

MARBEFES Project

Title: User manual for Multi-scale trait-based approaches to link biological structure to function
Project General Coordinator: Jan Marcin Węśławski Project Scientific Manager: Julie Bremner Project Manager: Joanna Potrykus potrykus@iopan.pl
Task Leader(s) Task leader: Marie Nordström, Åbo Akademi University Activity leader: Anna Törnroos, Åbo Akademi University
Author(s) Anna Törnroos, Marie C. Nordström, Riikka Puntila-Dodd and Tiina Salo, Åbo Akademi University
Version date: 28th of January 2025



MARBEFES project has received funding
 from the European Union's Horizon Europe research and
 innovation programme under Grant Agreement no 101060937
 and
 from UKRI under Grant Agreement no. 10048815





MARBEFES project has received funding
from the European Union's Horizon Europe research and
innovation programme under Grant Agreement no 101060937
and
from UKRI under Grant Agreement no. 10048815



User guide: Guidelines for application of the Multi-scale trait-based approach to link biological structure to ecosystem function

Table of contents

1. Introduction	4
1.1. Aims and structure of the user manual.....	4
2. WHAT ARE TRAIT-BASED APPROACHES? (part 1)	5
3. GUIDANCE on applying the Multi-scale BTA tool (part 2).....	8
3.1. Example 1: Assessment of traits on community level for different EUNIS-3 level habitats	8
3.1.1. General overview	8
3.1.2. Phase 1: Defining the study system.....	8
3.1.3. Phase 2: Assembling the macrofaunal data	9
3.1.4. Phase 3: Assessing community weighted mean trait structure and trait diversity in EUNIS-3 habitats	9
3.1.5. Step-by-step guideline to Example 1	10
3.2. Example 2: Trait biogeographies in a marine area.....	17
3.2.1. General overview	17
3.2.2. Phase 1: Defining the study system.....	17
3.2.3. Phase 2: Assembling the data	17
3.2.4. Phase 3: Assessing trait biogeographies.....	18
3.2.5. Step-by-step guideline to Example 2	19
3.2.6. Spatial extrapolation of data.....	25
4. REFERENCES	31



MARBEFES project has received funding
from the European Union's Horizon Europe research and
innovation programme under Grant Agreement no 101060937
and
from UKRI under Grant Agreement no. 10048815



1. Introduction

The guidance in this document have contributed to MARBEFES WP3 (Biodiversity and Ecosystem Tools), task 3.2, activity 3.2b), and specifically sub-task ii), for which descriptions are given below, following the project proposal:

WP3 Biodiversity and Ecosystem Tools

Task 3.2: Tools for exploring ecological phenomena (Lead **ABO**; IO PAN, LE ERIC, APN, IECS, KU, TU, CNR, UCD, GEOECOMAR, UC, FIHAC, CSIC, AWI, UEA, HCMR, NBE, BC3). This task will develop tools to explore ecological connectivity and the links between biodiversity and function across seascapes.

Activity 3.2b: Techniques for modelling ecological phenomena: It will compare models of seascape structure, function, stability and dynamics for better understanding the connections between biological entities. The models will be reviewed according to criteria on data requirements and capabilities, culminating in advice on which can be best used in different circumstances. T3.2c will further develop selected models to improve coverage of the river-sea continuum, benthic-pelagic coupling, population-level drivers (e.g. sex-determined space use and better species representation (microscopic organisms, plants, algae, invertebrates, fish, apex predators).

3.2bii: BROADSCALE biological trait methods: BBT methods consider a range of connections relevant to structure and functioning. They vary from single- to multi-traits and have the potential to assess functioning across groups since traits are a common currency. However, they are often implemented within a group rather than across, frustrating efforts to yield a broadscale picture. Critical challenges are to identify relevant and measurable traits and to elucidate synergies and trade-offs. Our solution is to formulate theoretical and mechanistic models of trait expression and to apply them holistically across groups. We will accomplish this by empirically assessing key traits across the BBT gradients, allowing for comparative studies that generate models, common guidelines and criteria for trait based assessments.

1.1. Aims and structure of the user manual

Here we describe two examples of applying the Multi-scale trait-based approach to link biological structure to ecosystem function (Multi-scale BTA) for characterising seascape diversity and function. The example has been applied in the MARBEFES project Broad Belt Transects (BBTs) areas. These instructions describe the approach, the focus, and the associated data needs to support the planning of data collation, sampling, and analysis in an area of interest.

These guidelines include a protocol for two ways of applying the Multi-scale BTA tool:



MARBEFES project has received funding from the European Union's Horizon Europe research and innovation programme under

Grant Agreement no 101060937



1) *Assessment of functional structure on community level for different EUNIS-3 level habitats*

2) *Trait biogeographies in a marine area*

Example 1) is an assessment of functional differences of habitats (EUNIS level-3) in a given area (in this case the BBTs) based on community level trait structure (functional indices). Example 2) aims to allow for investigations of single trait biogeographies in a given area (here again exemplified with a BBT), i.e. providing spatial distributions of a trait across the seascape in an area.

Here we also provide short guidance (a primer) on conducting trait-based assessments, compilation(s) of trait data (if not already existing for the area and/or taxa in question). The tool has been developed with other work in the MARBEFES project in mind (e.g. it builds on the Ecological Valuation Assessment (EVA) data collection and links to the functional habitat metric).

This document is structured into two parts: 1) a primer on trait-based approaches, with a focus on the biological trait approach as a general method applied to benthic habitats and organisms, but applicable also to other taxonomic groups, 2) a guidance section including detailed step-by-step instructions on the two examples.

2. WHAT ARE TRAIT-BASED APPROACHES? (part 1)

Traits are species attributes and can be morphological or physiological (e.g. size, body structure), reproductive (e.g. larval type, reproductive frequency), or behavioural (e.g. feeding ecology, movement method) (Table 1).

Table 1. Examples of invertebrate traits used in studies

Trait	Categories	Labels	Relevance
Body size (categorical)	small (1-5mm)	s	Productivity, feeding interactions, respiration, elemental cycling
	small-medium (5mm-1cm)	sm	
	medium (1-3 cm)	m	
	medium-large (3-5 cm)	ml	
	large (>5cm)	l	
OR			
Body size (continuous)	actual average value per species per sample		
Longevity	very short (<1yr)	vsho	Production, elemental cycling
	short (1-2 yr)	sho	
	long (2-5 yr)	lon	
	very long (5-10yr)	vlon	
Reproductive frequency	annual episodic	anep	Production
	annual protracted	anpr	
	semelparous	sem	
Living habit	attached	att	Elemental cycling, dispersal, foraging
	burrow dweller	budw	

	free tube dweller	free tub	
Movement type	no movement swimmer rafter-drafter crawler bussys threads tube dweller burrower	nom swim raft crawl byss tube burr	Mobility, dispersal, antipredatory behaviour
Feeding position	suspension feeder surface feeder sub-surface feeder selection feeder miner parasite	sus sur sub sel min par	Production, elemental cycling feeding mode
Feeding habit	detrivore omnivore herbivore carnivore scavenger parasite	detr omni herb car sca para	Production, trophic

Trait-based approaches aim to link community structure to ecological functions, and the most commonly used traits in marine studies are linked to mobility, feeding, and body size (Fig. 1).

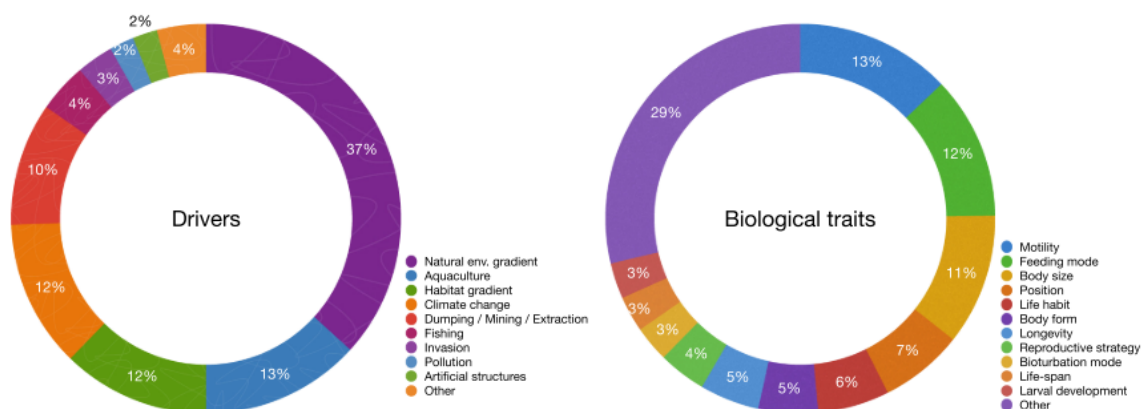


Figure 1. Some of the most assessed drivers affecting biological traits (left) and most commonly used traits in published studies (2003- 2021). From de Juan et al. (2022).

