User Manual BENMMI

D.J.J. Walvoort
Wageningen Environmental Research
Wageningen, The Netherlands; e-mail: dennis.walvoort@wur.nl

W.M.G.M. van Loon
Rijkswaterstaat Water, Transport and Living Environment; Department of Information Management
Lelystad, The Netherlands; e-mail: willem.van.loon@rws.nl

2019-01-10

Contents

1 Introduction

This tutorial provides a brief introduction to the **BENMMI**-package. This package facilitates the following types of benthos data analysis:

1. Quality control of benthos data
2. Single index and multimetric index (MMI) optimizations against a pressure gradient
3. Routine benthos assessments using an optimized index or MMI
4. Production of an analysis report and result files

The MMI is a weighted linear combination of the environmental quality ratio's (EQR) of a subset of the following indices:

- total abundance (N) and its natural logarithm (LNN);
- species richness (S);
- Margalef's diversity index (D);
- Rygg's index (SN, Rygg (2006); Van Loon *et al.*, 2017; Appendix H);
- Adjusted version of Rygg's index (SNA, Appendix H);
- Simpson's index (L);
- Hurlbert's Probability of Interspecific Encounter (PIE) (PIE = 1-L);
- Hill's diversity number N2 (N2 = 1/L);
- Shannon's index (H');
- AZTI Marine Biotic Index (AMBI);
- Infaunal Trophic Index (ITI);

The maximum number of indices which can be used in a BENMMI-session is three. The weights are optimized in order to maximize the correlation between the MMI and a user specified pressure index. Both measured pressure data (*e.g.*, fisheries pressure, oxygen saturation) or a estimated human pressure data (using *e.g.*, multiple information sources) can be used. It is even possible to use dates (YYYY-MM-DD, e.g. 2017-01-24) in the pressure column[1]. This gives the temporal trend of each index.
The package includes five additional optional features that enhance data preprocessing:

- data pooling: data from small samples are combined to bigger samples with a standardized size to (a) meet the data requirements of the AMBI, (b) generate comparable species richness values and (c) give a higher benthos signal to noise ratio.

[1] BENMMI will convert the date to a numerical value by using the following expression: year + (day_of_year - 0.5) / days_in_year. In case of date '2017-01-24', the numerical value is $2017 + \frac{24-0.5}{365} = 2017.064$

- species name conversion: the tool automatically converts the synonym names into standardized species names using a conversion table (Appendix E) which is based on the WoRMS species names list (www.marinespecies.org).
- genus to species conversion: taxa counts at the taxonomic genus level can optionally be converted to the species level. It is assumed that the unidentified taxa at the genus level can be proportionally distributed over the identified taxa at the species level in the same sample;
- Non-metric multidimensional scaling (Sammon, 1969) of the Bray-Curtis dissimilarity (distance). The results are presented in a plot that can be used for the detection of potential outliers.
- Quality control by fitting $S - 1$ versus $\log(N)$. Samples that deviate from this fit are potential outliers.

In addition, the BENMMI tool offers additional MMI optimization settings:

- choice of maximum three indices;
- setting for (a) optimization or (b) use of fixed index weight factors;
- setting of the confidence interval;
- description of the pressure type (e.g., fisheries pressure, oxygen saturation, sediment load, or even date) and pressure unit.

If you're not familiar with R, and don't know how to install R and the BENMMI-package, you should consult the R-website (www.R-project.org) or read the installation guide provided with the BENMMI-package.

2 Quick-start

The **BENMMI**-package is loaded by typing

```
library(BENMMI)
```

in the R-console followed by pressing the Return/Enter key.
The workhorse function of the **BENMMI**-package is called `BENMMI`. This function performs a full BENMMI-analysis. It reads all its inputs from, and stores all its outputs to files. By default, the BENMMI-package uses the directory structure in Figure 1.
The default directory structure in Figure 1 can be created and populated with sample files by means of the `BENMMIdir` function. This function starts after typing

```
BENMMIdir()
```

in the R-console. A directory selection dialogue starts to let you set an (existing, but empty!) working directory interactively. In non-interactive mode, the path to the (existing, but empty) working directory should be supplied as argument, *e.g.*:

```
BENMMIdir(path = "c:/myprojects/BENMMI/BENMMI_FILES")
```

Note that paths are separated by (forward) slashes. The working directory needs to be empty otherwise the function does not copy files.
After running `BENMMIdir` two directories have been created:

- INPUT-FILES: a directory containing BENMMI-input files. These files contain the number of taxa that have been found in each sample. See Appendix A for more information.
- REF-FILES: a directory containing BENMMI-reference files. These files contain information on OBJECTIDs/HABITATs (Appendix D) and species sensitivities (AMBI, Appendix B, ITI, Appendix C), and a recent copy of the Taxa Water management of the Netherlands (TWN) list. The latest version of this list can be downloaded from sofus.ecosys.nl/taxabase.htm

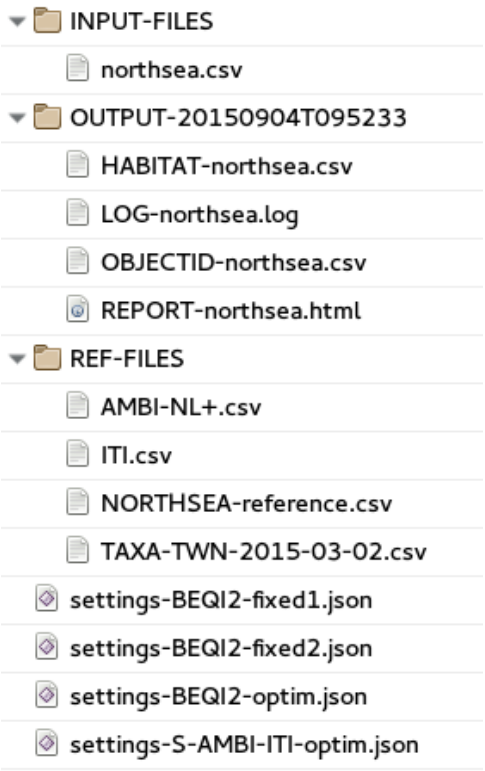


Figure 1: Default BENMMI-directory structure.

In addition, a set of files with extension '`json`' has been created. Each json-file contains the settings for a specific BENMMI-run. See Section 3.1 for more information.
The BENMMI-tool can be started by typing

```
BENMMI()
```

in the R-console. This launches an interactive file selection dialogue. The user is asked to select a settings (*.json) file.
Alternatively, one may also provide the name of the settings file as function argument:

```
BENMMI(filename = "c:/myprojects/BENMMI/BENMMI_FILES/settings-S-D-lin.json")
```

After the BENMMI-run is completed, all results are available in a directory with prefix 'OUTPUT' and the current date-time stamp as postfix. In addition, your default web-browser is launched showing the analysis report.

3 The BENMMI-package in more detail

This section gives an overview of all analysis performed by the BENMMI package. For convenience, the order of the subsections follow that in the BENMMI-report.

