judge how likely a model is to detect a particular class (or the correct prediction) and how specific the predictions are to the correct class. High sensitivity with a low specificity indicates a model that is overpredicting, whilst an underpredicting model shows high specificity and low sensitivity. Balanced Accuracy is the average of Sensitivity and Specificity and can give a much better estimate of model overall performance than the proportion of samples correctly classified in cases where the classes are unbalanced. The overall accuracy was additionally investigated using the Kappa statistic, a measure of performance which takes account of class imbalance, and two spatially derived statistics: Quantity Disagreement, the amount of difference between the observed and predicted proportions of the classes and Allocation Disagreement, the amount of difference between the observed and predicted data that is due to the spatial allocation of the classes, given the proportions. All accuracy statistics are presented as means and standard deviations of the scores from the 10 model runs.

4.1.5. Results

In keeping with earlier work for the UK (Cooper and Barry, 2017), the elbow plot from this study did not suggest an obvious clustering solution (Fig. 8a). We opted for 11 groups, as this number coincided with a levelling out of the plot. In addition, the associated dendrogram for 11 groups (Figure 8b) was broadly consistent with the one reported in Cooper and Barry (2017) involving a clustering of UK data. The addition of new data in the present study did not result in a markedly different clustering solution, suggesting the groups identified in Cooper and Barry (2017) are broadly representative of the wider study area, perhaps with the exception of the Norwegian Channel where data were sparse. For these reasons we adopted the same cluster group labelling and colouring.

a)



b)

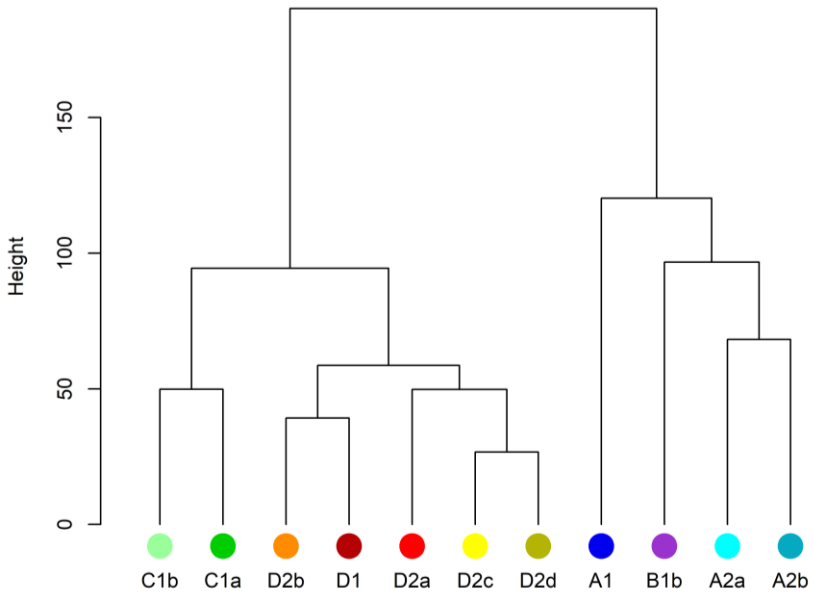


Figure 8. a) Elbow plot and b) dendrogram associated with the k-means clustering of selected macrofaunal data from OBGC.

The model outputs and results shown here are also available in the OneBenthic Layers Application (https://rconnect.cefas.co.uk/onebenthic_layers/). The assemblage cluster group model prediction map is accompanied by a corresponding confidence map along with additional information on the cluster groups, the model input data, model fit and performance statistics and the importance and effects of predictor variables in the model, to help the user interpret the model output.

The **Layer Info** and **Characteristics** tabs in the application contain basic information about the modelled layer and the macrofauna assemblage clusters it represents, including a table summarising the characteristic taxa, species richness and abundance in each group (Table 3).

Table 3. Biological characteristics of the macrofaunal assemblages identified through a k-means clustering of macrofaunal data.

		Richness		Abundance		
Cluster	n	Mean	s.d.	Mean	s.d.	Taxa
A1	288	70	14	1,081	767	Balanidae, Styelidae
A2a	523	52	14	1,052	1,238	Sabellariidae
A2b	814	59	12	384	285	Sabellariidae, Serpulidae, Syllidae, Terebellidae, Spionidae, Capitellidae, Polynoidae, Styelidae, Lumbrineridae, Porcellanidae, Amphiuridae, Cirratulidae, Verrucidae
B1b	926	58	13	222	135	Spionidae, Serpulidae, Syllidae, Galatheidae, Glyceridae, Terebellidae, Phyllodocidae, Amphiuridae, Polynoidae, Capitellidae, Nemertea, Scalibregmatidae, Fibulariidae, Eunicidae, Lumbrineridae
C1a	1799	32	9	147	197	Spionidae, Terebellidae, Serpulidae, Syllidae, Capitellidae, Lumbrineridae, Cirratulidae, Sabellariidae, Nemertea, Polynoidae, Phyllodocidae, Glyceridae, Maldanidae
C1b	1249	43	10	379	1,489	Spionidae, Capitellidae, Terebellidae, Lumbrineridae, Ampeliscidae, Nemertea, Cirratulidae, Semelidae, Ampharetidae, Phyllodocidae, Pholoidae
D1	1043	30	9	444	727	Lasaeidae, Amphiuridae, Spionidae, Semelidae, Nephtyidae, Nuculidae, Pectinariidae, Cirratulidae, Phoronidae, Pholoidae
D2a	2398	24	8	93	117	Spionidae, Glyceridae, Nemertea, Terebellidae, Capitellidae, Fibulariidae, Syllidae, Phyllodocidae, Cirratulidae, Opheliidae, Lumbrineridae, Goniadidae, Polynoidae, Nephtyidae, Dorvilleidae
D2b	1313	30	8	208	497	Oweniidae, Spionidae, Amphiuridae, Capitellidae, Ampharetidae, Thyasiridae, Lumbrineridae, Nemertea, Nephtyidae, Cirratulidae
D2c	5672	9	5	25	40	Nephtyidae, Spionidae, Opheliidae, Glyceridae, Bathyporeiidae, Nemertea, Terebellidae, Orbiniidae, Electridae, Urothoidae, Semelidae, Capitellidae, Ophiuridae, Cirratulidae, Mysidae, Mactridae, Phyllodocidae, Magelonidae, Lumbrineridae, Tellinidae
D2d	2323	18	6	96	112	Bathyporeiidae, Spionidae, Magelonidae, Nephtyidae, Tellinidae, Cirratulidae, Semelidae, Nemertea

The **Input data** tab shows (1) a bar graph illustrating the distribution of the model input data (total N = 17,842) over the 11 assemblage cluster groups (Figure 9) and (2) the spatial distribution of the input data (Figure 10). The most common cluster group, D2c, has 5,433 observations, almost 20 times that of the least common group A1, with only 286 observations and more than twice as many as the second most populous group D2d with 2,304 observations.

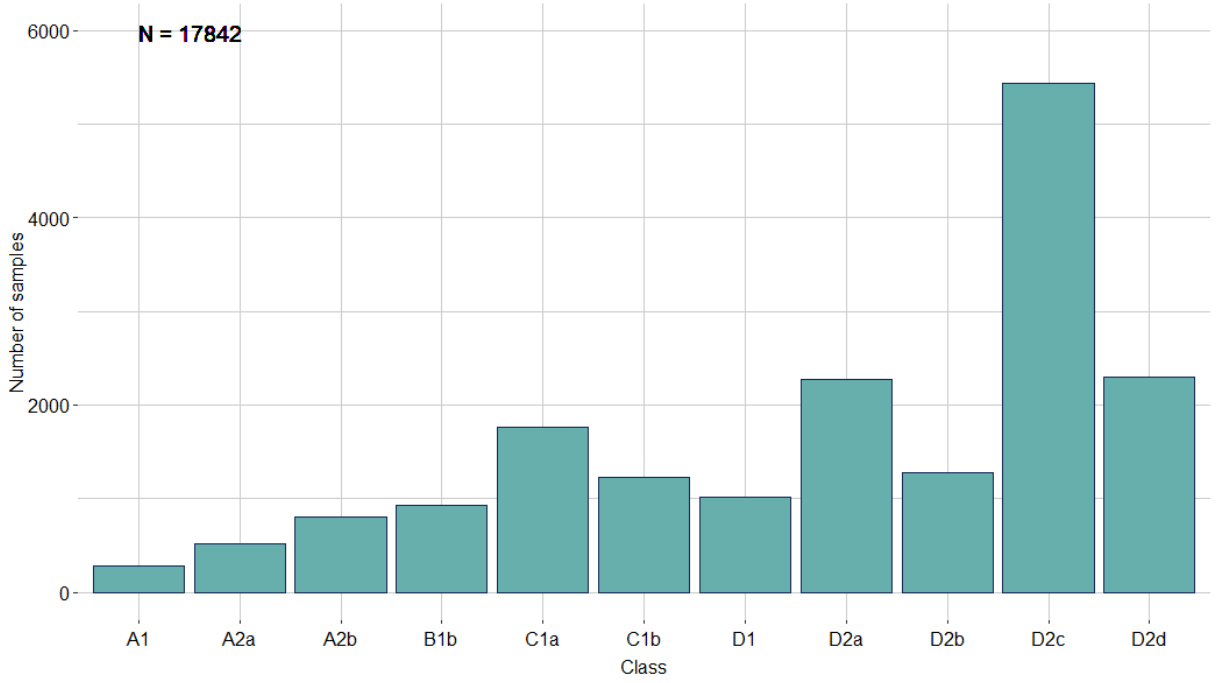


Figure 9. Macrofauna assemblage model input data distribution over cluster classes.

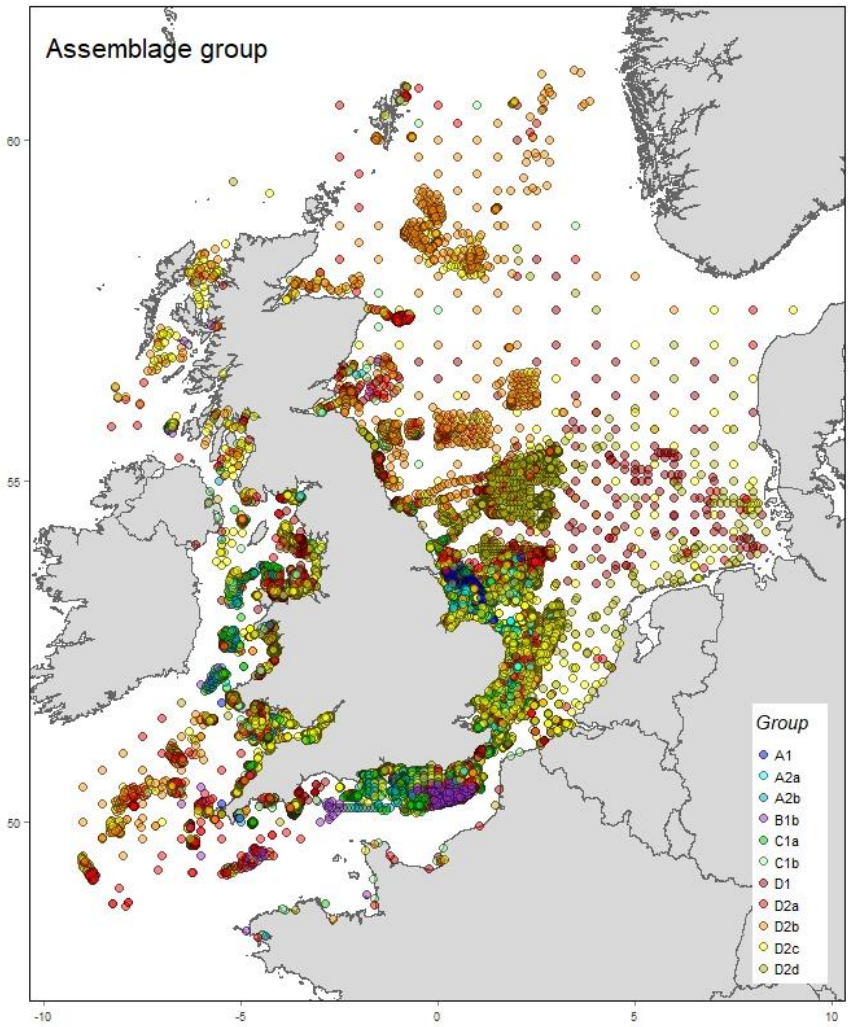


Figure 10. Spatial distribution of faunal cluster observations.