Traits can also be divided into different types based on what they describe, for example:

Ecological traits – the morphological, physiological, and phenological attributes of species

Response traits – how organisms respond to environmental factors

Effect traits – how species influence the ecosystem

Matching traits – how organisms interact with other species

Functional traits – phenotypic characteristics influencing ecosystem processes

Traits may be used to assess responses to different environmental changes (see examples in Fig. 1) of organisms, how community functioning varies between habitats or across a seascape, and resilience in the system, among others.

Community traits can be assessed in many ways, with functional diversity indices and community weighted means (CMW) being among the most common approaches. **Functional diversity indices (Activity 1)** include several complementary indices that are assessed in the multivariate community trait space (detailed descriptions in e.g. Villeger et al. 2008, Mouchet et al. 2010). Functional dispersion (FDis) describes the average distance from all taxa to the weighted centroid of the trait space and Rao's quadratic entropy (Rao's Q) assesses the dissimilarity between taxa in a community. Functional richness (FRic) indicates the number of functions (i.e. traits or trait categories) and the trait space that the taxa in a community occupy. Functional evenness (FEve) describes how evenly distributed the taxa in a community are in the trait space, with high values indicating even distribution in the niche space. Functional divergence (FDiv) is based on the distance between the taxa and the centre of the multivariate community trait space and gains a high value when the taxa with the most individuals are far from the centre of the multivariate trait space. Some indices consider the proportion of different taxa in the community (e.g. Rao's Q, FDis, FEve, FDiv) while others do not (e.g. FRic). **The community weighted means-approach (CMW, Activity 2)** is used to calculate the expression of a trait or trait category in a community and is weighted by the proportion of the taxa. Functional richness indices and CMW can be calculated using e.g. the FD package (Laliberte et al. 2022) in R.

The common workflow and data matrices needed to analyse trait-environment relationships are presented in the tutorial by Frelat et al. (2022): The tutorial includes guidance for (A.) data preparation for trait-environment relationship, e.g., how to prepare the data, which traits to choose and how to treat correlated traits. It also guides (B.) how to calculate community weighted mean traits and trait diversity indices. <https://github.com/rfrelat/TraitEnvironment>

Trait data sources span several trait databases and datasets, which can aid applications of trait-based analysis in new areas. For an exhaustive list of such sources, please see [Table S1 \(de Juan et al. 2022\)](#). This list contains information on existing open access and/or access upon registration trait databases and/or trait compilations of marine taxa. However, when applied to new areas, trait data is often supplemented with local or regional data and information.



3. GUIDANCE on applying the Multi-scale BTA tool (part 2)

3.1. Example 1: Assessment of traits on community level for different EUNIS-3 level habitats

3.1.1. General overview

OBJECTIVE: The aim is to assess functional differences between habitats within BBTs

INPUT: Essential information includes presence/absence, abundance or biomass data and biophysical data, with a minimum requirement being EUNIS habitat maps at EUNIS Level 3, and benthic macrofauna.

OUTPUT: The output is a spatial map of macrofaunal community-level functional diversity across EUNIS habitats within an area (here BBT).

AIM: assessment of functional diversity patterns between habitats.

3.1.2. Phase 1: Defining the study system

This first phase includes defining the spatial-temporal boundaries for the study area (here exemplified with a MARBEFES BBT) and the assessment units within it, identifying the relevant organism group with data available for the assessment, and the time for which the data were acquired.

In this activity, we suggest adopting the spatial designation following the procedure outlined in the Ecological Value Assessment (task 4.1) protocol: hexagonal spatial units with size (0.25-3 km wide) depending on the ecological component in question (see M39 EVA guidelines, page 14), and EUNIS Level 3 (as a minimum level adopted by the project). There is also a possibility of using higher EUNIS level categories of habitats (e.g. level 4-5) if data is sufficient for this (see phase 3). This will also be used in other work in the WP3 (e.g. WP3.2c) and WP4, enabling BBTs to harmonize task outputs across the project.

The choice of the taxonomic group(s) to include in the analysis should be based on the availability of biological data (presence/absence, abundance or biomass data) in relation to the coverage of the EUNIS-3 level habitats as well as trait data availability in the area of interest. However, based on the focus on EUNIS level 3 and indicated availability of datasets from partners and BBTs, the feasible option is to focus on one or several of the benthic-related taxonomic groups (e.g. macrofauna). It is preferable to have replicates ($n \geq 3-5$) for communities per habitat. **Please note that here we conduct the analysis on macrofaunal data (analyses of other taxonomic groups are optional).** The choice of traits is also dependent on the taxonomic group in question.



3.1.3. Phase 2: Assembling the macrofaunal data

The macrofaunal community data should include the list of taxa and presence/absence (minimum requirement), abundance (preferred) or biomass of each taxon in each community. Use only dominant taxa for this, i.e. taxa occurring in less than 5% of communities are considered rare (following WP4.1 EVA phase 2 limit for rarity). We recommend a maximum time span of 10 (most recent) years for the community data in the BBTs. In the data file, samples should be organized in rows and taxa in columns. As an example of a community data file, please see Frelat et al. (2022):

https://github.com/rfrelat/TraitEnvironment/blob/main/Data/Fish_Abundance.csv

Trait data can be found readily for some groups (for example see [Table S1](#) in de Juan et al. 2022 for benthic species). Instructions for how to assemble trait data matrices, i.e. traits x taxa matrices, follow standard ways for e.g. assessing trait diversity indices or community weighted mean (CWM), see detailed information below (3.1.5.2) or Frelat et al. (2022):

<https://rfrelat.github.io/CWM.html>.

3.1.4. Phase 3: Assessing community weighted mean trait structure and trait diversity in EUNIS-3 habitats

The aim to assess trait diversity, that is Rao's Q and Functional dispersion (FDis) through community weighted means (CWMs) in EUNIS 3 level habitats, is twofold: i) it allows the BBT to illustrate community-level trait (functional) structure across the seascape, and ii) it enables assessments of the degree of functional (trait) differences between EUNIS habitats, i.e. answering, for example, the question of whether EUNIS habitats also differ functionally in the BBT, which can provide e.g. insights for management.

The Rao's Q and Functional dispersion (FDis) were chosen because they can be calculated for areas with low taxa diversity (e.g. communities with less than three species, such as in parts of the Baltic Sea or disturbed areas). Both indices can be calculated using FD-package (Laliberté et al. 2022) in R. The package simultaneously calculates several diversity indices ranging from the number of species to the functional richness, divergence, evenness, and CWM values for the trait categories. Detailed instructions and more information in the Step-by-step guideline to Activity 1 below and in (Frelat et al. 2022): <https://rfrelat.github.io/CWM.html>.

The final step is to project the indices and CWM trait categories on maps using the spatial resolution chosen (e.g. EUNIS 3).



3.1.5. Step-by-step guideline to Example 1

Here is an example from the Archipelago Sea area (in MARBEFES: BBT4). The work consists of four steps:

Step 1: Define data and traits, harmonize data

Step 2: Run the FD analysis

Step 3: Import to GIS to extract data per habitat type

Step 4: Conduct basic summary statistics

In the example below, we work with monitoring data of benthic macrofauna in the Finnish Archipelago Sea. The data is point data from monitoring. We use the following software: Excel, R, ArcGIS. While other options are available, we can give support mainly in these. The BBT-specific choices and results are indicated in blue font to separate from the general instructions.

3.1.5.1. Step 1: Define data and traits, harmonize data

1. Define your data. We use abundance data and refer to this data file as “abundance”, but the analyses may also be based on presence/absence data or biomass data. The trait data can include several traits, and each of these may include different categories. Trait data can often be found in different trait repositories or literature. Note that the abundance and trait data need to match.

- a) abundance: organize your community data so that samples are on rows and different taxa in columns (Fig. 2). If you have trait data at the species level, you may use that level information. However, trait data is often only available for family or higher taxa levels (e.g. *Gammarus* sp. or Oligochaetes), and you must group the abundance information accordingly. Also, if you have some rare species (occurring in < 5% of communities), these should be excluded from the data. Also, groups where no trait data exists can be excluded, or one may use expert evaluation to estimate traits for these. Please keep a record of how you have proceeded with these so backtracking is possible.

In the example data we included benthic macrofauna samples from our BBT area between years 2013-2022 (10 years). This is the same data that we used in the Ecological Value Assessment, but rare taxa (i.e. taxa occurring in < 5% of communities) were removed. We also combined some species to family or higher level, as no trait information was available for all taxa at the species level. So, in the final data, the taxonomic resolution varies from species to family or group level.



The abundance data file should only have the row or sample number and the taxa information. Save this information in a separate file as a .csv so that you can later combine the obtained data with the spatial and other relevant information (e.g. coordinates, sampling year or gear used).

- b) traits: decide which traits you wish to include. In the example data, we included the traits and categories listed in Table 1. We used existing trait data (Törnroos & Bonsdorff 2012). In some cases, we evaluated that we could apply the same traits for some other, closely related species. However, as some small, cryptic species were lacking trait information, we excluded these from both data sets. For some taxa, we complemented the trait information from the literature (Salo et al. 2023).

