BEQI2: Introduction

Dennis Walvoort
Alterra – Wageningen University & Research Center
Wageningen, The Netherlands; e-mail: `dennis.walvoort@wur.nl`

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

2015-01-06

Contents

---

1 Introduction

This tutorial provides a brief introduction to the **BEQI2**-package. This package should facilitate the analysis of benthos data. It estimates the following benthic quality indices:

- total abundance of species (N);
- species richness (S);
- Shannon index (H');
- AZTI Marine Biotic Index (AMBI);

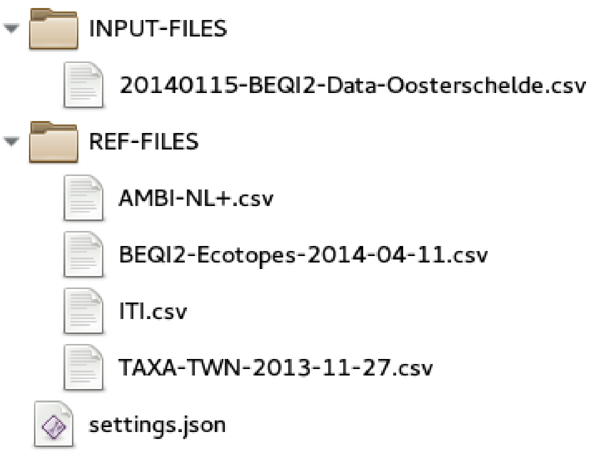
INPUT-FILES
20140115-BEQI2-Data-Oosterschelde.csv
REF-FILES
AMBI-NL+.csv
BEQI2-Ecotopes-2014-04-11.csv
ITI.csv
TAXA-TWN-2013-11-27.csv
settings.json

Figure 1: Default BEQI2-directory structure.

- of each indicator given above, an ecological quality ratio (EQR) is calculated using the indicator value, a ‘reference’ value and ‘bad’ value (Section 3.10);
- The BEQI2 index gives an EQR value, which is calculated as the average of the EQR values of species richness, Shannon and AMBI.

The package includes three 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.
- species name conversion: the tool automatically converts the synonym names into standardized species names using a conversion table 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;

If you’re not familiar with R, and don’t know how to install R and the BEQI2-package, you should consult the R-website (`www.R-project.org`) or read the installation guide provided with the BEQI2-package.

2 Quick-start

The workhorse function of the **BEQI2**-package is called `BEQI2`. This function performs a full BEQI2-analysis. It reads all its inputs from, and stores all its outputs to files. Storing these files in a structured way is therefore highly recommended. By default, the BEQI2-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 `BEQI2dir` function. This function starts after typing

```
BEQI2dir()
```

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.*:

```
BEQI2dir(path = "c:/myprojects/BEQI2/BEQI2_FILES")
```

Note that paths are separated by (forward) slashes. In both cases, the working directory needs to be empty to avoid overwriting existing data.

After running `BEQI2dir` two directories have been created:

- INPUT-FILES: a directory containing BEQI2-input files. These files contain the number of taxa that have been found in each sample. Consult Appendix A for more information.
- REF-FILES: a directory containing BEQI2-reference files. These files contain information on waterbodies/ecotopes (Appendix C) and species sensitivities (AMBI, Appendix B), 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`

In addition, a file with the name '`settings.json`' has been created. This file contains all settings specified by the user to run the BEQI2-tool. See Section 3.1 for more information.
The BEQI2-tool can be started by typing

```
BEQI2()
```

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

```
BEQI2(filename = "c:/myprojects/BEQI2/BEQI2_FILES/settings.json")
```

After the BEQI2-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 BEQI2-package in more detail