3.1 Data Files and Settings

A BENMMI-run is entirely specified by the contents of a JSON-file (see '`*.json`' in Figure 1). The format of this file is JavaScript Object Notation (JSON). This is a well structured, human-readable, open standard format (`www.json.org`). To improve readability, comments are allowed as an extension to the JSON-standard. Text after two (forward) slashes (//) is interpreted as comments and will be ignored. The figure below gives an example of a settings file.
Most editors support editing JSON files, including the built-in editor of the Rgui for MS-Windows (see main menu: `File | Open script...`).

```
{
    "title": "MMI analysis 'North Sea'",
    "user": "Jan de Ruiter",
    "date": "2016-12-21",
    "files": {
        "benthos": "INPUT-FILES/DATA-SNS-NL-1000-surface-OBJECTID-YEAR.csv",
        "taxa": "REF-FILES/TAXA-BE-DE-NL-UK-2017-01-06.csv",
        "groupsToExclude": "REF-FILES/TAXONOMIC-GROUPS-EXCLUDED.csv",
        "Habitats": "REF-FILES/AREAS-HABITATS-SNS-2016-11-27.csv",
        "AMBI": "REF-FILES/AMBI-NL+.csv",
        "ITI": "REF-FILES/ITI+carnivores-2015-10-23.csv"
    },
    "indices": ["S", "D"],
    "weights": null,
    "confidenceLevel": 0.90,
    "pressure": {
        "name": "Fishing activity",
        "unit": "subsurface, sweeps/year"
    },
    "months": [1,12],
    "pooling": {
        "enabled": false,
        "randomSeed": 314,
        "targetArea": [0.09, 0.11]
    },
    "genusToSpeciesConversion": true,
    "model": "linear",
    "legendText": "normalized"
}
```

The list below briefly describes each key in the JSON-file:

- *title*: the title of the BENMMI run;
- *user*: name(s) and affiliation(s) of the analyst(s);
- *date*: date when the JSON-file was written;
- *files*: the paths to each input file:
  - *benthos*: the benthos-input file (see Appendix A for details);
  - *taxa*: Taxa Water management of the Netherlands (TWN) list, which is fully based on the WoRMS species list. A recent version of this list is included in the BENMMI-package. The latest version of this list can be downloaded from `sofus.ecosys.nl/taxabase.htm`;
  - *groupsToExclude*: Taxonomic groups that should be removed from the benthos-input file (see Appendix E for details);
  - *Habitats*: Habitats Reference file (see Appendix D for details);
  - *AMBI*: optional user defined AMBI-file (see Appendix B for details);
  - *ITI*: optional user defined ITI-file (see Appendix C for details);

  Note: optional files can be excluded from analysis by setting its value to null (lower-case!), or by removing the line from the JSON file, or by commenting this line out by C-style comments (//);
- *indices*: the indices to include in the MMI (a subset of N, LNN, S, D, SN, SNA, L, PIE, N2, H, AMBI and ITI). Maximum three indices are allowed. Note: although key-name 'indices' is recommended, the key-name 'indicators' is still allowed for backward compatibility;
- *weights*: setting weights to null, *i.e.*, `"weights": null` (see first JSON-file), the weight factors of the MMI will be optimized, which is the main aim of BENMMI. In addition, BENMMI can use fixed weight factors for each selected index. Weights have to be given as numeric values like `[0.333, 0.333, 0.333]`, or as fractions `["1/3", "1/3", "1/3"]` (see second JSON-file). Note that fractions should always be quoted. The number of weights should be identical to the number of indices;
- *confidenceLevel*: The BENMMI tool will also compute confidence intervals. `confidenceLevel` gives the confidence level. This should be a numeric value between 0.5 and 0.99;
- *pressure*: the name and unit of the pressure in the BENMMI-file (Section A);

- *months*: integer vector of length 2 containing the first and last months to analyse. *E.g.*, [6, 10] means: 'analyse all data from June to October, inclusive';
- *pooling (see Section 3.8)*:
  - *enabled*: is pooling enabled? [true, false];
  - *RandomSeed*: seed to initialize the pseudo random number generator;
  - *TargetArea*: samples are combined until the total area is within this range (numeric vector of length 2; units: m^2);
- *GenusToSpeciesConversion* (see Section 1): is genus to species conversion enabled? [true, false].
- *model*: type of model to fit to the index-pressure relation. Two models are implemented
  - 'linear': simple linear regression model
  - 'exponential': exponential model
- *legendText*: Text to be used in figures: either 'EQR' or 'normalized'

3.2 Selection of benthos records

After the tool has been started with function `BENMMI()`, the BENMMI-file (Section A) is read and records outside the specified time-frame (see 'months'-key in the JSON-file), and non-endofauna records are removed.

3.3 Taxonomic groups

In this section of the BENMMI-report, the average abundance (both as counts and as percentages) of taxa in dominant taxonomic groups are graphically presented for each combination of OBJECTID-HABITAT-YEAR.

3.4 Conversion of species names

All taxa in the BENMMI-file are looked up in the species names file. Species not found are reported in a table. Occasionally it is possible to suggest a taxon name that is very similar (but not identical) to the names in the species names file. In that case, this name is reported as suggested name. The analyst can use this information to correct potential typing errors.

3.5 Species sensitivity values

Species sensitivity values are read, first from the user supplied species sensitivity file (Sections B and C). Species sensitivity values that are still missing are taken from Borja et al., (2000). The tool reports for which taxa species sensitivity values are missing. The species sensitivity values are used to compute the AMBI-index (Borja *et al.*, 2000) or the Infaunal Trophic Index (ITI)-index (Van Loon *et al.*, 2015).

3.6 OBJECTID-HABITATs and sample areas

For each combination of OBJECTID-HABITAT, an overview of available sampling areas is tabulated. An OBJECTID usually is a national area, *e.g.*, the Dutch Borkum reef area. This table can be used to judge the efficacy of pooling.

3.7 Conversion of genus to species within a single sample

If enabled in the JSON-file, the tool converts taxa at the taxanomic genus level to the species level. It is assumed that the unidentified taxa at the genus level can be proportionally distributed over the identified taxa at the species level in the same sample.