Organize your trait data file so that species (grouped and typed exactly as in the abundance data file and in the same order) are represented in different rows and trait categories in different columns (Fig. 2). We suggest assessing species either values 0 (not expressed) or 0.99 (previously 1) (trait expressed). If taxa express several categories for one trait, the traits should be standardized so that the sum of all categories for a taxon is 1 for each trait. For example, if the movement method for species a is swimmer 1, crawler 1, then the sum is 2, and the value for swimmer will be $\frac{1}{2} = 0.5$ and for crawler $\frac{1}{2} = 0.5$. Species b is only a swimmer 1, with the sum of categories = 1, thus swimmer will be given the value 0.99 and crawler the value 0. Standardization is also needed if one uses weighted trait values (e.g. values between 0-3) expressing the probability or commonness of certain categories.

The trait data file should only have the taxa name and the trait categories (remove the uppermost row with the trait name). Save this as a .csv file. Remember that the taxa names should be identical and in the exact same order in both trait and abundance data files (capital letters matter)! Add a new column (A) and add row numbers to it. Save these files as csv.



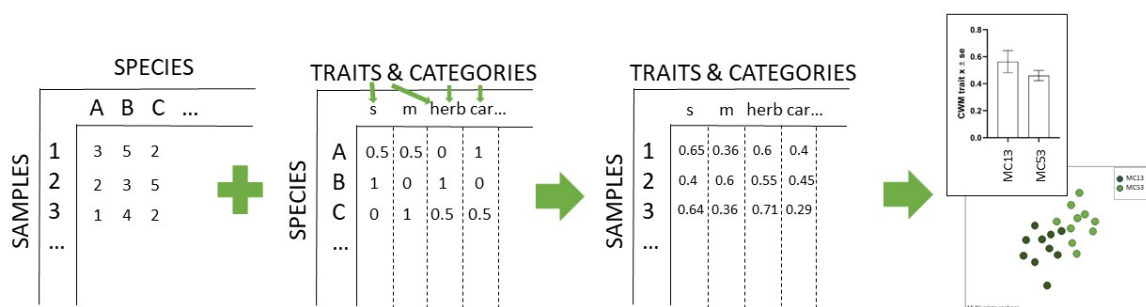


Fig. 2. The matrices required to calculate functional diversity indices include community data (left) and trait data (middle). The resulting data matrix (right) can then be used to conduct comparisons between the habitats, among others.

3.1.5.2. Step. 2: Run the FD analysis

1. Open R and download the package “FD”
2. Import the “abundance” and “trait” files.
3. Calculate the functional indices and community weighted means for different trait categories. Below is the r-code used to calculate the example data.

```

***CODE START***
#Marbefes WP3 trait calculations
setwd("C:/Users/.../MARBEFES/WP3/Traits/trait calc")

#install package FD
install.packages ("FD")
library (FD)

#abundance data (the .csv data file was called "abundance")
Abundance<-abundance
Abund2 <- Abundance[,-1]
rownames(Abund2) <- Abundance[,1]
View(Abund2)

#trait data (the .csv data file was called "traits")
Trait<- traits
Trait2 <- Trait[,-1]
rownames(Trait2) <- Trait[,1]
View(Trait2)

```

#the FD calculation:

```
ez4<-dbFD(Trait2, Abund2, w.abun=TRUE,corr = "none")
```

```
View(ez4)
```

#Write CSV file to Excel

```
write.csv(ez4, file = "CMW_WP3_test.csv")
```

#Note that if there are <3 unique species in a community, the program gives warnings and functional #richness, functional evenness, and functional divergence cannot be calculated (but calculates the #other things).

#If empty communities or trait categories do not match any fauna, the program gives warnings.

#Run the analyses again after removing the empty rows or columns.

***R CODE ENDED**

3.1.5.3. Step 3: Import to GIS-software to extract data per habitat type.

1. Combine the obtained data with the spatial information and other relevant information in e.g. Excel.

2. Import the data to your GIS project. Do any required pre-selection that you find appropriate.

For example, in the example data we calculated the functional diversity indices and traits for a slightly larger data set and sorted the data geographically (the BBT area) and temporally (10 years) in the ArcGIS (Fig. 3).



MARBEFES project has received funding from the European Union's Horizon Europe research and innovation programme under

Grant Agreement no 101060937



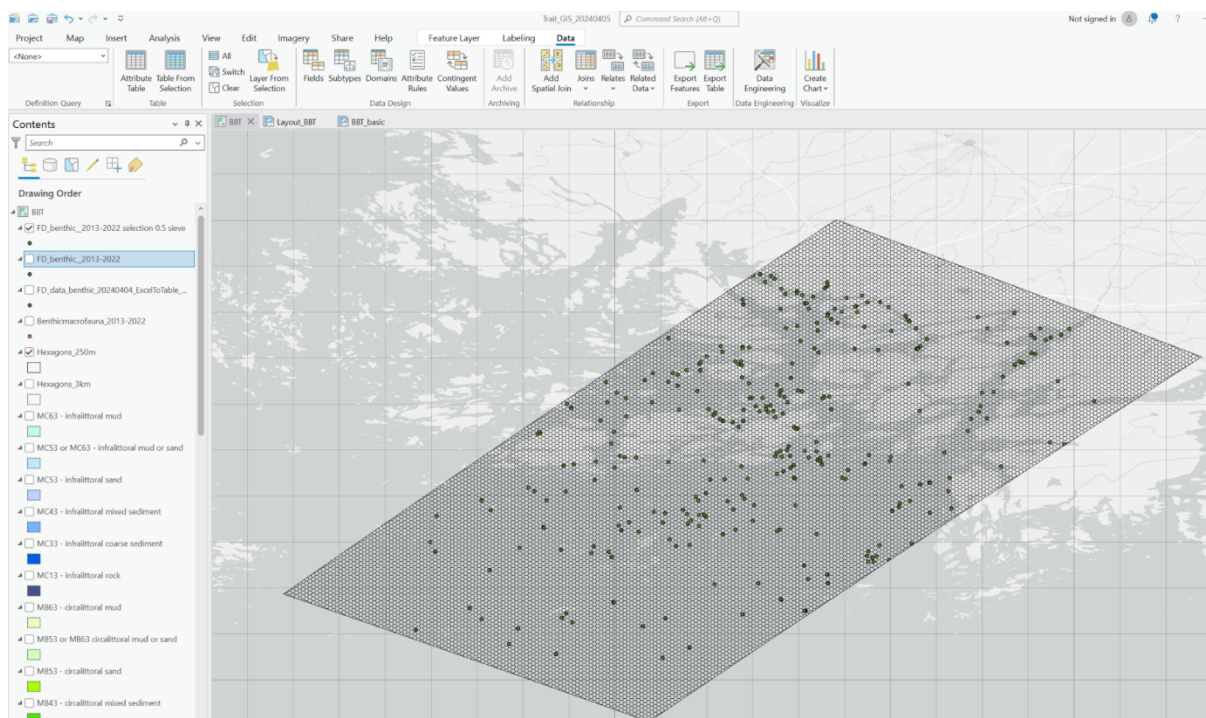


Fig. 3. The example benthic BBT data (Archipelago Sea BBT). The gridlines are the 250m hexagonal grids also used in the EVA.

3. Summarize the information per EUNIS habitat.

In ArcGIS, this can be done e.g. by using the summarize within tool (Analysis->Tools->Geoprocessing->). If the habitats in the area consist of large areas, it is recommended to extract all functional community metrics for the area, i.e. all FD data points per habitat. The data for each habitat type should then be exported to Excel. Extract also the data per hexagon and illustrate it on the map (Fig. 4). Please use a continuous scale to illustrate the diversity index on the map.

In the example data, the average Rao's Q and FDis were calculated separately for each EUNIS level 3 habitat area in the BBT area. As this area has high mosaicism of EUNIS-habitats (Fig. 5), the average value and n for the functional community metrics were calculated per habitat patch. Please note down if you deviate from the instructions. We then summarized these functional community metrics for the 250m hexagons to see any general spatial patterns in the functional diversity (Fig. 5: Rao's Q).