3.1 Data Files and Settings

A BEQI2-run is entirely specified by the contents of a json-file (see '`settings.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": "BEQI-2 analysis 'Oosterschelde'",
    "user": "Jan de Ruiter",
    "date": "2014-12-24",
    "files": {
        "BEQI2": "INPUT-FILES/20140115-BEQI2-Data-Oosterschelde.csv",
        "SpeciesNames": "REF-FILES/TAXA-TWN-2013-11-27.csv",
        "Ecotopes": "REF-FILES/BEQI2-Ecotopes.csv",
        "AMBI": "REF-FILES/AMBI-NL+.csv"
    },
    "months": [1,12],
    "pooling": {
        "enabled": true,
        "randomSeed": 314,
        "targetArea": [0.09, 0.11]
    },
    "genusToSpeciesConversion": true
}
```

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

- *title*: the title of the BEQI2 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:
    - *BEQI2*: the BEQI2-input file (see Appendix A for details);

  - *SpeciesNames*: Taxa Water management of the Netherlands (TWN) list. A recent version of this list is included in the BEQI2-package. The latest version of this list can be downloaded from `sofus.ecosys.nl/taxabase.htm`;
  - *AMBI*: optional user defined AMBI-file (see Appendix B for details);
  - *Ecotopes*: ecotopes reference file (see Appendix C for details);

  Note: optional files can be excluded from analysis by setting its value to 'null', or '' (empty string), or by removing the line from the JSON file, or by commenting this line out by C-style comments (//);

- *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 3.7)*:
  - *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].

3.2 Selection of benthos records

After the tool has been started with function `BEQI2()`, the BEQI2-file (Section A) is read and records outside the specified time-frame (see 'months'-key in the JSON-file), non-endofauna records, and records that are hard to identify by benthos laboratories (labelled 'REST') are removed.

3.3 Conversion of species names

All taxa in the BEQI2-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.

3.4 Species sensitivity values

Species sensitivity values are read, first from the user supplied species sensitivity file (Section B). 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-indicator (Borja *et al.*, 2000).

3.5 Waterbody-ecotopes and sample areas

For each combination of waterbody-ecotope, an overview of available sampling areas is tabulated. This table can be used to judge the efficacy of pooling.

3.6 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.

3.7 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.

3.8 Indicator calculation

After all steps above are carried out, four marine benthos indicators are estimated, *i.e.*, total abundance (N), species richness (S), Shannon index (H'), and the AMBI marine biotic index. The last three indicators are part of the BEQI2.

3.9 Indicator percentile values

To get information on the distribution of each indicator in each waterbody-ecotope, percentile values are given. These percentiles may also be useful to set the reference values of the Indicator Ecological Quality Ratios (Section 3.10).

3.10 Indicator Ecological Quality Ratios

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

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

where I_{ass} is the estimated indicator 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 Ecotope reference file (Appendix C). Depending on 'bad' and 'ref', the EQR usually (but not necessarily!) varies between 0 (bad ecological quality) and 1 (reference ecological quality).
The EQR(BEQI2) is the average EQR for AMBI, Shannon index H' and species richness S.

3.11 Results

After running the tool, the following results are available:

- a report in HTML-format, containing all sections given above;