The **Performance** tab contains two tables summarising the mean and standard deviation of accuracy statistics calculated on the test data sets from the 10 split sample cross-validation runs (Tables 4 and 5) and a visualisation combining the confusion matrices from the 10 model runs into one matrix of boxplots for each class combination and the user's and producer's accuracies for each class (Figure 11). The very low variability (in many cases rounded to zero) in the accuracy statistics indicates that the subsets of data converge on very similarly performing models. The first table (Table 4) shows the total number of samples used in the training data set and the full model Sensitivity, Specificity and Balanced Accuracy, Kappa statistic, Quantity Disagreement and Allocation Disagreement. The second table (Table 5) shows class specific Sensitivity, Specificity and Balanced Accuracy scores. The performance statistics here indicate a moderately well performing model, which tends to somewhat underpredict the distribution of the correct cluster classes, whilst the most common class observed in the grab dataset (D2c) is somewhat overpredicted. The spatial area of each class predicted corresponds well to the proportions observed in the test dataset. Classes that have the highest overall rate of correct prediction are B1b and D2b, whilst D2a, C1b and C1a have the lowest.

Table 4. Mean and standard deviation of model validation statistics over 10 random split sample runs.

Statistic	Mean ± SD
N	13,386
Sensitivity	0.61 ± 0.00
Specificity	0.96 ± 0.00
Kappa	0.53 ± 0.00
Balanced Accuracy	0.78 ± 0.00
Quantity Disagreement	0.14 ± 0.01
Allocation Disagreement	0.23 ± 0.01

Table 5. Class-specific performance.

Cluster	N	Sensitivity	Specificity	Balanced Accuracy
A1	215	0.58 ± 0.05	0.99 ± 0	0.79 ± 0.03
A2a	391	0.41 ± 0.05	0.99 ± 0	0.7 ± 0.02
A2b	606	0.47 ± 0.03	0.98 ± 0	0.73 ± 0.02
B1b	693	0.79 ± 0.02	0.98 ± 0	0.89 ± 0.01
C1a	1319	0.4 ± 0.03	0.95 ± 0	0.67 ± 0.01
C1b	923	0.38 ± 0.02	0.97 ± 0	0.67 ± 0.01
D1	767	0.59 ± 0.02	0.98 ± 0	0.78 ± 0.01
D2a	1709	0.37 ± 0.02	0.93 ± 0	0.65 ± 0.01
D2b	960	0.79 ± 0.03	0.98 ± 0	0.88 ± 0.01
D2c	4075	0.76 ± 0.01	0.82 ± 0.01	0.79 ± 0
D2d	1728	0.68 ± 0.02	0.95 ± 0.01	0.82 ± 0.01

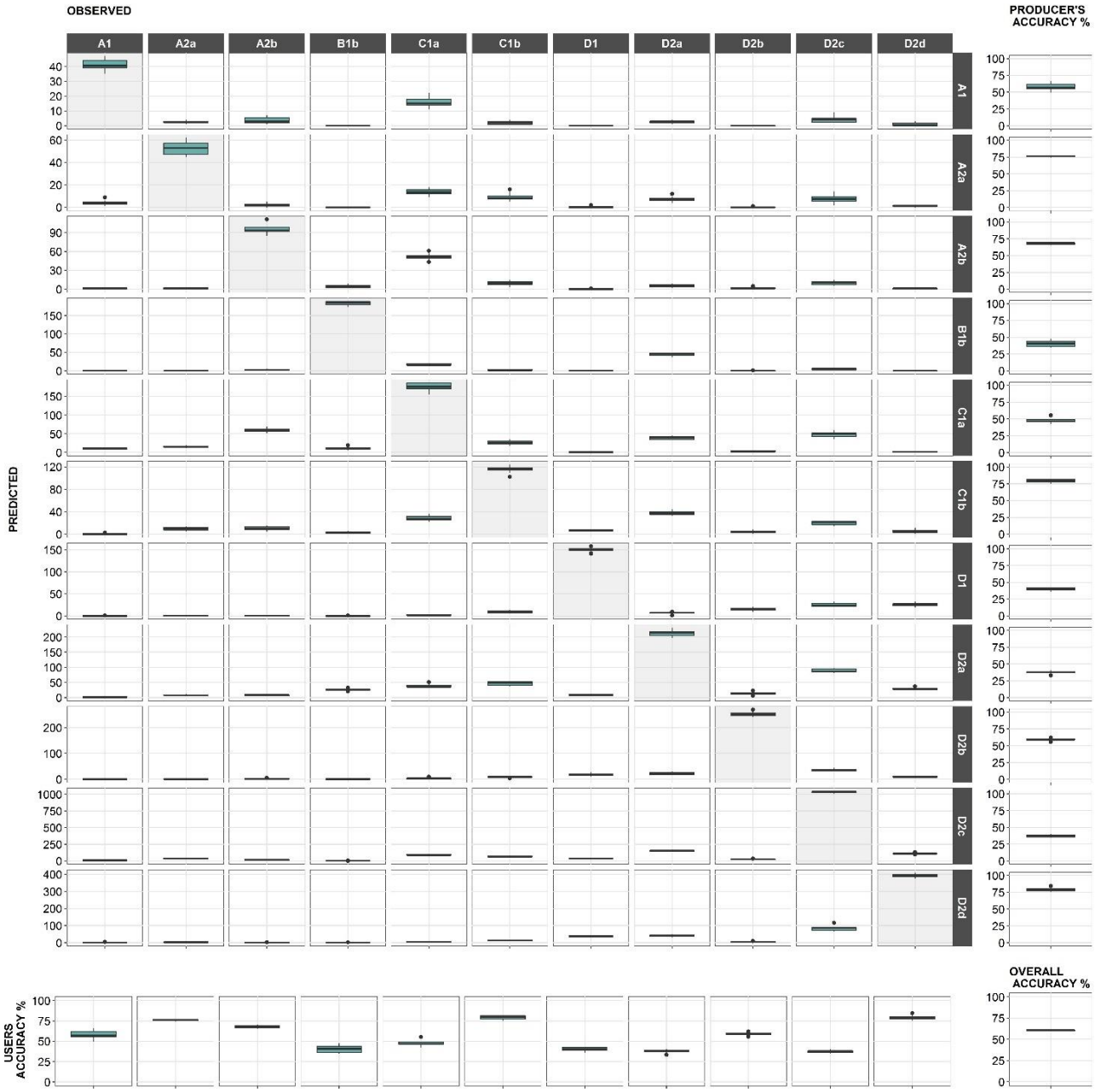


Figure 11. Error matrix for observed versus predicted classes in the faunal assemblage map. Boxplots represent the variation observed from the 10 iterations of the model with average number of observations per class represented in the centre of each cell. Overall accuracy, user's and producer's accuracy are expressed as percentages.

The **Variable contribution** tab contains a plot showing the ranked importance of predictor variables in the Random Forest model (Figure 12). Variable importance is a measure of how much information the predictor variable contributes to the model, in the presence of all of the other variables and is calculated by Monte Carlo permutation inside the model algorithm. The importance is presented as the mean decrease in the Gini coefficient (see https://en.wikipedia.org/wiki/Gini_coefficient) when a variable is permuted. The bars correspond to the mean model variable importance and error bars show standard error.

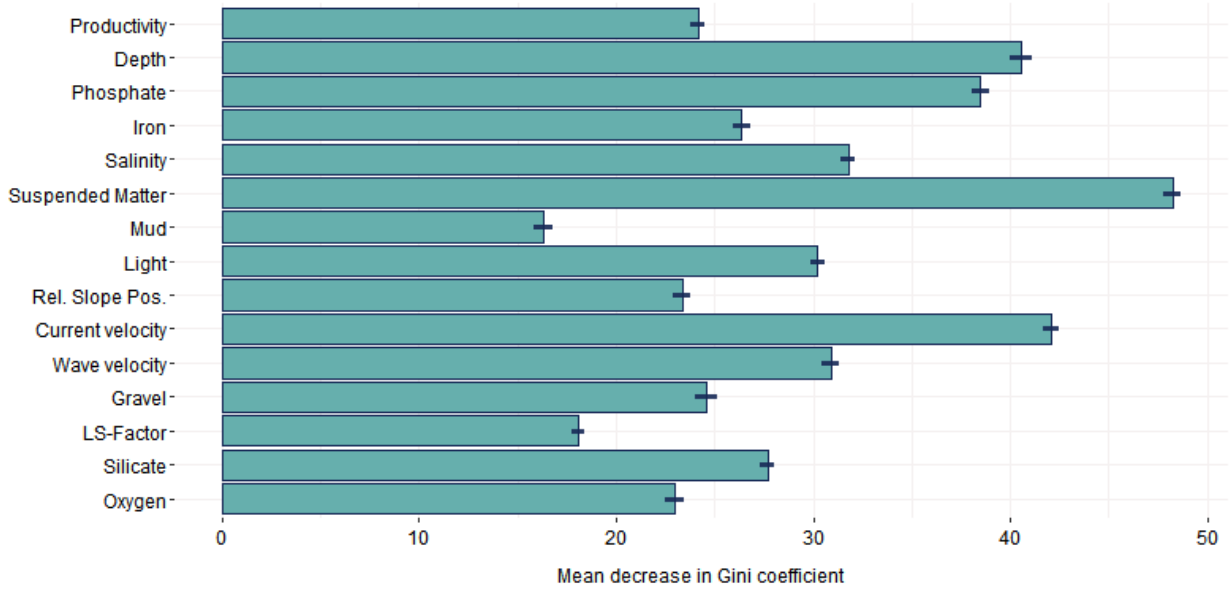


Figure 12. Mean variable importance with error bars showing standard error over 10 random split sample random forest model runs for macrofaunal assemblage (biotope) groups.

The **Partial Response Curves** tab shows plots of the effect each predictor variable has on the probability of each assemblage cluster group class being present (Figure 13). The curves are produced by varying each predictor variable over its range in turn and predicting each class probability at a time with the model. All other predictor variables are held constant at their mean (or in the case of a factor variable, most common) value. The specific condition of holding all other variables at a constant value is important to consider when interpreting the plots. Classification trees incorporate interaction structure, and hence the curves can look different at different values of the other variables. The effect of prevalence on the most likely class can also be seen in the plots, where the most common class has the highest probability. The plot should be interpreted as a general guide of the direction of influence for each predictor variable, but they are not directly indicative of a predictor variable's overall effect. The plots show the response as a loess smooth fitted (based on all cross-validation runs) to predictions made at 100 evenly spaced values along the range of the predictor variable.