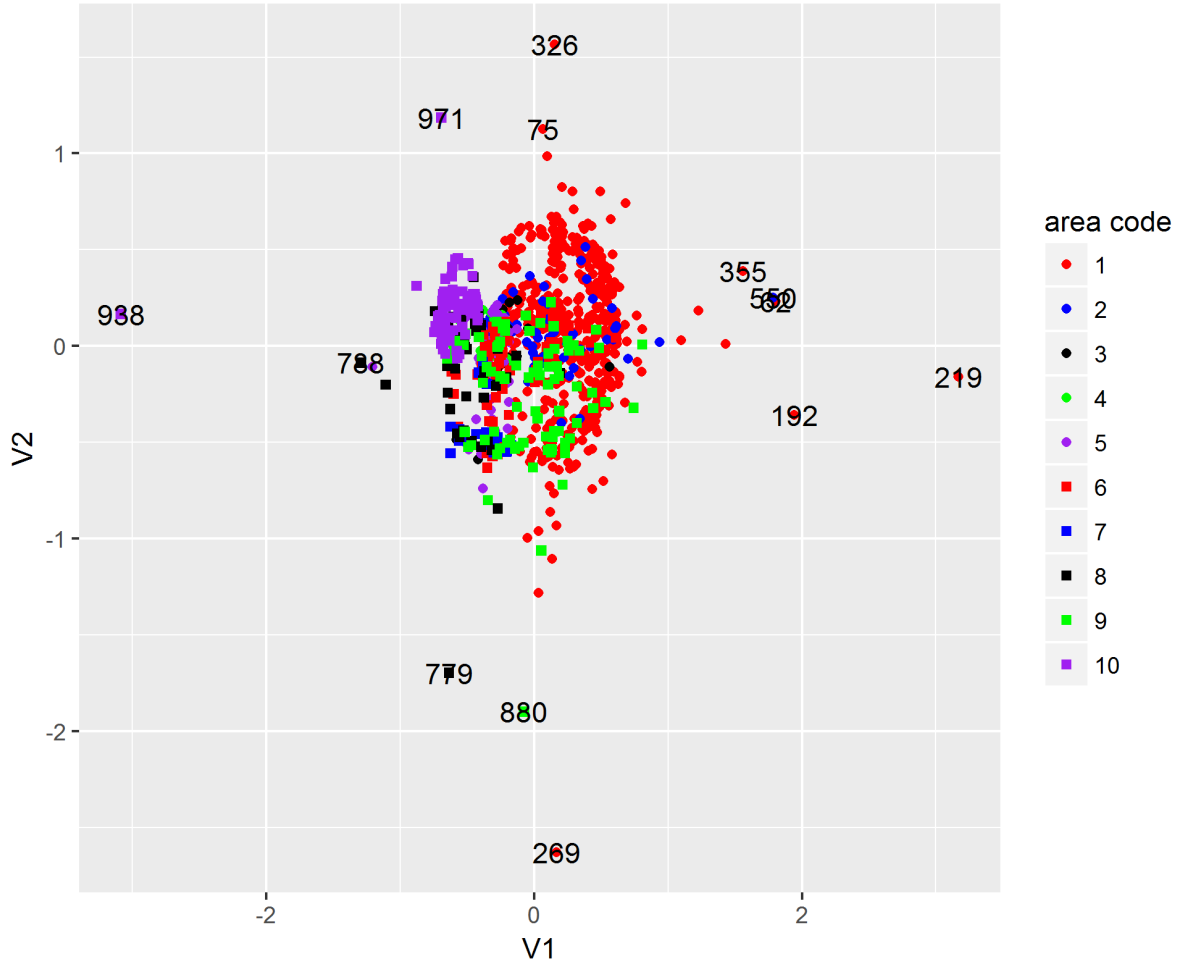


Figure 2: Example of a BENMMI multidimensional scaling plot with area codes: 1 = BE_NorthSea-Coarse; 2 = BE_NorthSea-Sand; 3 = DE_Coastal-Sand; 4 = DE_SylterOuterReef-Coarse; 5 = DE_SylterOuterReef-Sand; 6 = NL_CoastalZone-Sand; 7 = NL_DoggerBank-Sand; 8 = NL_FrysianFront-Sand; 9 = NL_Offshore-Sand; 10 = NL_OysterBanks-Mud. The labels in the plot refer to individual samples.

3.8 Data pooling

If enabled in the JSON-file, data from small samples are combined to bigger samples with a specified size to (a) meet the data requirements of the AMBI, (b) generate comparable species richness values and (c) give a higher benthos signal to noise ratio.

3.9 Multidimensional scaling (MDS)

This section provides a non-metric multidimensional scaling plot of the Bray-Curtis dissimilarity (distance). The algorithm used is that of Sammon (1969). The MDS-plot can be used to identify potential outlier samples. A table with information on potential outlier samples is given for convenience. MDS is only applied if the pooling option is disabled. Note that MDS is slow for large data sets. An example of an MDS-plot is give in Figure 2.

3.10 Index calculation

After all steps above are carried out, a maximum of three user-specified marine benthos indices are estimated.

3.11 Index percentile values

To get information on the distribution of each index in each OBJECTID-HABITAT, percentile values are given. These percentiles may also be useful to set the reference values of the Index Ecological Quality Ratios (Section 3.12).

3.12 Index Ecological Quality Ratios (normalization)

Finally, the index values are converted to ecological quality ratio's (EQR). The general EQR-formula for index I is:

$$\mathrm{EQR}(I) = \frac{I_{\mathrm{ass}} - I_{\mathrm{bad}}}{I_{\mathrm{ref}} - I_{\mathrm{bad}}}$$

where I_{ass} is the estimated index value, and I_{ref} and I_{bad} are its values for a reference status and a bad status respectively, which need to be specified in the habitat reference file (Appendix D). To estimate reference values, BENMMI should be run with best guess values first. After that, the reference values can be updated by means of the percentiles given in the BENMMI-report (see Van Loon et al., 2015 and 2017 for details about this procedure).
Depending on 'bad' and 'ref', the EQR usually (but not necessarily!) varies between 0 (bad ecological quality) and 1 (reference ecological quality).

3.13 Results: optimization and aggregation

This section contains all the results.

3.13.1 Study area

This section contains a map of the sampling locations in the study area.

3.13.2 Correlations

This section contains scatter plots for all pairs of indices. An example is given in Figure 3. This figure gives the relation between Margalef's diversity index (D) and species richness (S).

3.13.3 Optimization

The model we want to optimize is:

$$\mathrm{MMI}^* = f(P_1, P_2, \ldots)$$

Where MMI* is the EQR (or normalized version) of the multimetric index MMI, P_1, P_2 are pressures, and $f(\cdot)$ means 'is a function of'.
If we assume a linear model and only one pressure P_1 this model simplifies to

$$\mathrm{MMI}^* = b_0 + b_1 \times P_1$$

Where b_0 and b_1 are the intercept and the slope respectively.
MMI* is a weighted linear combination of the benthic indices:

$$\mathrm{MMI}^* = w_1 \times index1^* + w_2 \times index2^* + \ldots$$

where the weights are nonnegative ($w_i \geq 0$) and sum to one ($\sum w_i = 1.$)
Suppose the number of indices in the MMI and therefore the number of weights equals 3. Then the aim is to find that set of weights w_1, w_2, w_3, b_0, b_1 that maximizes r^2_{adj} (explained variance of the model), given the constraints that the weights w_i are nonnegative and sum to one.
For a single metric index, ordinary least squares is used for optimization. For a bimetric index, Brent's method (combination of golden section search and successive parabolic interpolation) is used, and for a trimetric index the Downhill simplex method by Nelder & Mead is used. See www.nr.com for more information on these optimizaton methods.
In case an exponential model is specified in the settings file, the following model will be fitted by means of non-linear least squares (Gauss-Newton algorithm as implemented in the nls-function of R):

$$\mathrm{MMI}^* = b_0 + b_1 \times \exp(b_2 \times P_1)$$

The estimation of the weights is identical to that of fitting a linear model.