- output files aggregated to the ECOTOPE or OBJECTID (*e.g.*, waterbody) levels;
- log-file with informative, warning, and error messages;
- and an optional file with pooling information.

A BEQI2 input file

The format of the BEQI2 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 compulsory: OBJECTID, ECOTOPE, SAMPLEID, TAXON, CHAR, SAMPDEV, AREA, DATE, and VALUE. Azoic samples must have a value of 0.

Item (header in data table)	Comments	Examples
Measuring object id (OBJECTID)	The WFD water body code is used here	NL91_veersmr
Ecotope code (ECOTOPE)	The ecotope code of the sample can be constituted freely, for example based on the intertidal/subtidal position, the salinity zone and the sediment composition.	MesohalineIntertidal
Sample-id (SAMPLEID)	Location code OR Transect station code OR Sample number Note1: the BEQI2 tool combines the sample code with the date to ensure a unique sample code. Note 2: replicas should have different sample codes, e.g. by adding an additional letter a, b,etc.	North Sea: e.g. NOORDWK2; Dollard: Transect-station code; Delta: 1304
Parameter (PAR)		COUNT
Taxon name (TAXON)	Standardized taxon code (see WoRMS-website) In NL: the TWN code	
Characteristic (CHAR)	Additional aspect of the taxon; e.g. Juvenile (JUVNL) or quality code (e.g REST). Note 1: different annual classes of specific taxa are combined for BEQI2 analysis.	
Compartment code (COMP)	Relevant compartment codes can be chosen freely.	BS = Bottom Sediment
Samping device (SAMPDEV)	The code of the sampling device.	e.g., BOXCRR of VEENGB
Sampled area (AREA)	The sampling area is necesarry for the data pooling process.	E.g. 0.015 m2.
Value processing code (WBEW)		NA = Not Applicable
Value determination code (WBEP)	Benthic sampling and analysis method code	e.g., A2.107 or Essink1991
Date (DATE)	ISO format: eejj-mm-dd is mandatory.	2001-03-05
Numerical value (VALUE)	Numerical value. Decimals are allowed. Note: the decimal sign must be a point (.)	20.5

B AMBI file

The 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.

SUBMITTED.NAME	AMBI
Pectinaria californiensis	II
Apohyale prevostii	I
Astacilla gorgonophila	I
Priapulopsis bicaudatus	III
Paratanais sp.	I
Liljeborgia geminata	I
Grubeosyllis rugulosa	II
Polydora hoplura	IV
Athanas amazone	I
Oriopsis eimeri	II
Iphimedia minuta	I
Chlamys glabra	I
Cyanophthalma obscura	III
Laticorophium baconi	II
Isolda sp.	III
Ampharete grubei	I
Paramphitrite tetrabranchia	I
Exogone longicirris	II
Cuspidaria jugosa	I
Tetrastemma vermiculus	III
Polycirrus denticulatus	IV
Ophiopholis mirabilis	II
Angulus donacinus	I
Euclymene collaris	I
Sertularella mediterranea	II

C Ecotope reference file

The Ecotope Reference file contains meta-information about each waterbody (OBJECTID) and ecotope.

OBJECTID	DESCRIPTION	ECOTOPE	RELAREA	SREF	SBAD	HREF	HBAD	AMBIREF	AMBIBAD
nl89_westsde	Westerschelde	Mesohaline-Intertidal	0.11	29	0	3.30	0	0.57	6
nl89_westsde	Westerschelde	Mesohaline-Subtidal	0.17	22	0	3.20	0	0.54	6
nl89_westsde	Westerschelde	Polyhaline-Intertidal	0.18	41	0	3.60	0	1.20	6
nl89_westsde	Westerschelde	Polyhaline-Subtidal	0.54	31	0	3.80	0	0.63	6
nl89_oostsde	Oosterschelde	Polyhaline-Intertidal	0.26	45	0	3.70	0	0.54	6
nl89_oostsde	Oosterschelde	Polyhaline-Subtidal	0.74	67	0	5.10	0	0.50	6
nl89_veersmr	Lake Veere	Polyhaline-Subtidal	1.00	30	0	3.80	0	0.52	6
nl89_grevlemr	Lake Grevelingen	Polyhaline-Subtidal	1.00	44	0	4.20	0	0.54	6
NL81_2	Dollard	Mesohaline-Intertidal	1.00	22	0	3.30	0	0.54	6
NL81_1	Wadden Sea	Polyhaline-Intertidal	0.59	29	0	3.60	0	0.54	6
NL81_1	Wadden Sea	Polyhaline-Subtidal	0.41	23	0	3.50	0	0.50	6
NL95_1A2A	Delta Coast	Polyhaline-Subtidal	1.00	33	0	3.80	0	0.01	6
NL95_3A	Dutch Coast	Polyhaline-Subtidal	1.00	30	0	4.00	0	0.16	6
NL95_4A	Wadden Coast	Euhaline-Subtidal	1.00	30	0	3.70	0	0.12	6