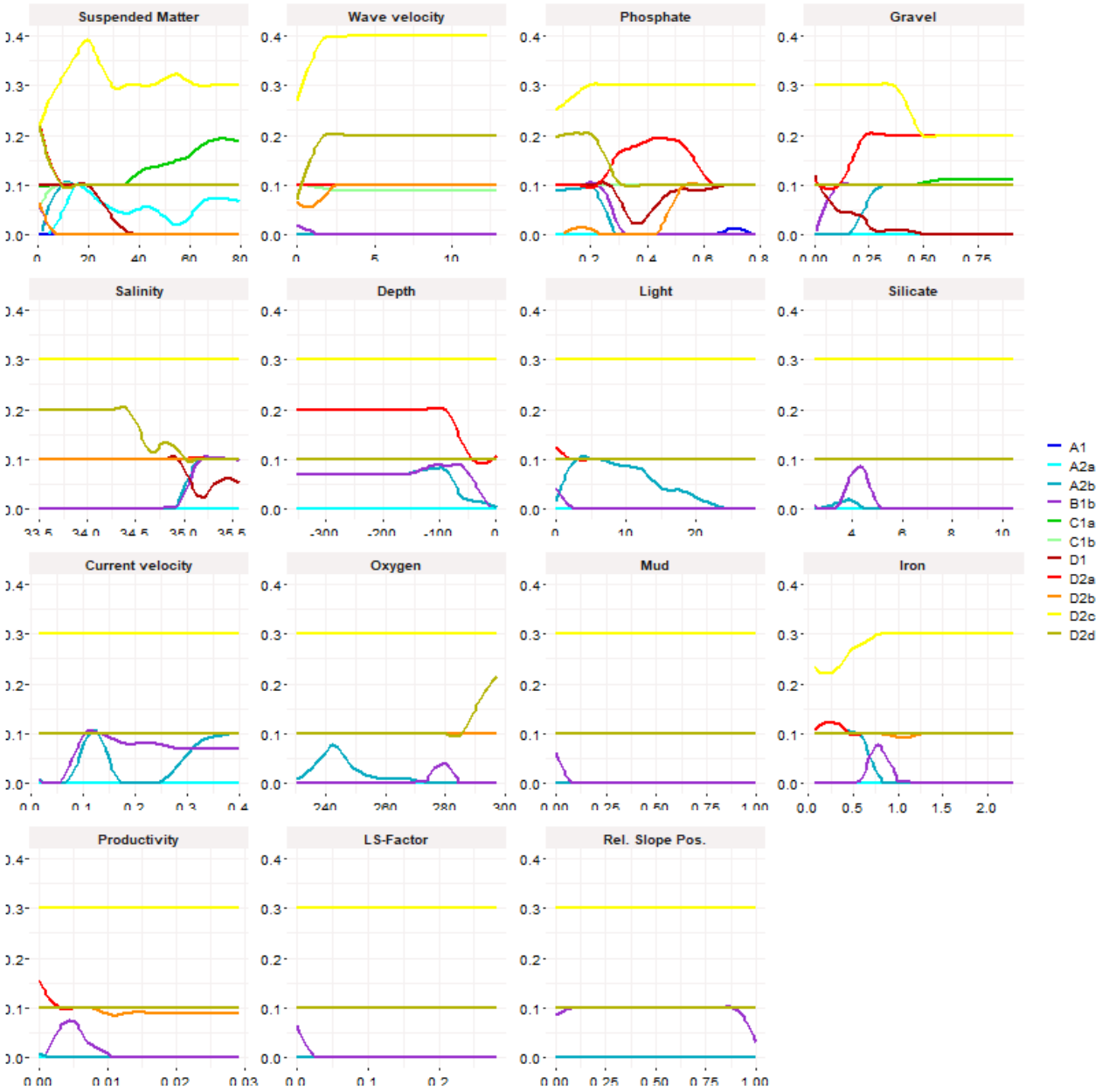


Figure 13. Partial response curves derived from 10 random split sample random forest model runs for macrofaunal assemblage (biotope) groups with a LOESS fit line through all data points. The y-axis is the effect of that variable on the probability of the class I question, the x-axis is the value of the predictor variable.

The **output maps** from the model consist of the predicted distribution of each assemblage cluster class (derived from a majority vote of 10 model runs each indicating the most likely class) and a confidence map layer calculated by multiplying the frequency of the most common class by its average probability over the 10 model runs (Figure 14). In this map the high values represent high confidence.

Both the assemblage (biotope) model and associated confidence layer can be downloaded in vector (shapefile) format from the Cefas Data Hub https://doi.org/10.14466/cefasdatahub.125 (Cooper et al, Cefas, 2022).

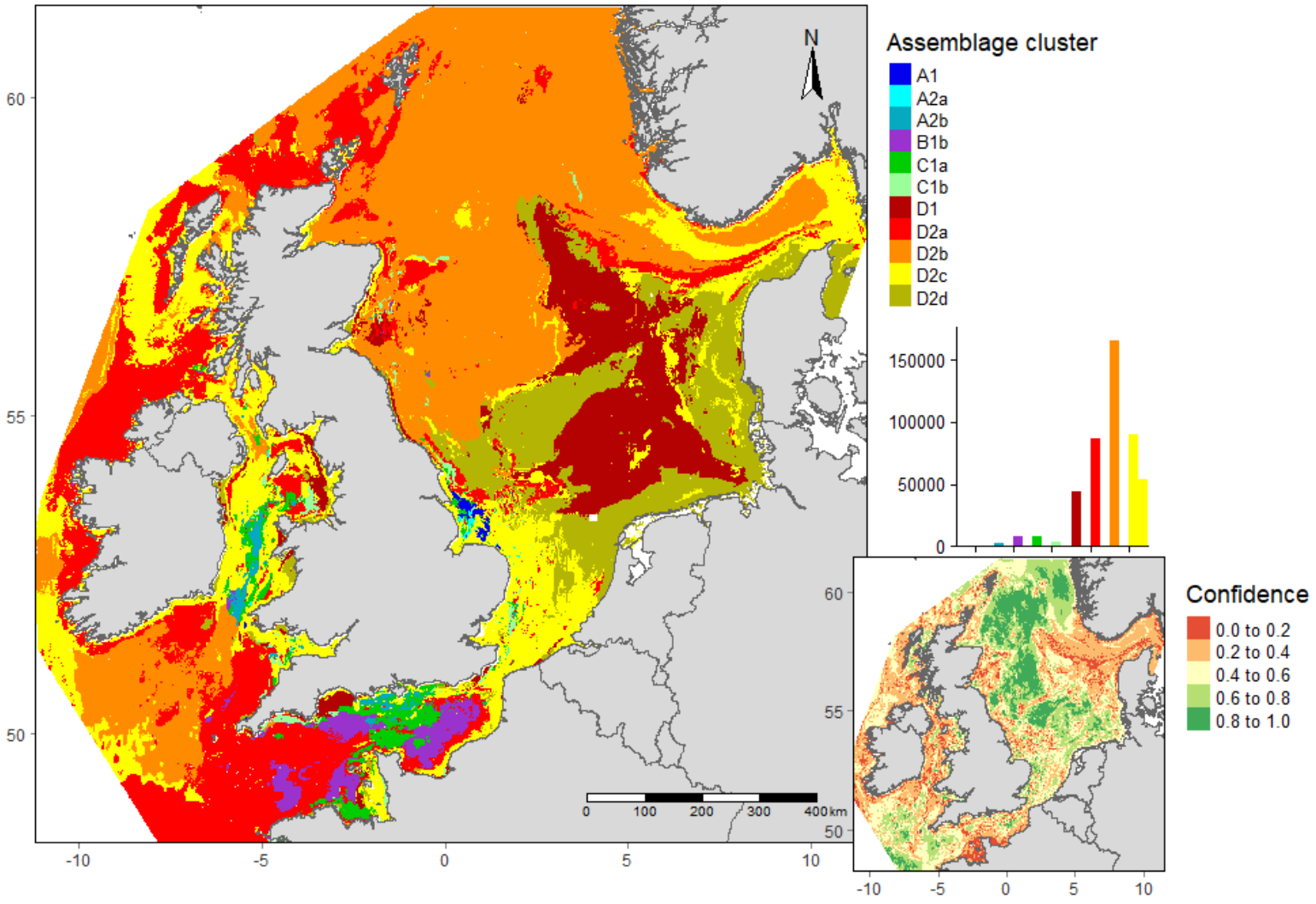


Figure 14. Faunal biotope map and accompanying confidence plot. Note assemblage group A1 does not appear to be found outside UK waters.

4.2. Species Distribution Models (Objective 6)

Species distribution models were created for eight taxa. Here we describe the methodology used for modelling, and the results for one taxon, *Modiolus modiolus*. Models for other species can be viewed in the OneBenthic Layers tool (https://rconnect.cefas.co.uk/onebenthic_layers/).

4.2.1. Taxa Selection

Eight benthic taxa were selected for species distribution modelling. The process for selection involved sending the combined taxon list for OBT and OBGC to VLIZ who identified species of importance to society, using a list available from the World Register of Marine Species (WoRMS) http://www.marinespecies.org/traits/wiki/Traits:Species_Importance_To_Society. Based on the returned list, feedback from project stakeholders and considerations regarding data availability, a number of taxa were taken forward for modelling. These included several ecologically important reef/matt forming species (the polychaete worm *Sabellaria spinulosa*, the large bivalve *Modiolus modiolus* and the tube dwelling shrimp *Ampelisca spinipes*), other species of conservation importance (the long-lived ocean quahog *Arctica islandica*), non-native / naturalised species (the slipper limpet *Crepidula fornicata* and the polychaete worm *Goniadella gracilis*) and species which are important prey items for demersal fish (the thin shelled bivalve *Abra alba* and the tube dwelling trumpet worm *Lagis koreni*). Unfortunately, it was not possible to model epifaunal species from OBT as the database was not ready at the time the modelling work was undertaken.

4.2.2. Data Preparation

Macrofauna abundance data were extracted from the OBGC database. At the time of extraction (October 2021) the database contained taxon abundances observed in 41,197 grab and core samples with an area of 0.1 m^2 sieved over a 0.5 or 1 mm sieve, with fauna identified to the lowest possible taxonomic level. A subset of the core samples came from a source where the sampler area was not included in the metadata, but it is assumed here it was 0.1 m^2. The spatial distribution of the records is shown in Figure 15. Table 6 shows the list of taxa chosen for modelling with summary statistics for all 41,197 records, as well as the number of and summary statistics for non-zero records.