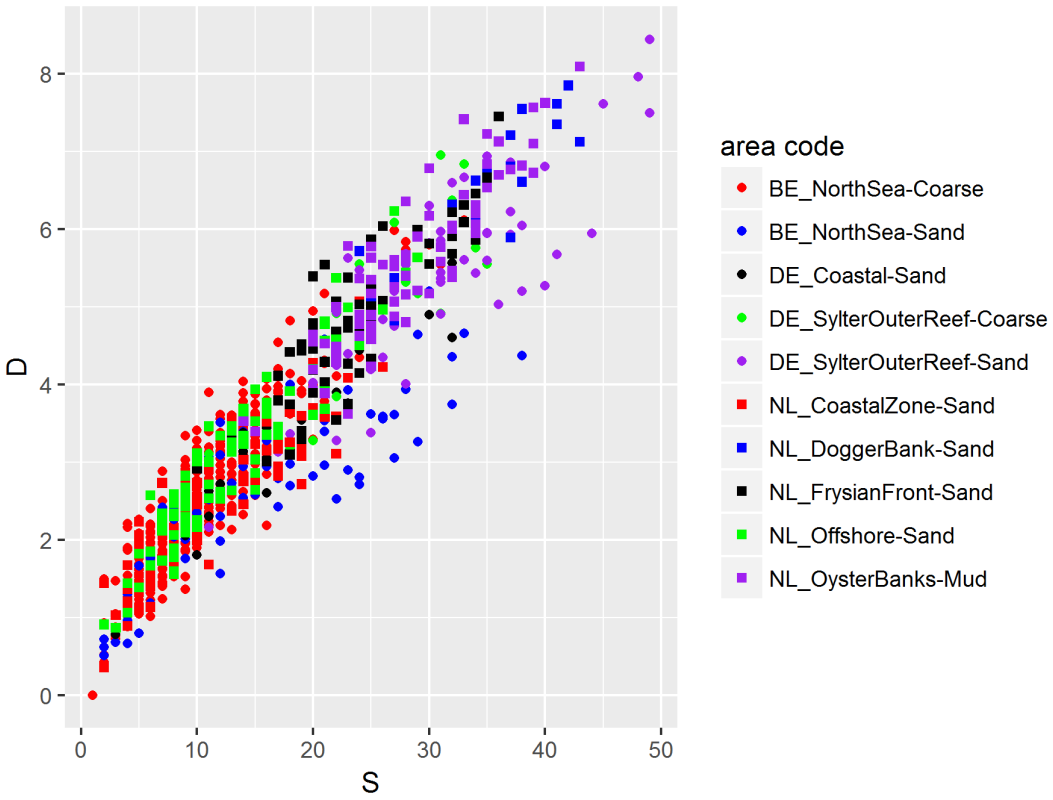


Figure 3: Scatter plot of Margalef's diversity index (D) versus species richness (S). The colours refer to different areas.

3.13.4 Aggregation

The final section gives averaged indices for all combinations of OBJECTID-HABITAT-YEAR including confidence limits.

3.14 Appendices

The appendices are described below.

3.14.1 Quality control plot based on Margalef D

In this section, plots are provided of $S-1$ as function of $\log(N)$. See Figure 4 for an example. These plots can be used to identify potential outliers. The assumption is that the relation between $S-1$ and $\log(N)$ is approximately linear. Note that the ratio of $S-1$ and $\log(N)$ is Margalefs index of diversity.

3.14.2 Confidence interval versus sample size

Plots of precision as function of sample size are provided for each combination of OBJECTID and HABITAT (following the method outlined in Knotters and Brus, 2012). These plots and accompanying table provide the user an estimate of the minimum sample size needed to obtain an estimate of the mean EQR with a required confidence interval.

3.15 Output files

After running the tool, the following results are available:

1. a report in HTML-format, containing all sections given above (file prefix: 'REPORT');
2. a file with prefix 'TIDY', which is a version of the benthos-input file (Appendix A) that has been cleaned up by BENMMI prior to calculating indices. Cleaning up includes:

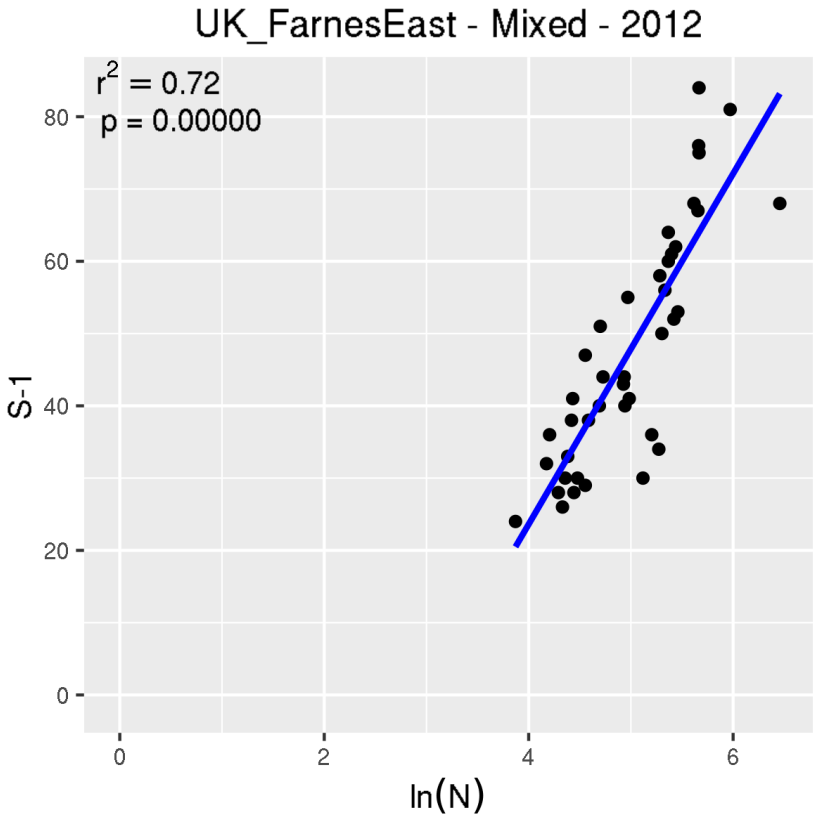


Figure 4: Example of a quality control plot based on Margalef D.

(a) removing months outside the period of interest (Section 3.2);

(b) removing non-endofauna taxa (Section 3.2);

(c) making taxon names WoRMS-compliant (Section 3.4);

(d) optionally: genus to species conversion (Section 3.7);

(e) optionally: data pooling (Section 3.8). In case data pooling has be enabled, two additional columns will be added:

i. POOL_RUN giving the identifier of the pool run (1...10) and

ii. POOL_ID giving the identifier of a pool within a POOL_RUN.

3. output files with results at three spatial scales:

(a) sampling unit level (file prefix: 'SAMPLE')

(b) habitat level (file prefix: 'HABITAT')

(c) area level (file prefix: 'OBJECTID')

4. a log-file with informative, warning, and error messages (file prefix: 'LOG');

5. an optional file with pooling information (file prefix: 'POOLING');

4 References

1. David Ardia, Katharine M. Mullen, Brian G. Peterson, Joshua Ulrich (2016). 'DEoptim': Differential Evolution in 'R'. version 2.2-4. https://CRAN.R-project.org/package=DEoptim

2. Borja, A., J. Franco and V. Pérez, 2000. A Marine Biotic Index to Establish the Ecological Quality of Soft-Bottom Benthos Within European Estuarine and Coastal Environments. Marine Pollution Bulletin 40:1100-1114

3. Knotters, M. and D.J. Brus, 2012. KRW-monitoring rivieren: analyse van het benodigde aantal macrofaunamonsters. Report 2368, Alterra, Wageningen University and Research Centre, The Netherlands.

4. Rygg, B. (2006). Developing indices for quality-status classification of marine soft-bottom fauna in Norway. Norwegian Institute for Water Research, Oslo, Norway. NIVA Report SNO 5208-2006.

5. Sammon, J. W., 1969. A non-linear mapping for data structure analysis. IEEE Trans. Comput., C-18 401-409.

6. Storn, R. and Price, K. (1997) Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces, Journal of Global Optimization, 11:4, 341-359.