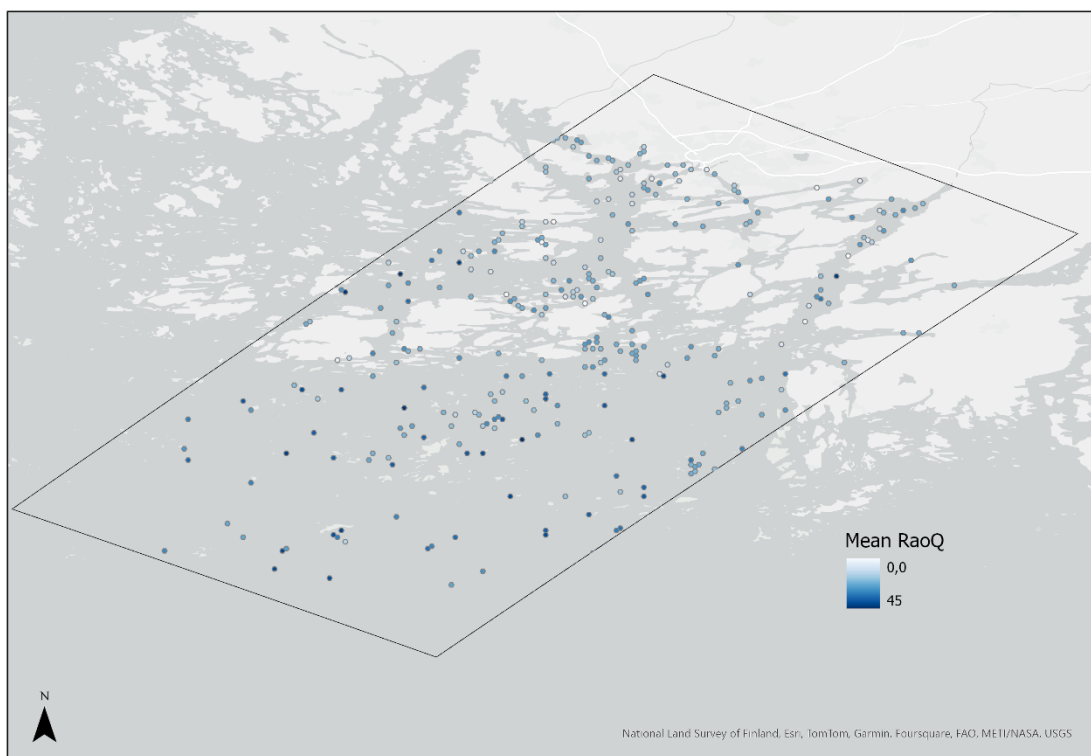


Fig. 4. The average Rao's Q of benthic communities on the BBT spatial scale (250 m hexagons). The darker the shade, the higher the diversity index.

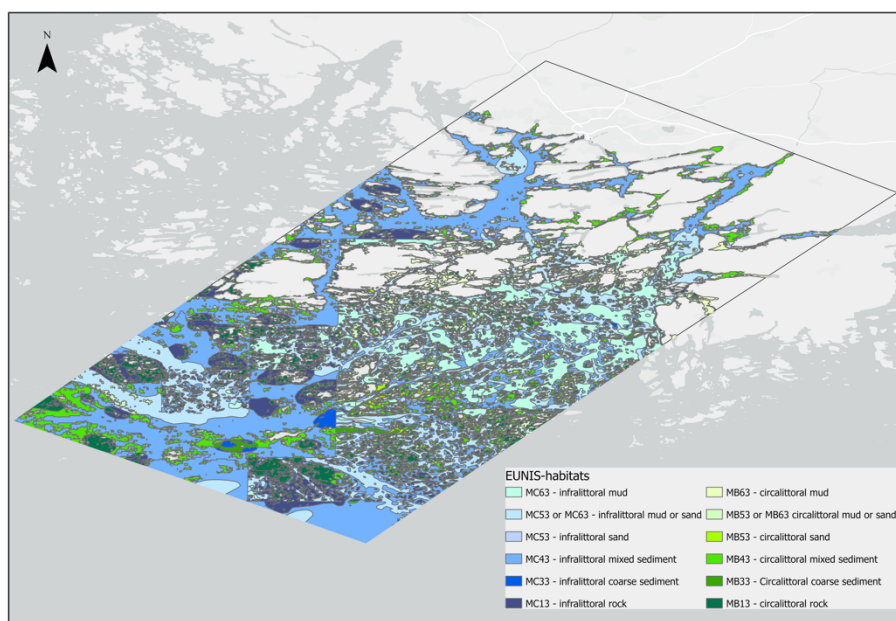


Fig. 5. The EUNIS level 3 -habitat mosaicism in the example BBT. Due to the high mosaicism, the individual habitat patches were used as replicates for the statistical assessments.

3.1.5.4. Step 4: Conduct basic summary statistics

The extracted and exported data should then be used to assess statistical differences in functional diversity across habitat types. The basic summary statistics (x, n, SE or SD per EUNIS habitat) may be calculated in e.g. Excel. Further analyses by the BBTs are optional. The statistical comparisons of the indices across different EUNIS habitats can be done e.g. by using Analysis of Variance (ANOVA) or PERMANOVA, where the habitat type should be a fixed factor.

In the example data, permutational univariate analysis of variance (PERMANOVA) illustrated that Rao's Q significantly differs across EUNIS level 3 –habitats (Table 2, Fig. 6). Pairwise comparisons illustrated that these differences were due to higher Rao's Q in harder sediments (significant differences e.g. between mud vs sand, mixed sediment, and rock, mud or sand vs mixed sediment and rock, as well as between mixed sediment and infralittoral rock (statistics not shown here)).

Table 2. Results from the PERMANOVA for the example data assessing how Rao's Q varies across habitats.

Source	df	SS	Pseudo-F	P(perm)
habitat	10	4369.8	6.0145	0.0001
residual	152	11044		

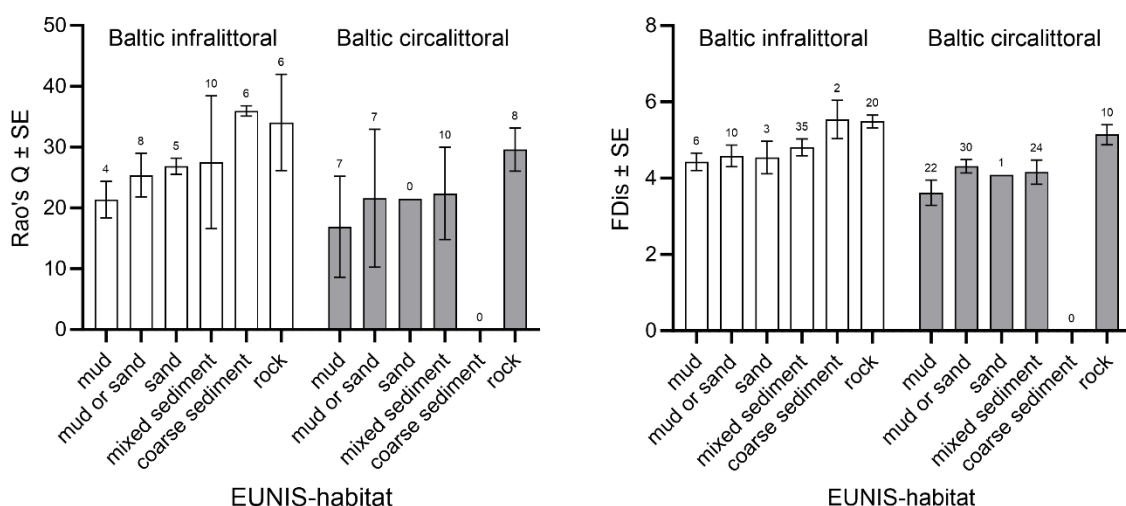


Fig. 6. The average Rao's Q (above) and functional dispersal (below) in the different EUNIS level 3 –habitats. The values above bars indicate n. (Please note that the results for the FDis statistical analysis are not given in this document to maintain the guidelines concise).

3.2. Example 2: Trait biogeographies in a marine area

3.2.1. General overview

OBJECTIVE: The aim is to assess the spatial distribution of individual traits in a marine area

INPUT: Essential information includes presence/absence and/or biomass/abundance data from one or several different taxonomic groups (e.g. macrofauna, fish, phyto- or zooplankton, macrophytes). Maps of the essential environmental parameters shaping the species assemblages in the area (e.g., temperature (max or min), salinity, primary production (chl_a), benthos type, depth, slope...).

OUTPUT: Generate spatial maps of trait biogeographies (distributions) per taxonomic group or trait.

AIM: assessment of trait distributions within BBTs (trait maps).

3.2.2. Phase 1: Defining the study system

In accordance with Example 1 and other guidelines, we suggest using the hexagonal spatial units with size (0.25-3km wide) following the EVA guidelines) or point data (i.e. stations) to define the spatial resolution in the study area (BBT). Regarding replication and timescale the same applies here, as in Activity 1.

In this activity, several taxonomic groups can preferably be considered, especially when the aim is to assess a trait across taxonomic levels. For such a question, a focus on traits “size”, “longevity” and “movement”. BBTs may even assess additional traits such as “reproductive frequency” and “trophic level”.

3.2.3. Phase 2: Assembling the data

The community and trait data, as well as formats of the data, follow that in Example 1, that is:

- **The community data** should include the list of taxa and the presence/absence or abundances or the biomasses of each taxon in each community. In the data file samples should be organized in rows and taxa in columns. As an example of a community data file, please see Frelat et al. (2022):
https://github.com/rfrelat/TraitEnvironment/blob/main/Data/Fish_Abundance.csv
- **Trait data matrices** follow standard ways for e.g. assessing trait diversity indices or community weighted mean (CWM), see guidance in Activity 1 or (Frelat et al. 2022):
<https://rfrelat.github.io/CWM.html>

3.2.4. Phase 3: Assessing trait biogeographies

The trait biogeography analysis consists of two steps (Fig. 7). First, trait information and field observations on the EUNIS 3 level or point/station level are combined to produce Community Weighted Means (CWMs), for detailed instructions and more information see Activity 1 above. Then, either commonly used interpolation methods such as inverse distance weighting (IDW) or ordinary kriging (OK) could be applied, or statistical models (e.g. Brun et al. 2016) to investigate the spatial (or spatiotemporal) patterns of trait distributions. In section 3.2.6. we provide instructions for producing spatial extrapolation maps and in section 3.2.7. we provide instructions for a spatial model using GAM and Random Forest (RF) regression.