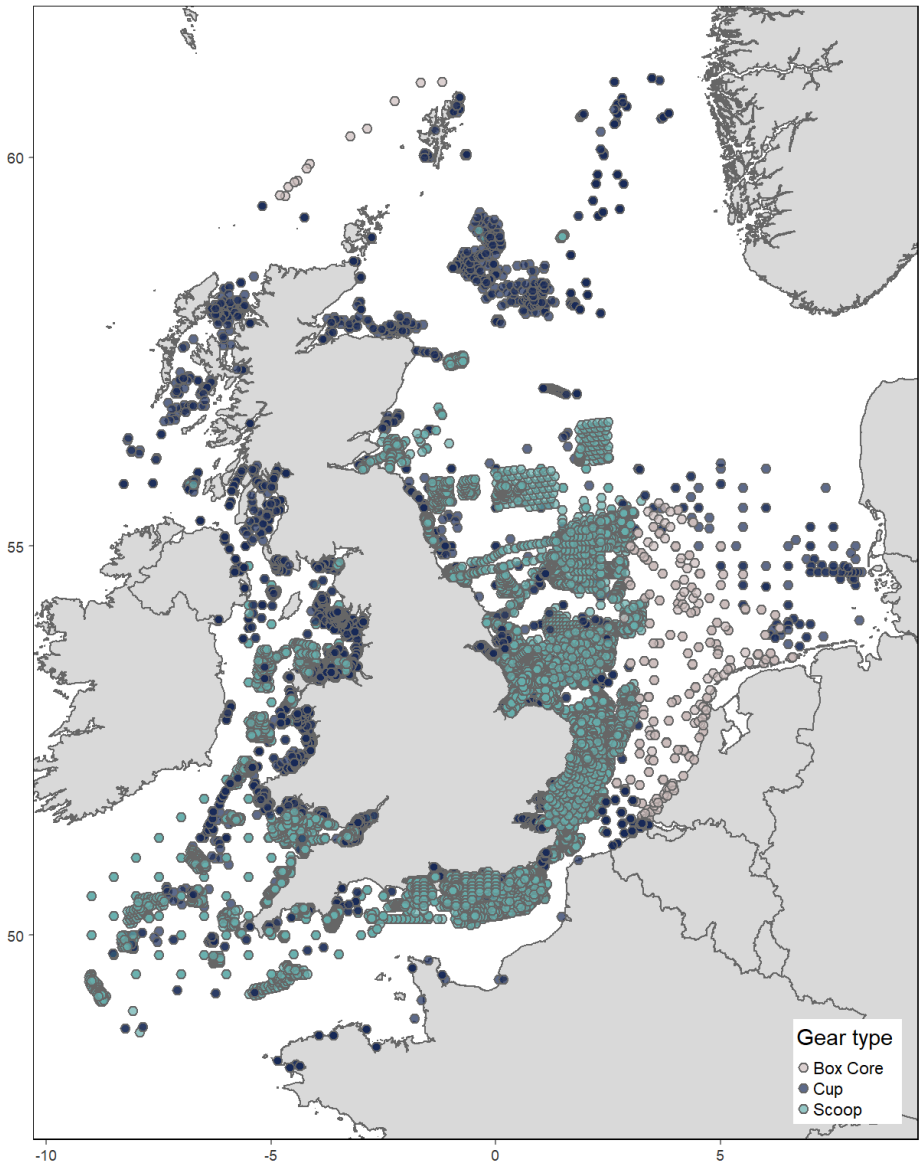


Figure 15. All 0.1 m2 grab and core sampling stations extracted from the OBGC database, which overlap with the environmental predictor variable raster layers used in modelling. Gear type is shown simplified to Box Core, Cup (including Van Veen, Day and Smith-McIntyre grabs) and Scoop (0.1m² Hamon grab).

Table 6. Summary Statistics for modelled taxa.

	N	N>0	Min	Mean	Median	Min>0	Mean>0	Median>0	Max
Abra alba	37378	7063	0	3.80	0	1	20.13	4	1419
Ampelisca spinipes	37378	5500	0	0.66	0	1	4.46	2	346
Arctica islandica	37378	619	0	0.03	0	1	1.99	1	16
Atrina fragilis	37378	13	0	0.00	0	1	1.08	1	2
Crepidula fornicata	37378	2061	0	1.09	0	1	19.80	3	744
Goniadella gracilis	37378	927	0	0.07	0	1	2.88	2	26
Lagis koreni	37378	4764	0	5.19	0	1	40.74	2	9788
Modiolus modiolus	37378	392	0	0.06	0	1	6.16	1	102
Pennatula phosphorea	37378	80	0	0.00	0	1	1.21	1	4
Sabellaria spinulosa	37378	6315	0	11.12	0	1	65.85	5	9596

4.2.3. Environmental Variables

See section 3.1.3.

4.2.4. Modelling Methodology

Data sets for each modelled species were extracted from the full data set, including all abundance observations and a subset of all sampled locations where the species was absent. Zero observations were subset to reduce zero-inflation in the data. The zero records were first filtered to remove all samples with a presence of the target species' genus or family and then randomly sampled to extract 10% of the number of the species abundance records to make up the full model data set. Abundance values were then square-root transformed.

The abundance of the target species was modelled using Random Forests. Regression trees are used for response variables consisting of continuous data, such as the abundance data used here. In regression tree models, predictions are based on averages from all trees. The models were built in R (v.4.0.2, R Development Core Team, 2020), using the 'randomForest' implementation of Random Forests in the randomForest package (Liaw and Wiener, 2002). The models were run on the default settings, using 1000 trees. Preliminary single models using all environmental variables were run first to select best variables and remove variables with high covariance. Variables were dropped from the models based on redundancy (high correlation with a more important variable), or poorly defined relationship with the response variable (based on the response curves).

The final random forest models were built with the subset of predictor variables defined in the previous step. Gear type was included as a factor predictor variable in all models with the grab / core samplers simplified to three types: Box Core, Cup (including Van Veen, Day and Smith-McIntyre grabs) and Scoop (0.1m^2 Hamon grab). Variable importance was determined through a multiple permutation procedure during the model run. Cross-validation via repeated sub-sampling was done to evaluate the robustness of the model estimate and predictions to data sub-setting. The cross-validation was done using 10 random split samples with 75% used to train and 25% to test models. The final model outputs were plotted as the mean of all 10 runs. A confidence map layer consisting of the coefficient of variation (CV, 10 run standard deviation/mean) was also produced.

Model performance was evaluated using both R^2 values and Root Mean Squared Error (RMSE), which for convenience of interpretation was also calculated as proportion of the range of values in the input data. All accuracy statistics are presented as means and standard deviations of the scores from the 10 model runs.

4.2.5. Model Results (*Modiolus modiolus*)

This section illustrates the model diagnostic and map outputs that are shown for each species in the OneBenthic Layers Application, presenting the results for the horse mussel *Modiolus modiolus* as an example. Each model prediction map is accompanied by a corresponding confidence map along with additional information on the map, the input data, model fit and performance statistics and the importance and effects of predictor variables in the model, to help the user interpret the model output.

The **Layer Info** tab provides a short summary of the target of the model and the modelling method used, and a description of the map units to support interpretation, respectively.

The **Input data** tab shows (1) the spatial distribution of the input data (Figure 16), and (2) a histogram illustrating the distribution of the square-root transformed abundance of the modelled species, with an insert summary table stating the number of observations (N) and the minimum, maximum, mean and median values in the data (Figure 17). The *M. modiolus* data set used for the model consists of 431 records with square-root transformed abundance ranging from 0 to 10.1 and a mean of 1.7 and median of 1. The majority of records are from the Irish Sea and Southern North Sea particularly around the Wash, but occasional records are also present in the English Channel, and around the coast of Scotland (Figure 16).

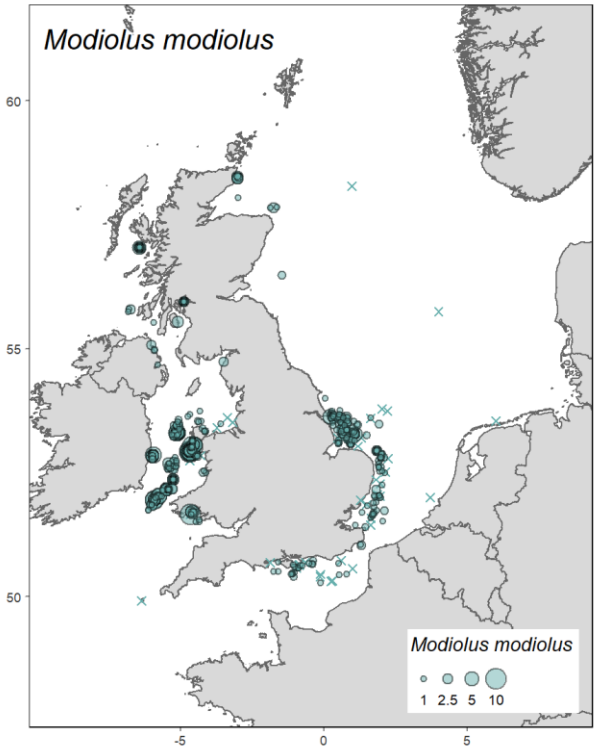


Figure 16. Spatial distribution of Modiolus modiolus square-root transformed abundance observations. The crosses represent zeros.

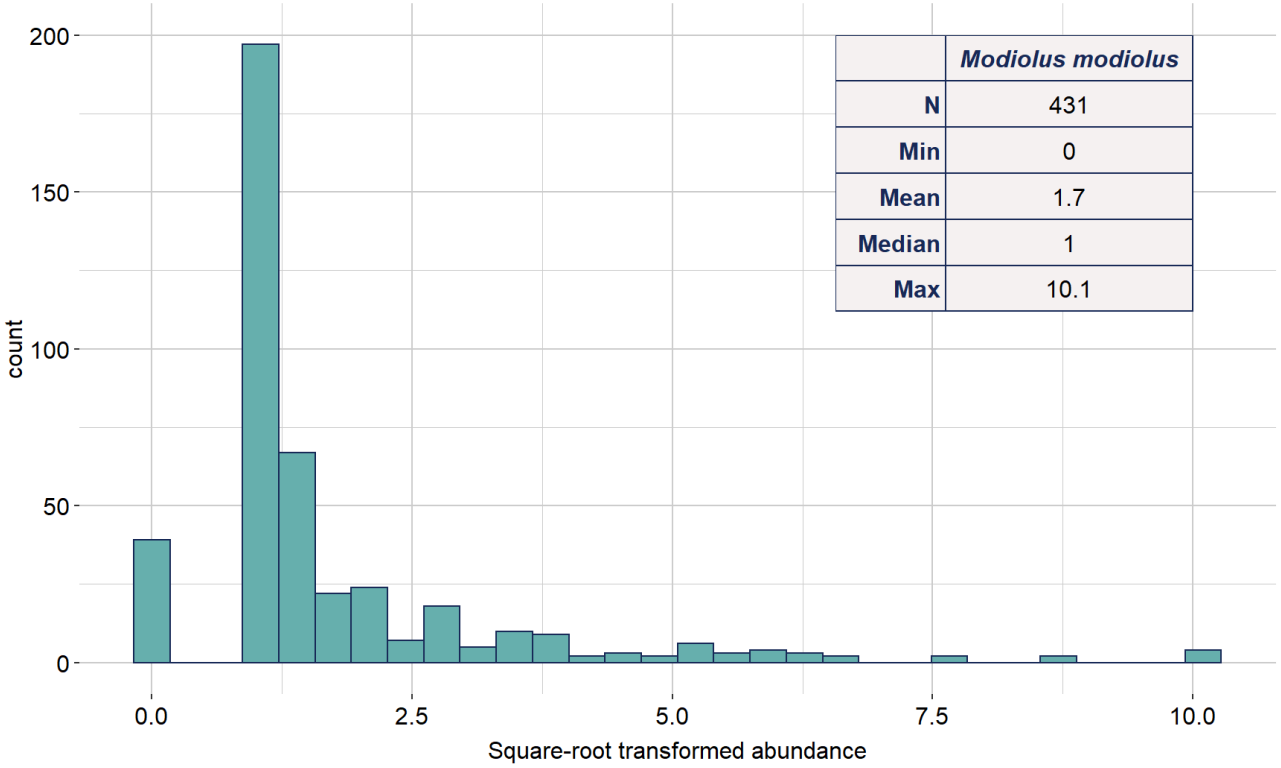


Figure 17. Histogram illustrating the distribution of Modiolus modiolus square-root transformed abundance.