7. Van Loon, W.M.G.M. A.R. Boon, A. Gittenberger, D.J.J. Walvoort, M. Lavaleye, G.C.A. Duineveld, A.J. Verschoor, 2015. Application of the Benthic Ecosystem Quality Index 2 to benthos in Dutch transitional and coastal waters. Journal of Sea Research 103:1-13

8. Willem van Loon, Dennis Walvoort, Gert van Hoey, Christina Vina-Herbon, Abigayil Blandon, Roland Pesch, Petra Schmidt, Jorg Scholle, Karin Heyer, Marc Lavaleye, Graham Philips, Gerard Duineveld, Mats Blomqvist, 2017. A regional benthos assessment method for the Southern North Sea using Margalef diversity and reference value modeling. Accepted for publication by Ecological Indicators.

A BENMMI input file

The format of the BENMMI input file has been specified in the table below. The format is the so called comma-separated values format (CSV) with the following characteristics:

- decimal separator: period (.)
- column separator: comma (,)
- text values are preferably quoted

The following columns are present in the data format: OBJECTID, SAMPLEID, HABITAT, LAT, LONG, TAXON, SAMPDEV, AREA, DATE, VALUE, and PRESSURE. SAMPDEV and PRESSURE are optional, the other columns are required. The order of the columns is invariant. In addition, the use may also add other columns containing additional data (like meta-information). These columns will be ignored by BENMMI.
NB: It should be stressed that the VALUE column should contain taxa-counts and *not* taxa densities.

Item (header in data table)	Comments	Examples
Measuring object id (OBJECTID)		Dogger_bank
Sample-id (SAMPLEID)	Location code or Sample number or transect station etc. Note1: the BENMMI tool combines the sample code with the date to ensure a unique sample code.	North Sea: e.g. NOORDWK2; Dollard: Transect-station code; Delta: 1304
Habitat code (HABITAT)	For the North Sea project EUNIS 3 codes are used, e.g., Sand, Mud, Coarse, Mixed. However, other global habitat descriptions including e.g. salinity class can be used.	Sand
LAT	latitude (decimal degrees)	52.31
LONG	longitude (decimal degrees)	3.17
Taxon name (TAXON)	Standardized taxon code (see WoRMS-website) In NL: the TWN code	
Sampling device (SAMPDEV)	The code of the sampling device.	e.g., BOXCRR of VEENGB
Sampled area (AREA)	The sampling area is necesarry for comparing sample areas and the data pooling process. The unit is m2.	E.g. 0.015 m2.
Sieve mesh size (MESH)	Mesh size of the sieve that has been used to filter sediment/benthos. The mesh size needs to be a unique numerical value. The unit is micrometer.	1000 (commonly used) or 500
Date (DATE)	ISO format: YYYY-MM-DD is mandatory.	2001-03-05
Numerical value (VALUE)	Numerical value representing the number of individuals (NB: not the density!). Decimals are allowed to indicate fractions of individuals..	20.5
pressure (PRESSURE)	Numerical value.	6.23

B AMBI file

This species sensitivity file consists of two columns, and is stored in comma separated file format (CSV). The first column contains the taxa, the second column the corresponding sensitivity classes. The table below gives an example of (part of) a species sensitivity file.

TAXON	GROUP
Hemimysis lamornae	II
Phyllodoce groenlandica	III
Arctica islandica	II
Neomysis integer	III
Molgula manhattensis	III
Upogebia deltaura	II
Pisione remota	II
Einfeldia carbonaria	III
Urothoe poseidonis	II
Alvania lactea	I
Spirorbis tridentatus	II
Liocarcinus marmoreus	II
Aeolidia papillosa	II
Scoloplos armiger	II
Hemigrapsus sanguineus	II
Golfingiidae	II
Bathyporeia elegans	II
Microprotopus maculatus	II
Propebela turricula	II
Caprella mutica	II
Echinocardium cordatum	II
Golfingia	II
Aora gracilis	II
Pilumnus hirtellus	II
Ensis minor	II

For most of the taxa that are not present in the AMBI user file, the (built-in) default AMBI file from AZTI will be used. Users may append new taxa to the AMBI user file (AMBI-NL+.csv or user-made file), or modify classes of existing taxa using the external AMBI-NL+ or user file.

C ITI file

This species sensitivity file consists of three columns, and is stored in comma separated file format (CSV). The first column contains the taxa, the second column the corresponding sensitivity classes, and the third column specifies if a taxon is a carnivore or not. The table below gives an example of (part of) a species sensitivity file.

TAXON	GROUP	CARNIVORE
Spio	II	FALSE
Gammaropsis maculata	II	FALSE
Glycera capitata	IV	TRUE
Myrianida prolifera	IV	TRUE
Glycera lapidum	IV	TRUE
Cirratulus	III	FALSE
Upogebia deltaura	I	FALSE
Arenicola marina	IV	FALSE
Astropecten irregularis	IV	TRUE
Chaetopterus	I	FALSE
Lanice conchilega	III	FALSE
Glycera unicornis	IV	TRUE
Sipuncula	III	FALSE
Dipolydora caulleryi	II	FALSE
Diastylis bradyi	II	FALSE
Pontocrates arenarius	II	FALSE
Pinnotheres pisum	IV	TRUE
Pisidia longicornis	IV	TRUE
Westwoodilla caecula	II	FALSE
Amphiura	I	FALSE
Harpinia antennaria	IV	FALSE
Astarte montagui	I	FALSE
Tanaidacea	III	FALSE
Pilumnus hirtellus	IV	TRUE
Asteroidea	IV	TRUE

Users who want to add additional taxa to the ITI file are requested to do this via Willem van Loon, in order to keep a standardized ITI file in BENMMI. It is recommended not to change the ITI classifications in the BENMMI ITI file, apart from experimental analyses, in order to maintain comparable ITI results between the users of this tool, in at least the North Sea region. Note that the classification of carnivores has been added purely for information, and that carnivores are not used in the standard ITI calculation.

D Habitat Reference file

The Habitat Reference file contains meta-information about each OBJECTID and HABITAT. In particular, it lists the REF and BAD values for each index for computing the EQRs (see Section 3.12). At least the reference data for the selected indices in the settings-file (JSON-file) should be given. Note: estimated reference values (using the 99-percentile method; Van Loon et al., 2015, 2017) have been added to this file.

OBJECTID	RELAREA	HABITAT	DREF	DBAD	SREF	SBAD	AMBIREF	AMBIBAD	ITIREF	ITIBAD
BE_NorthSea	1	Coarse	5.48	0.00	26.80	0.00	0.39	6	57.00	0
BE_NorthSea	1	Sand	5.30	0.00	37.70	0.00	0.33	6	90.40	0
DE_Coastal	1	Sand	4.84	0.00	29.40	0.00	0.36	6	78.70	0
DE_ElbeUrstromValley	1	Mud	6.20	0.00	30.90	0.00	0.63	6	81.80	0
DE_SyltOuterReef	1	Coarse	6.46	0.00	35.30	0.00	0.13	6	91.20	0
DE_SyltOuterReef	1	Sand	6.75	0.00	39.80	0.00	0.49	6	76.20	0
NL_CoastalZone	1	Sand	5.20	0.00	29.80	0.00	0.33	6	83.90	0
NL_DoggerBank	1	Sand	7.60	0.00	42.00	0.00	0.54	6	85.90	0
NL_FrysianFront	1	Sand	7.60	0.00	40.00	0.00	0.68	6	86.80	0
NL_Offshore	1	Sand	6.26	0.00	31.80	0.00	0.67	6	79.70	0
NL_OysterBanks	1	Mud	7.68	0.00	37.00	0.00	1.09	6	85.70	0
UK_DoggerBank	1	Sand	5.38	0.00	27.00	0.00	0.88	6	76.90	0

The column RELAREA is the spatial area of each OBJECTID/HABITAT. This area may be given either in km^2 or as a fraction (but not both). BENMMI uses the relative area to estimate the HABITAT area weighted average for each OBJECTID and YEAR. In the table above, each OBJECTID/HABITAT is equally weighted.