MARBEFES project has received funding from the European Union's Horizon Europe research and innovation programme under

Grant Agreement no 101060937



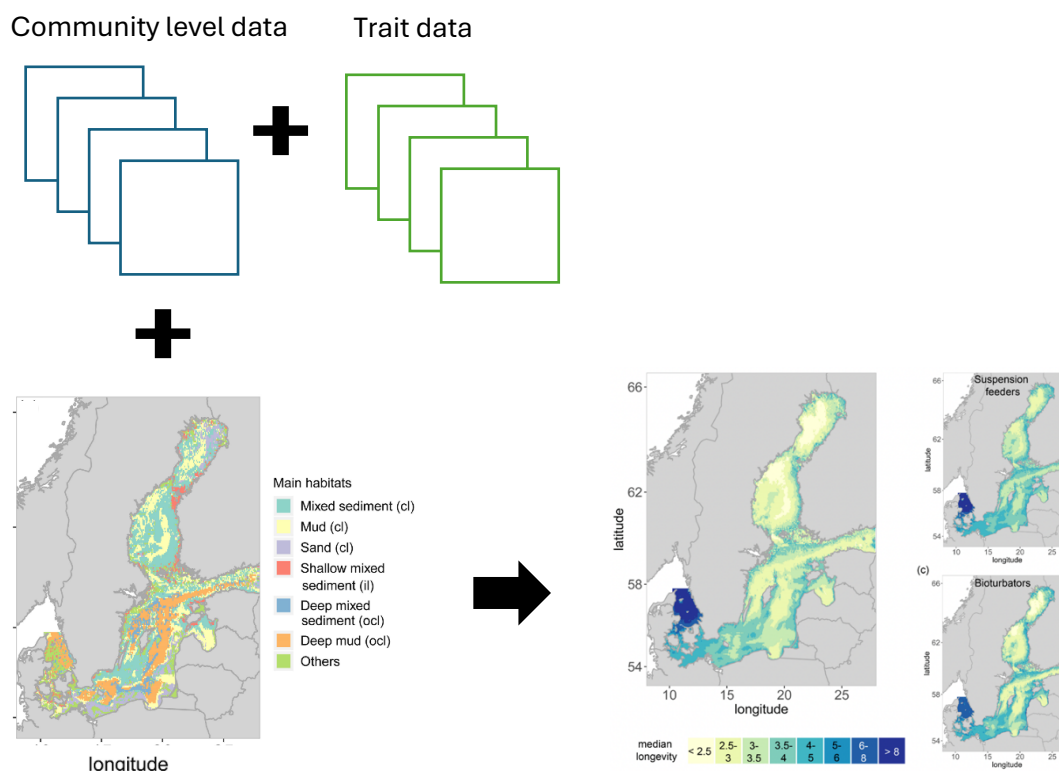


Figure 7. A conceptual figure of trait biogeography analysis. Community-level data are combined with trait data to produce e.g. CWMs, and then combined with EUNIS level 3 habitats to produce trait biogeography maps. The maps stem from van Denderen et al. (2019) and are here provided as an example illustration of the outcome (trait maps) rather than the modelling approach applied.

3.2.5. Step-by-step guideline to Example 2

Step 1: Define data and traits, harmonize data

Step 2: Run the FD analysis

Step 3: Import to GIS to visualize the areal expression of traits

In the example below we work with monitoring data on benthic macrofauna in the Finnish Archipelago Sea area (in MARBEFES: BBT4). The data is point data from monitoring. We use the following software: Excel, R, and ArcGIS. While other options are available, we can give support mainly in these.

3.2.5.1. Step 1: Define data and traits, harmonize data

BBTs should include benthic macrofauna but can freely decide which other taxonomic groups they use (e.g. fish, phytoplankton or zooplankton, macrophytes). This decision should be based on the available community data information in the BBT (and interest by BBT) and the trait information available for **3 key traits** (size, longevity, mobility).

Use only dominant taxa for this, i.e. taxa occurring in less than 5% of communities are considered rare (same as in WP4.1 EVA). Define the time scale for the data. It is suggested that the same time scale is used as in Activity 1 (i.e., max 10 years).

Organize the community and trait data files following the instructions in Activity 1. If trait information for certain species is not found for the key traits, note this down and leave taxa out of the collection.

In the example data, we included benthic macrofauna samples from our BBT area between 2013-2022 (10 years). We removed rare taxa occurring in <5% of the samples. We combined some species to family or higher level, as no trait information was available for all taxa at the species level. Otherwise, this is the same data that we used in the Ecological Value Assessment. So, in the final data, the taxonomic resolution varies from species to family or group level. We used existing trait data (Törnroos & Bonsdorff 2012, Salo et al. 2023). In some cases, we evaluated that we could apply the same traits even for some closely related species.

Key traits for adult individuals:

- Maximum size (currently different scale for different taxa groups, but this will be harmonized across taxa groups later):
 - phytoplankton:
 - nano
 - micro
 - macro
 - zooplankton:
 - TBD
 - invertebrates:
 - small (0.1-0.5 cm)
 - small-medium (0.5-1 cm)
 - medium (1-3 cm)
 - medium-large (3-5 cm)
 - large (>5 cm)
 - macrophytes
 - very small (>1 cm)
 - small (1-5 cm)
 - small-medium (5-10cm)
 - medium (10-50cm)

- medium-large (50cm-1m)
- large (1-1.5m)
- very large (>1.5m)
- fish
 - TBD (often continuous, e.g. mean/common size, see for example fishbase.se, or if measured during e.g. surveys in an area)
- Longevity: extremely short (<1week), very short (< 1yr), short (1-2 yrs), long (2-5 yrs), very long (>5yrs)
- Mobility: no movement/sessile, passive, active movement <1m, active movement <1km, active movement >1km

If the BBTs wish, they can also assess other additional traits for the different taxonomic groups based on the interest and data availability. This will be discussed during the workshop and general assembly, and BBTs are welcome to contact the leads (Anna Törnroos and Tiina Salo) regarding this at any point.

3.2.5.2. Step 2: Run the FD analysis

Follow the instructions given in Activity 1 to calculate the CMW traits using the FD package in R.

3.2.5.3. Step 3: Import to GIS-software to visualize the areal expression of traits

1. Combine the obtained data with the spatial information and other relevant information.
2. Import the data to your GIS project. Do any required pre-selection that you find appropriate.

For example, in the example data, we calculated the CMW traits for data from a slightly larger geographic area and sorted the data geographically (the BBT area) and temporally (10 years) in ArcGIS.

3. Combine values for trait categories for a trait to one value. This should be done in a standardized manner across BBTs. Calculation should be done by multiplying trait category 1 CMW value with one and trait category 2 CMW value with 2 ... category n CMV value with n, summing these up and then dividing with the number of categories (n) in the guidelines for that specific trait. For example, for macrozoobenthos size the equation is $(CMW.s*1+CMW.sm*2+CMW.m*3+CMW.ml*4+CMW.l*5)/5$. This results in comparable values across traits per taxa.

The total number of categories should be used even if a trait category is not present in the data. E.g. if small individuals are missing in the area, then the formula would be $(CMW.sm*2+CMW.m*3+CMW.ml*4+CMW.l*5)/5$. The resulting values for each trait should not exceed 1.

4. Summarize and visualize the spatial information per hexagons (for all BBTs)

Summarize the resulting CMW traits per hexagon (250m or 3km depending on the taxa, following the EVA guidelines) to visualize the distribution of different trait categories across the



BBT (Fig. 8). In ArcGIS this can be done e.g. by using the “summarize within tool” (Analysis-Tools-Geoprocessing-). The visualization should, at this point, be conducted separately for different trait categories and organism groups (ecosystem components). Please use a continuous scale for the colouring (with values ranging from 0.001 to 1) instead of e.g. categories.

5. To combine trait information from ecosystem components with spatially varying coverage and to produce more advanced rasterised output maps, please see section 3.2.6 and 3.2.7 below. This approach includes spatial modelling and will require expert knowledge in species distribution modelling.