The **Performance** tab shows a table summarising the number of samples used in the training data set, and the mean and standard deviation of the accuracy statistics calculated on the test data sets from the 10 split sample cross-validation runs (Table 7). RMSE and the relative RMSE represent the deviation of predicted values from those observed on absolute terms and in proportion to the range of observed values. The R^2 in turn indicates how well the observed and predicted values correlate. A relative RMSE of < 25% and an R^2 >0.3 should be attained for a model to be judged fairly good. A model with a low RMSE and low R^2 can result from large datasets with unbalanced observations, where many of the small values are correctly predicted but the model performs poorly on larger values, leading to poor correlation between predicted and observed values. Model performance is further visualised in a figure plotting predicted against observed values (Figure 18), where the linear fit through the scatterplot visualises R^2, whilst the spreads of the points around the fit visualises the RMSE. The *M. modiolus* model shows a fair RMSE in relation to the range of abundance values in the data set (12%) and fair correspondence between the observed and predicted values (R^2 >0.3). Whilst high observed values are under-predicted by the model, the trend is correct (Table 7, Figure 18).

Table 7. Mean and standard deviation of model validation statistics over 10 random split sample runs.

Statistic	Mean ± SD
N	324
RMSE	1.2 ± 0.09
Relative RMSE	0.12 ± 0.01
R^2	0.46 ± 0.13

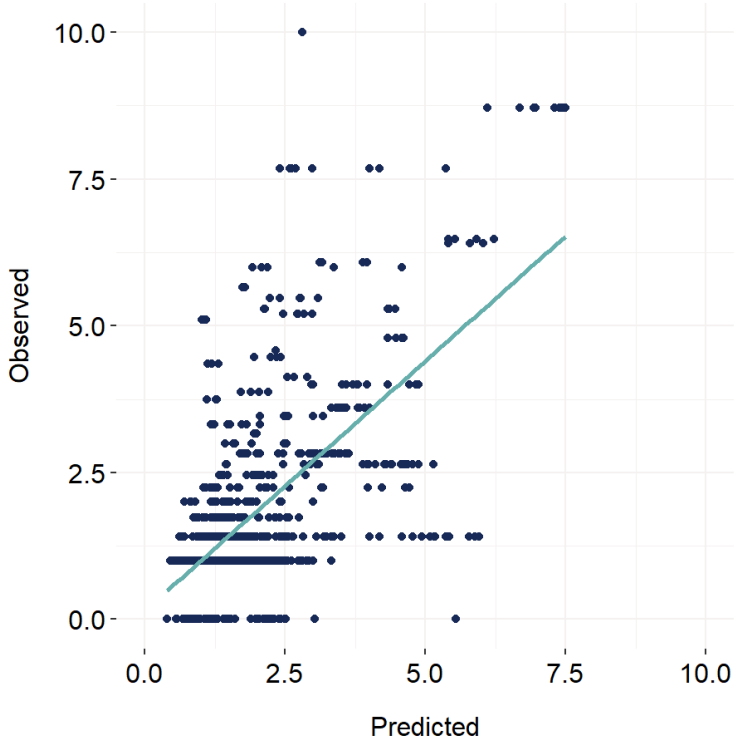


Figure 18. Combined plot of predicted vs. observed values on test data from all 10 model runs with a line of best fit.

The **Variable contribution** tab contains a plot showing the ranked importance of predictor variables in the Random Forest model (Figure 19). Variable importance is a measure of how much information the predictor variable contributes to the model, in the presence of all other variables and is calculated by Monte Carlo permutation inside the model algorithm. The importance is presented as the percent increase in mean square error (% MSE) when a variable is permuted. The most important variable in the *Modiolus modiolus* models was gear type. This may be due to the areas each gear type was most prominently deployed in and warrants further investigation. Of environmental variables, the most important were the proportion of gravel in the substrate, suspended matter in the water column, current velocity and silicate concentration (Figure 19).

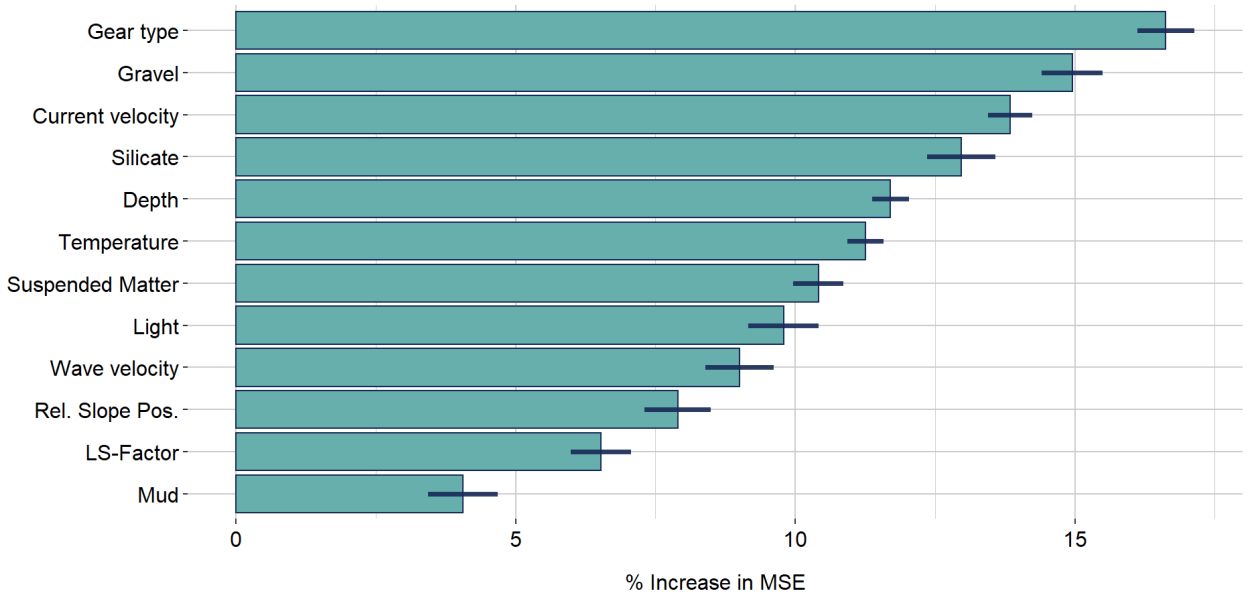


Figure 19. Mean variable importance with error bars showing standard error over 10 random split sample random forest model runs for Modiolus modiolus square-root transformed abundance.

The **Partial Response Curves** tab shows plots of the effect each predictor variable has on the response variable (Figure 20). The curves are produced by varying each predictor variable over its range in turn and predicting with the model, whilst all other variables are held constant at their mean (or in the case of a factor variable, most common class) value. The specific condition of holding all other variables at a constant value is important to consider when interpreting the plots. Regression trees incorporate interaction structure, and hence the curves can look different at different values of the other variables. The plot should be interpreted as a general guide of the direction of influence for each predictor variable, but they are not directly indicative of a predictor variable's overall effect. The plots show the response as a loess smooth fitted to predictions made at 100 evenly spaced values along the range of the predictor variable for each cross validation run (grey lines) and one smooth based on all cross-validation. The response curves for *M. modiolus* show it is predicted to be more abundant in higher current and wave velocity with low gravel content (Figure 20). The response plot for gear type also clearly shows higher abundance of *M. modiolus* is caught by the cup type sampling gears.

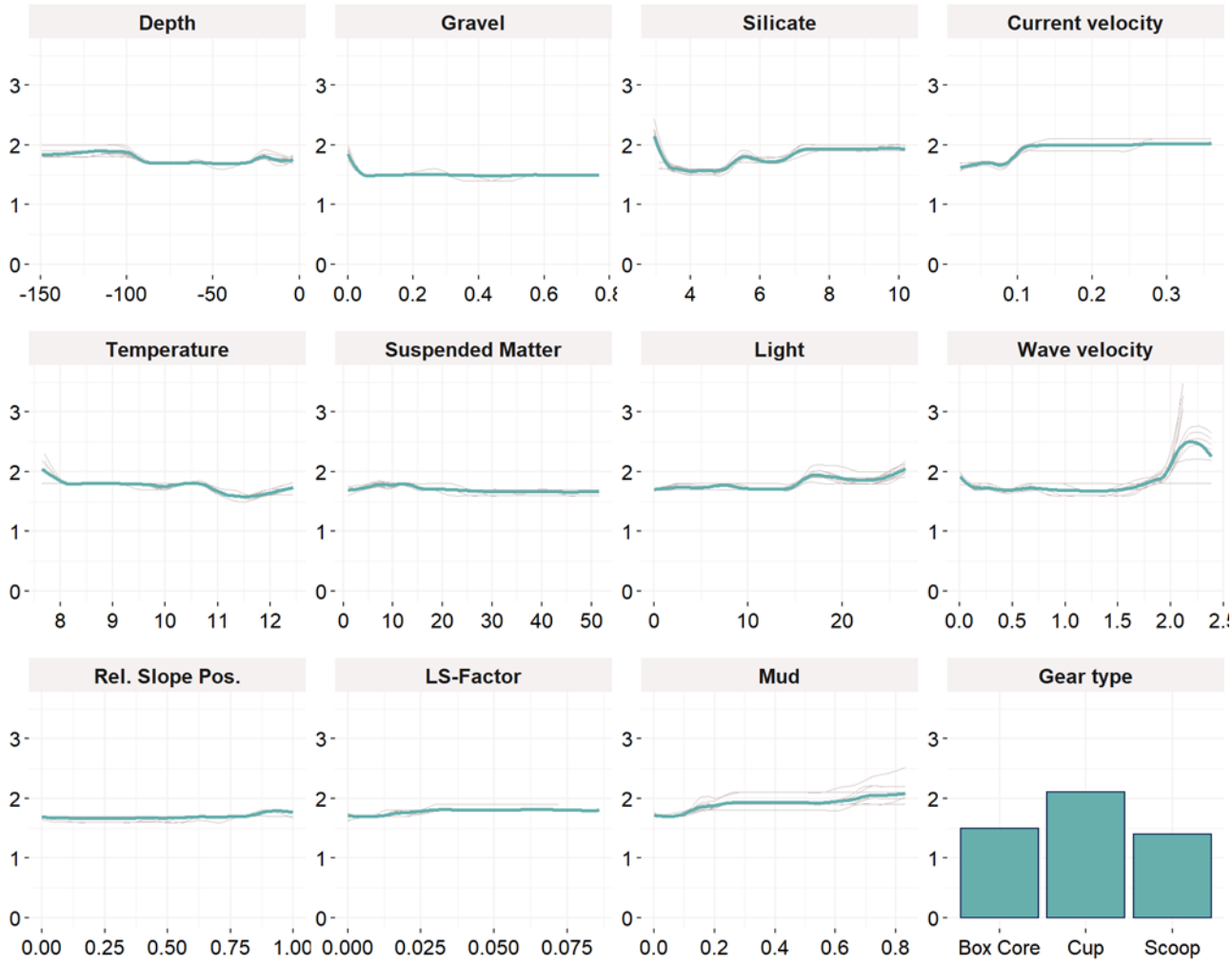


Figure 20. Partial response curves (grey) derived from 10 random split sample random forest model runs for Modiolus modiolus square-root transformed abundance with a LOESS fit line through all data points (green line).