E Taxa-list and taxonomic groups to exclude

File 'TAXONOMIC-GROUPS-EXCLUDED.csv', is located in the REF-FILES directory and contains all taxonomic groups that should be excluded from analysis. This file should contain zero or more records. The default table is given below.

GROUP	DESCRIPTION
CRDEC	Crustacea - Decapoda
CRMYS	Crustacea - Mysida

All taxonomic groups in this table will be removed from the BENMMI-input file (Appendix A) prior to analysis.
The taxonomic group (GROUP-column) in the table above should correspond to the GROUP-column in the taxa-list which is also in the REF-files directory (file name: 'TAXA-BE-DE-NL-UK-2017-01-06.csv'). The first 10 records of the taxa-list are given below.
The taxa list is a simple conversion table, based on WoRMS, that converts provided taxa (see TAXON-column in the BENMMI-input file (Appendix A)) to accepted taxa. This table also contains the taxonomic groups, the taxonomic levels, and a quality code (*i.e.*, 10 = 'Preferred name', 20 = 'Synonym', 80 = 'Non-taxonomic species group').

group	provided	accepted	level	quality_code
MAREM	Anopla	Anopla	Classis	10
MAREM	Anthozoa	Anthozoa	Classis	10
APOLI	Aphanoneura	Aphanoneura	Classis	10
ARACH	Arachnida	Arachnida	Classis	10
MAREM	Ascidiacea	Ascidiacea	Classis	10
ECHIN	Asteroidea	Asteroidea	Classis	10
MOBIV	Bivalvia	Bivalvia	Classis	10
APOLI	Branchiobdellea	Branchiobdellea	Classis	10
BRHYP	Calcarea	Calcarea	Classis	10
MOREM	Caudofoveata	Caudofoveata	Classis	10

F Distribution of taxa over taxonomic groups file

The taxonomic group file in the OUTPUT directory gives for each sample the percentage of taxa that are in each taxonomic group. An example of a small part of this table is given below

OBJECTID	SAMPLEID	HABITAT	DATE	YEAR	N	APOLI	APPOL	APTUR	BRHYP	CRAMP
NL_CoastalZone	BREEVTN21	Sand	2009-04-02	2009	7	0	43	0	0	29
NL_CoastalZone	BREEVTN21	Sand	2010-03-01	2010	23	0	39	0	0	9
NL_CoastalZone	BREEVTN21	Sand	2012-03-26	2012	8	0	88	0	0	0
NL_CoastalZone	BREEVTN24	Sand	2009-03-10	2009	7	0	71	0	0	29
NL_CoastalZone	BREEVTN24	Sand	2010-03-02	2010	82	0	1	0	0	38
NL_CoastalZone	BREEVTN24	Sand	2012-03-26	2012	20	0	25	0	0	55
NL_CoastalZone	EGMAZE1	Sand	2009-03-02	2009	17	0	94	0	0	6
NL_CoastalZone	EGMAZE1	Sand	2010-03-11	2010	384	0	96	0	0	1
NL_CoastalZone	EGMAZE1	Sand	2012-03-28	2012	24	0	75	0	0	8
NL_CoastalZone	HOLLSKT02	Sand	2009-03-03	2009	30	0	57	0	0	13
NL_CoastalZone	HOLLSKT02	Sand	2010-03-11	2010	348	0	12	0	0	1
NL_CoastalZone	HOLLSKT02	Sand	2012-03-29	2012	75	0	19	0	0	0
NL_CoastalZone	HOLLSKT03	Sand	2009-03-02	2009	153	0	48	0	0	20
NL_CoastalZone	HOLLSKT03	Sand	2010-03-11	2010	282	0	78	0	0	12
NL_CoastalZone	HOLLSKT03	Sand	2012-03-09	2012	149	0	37	0	0	15
NL_CoastalZone	HOLLSKT04	Sand	2009-03-02	2009	156	0	18	0	0	67
NL_CoastalZone	HOLLSKT04	Sand	2010-03-11	2010	128	0	31	0	0	45
NL_CoastalZone	HOLLSKT04	Sand	2012-03-28	2012	117	0	18	0	0	76
NL_CoastalZone	NOORDWK10	Sand	2009-03-06	2009	29	0	41	0	0	38
NL_CoastalZone	NOORDWK10	Sand	2010-03-15	2010	73	0	88	0	0	11

G Distribution of taxa over ITI-groups

The ITI-group file in the OUTPUT directory gives for each sample the percentage of taxa that are in each ITI-group. An example of a small part of this table is given below

OBJECTID	SAMPLEID	HABITAT	DATE	YEAR	N	carnivore	I	II	III	IV	NA
NL_CoastalZone	BREEVTN21	Sand	2009-04-02	2009	7	57	14	14	0	14	0
NL_CoastalZone	BREEVTN21	Sand	2010-03-01	2010	23	43	48	0	0	9	0
NL_CoastalZone	BREEVTN21	Sand	2012-03-26	2012	8	62	12	12	0	12	0
NL_CoastalZone	BREEVTN24	Sand	2009-03-10	2009	7	71	0	14	0	14	0
NL_CoastalZone	BREEVTN24	Sand	2010-03-02	2010	82	52	0	0	0	46	1
NL_CoastalZone	BREEVTN24	Sand	2012-03-26	2012	20	25	15	10	0	50	0
NL_CoastalZone	EGMAZE1	Sand	2009-03-02	2009	17	94	0	0	0	6	0
NL_CoastalZone	EGMAZE1	Sand	2010-03-11	2010	384	7	2	2	89	1	0
NL_CoastalZone	EGMAZE1	Sand	2012-03-28	2012	24	42	8	21	21	8	0
NL_CoastalZone	HOLLSKT02	Sand	2009-03-03	2009	30	37	0	27	20	17	0
NL_CoastalZone	HOLLSKT02	Sand	2010-03-11	2010	348	6	85	4	5	1	0
NL_CoastalZone	HOLLSKT02	Sand	2012-03-29	2012	75	8	69	5	16	1	0
NL_CoastalZone	HOLLSKT03	Sand	2009-03-02	2009	153	13	24	14	16	29	3
NL_CoastalZone	HOLLSKT03	Sand	2010-03-11	2010	282	5	6	1	9	79	0
NL_CoastalZone	HOLLSKT03	Sand	2012-03-09	2012	149	6	36	22	16	19	0
NL_CoastalZone	HOLLSKT04	Sand	2009-03-02	2009	156	5	6	3	14	72	0
NL_CoastalZone	HOLLSKT04	Sand	2010-03-11	2010	128	10	15	3	25	44	3
NL_CoastalZone	HOLLSKT04	Sand	2012-03-28	2012	117	8	3	2	11	76	0
NL_CoastalZone	NOORDWK10	Sand	2009-03-06	2009	29	31	17	7	3	38	3
NL_CoastalZone	NOORDWK10	Sand	2010-03-15	2010	73	40	1	4	44	5	5