MARBEFES project has received funding from the European Union's Horizon Europe research and innovation programme under

Grant Agreement no 101060937



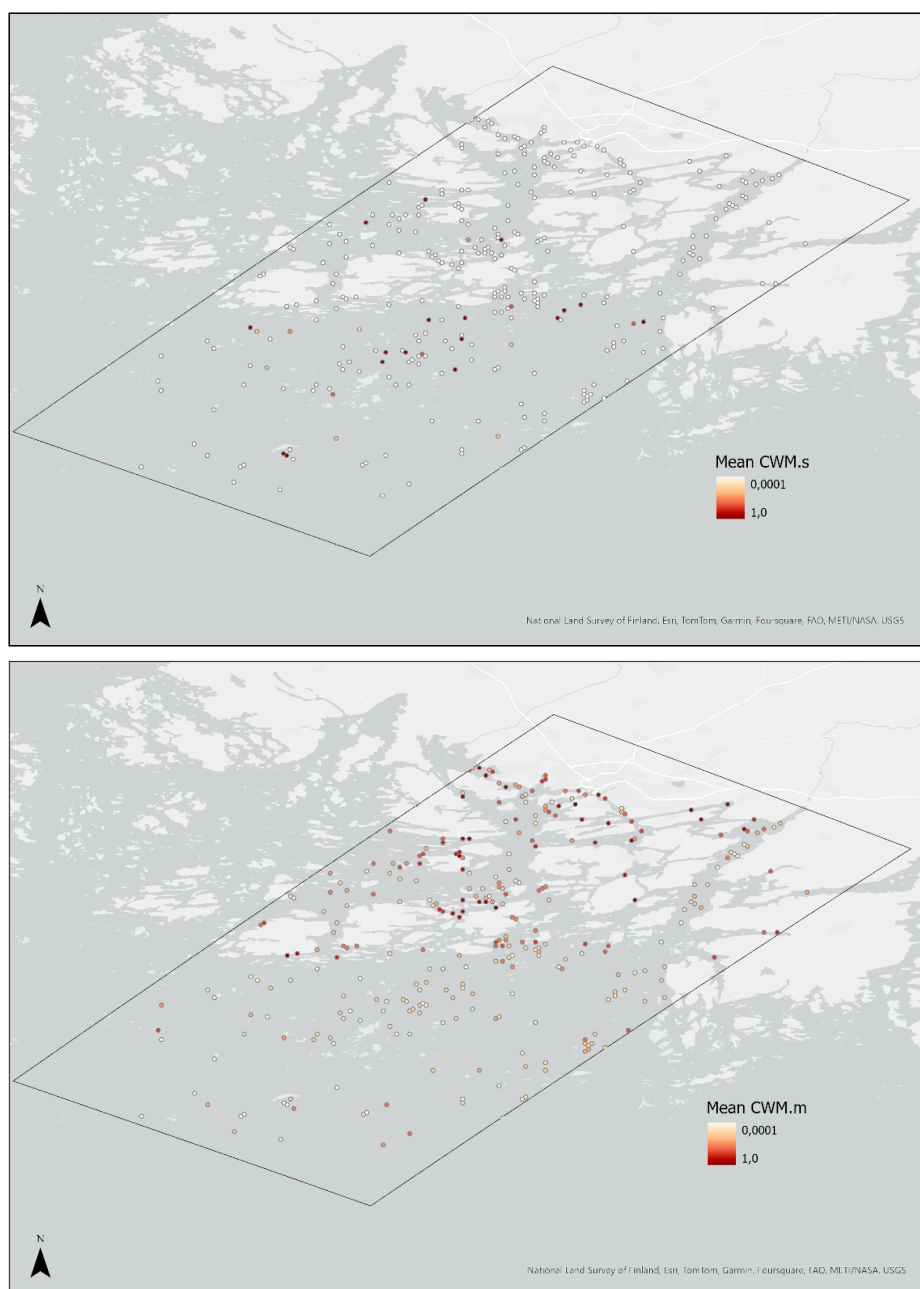


Figure 8. The expression of CMW trait size categories small (above) and medium (below) in benthic faunal communities in the Archipelago Sea BBT. The lighter (darker) the colour of the 250 m hexagons, the lower (higher) average CWM trait values. Not all categories are expressed in all communities.

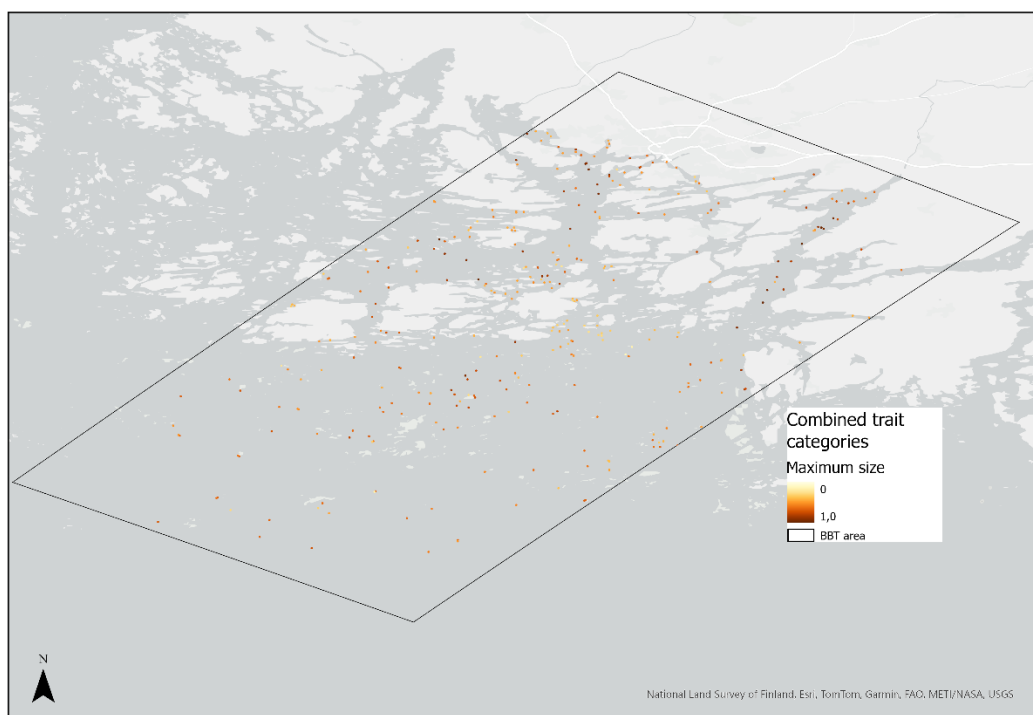


Figure 9. Combined trait categories for trait “size”. The darker the colour, the larger the CWM size in the area.

3.2.6. Spatial extrapolation of data

3.2.6.1. *Spatial extrapolation*

The simplest method to produce trait biogeography maps is to use a raster interpolation method, for example, the inverse distance weighting (IDW) method (Shepard 1968). IDW is a widely utilized spatial interpolation technique. In short, it is a deterministic method for estimating unknown values at specific locations using information from nearby known values. The fundamental principle in IDW is that spatial proximity plays a crucial role in data-point similarity: closer points influence the estimated value more than those farther away.

The method, in its simplest form, does not take into account any information about underlying factors (e.g., environment or geology), but simply uses information from surrounding points. If there are enough data points, the method likely yields quite reliable maps. Below, you can find step by step instructions on how to perform the calculations in QGIS.

Step 1. Preparatory steps in QGIS

Import the FD file, where mean traits have been calculated using the instructions in section 3.2.5.3, into your GIS software. Layer -> Add layer -> Add Delimited Text Layer. Save the points as a shapefile. Also import your BBT hexagon shapefile. You will use the corner coordinates from the file to write a raster of the IDW results. Optimally, you'll have data points outside the boundary of your BBT to avoid edge effects (Fig. 10).

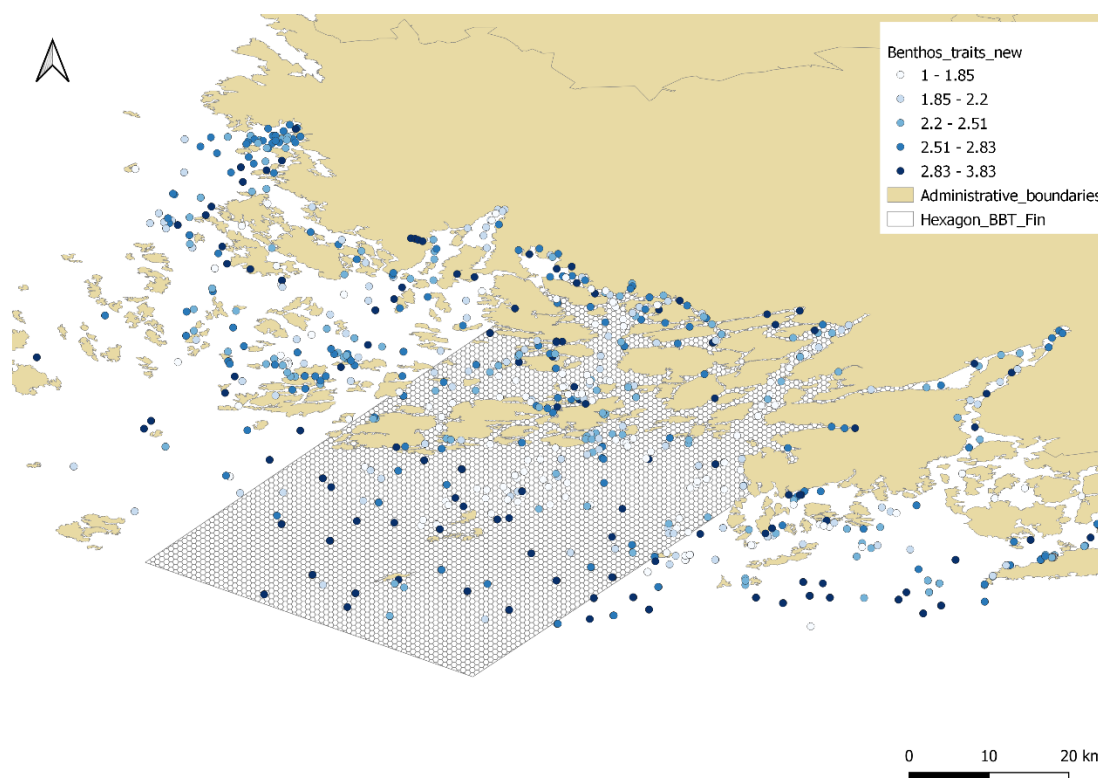


Figure 10. Benthos data points from the FD output file. Make sure you have datapoints also outside the BBT area (also shown on the map).