The **output maps** from the model consist of the predicted abundance over the study area as mean square-root transformed counts and the coefficient of variation (CV), both calculated over the ten split sample cross-validation runs (Figure 21). The model predicts patches of high abundance mainly in the Irish Sea and the Northern North Sea. High abundance predicted in the deep waters on the shelf break and in the Norwegian Trench are expected to be model artefacts as there is no data supporting the predicted distribution in those areas (Figure 21). The coefficient of variation (CV) of predictions from the 10 cross-validation runs stays mostly below 25% with patches of highly variable predictions (up to 60% of the mean prediction).

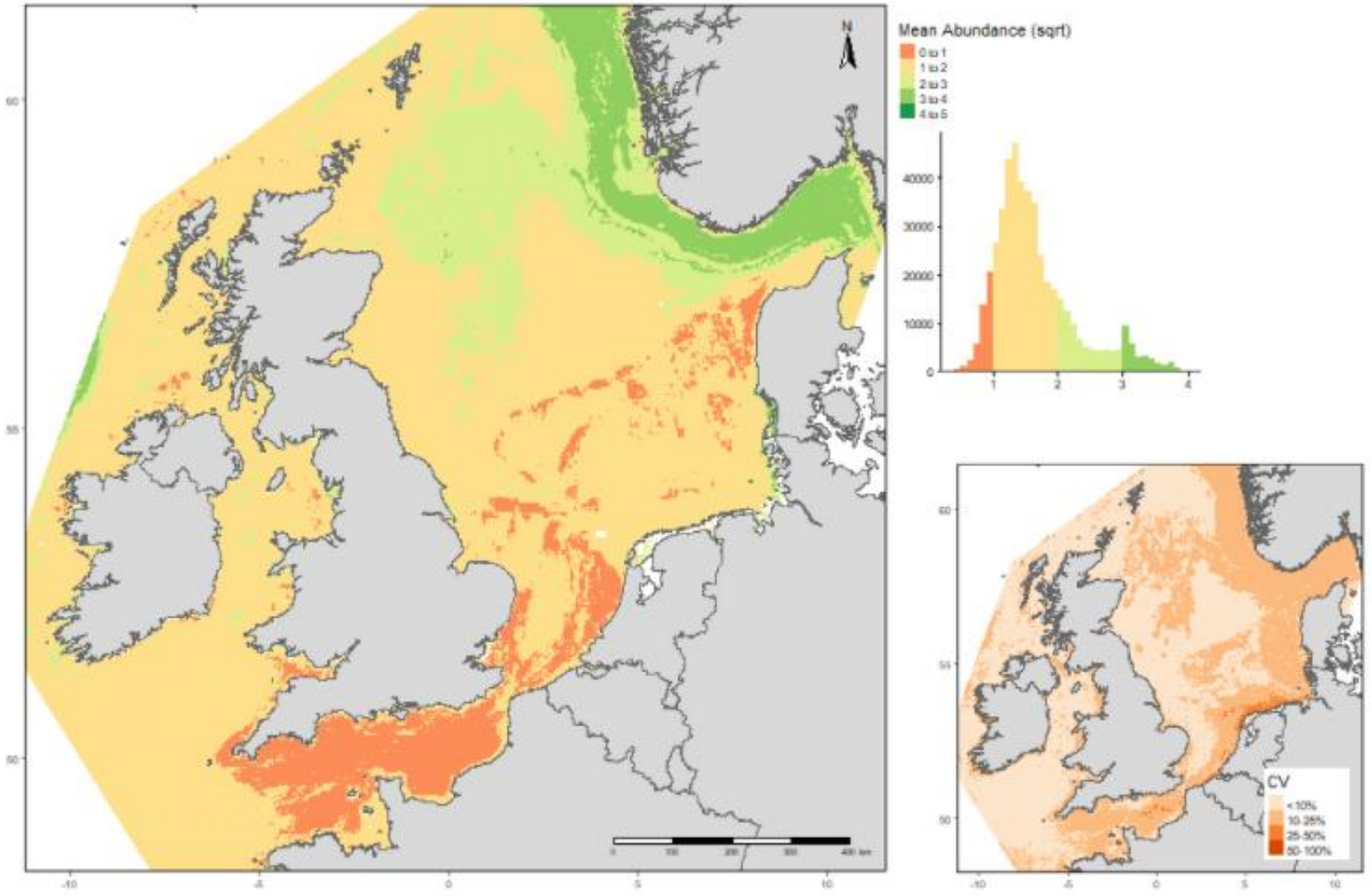


Figure 21. Figure 10 Mean square-root transformed abundance from 10 random split sample random forest model runs for Modiolus modiolus with the coefficient of variation (CV) to illustrate model confidence. Note that high values of CV indicate high confidence in the model.

5. Conclusions and Recommendations

The aim of this project was to lay foundations to improve understanding of benthic biodiversity across the Greater North Sea (and beyond). This is important to support decisions around continued expansion of the OW industry, including those related to NG, NNL and NID. This was achieved through the further development of the OB big data infrastructure, and use of an expanded transboundary dataset. As noted by Runting et al. (2020) and demonstrated through various OB initiatives (https://sway.office.com/HM5VkWvBoZ86atYP?ref=Link), big data approaches offer new insights, helping to improve sustainability.

Two international partnerships were forged during this project. The first of these was with VLIZ, allowing selected datasets from the EurOBIS data repository to be incorporated into OB, expanding geographic scope and relevance of data products, and helping to providing vital context. Work is now underway to help establish a data flow from OneBenthic to OBIS UK (hosted & managed at Marine Biological Association), EurOBIS, and from there on to global biodiversity initiatives including OBIS (https://obis.org/) and GBIF (https://www.gbif.org/) (see Figure 22). Thus OB is an important part of European and global efforts to improve data sharing and accessibility, particularly for UK industry data. The second partnership was with the RNS initiative, providing a valuable link to the OW industry, helping to ensure the relevance of project outputs, particularly in the context of NID. Furthermore, the involvement of UK statutory nature conservation bodies (Natural England and the Joint Nature Conservation Committee) on the project steering group provided a helpful perspective from the regulatory viewpoint.

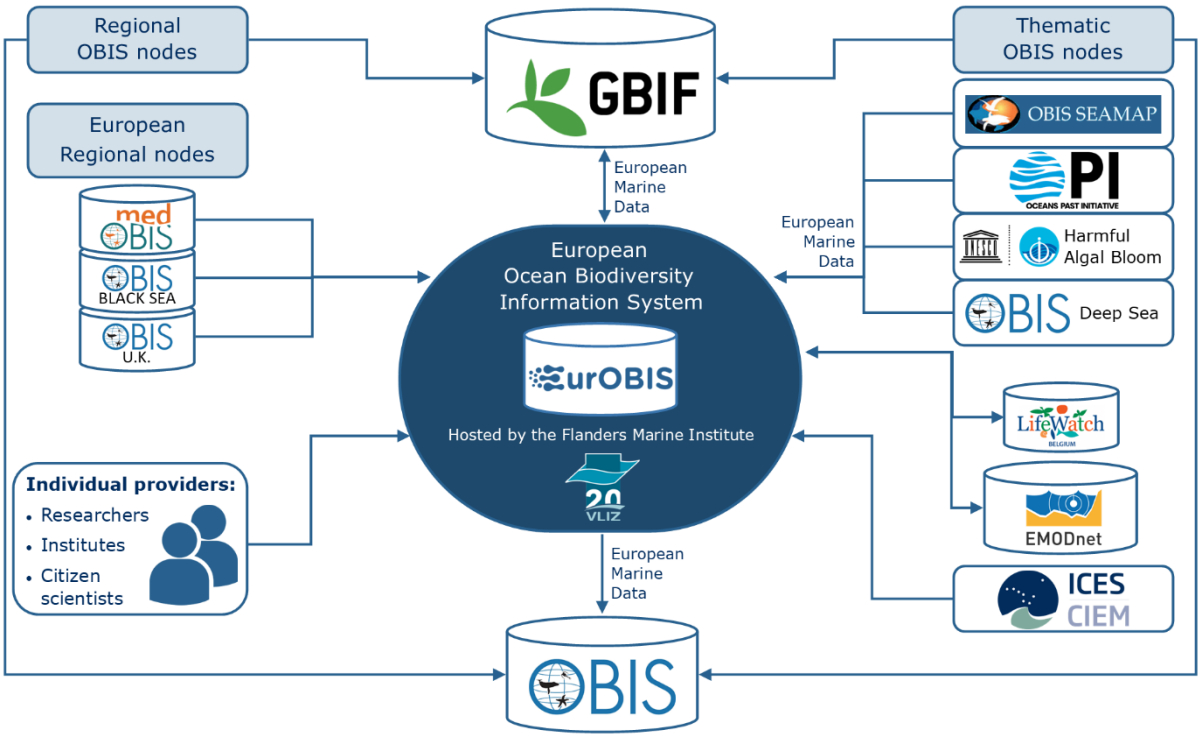


Figure 22. EurOBIS – data flows & connections at several levels (Figure supplied by Flanders Marine Institute).

This project has enhanced the existing OB infrastructure, through development of a trawl sample database, expansion of the existing OB dataset (addition of 7,456 samples from grabs, cores and trawls from countries including France, Belgium, The Netherlands, Germany, Denmark and Norway), and development of web applications for sharing data (https://rconnect.cefas.co.uk/onebenthic_dataextractiontrawl/) and data products (https://rconnect.cefas.co.uk/onebenthic_layers/). The new trawl sample database contains 1,528 samples, and access to these data, and those available in the OBGC database (45,519 samples), can help developers by potentially reducing the need for collection of new characterisation samples, further contributing to Net Zero efforts. In addition, big data can help facilitate new ways of monitoring that both improve understanding and help lower costs (see aggregate industry example from Cooper and Barry, 2017).

The biodiversity layers generated in this project include biotopes (macrofaunal assemblages), one important aspect of benthic biodiversity, and SDMs for a variety of taxa. These maps, and others currently in production, provide, for the first time, high-resolution broadscale maps of benthic biodiversity that can help reduce reliance on physical habitat proxies. Biodiversity maps can assist the OW sector by highlighting areas of high ecological importance, thereby reducing possible consenting risk, and species which might benefit locally from NID initiatives.

The biotope map produced in this study is broadly similar to that produced by Cooper et al. (2019) for UK waters, although there are some local differences. These differences stem from the larger

number of samples and associated taxa involved in the present study, the reclustering of data, and the greater number of raster predictor layers used in modelling (15 versus 9). Furthermore, the final model outputs presented here represent a consensus view from multiple runs, providing greater confidence in the results. The development of R code and database infrastructure means that it is now a relatively simple process to generate further biodiversity models (e.g. new taxa or biodiversity metrics), or to update existing models as new data become available. For instance, the expanded dataset from this study is already being used, in other projects, to generate further benthic biodiversity layers (e.g. taxon richness; abundance; taxonomic distinctness (Clarke & Warwick, 1998); Pileou's evenness; alpha, beta and gamma diversity (Whittaker, 1972)) and for biological traits (see Bolam et al. 2016) which better relate to functional properties of the benthos. These new layers, which will also be made available via the OneBenthic Layers tool (https://rconnect.cefas.co.uk/onebenthic_layers/), can complement existing approaches such as EUNIS (Davies, Moss, & Hill, 2004), providing an enhanced understanding of benthic biodiversity. Furthermore, additional benthic data will be collected in the coming years under the new OWEC project, POSEIDON (https://www.thecrownestate.co.uk/en-gb/media-and-insights/news/the-crown-estate-invests-over-12million-in-new-research-to-help-protect-the-uk-marine-environment/) which *inter alia* aims to address gaps in existing benthic sample coverage, and to update spatial models for biodiversity.