H Rygg's index of diversity

Species richness S is strongly dependent on sampling size. Rygg's index of diversity (SN) takes sampling size into account (Rygg, 2006). It is given by

$$SN = \frac{\ln S}{\ln(\ln(N))}$$

where N is the total abundance, *i.e.*, the total number of individuals in the sampling unit.
Rygg's index shows some inconsistencies for small N and S (notably, for $(S = 2, N = 2)$, $(S = 2, N = 3)$ and $(S = 3, N = 3)$; see the package vignette of the **benthos**-package for more details). To correct for this, we have introduced an *adjusted* version of Rygg's index, SNA. It is given by:

$$SNA = \frac{\ln S}{\ln(\ln(N + 1) + 1)}$$

I Fitting a generalized logistic function

In Van Loon *et al.* (2017, submitted), a generalized logistic function is fitted to index values as function of time. In this Appendix, we will demonstrate how to fit this function to Margalef's diversity index. All source code below can be copied to the R-console and executed.
A generalized logistic function is given by

$$A + \frac{K - A}{1 + \exp\left(-B(x - M)\right)}$$

where A is the lower asymptote, K is the upper asymptote, B is the growth rate, M is a starting time, and x is time.
A generalized logistic function can be written as an R function:

```
f_gl <- function(x, theta) {
    A <- theta[1]
    K <- theta[2]
    B <- theta[3]
    M <- theta[4]
    A + (K - A) / (1 + exp(-B*(x-M)))
}
```

We will use the 'Saltkallefjord' data in Van Loon *et al.* (2017, submitted) as an example. These data reside in the vignette directory of the BENMMI-package and can be loaded as follows:

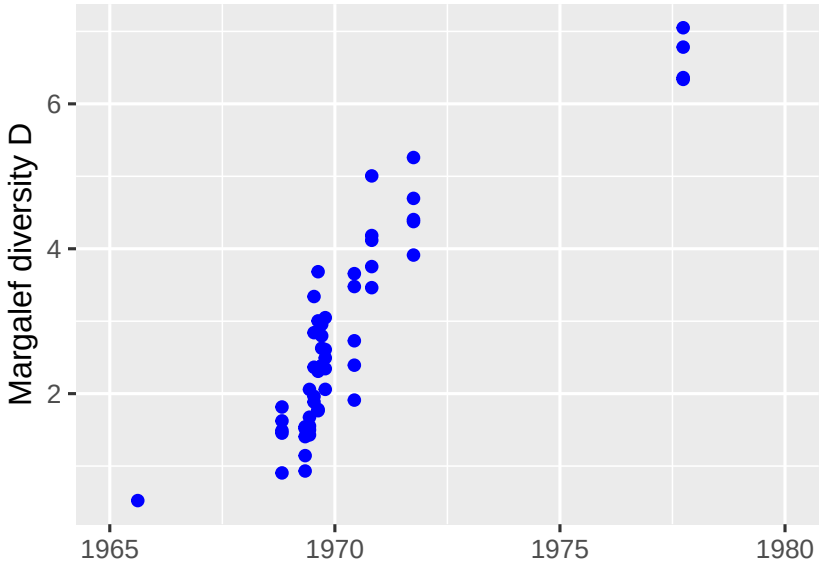
```
# read data
d_obs <- read.csv("./data/saltkallefjord-D.csv", as.is = TRUE)

# coercion from character to Date-object
d_obs$date <- as.Date(d_obs$date)

# print head of these data
head(d_obs)
```

```
        date         D
1 1965-08-16 0.5238484
2 1968-10-29 1.8154095
3 1968-10-29 0.9052148
4 1968-10-29 1.4905530
5 1968-10-29 1.6236225
6 1968-10-29 1.4546409
```

The data are given in the plot below:



To use the `date`-column of these data in our computations, we need to convert it to a numeric value. That can be done as follows:

```
d_obs$ndate <- as.numeric(d_obs$date)
```

Next we have to define an objective function. The objective function describes what exactly we want to optimize and under what constraints.

```
f_obj <- function(theta, data) {
    # constraint: lower asymptote should be nonnegative
    if (theta[1] < 0) {
        return(Inf)
    }

    # predict Margalef diversity by means of the generalised logistic function
    D_hat <- f_gl(data$ndate, theta)

    # difference between Margalef diversity based on
    # observations and the generalised logistic function
    error <- d_obs$D - D_hat

    # our objective to minimize: the sum of squared errors
    sum(error * error)
}
```

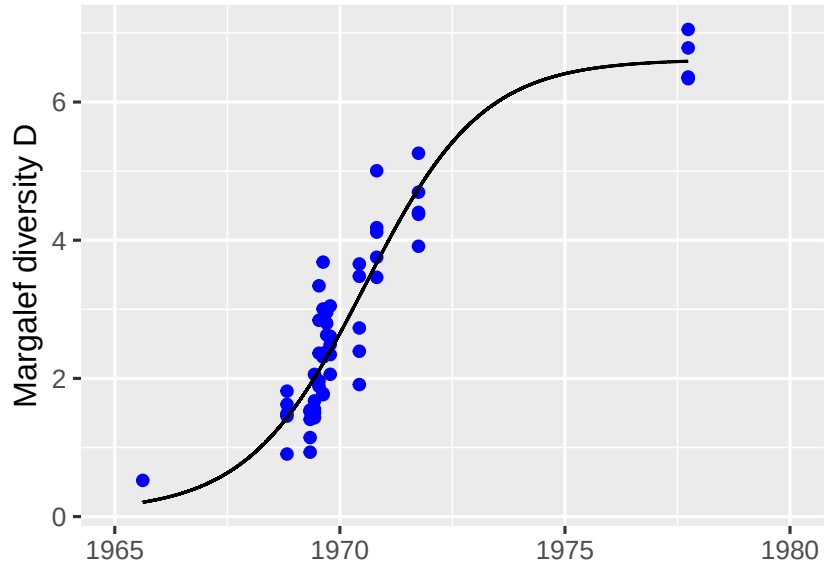
We need an optimization method to minimize this objective function. We found that, for our data, simple general purpose optimization methods like 'Nelder-Mead' were susceptible to local optima. Therefore, we used the global optimization method 'differential evolution' (Storn & Price, 1997, Ardia et al., 2016) instead:

```
# define lower and upper bounds for each parameter
lower_bounds <- c(0, 5, 0.0001, -1000)
upper_bounds <- c(5, 9, 1.0000,  1000)

# DE-optimization
opt <- DEoptim(
    fn = f_obj,
    lower = lower_bounds,
```

```
    upper = upper_bounds,
    data = d_obs,
    control = DEoptim.control(NP = 100, itermax = 100, trace=FALSE)
)
```

The optimal fit is given below:



J Overview of quality checks in BENMMI

The table below gives the most important quality checks implemented in BENMMI. This table focuses on the main user functions of BENMMI: `BENMMI` and `BENMMIdir`. This table is not exhaustive as it doesn't contain the numerous checks in packages it depends on. This is in particularly true for the **benthos**-package. Neither does this table contain the checks that are already documented in the BENMMI-report (this report is the result of calling the function `BENMMI`). Examples are the compliance check with accepted names in the WoRMS-database (`www.marinespecies.org`), the cumulative distributions of species abundances, or plots of species richness versus species abundance.
The table has the following columns:

Topic: The name of a user-function in the **BENMMI**-package, or the name of a BENMMI-file;

Check: A brief description of the quality check;

Action: The action BENMMI takes if the quality check fails: error (execution stops), warning, or informative message;

Comment: Additional information on the quality check.