Step 2. Perform the Inverse Distance Weighting

From the qGIS toolbox, search for IDW (under “Interpolation” or use the search function). In the pop-up window, choose the point data as a “Vector layer”, and choose the trait you want to interpolate (Mean_size, for example). Press the green + symbol. In Extent, choose “Calculate from layer” and the BBT hexagon as the layer. This will pick the bounding coordinates from the BBT hexagon data. At this stage, it is a good idea to manually increase the bounding coordinates by a few degrees to ensure that your raster is larger than the BBT area. To avoid edge effects, the resulting raster must be larger than the BBT area.

Finally, define the resolution for your resulting raster in “Pixel size X” and “Pixel size Y”. Minimum resolution (or the largest grid size possible) is 0.0116 degrees, as it is the resolution of the finest hexagon grid. Once the IDW is finished, save the raster layer (export layer)(output shown in Fig. 11).

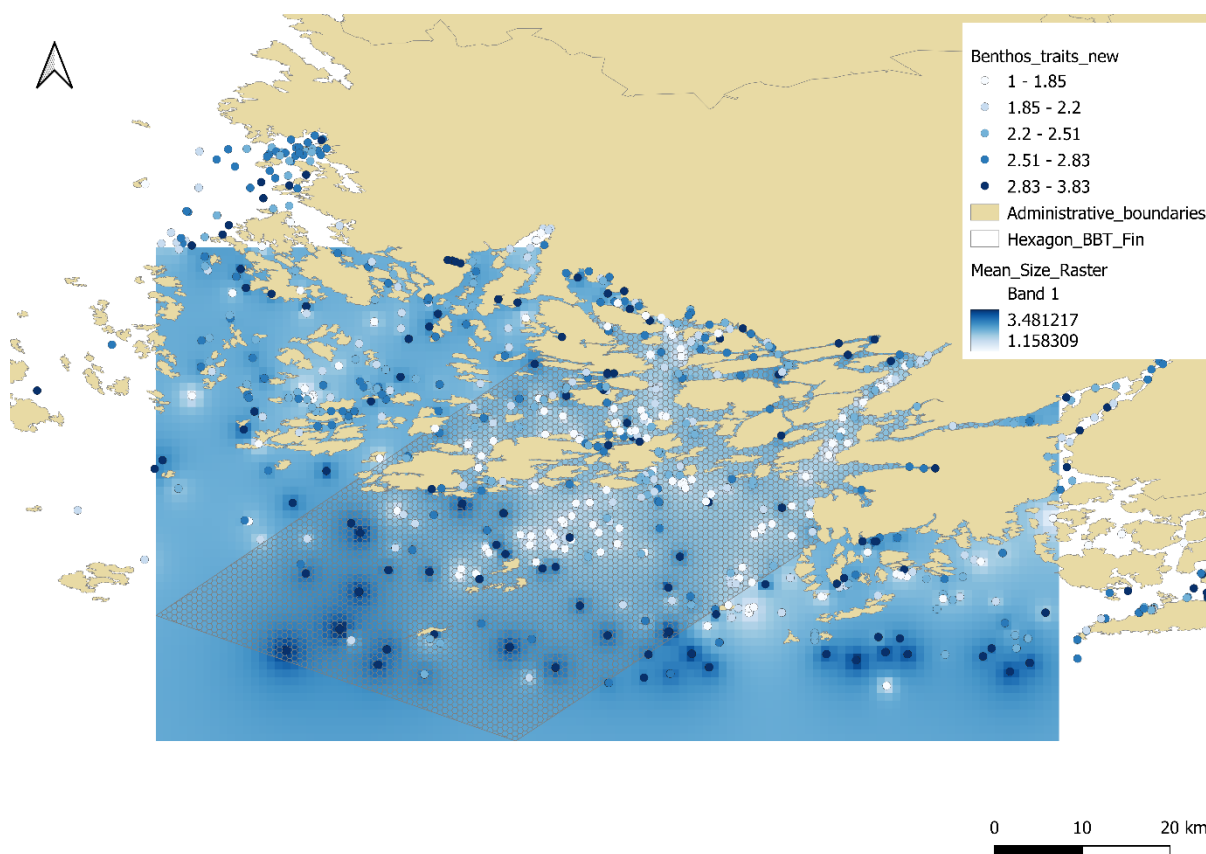


Figure 11. Raster resulting from IDW visualized with the point data and BBT hexagon

Step 3. Calculate the values for the BBT hexagon grid

Calculate the final mean trait values for the BBT hexagon grid (Fig. 12). This can be done using the Zonal Statistic tool in the qGIS toolbox (“Raster analysis – Zonal statistics”). In the pop-up window, choose the BBT layer as “Input layer” and the raster layer you created in step 2 as the “Raster layer”. For the “Statistics to calculate”, unclick other than “Mean”. This will produce a shapefile in which hexagons contain the mean values from the raster cells. You can modify the appearance of the shapefile by using symbology (Instead of a Single symbol, choose “Gradual” and the mean value as “Value”. You may also alter the borders of the hexagon to grey and a narrower stroke). The result will be a smoothed hexagon “raster” representing the trait in question (Fig 12). Repeat all steps to produce maps for all traits for which you have data.

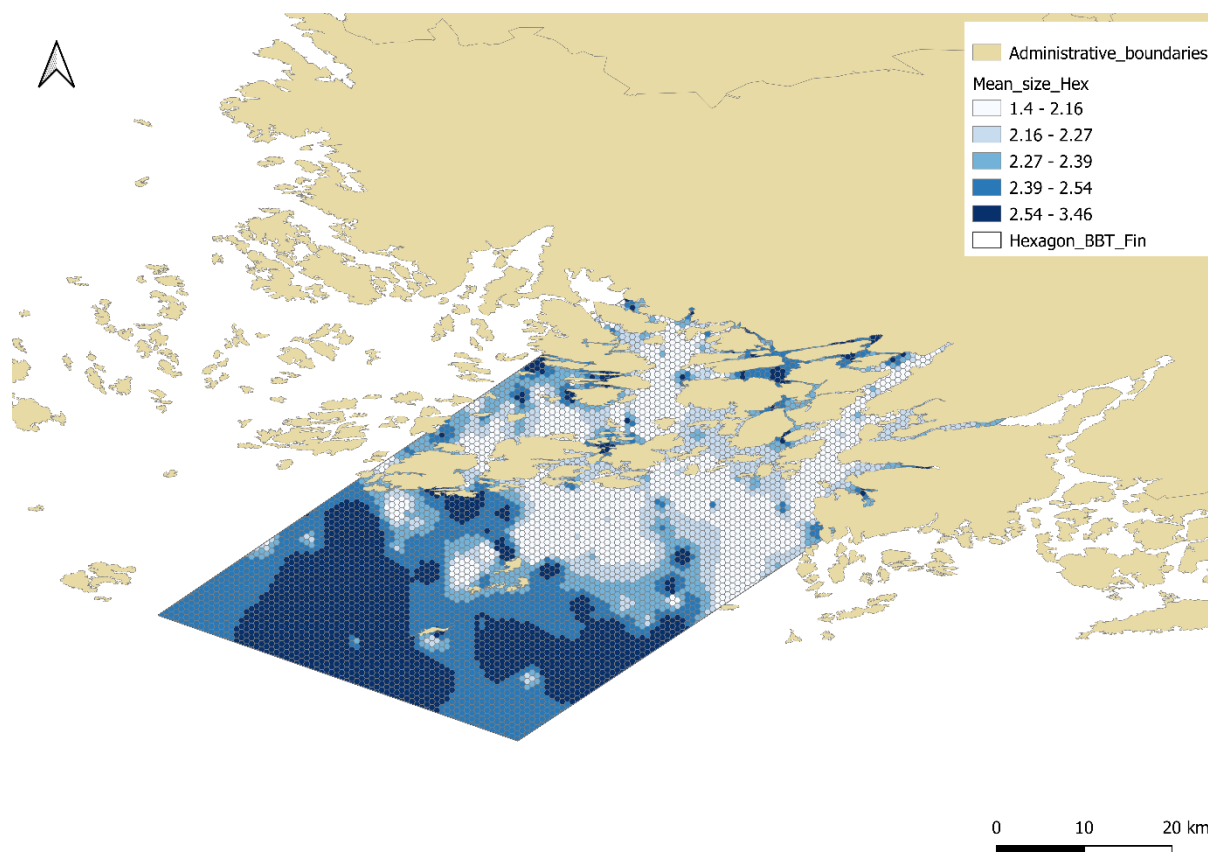


Figure 12. Trait Mean size illustrated in the BBT hexagon

3.2.6.2. Spatial model using Random Forest Regression and GAMs

Random forest regression and GAM models are widely used and powerful tools in species distribution modelling, and it is advisable to use a combination of models (Kosicki, 2020, <https://doi.org/10.1007/s10651-020-00445-5>). In most cases, results are expressed as the likelihood of a certain species' presence in a raster cell. They can, however, also be used to model trait indicators or mean traits (see, for example, Nemani et al. 2024, <https://doi.org/10.3389/fmars.2024.1141737>). In this analysis, we use continuous environmental layers to explain trait indices or mean traits at observed points and extrapolate

this information to a wider area using both Random Forest Regression and the GAM model. In this process, we also identify the most important environmental drivers explaining the traits, as well as produce standard diagnostics for our models (spatial cross-validation and illustrating the differences in performance parameters between models). In this work, you will need the FD output file (with the mean traits calculated) and relevant environmental layers. You can download data, for example, from Emodnet and Copernicus. Similar to the IDW method, you should model a larger area than the BBT to avoid edge effects.