The North Sea being is one of the first regions in the world to develop OWFs at scale. For this reason, the development of scientific evidence and tools to understand the effects on biodiversity are likely to be of global interest, and findings can help ensure the successful roll out of OWFs in other locations. For this reason it is important that steps are taken now to ensure that change in benthic biodiversity can be effectively assessed. This study makes an important contribution in this regard.

A number of project recommendations are made:

Project Recommendations

- Infrastructure and processes should be maintained to facilitate continued flow of data into OneBenthic from sources including industry and the appropriate Data Archiving Centre (DASSH), as well as enabling the flow of data internationally to EurOBIS via DASSH. The functionality of OneBenthic should be built upon to enhance its analytical capabilities, enabling big data approaches which have huge potential to improve sustainability by providing new scientific insights.
- The data layers produced as part of this project complement existing approaches for characterising offshore wind development areas by providing an enhanced understanding of benthic biodiversity. "Nature knows no boundaries", and within environments such as the North Sea, biodiversity should be considered at an ecologically relevant scale using datasets that don't stop at national borders. We recommend future consideration of international data set development and use to drive this agenda forward.
- A wealth of data collected via industry and other programmes is now available that hasn't yet been incorporated into the current benthic mapping systems used in UK (or EU) decision making. This project demonstrates the potential to harness these data to provide new insights into benthic biodiversity. We recommend that further work is required to consider how these data and new layers could be incorporated into the decision-making process. In the UK, the OWEC POSEIDON and Defra's marine Natural Capital and Ecosystem Assessment (mNCEA) projects offer opportunities to address this.
- It is recommended that consideration should be made of the Rich North Sea's approach to Nature Inclusive Design and how this could be adopted in UK waters to contribute to biodiversity enhancement. The layers produced as part of this project can be used to help inform spatial decisions on where Nature Inclusive Design initiatives might be most appropriate.
- Outputs from this project can help identify areas of seabed that are less well characterised (e.g. Celtic Sea) and hence where future sampling can benefit understanding. This is a key goal of the OWEC POSEIDON project (see https://www.thecrownestate.co.uk/en-gb/media-and-insights/news/the-crown-estate-invests-over-12million-in-new-research-to-help-protect-the-uk-marine-environment/).

6. References

ABPmer, (2019). Marine Net Gain, Moving towards a practical framework and metric for the marine environment. ABPmer for White Paper, July 2019.

Akhtar, N., Geyer, B., Rockel, B., Sommer, P.S. & Schrum, C. 2021. Accelerating deployment of offshore wind energy alter wind climate and reduce future power generation potentials. Scientific Reports, 11:11826 https://doi.org/10.1038/s41598-021-91283-3

Assis, J., Tyberghein, L., Bosh, S., Verbruggen, H., Serrão, E. A., & De Clerck, O. 2017. Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling. Global Ecology and Biogeography.

Barnosky, A.D., Matzke, N., Tomiya, S. et al (2011). Has the Earth's sixth mass extinction already arrived? Nature 471:51–57. https://doi.org/10.1038/nature09678

Böhner, J. & Selige, T. 2006. Spatial prediction of soil attributes using terrain analysis and climate regionalisation. SAGA - Analysis and Modelling Applications. 115. 13-27.

Bolam, S.G., McIlwaine, P., Garcia, C. 2016. Application of biological traits to further our understanding of the impacts of dredged material disposal on benthic assemblages. Marine Pollution Bulletin 105(1):180-92. doi: 10.1016/j.marpolbul.2016.02.031.

Breiman, L. 2001. Random Forests. Machine Learning, 45, 5–32. https:// doi.org/10.1023/A:1010933404324

Cardinale, B.J., Duffy, J.E., Andrew Gonzalez,David U. Hooper, Charles Perrings, Patrick Venail,Anita Narwani, Georgina M. Mace, David Tilman,David A.Wardle, Ann P. Kinzig, Gretchen C. Daily,Michel Loreau, James B. Grace, Anne Larigauderie,Diane S. Srivastava & Shahid Naeem. Biodiversity loss and its impact on humanity. Nature 486, 59–67 (2012); doi:10.1038/nature11148

Clarke, K.R. & Warwick, R.M. 1998. A taxonomic distinct-ness index and its statistical properties. Journal of Applied Ecology, 35, 523 –531.

Coates, D.A., Kaapasakali, D-A., Vincx, M., Vanaverbeke, J. 2016. Short-term effects of fishery exclusion in offshore wind farms on macrofaunal communities in the Belgian part of the North Sea. Fisheries Research 179: 131-138. https://doi.org/10.1016/j.fishres.2016.02.019

Cochrane, S.K.J., Andersen, J.H., Berg, T., Blanchet, H., Borja, A., Carstensen, J., Elliott, M., Hummel, H., Niquil, N., Renaud, P.E. 2016. What Is Marine Biodiversity? Towards Common Concepts and Their Implications for Assessing Biodiversity Status. Frontiers in Marine Science 3: 248. https://www.frontiersin.org/article/10.3389/fmars.2016.00248

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J. 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geoscientific Model Development Discussions. 8. 2271-2312. 10.5194/gmdd-8-2271-2015.

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

Cooper K.M., Barry, J. 2020. A new machine learning approach to seabed biotope classification. Ocean and Coastal Management 198, 105361. https://doi.org/10.1016/j.ocecoaman.2020.105361

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

Cooper K.M., Barry, J. 2020. A new machine learning approach to seabed biotope classification. Ocean and Coastal Management 198, 105361. https://doi.org/10.1016/j.ocecoaman.2020.105361

Cooper et al, Cefas (2022). Biotope (macrofaunal assemblage) map and associated confidence layer based on grab and core data from 1976 to 2020. Cefas, UK. V1. doi: https://doi.org/10.14466/CefasDataHub.125

Cutler, D., Edwards, T., Beard, K., Cutler, A., Hess, K., Gibson, J., Lawler, J. 2007. Random Forests for Classification in Ecology. Ecology. 88. 2783-92. 10.1890/07-0539.1.

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. Retrieved from http://www.eea.europa.eu/themes/biodiversity/eunis/eunis-habitat-classification#tab-documents

Desmet, P.J.J. & Govers, G. 1996. A GIS procedure for automatically calculating the USLE LS factor on topographically complex landscape units. Journal of Soil and Water Conservation. 51. 427-433.

Energy & Climate Intelligence Unit. 2021. Net Zero Scorecard. https://eciu.net/netzerotracker December 2021.

Environment Act *2021* c. 30 (UK).

European Commission, 2020. Communication from the commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. EU Biodiversity Strategy for 2030 Bringing nature back into our lives. COM/2020/380 final

Gamfeldt, L.; Lefcheck, J.S.; Byrnes, J.E.K.; Cardinale, B.J.; and Duffy, J.E., 2015. Marine biodiversity and ecosystem functioning: what's known and what's next? Oikos, 124(3), 252-265. 10.1111/oik.01549

Hermans, A., Bos, O. and Prusina, I. (2020). Nature-Inclusive Design: a catalogue for offshore wind infrastructure. 10.13140/RG.2.2.10942.02882.

Herzog, T. World Greenhouse Gas Emissions in 2005. WRI Working Paper. World Resources Institute. Available online at http://www.wri.org/publication/navigating-the-numbers.

HM Government (2018). A Green Future: Our 25 Year Plan to Improve the Environment. 151pp.

Liaw, A. & Wiener, M. 2001. Classification and Regression by Random Forest. Forest. 2/3. https://cogns.northwestern.edu/cbmg/LiawAndWiener2002.pdf

MacQueen, J. Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, eds L. M. Le Cam & J. Neyman, 1, pp. 281–297 Berkeley, CA: University of California Press (1967).

Mitchell, P.J., Aldridge, J. and Deising, M. 2019. Legacy Data: How Decades of Seabed Sampling can Produce Robust Predictions and Versatile Products. Geosciences 9(4), 182; https://doi.org/10.3390/geosciences9040182

Mitchell, P.J., Downie, A.-L., Diesing, M. How good is my map? 2018. A tool for semi-automated thematic mapping and spatially explicit confidence assessment. Env. Model. Softw. 108, 111–122. https://doi.org/10.1016/j.envsoft.2018.07.014

Naeem, S., Chazdon, R., Duffy, J. E., Prager, C. and Worm, B. 2016. Biodiversity and human well-being: an essential link for sustainable development. Proceedings of the Royal Society B. 283: 20162091. https://doi.org/10.1098/rspb.2016.2091

National Infrastructure Commission (2021). Natural capital and environmental net gain: A discussion paper. 31pp. https://nic.org.uk/app/uploads/Updated-Natural-Capital-Paper-Web-Version-Feb-2021.pdf

Oksanen, J. et al. vegan: community ecology package. R package version 2.3–4 https://CRAN.R-project.org/package=vegan (2016).

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M. and Rigol-Sanchez, J.P. 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing 67, 93–104.

Runting, R.K., Phinn, S., Xie, Z., Venter, O. & Watson, J.E.M. Opportunities for big data in conservation and sustainability. Nature Communications 11, 2003. https://doi.org/10.1038/s41467-020-15870-0

Seaman, W., Lindberg, W.J., 2009. Artificial Reefs (Ed: Steele, J.H.) in Encyclopedia of Ocean Sciences (Second Edition), Academic Press, pp 226-233, ISBN 9780123744739.https://doi.org/10.1016/B978-012374473-9.00668-8.

Strobl, C., Malley, J. & Tutz, G. 2009. An Introduction to Recursive Partitioning: Rationale, Application, and Characteristics of Classification and Regression Trees, Bagging, and Random Forests. Psychological methods. 14. 323-48. 10.1037/a0016973.

Thompson, M.S.A., Couce, E., Webb, T., Grace, M., Cooper, K.M., Schratzberger, M. 2020. What's hot and what's not: Making sense of biodiversity 'hotspots'. Global Change Biology 27, 3: 521-535. https://doi.org/10.1111/gcb.15443

Thorndike, R. L. (1953). Who belongs in the Family? Psychometrika, 18: 267–276. https://doi.org/10.1007/BF02289263

WindEurope. 2019. Offshore wind in Europe: Key trends and statistics 2019.