	Topic	Check	Action	Comment
1	BENMMI	Check the existence of the settings-file.	error	
2	BENMMI	Check if the settings-file contains valid JSON.	error	
3	BENMMI	Check the validity of the JSON-keys in the settings-file.	error	Required versus optional keys.
4	BENMMI	Check the consistency of the JSON-keys in the settings-file.	error	Either indices or indicators, but not both.
5	BENMMI	Check index names in the settings-file.	error	
6	BENMMI	Check the number of indices in the settings-file.	error	
7	BENMMI	Check the format of the weights in the settings-file.	error	Floating point format (0.333) and rational format (1/3) are both allowed.
8	BENMMI	Check the consistency between the weights and the indicators in the settings-file.	error	
9	BENMMI	Check the range of the weights in the settings-file.	error	Should be non-negative.
10	BENMMI	Check the range of the confidence level in the settings-file.	error	Should be in [0.50, 0.99]
11	BENMMI	Check consistency between weights and pressure in settings-file.	error	Weights should be given if pressure is missing.
12	BENMMI	Check if the name and the unit of the pressure are given in the settings-file (if the pressure column has been found in the benthos-file).	warning	BENMMI issues a warning and continues without pressure optimization.
13	BENMMI	Check the range, the order, and the format of the months to analyse in the settings-file.	error	Range should be in {1, 2, ..., 12}, first month <= second month, and format should be an integer vector of length 2.

Continued on next page

	Topic	Check	Action	Comment
14	BENMMI	Check the format of Boolean (true/-false) variables in the settings-file.	error	E.g., enable data pooling, enable genus to species conversion.
15	BENMMI	Check the format and the length of the seed to initialize the pseudo random number generator (needed for pooling).	error	Format should be an integer of length 1
16	BENMMI	Check value of optional key 'legendText'.	error	should be either EQR or normalized.
17	BENMMI	Check if all JSON-keys have been used.	warning	Prevent typographical errors in optional keys.
18	BENMMI	Check the existence of the benthos-file.	error	
19	BENMMI	Check the column names of the benthos-file.	error	
20	BENMMI	Check the uniqueness of the coordinates in the benthos-file.	error	
21	BENMMI	Check the uniqueness of the sieve mesh in the benthos-file.	error	
22	BENMMI	Check if BENMMI can be launched by means of a graphical user's interface.	error	GUIs are usually only implemented on desktops and not on servers.
23	BENMMI	Check if the user presses the CANCEL-button or the ESC-button to cancel execution.	message	
24	BENMMI	Check if the settings-file exists.	error	
25	BENMMI	Check the existence of the species-file.	error	
26	BENMMI	Check if the species-file can be read.	error	
27	BENMMI	Check the existence of the 'taxon-groups-to-exclude file'.	error	
28	BENMMI	Check if the 'taxon-groups-to-exclude file' can be read.	error	
29	BENMMI	Check the existence of the 'benthos-file'.	error	
30	BENMMI	Check if the 'benthos-file' can be read.	error	
31	BENMMI	Check the consistency between weights and pressure.	error	Weights should be given when PRESSURE column is missing.
32	BENMMI	Check the situtation when pressure has been specified in the settings-file but is missing in the benthos-file.	error	
33	BENMMI	Check if the months specified in the settings-file are also in the benthos file.	error	
34	BENMMI	Check the existence of the AMBI-file (if specified).	error	
35	BENMMI	Check if the AMBI-file can be read.	error	

Continued on next page

	Topic	Check	Action	Comment
36	BENMMI	Check if the AMBI-file is consistent: TAXON-AMBI class combinations should be unique after conversion to WoRMS.	warning	Only the first TAXON-AMBI combination will be used.
37	BENMMI	Check the existence of the AMBI-file (if specified).	error	
38	BENMMI	Check if the ITI-file can be read.	error	
39	BENMMI	Check if ITI-data are still consistent after conversion to WoRMS.	error	
40	BENMMI	Check the existence of the habitat-reference file.	error	
41	BENMMI	Check if the habitat-reference-file can be read.	error	
42	BENMMI	Check if reference data are available for all records in the benthos-file.	error	
43	BENMMI	Check if the BENMMI-report can be generated without errors.	error	
44	BENMMIdir	Check if BENMMIdir can be launched by means of a graphical user's interface.	error	GUIs are usually only implemented on desktops and not on servers.
45	BENMMIdir	Check if the user presses the CANCEL-button or the ESC-button to cancel execution.	message	
46	BENMMIdir	Check if the BENMMI-direcory exists.	error	
47	BENMMIdir	Check if the BENMMI-direcory is empty.	error	We don't want to accidentally overwrite existing files.
48	benthos-file	Check if the following column names exists in the benthos-file: OBJECTID, HABITAT, SAMPLEID, TAXON, AREA, DATE, VALUE, LAT, LONG, MESH.	error	
49	benthos-file	Check if VALUE column in benthos-file is non-negative.	error	
50	benthos-file	Check date-format.	error	All dates should adhere to ISO 8601 (YYYY-mm-dd)
51	benthos-file	Check that all AREAs are unique for each sample (i.e., OBJECTID/SAMPLEID/DATE-combination).	error	
52	benthos-file	Check on duplidated records.	error	
53	benthos-file	Check on azoic samples with non-zero VALUE-column.	warning	These abundances will be set to zero.
54	benthos-file	Check if benthos file has still any records left after all checks.	error	
55	benthos-file	Check if VALUE column contains integer values (abundances).	warning	Fractional data may indicate fragments of species, but also species densities.
56	taxa-file	Check if the taxa-file exists.	error	

Continued on next page

	Topic	Check	Action	Comment
57	taxa-file	Check if the following column names exists in the taxa-file: group, provided, accepted, level.	error	
58	taxa-file	Check if all taxa can be converted from provided to accepted names.	error	No missing values are allowed in the taxa-file.
59	taxa-file	Check if each record contains a taxonomic group.	warning	An informative warning will be issued.
60	taxa-file	Check if the taxa in column 'accepted' belongs to only one taxonomic group.	error	
61	AMBI-file	Check if the AMBI-file exist.	error	
62	AMBI-file	Check if the AMBI-file can be read.		
63	AMBI-file	Check if the AMBI-file contains columns TAXON and GROUP.	error	
64	AMBI-file	Check on duplicated records.	message	Duplicated records will be removed.
65	AMBI-file	Check if AMBI-classes are permissible (I, II, III, IV, V).	error	
66	ITI-file	Check if the ITI-file exist.	error	
67	ITI-file	Check if the ITI-file can be read.	error	
68	ITI-file	Check if the ITI-file contains columns TAXON and GROUP.	error	
69	ITI-file	Check on duplicated records.	message	Duplicated records will be removed.
70	ITI-file	Check if ITI-classes are permissible (I, II, III, IV).	error	
71	REF-file	Check if the habitat-reference-file exists.	error	
72	REF-file	Check if the habitat-reference-file can be read.	error	
73	REF-file	Check if any indices are specified. If that's not the case an error will be raised.	error	
74	REF-file	Check if all indices are valid. Valid indices are N, LNN, S, D, SN, SNA, H, L, AMBI, ITI, PIE, N2.	error	
75	REF-file	Check if all compulsory column names are available. Compulsory names are: OBJECTID, RELAREA, HABITAT, and BAD*, REF* where * denotes an index.	error	
76	REF-file	Check if duplicated records are available.	warning	Duplicated records will be removed.