Step 1. Produce the environmental raster files

Use the script `Processing rasters.R`. The script should be generic enough, but you will likely have to amend it based on your own data. Furthermore, you will need to choose the relevant environmental layers that are likely to affect each of your organism groups. These can include geographical parameters (depth, slope, distance to shore), environmental parameters (temperature, salinity, oxygen concentration), or biological parameters (chl_a, primary productivity, presence of habitat-forming species). The raster files can include continuous variables or categorical variables. In the latter case, the layer may require some post-processing (see `RF vs GAM new.R`).

The raster files produced will all have the same CRS and resolution, which is important in the next step. If necessary, the files can be reprojected or resampled to the correct CRS and resolution.

Step 2. Build and validate the RF and GAM models and create spatial predictions

Use the script `RF vs GAM new.R`. The script will guide you through the steps to import the FD data, the necessary environmental layers, as well as all validation steps. The resulting raster files (`rf_prediction_[trait].tif` and `gam_prediction_[trait].tif`) will be saved in the outputs folder. The procedure should be repeated for all traits. Based on the diagnostics, you can choose the modelling technique (GAM or RF) that best suits your case.

Step 3. Calculating the mean trait biogeography for the BBT hexagon grid

Import the raster layers you produced to QGIS and repeat step 3 from the IDW section. This should produce a trait distribution map based on the model output (e.g., Fig. 13). The steps should be repeated for all traits and organism groups.



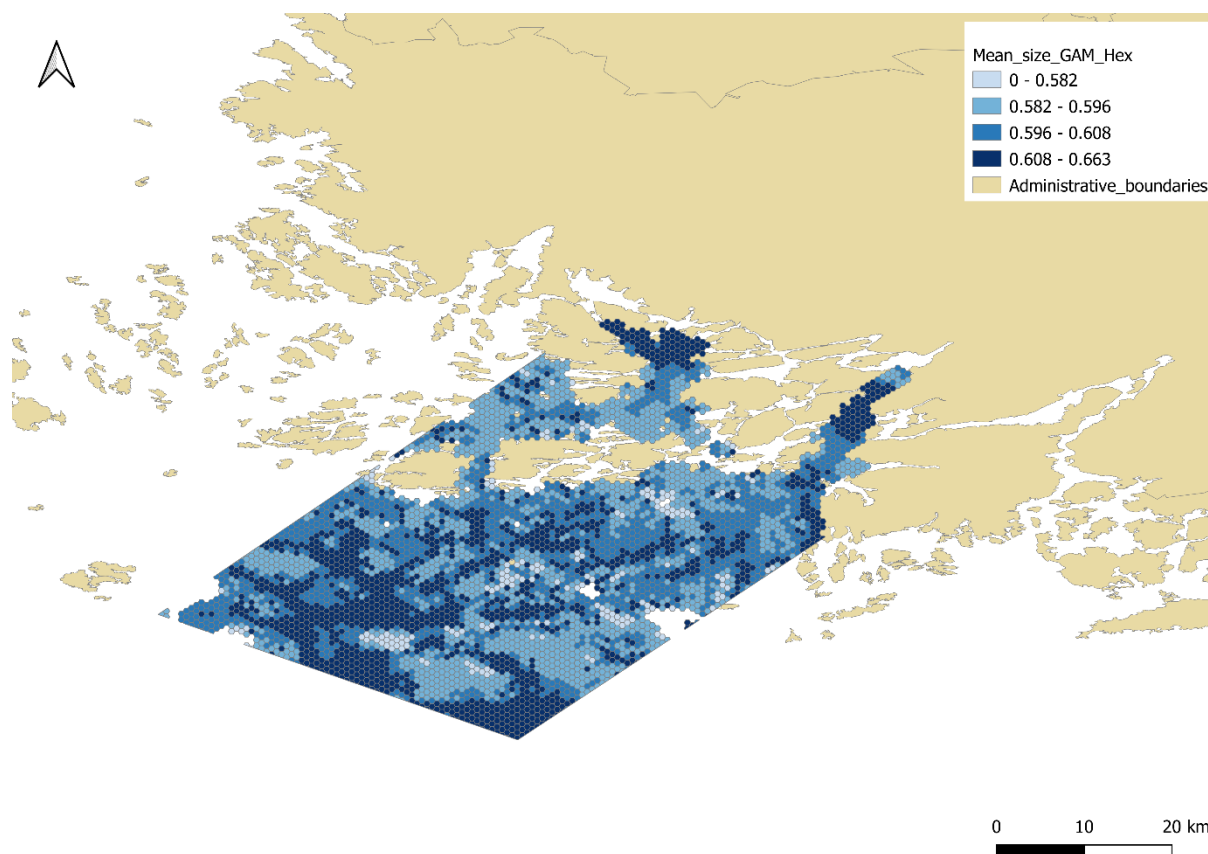


Figure 13. The GAM model output shown using the BBT hexagon. Note that the empty cells are a result of the environmental layers used.



MARBEFES project has received funding from the European Union's Horizon Europe research and innovation programme under

Grant Agreement no 101060937



4. REFERENCES

- Brun, P., Payne, M.R., Kiørboe, T. (2016). Trait biogeography of marine copepods – an analysis across scales. <https://doi.org/10.1111/ele.12688>
- Frelat, R., Beukhof, E., Weigel, B. & Lindegren, M. (2022). *Tutorial for analysing trait-environment relationships*, Zenodo, [10.5281/zenodo.6712534](https://doi.org/10.5281/zenodo.6712534).
- de Juan, S., Bremner, J., Hewitt, J., Törnroos, A., Mangano, M.C., Thrush, S., Hinz, H. (2022). Biological traits approaches in benthic marine ecology: Dead ends and new paths. *Ecol. Evol.* 12:e9001.
- Laliberté, E., Legendre, P., Shipley, B. (2022). Measuring functional diversity (FD) from multiple traits, and other tools for functional ecology. pp 0-12, package version 1.0-12.1.
- Mouchet, M.A., Villéger, S., Mason, N.W.H., Mouillot, D. (2010). Functional diversity measures: an overview of their redundancy and their ability to discriminate community assembly rules. *Funct. Ecol.* 24: 867–876.
- Salo, T., Rinne, H., Rancken, E., Blanc, J.-F., Salovius-Laurén, S., Nordström, M.C. (2023). Environment- and scale-dependent changes in the functioning of invertebrate communities associated with *Fucus vesiculosus*. *Est. Coast. Self. Sci.* 290: 108411.
- Shepard, Donald (1968). "[A two-dimensional interpolation function for irregularly-spaced data](#)". *Proceedings of the 1968 ACM National Conference*. pp. 517–524. [doi:10.1145/800186.810616](https://doi.org/10.1145/800186.810616).
- Törnroos, A., Bonsdorff, E. (2012). Developing the multitrait concept for functional diversity: Lessons from a system rich in functions but poor in species. *Ecol. Applic.* 22:2221-2236.
- van Denderen, P.D., Bolam, S.G., Friedland, R., Hiddink, J.G., Norén, K., Rijnsdorp, A.D., Sköld, M., Törnroos, A., Virtanen, E.A., Valanko, S. (2019). Evaluating impacts of bottom trawling and hypoxia on benthic communities at the local, habitat, and regional scale using a modelling approach. *ICES Journal of Marine Science*. [doi:10.1093/icesjms/fsz219](https://doi.org/10.1093/icesjms/fsz219)
- Villéger, S., Mason, N.W.H., Mouillot, D. (2008). New multidimensional functional diversity indices for a multifaceted framework in functional ecology. *Ecology* 89: 2290–2301.



SUPPLEMENTS

R scripts with a link to Nextcloud....



MARBEFES project has received funding
from the European Union's Horizon Europe research and
innovation programme under Grant Agreement no 101060937
and
from UKRI under Grant Agreement no. 10048815





MARBEFES project has received funding
from the European Union's Horizon Europe research and
innovation programme under Grant Agreement no 101060937
and
from UKRI under Grant Agreement no. 10048815