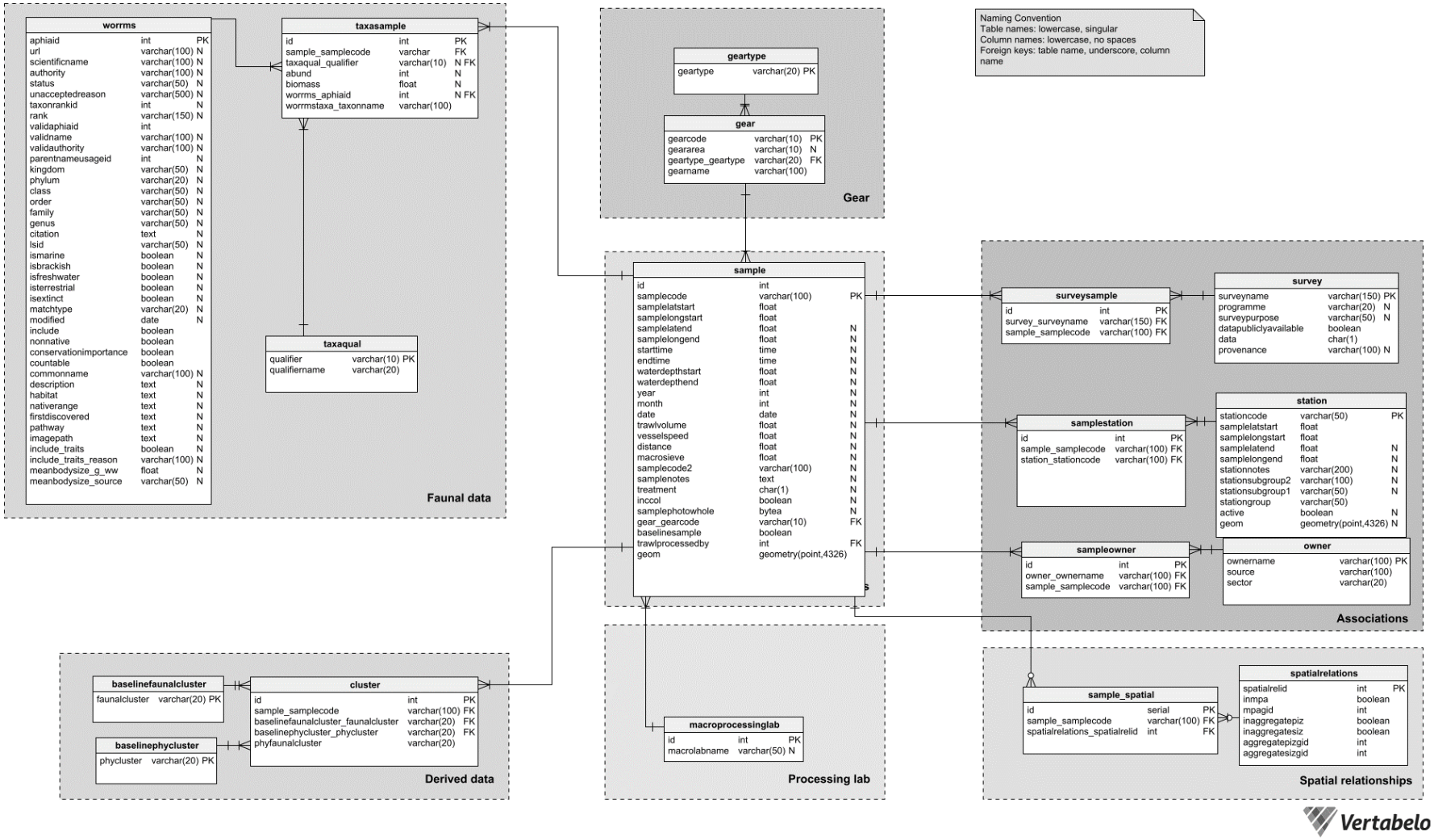
Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F., De Clerck, O. 2012. Bio-ORACLE: A global environmental dataset for marine species distribution modelling. Global Ecology and Biogeography, 21, 272–281.

van der Reijden, K.J., Koop, L., Mestdagh, S., Snellen, M., Herman, P.M.J., Olff, H., Govers L.L. 2021. Conservation Implications of Sabellaria spinulosa Reef Patches in a Dynamic Sandy-Bottom Environment. Frontiers in Marine Science 8, 362pp. https://www.frontiersin.org/article/10.3389/fmars.2021.642659

Whittaker, R. H. (1972). Evolution and Measurement of Species Diversity. Taxon, 21(2/3), 213–251. https://doi.org/10.2307/1218190

7. Appendices

Appendix 1: Database model for OBT



Appendix 2: Details for the data uploaded to OBT.

Source	Survey Name	Gear*	Year	No. samples
Cefas	A0908 Hastings Shingle Bank 2m beam trawl survey_2000	2BT	2000	12
Cefas	A0908 Shoreham 2m beam trawl survey_2000	2BT	2000	18
Cefas	AE0916 EEC 2004	2BT	2004	6
Cefas	Area 473 - Grenwich Light East_2002	2BT	2002	8
Cefas	Areas 458 & 464 West Bassurelle_2002	2BT	2002	6
Cefas	BEEMS-Hinkley power station studies - subtidal_2008	2BT	2008	54
Cefas	BEEMS-Hinkley power station studies - subtidal_2009	2BT	2009	17
Cefas	BEEMS-Hinkley power station studies - subtidal_2010	2BT	2010	76
Cefas	BEEMS-Hinkley power station studies - subtidal_2011	2BT	2011	34
Cefas	BEEMS Bradwell power station studies- subtidal_2008	2BT	2008	57
Cefas	BEEMS sizewell power station studies -subtidal_2008	2BT	2008	75
Cefas	BEEMS sizewell power station studies -subtidal_2009	2BT	2009	18
Cefas	BEEMS sizewell power station studies -subtidal_2010	2BT	2010	19
Cefas	BEEMS sizewell power station studies -subtidal_2011	2BT	2011	24
Cefas	BEEMS sizewell power station studies -subtidal_2012	2BT	2012	24
Cefas	BEEMS sizewell power station studies -subtidal_2014	2BT	2014	17
Cefas	C1103 Beam Trawl data Area 222 - Area 408 and Hastings Shingle Bank_2002	2BT	2002	43
Cefas	C2228 Area 222 2004 BT	2BT	2004	12
Cefas	C2228 Area 408 2004 BT	2BT	2004	12
Cefas	C2228 Hastings Area X 2004 BT	2BT	2004	4
Cefas	C2228 Hastings Area Y 2004 BT	2BT	2004	12
Cefas	C2474 Area 107_2006	2BT	2006	2
Cefas	C2474 Hastings Shingle Bank_2006	2BT	2006	1
Cefas	Celtic Sea epifauna surveys 2000-2009_2000	2BT	2000	61
Cefas	Celtic Sea epifauna surveys 2000-2009_2001	2BT	2001	1
Cefas	Celtic Sea epifauna surveys 2000-2009_2009	2BT	2009	92
Cefas	CSEMP_2001	2BT	2001	4
Cefas	East Coast REC_2009	2BT	2009	128
Cefas	Eastern English Channel_2002	2BT	2002	16
Cefas	EEC MEPF Beam Trawl samples 2006_2006	2BT	2006	32
Cefas	EEC MEPF Beam Trawl samples_2005	2BT	2005	40
Cefas	International Bottom Trawl Survey 2000	2BT	2000	270
Cefas	North Norfolk SandBanks Survey 2016	2BT	2016	66
MDE	Rampion Offshore Wind Farm Zone Benthic Characterisation Survey 2011	2BT	2011	18
Cefas	Regional Environmental Characterisation of the Isle of Wight Sea Area_2007	2BT	2007	23
Cefas	Rehab project C1103 beam trawls_2003	2BT	2003	40
Cefas	River Crouch Epifauna studies - corrected to 250m tow length_1997	2BT	1997	21
Cefas	River Crouch Epifauna Studies 1987 corrected to 250m tow	2BT	1987	17
Cefas	River Crouch Epifauna Studies 2005	2BT	2005	21
Cefas	River Crouch Epifauna studies 2005 corrected to 250m tow length	2BT	2005	21
Cefas	River Crouch Epifaunal studies 1988 corrected to 250m tow length	2BT	1988	21
Cefas	River Crouch Epifaunal studies 1989 corrected to 250m tow length	2BT	1989	21
Cefas	River crouch epifaunal studies 1992 corrected to 250m tow length	2BT	1992	21
MDE	Round 3 Hornsea Offshore Wind Farm Subzone 1 Benthic Ecology Survey (2010)	2BT	2010	41
Cefas	Sewage Sludge_2001	2BT	2001	2
TOTAL				**1528**

* 2BT: 2m Beam Trawl

Appendix 3: Details for the data uploaded to OBGC.

Source	Survey Name	Gear*	Year	Count
EurOBIS	REBENT: Benthic Networking (https://www.emodnet-biology.eu/portal/index.php?dasid=4412)	VV, BC, SM	2003-2013	1017
EurOBIS	North Sea Benthos Survey (ipt.vliz.be/eurobis/r=NSBS)	VV/C	1985-1986	586
EurOBIS	Macrobenthos from the Norwegian waters (http://ipt.vliz.be/eurobis/resource?r=nsbp_cochrane)	VV	2000	293
EurOBIS	Macrozoobenthos data from the southeastern North Sea in 2000 (http://ipt.vliz.be/eurobis/resource?r=nsbp_rachor)	VV	2000	342
EurOBIS	Dutch long term monitoring of macrobenthos in the Dutch Continental Economical Zone of the North Sea (http://ipt.vliz.be/eurobis/resource?r=deltaresbenthos)	BC	1991-2012	2196
EurOBIS	Macrobenthos monitoring at long-term monitoring locations, period 2001-ongoing (ipt.vliz.be/eurobis/r=long_term_2001forward)	VV	2001-2017	1494
TOTAL				**5928**

* VV: 0.1m^2 Van Veen grab, BC: 0.1m^2 Box Core, SM: 0.1m^2 Smyth McIntyre grab, VV/C: 0.1m^2 Van Veen or Core (not stipulated).

8. Acknowledgements

This study was funded by The Crown Estate's Offshore Wind Evidence and Change (OWEC) programme (Project C8210, North Sea Net Gain), with support from The Rich North Seas programme who funded the data harvesting undertaken by the Flanders Marine Institute (VLIZ) from the EurOBIS data repository. We are grateful to members of the project advisory group for their input throughout the duration of the project

The OWEC programme was established by The Crown Estate in December 2020 and aims to facilitate the sustainable and coordinated expansion of offshore wind to help meet the UK's commitments to low carbon energy transition whilst supporting clean, healthy, productive and biologically diverse seas. It is a collaborative programme led by The Crown Estate, together with its programme partners, the Department for Business, Energy and Industrial Strategy (BEIS) and Defra. It is being delivered in collaboration with devolved government bodies and organisations from across the UK that have an interest in planning for the future of offshore wind.

Centre for Environment Fisheries & Aquaculture Science

Cefas

World Class Science for the Marine and Freshwater Environment

We are the government's marine and freshwater science experts. We help keep our seas, oceans and rivers healthy and productive and our seafood safe and sustainable by providing data and advice to the UK Government and our overseas partners. We are passionate about what we do because our work helps tackle the serious global problems of climate change, marine litter, over-fishing and pollution in support of the UK's commitments to a better future (for example the UN Sustainable Development Goals and Defra's 25 year Environment Plan).

We work in partnership with our colleagues in Defra and across UK government, and with international governments, business, maritime and fishing industry, non-governmental organisations, research institutes, universities, civil society and schools to collate and share knowledge. Together we can understand and value our seas to secure a sustainable blue future for us all, and help create a greater place for living.

Pakefield Road, Lowestoft, Suffolk, NR33 0HT

The Nothe, Barrack Road, Weymouth DT4 8UB

www.cefas.co.uk | +44 (0) 1502 562244

disability confident EMPLOYER

CERTIFIED ISO 9001

CERTIFIED OHSAS 18001

ISO SGS