ANERIS

Operational Sensing Life Technologies for Marine Ecosystems

`

# Milestone 6. Testing new research products – DNA-based indicators

Lead Beneficiary: LifeWatch ERIC

Author/s: Christos Arvanitidis, Panagiotis Kasapidis, Cristina Huertas

17/12/2024

Funded by
the European Union

# Table of Contents

## Preface

This Milestone refers to new research products to be developed by the ANERIS project. It provides the methodology for the development of three DNA-based indicators as well as the next steps needed for their testing and application.

## Summary

Milestone 6: Testing new research products – DNA based indicators, refers to new research products to be developed by the ANERIS project. Here, special emphasis is given to the type of data to be used by the proposed indicators as well as the standardization of the methodology towards their release as new research products and in production mode.

The methodology for their development follows best practices established by relevant EU funded projects, such as DEVOTES (2) (DEVelopment Of innovative Tools for understanding marine biodiversity and assessing good Environmental Status) and GES4SEAS (4) (Achieving Good Environmental Status for maintaining ecosystem SErvices, by ASsessing integrated impacts of cumulative pressures). This has been deemed by the ANERIS Consortium to be the best approach to develop research products, which are directly comparable to those already existing. At the same time, the ANERIS research products aim to provide information, which is additional, and perhaps complementary, to that produced by the existing products.

In the next sections, the concept, the preparatory actions and the framework for the development of the new research products are provided. The entire approach for the development and testing of these ANERIS will be completed and presented in the deliverable D2.3: Development of new DNA-based indicators for pelagic habitats environmental status monitoring, which is due to month 28 of the project (30 March 2025).

## List of Abbreviations

**ANERIS:** operAtional seNsing lifE technologies for maRIne ecosystemS

**DNA**: Deoxyribonucleic Acid.

**DPSIR**: Drivers, Pressures, State, Impacts, Responses.

**EMOBON**: European Marine Omics Biodiversity Observation Network.

**EU:** European Union

**FAIR:** Findable, Accessible, Interoperable, and Reusable (Data Principles)

**GES4SEAS:** Achieving Good Environmental Status for Maintaining Ecosystem Services by Assessing Integrated Impacts of Cumulative Pressures.

**iBOL:** International Barcode of Life.

**M-AMBI:** Multivariate AZTI Marine Biotic Index.

**NEAT:** Nested Environmental Status Assessment Tool.

**PCA:** Principal Component Analysis.

**WFD:** Water Framework Directive.

# 1. Concept

The concept of the following proposed DNA-based indicators for monitoring the status of the pelagic habitats is based on existing ones. The innovation here lies in the type of information to be used: in the existing indices, the data used come from the conventional measurements of species occurrence, abundance, biomass or productivity while in the proposed indices they come from DNA metabarcoding. This type of data provides additional properties to the effort of assessing the ecological status of the habitats in question.

## 1.1 Criteria adopted

The primary criteria set for the adoption of the DNA-based indices proposed are provided below. However, through the process of their standardized description and publication, additional criteria will need to be set the performance of the proposed indices will be checked against them.

Simplicity: The ANERIS Consortium has adopted this criterion in all of its activities based on the principle that when something is simple it’s easier to be both understood and implemented, especially when it is to be used by non-scientists (e.g., citizen scientists or the public at large). Accordingly, if the indices are simple in their implementation, they will also be simple in the interpretation of their values and the establishment of their value ranges, which correspond to certain ecological status levels.

General use: Generality follows simplicity in that the simpler an index has the greater chances that the types of data it requires to be easily collected by a wide range of habitats and ecosystems. This is particularly useful for habitats like the pelagic one, which is one of the largest and unpredictable of all marine habitats. Indeed, although the DNA-based indices are proposed in this Milestone for the pelagic habitat, there isn’t anything to prevent their use in other habitats whenever the conditions for their application exist in those habitats.

Comparability: This criterion set so that the results of the proposed DNA-based indices should be directly comparable to those obtained by the same indices when conventional data are used. This way another angle of the ecological status of the habitats in question can be obtained.

Sample-size, sample-effort free: It ensures that theoretically every effort to collect data from any habitat, including the pelagic one, can contribute to the assessment of its ecological status. In addition, efforts even from non-scientists, for example from citizen scientists, can provide valuable data and information towards the ecosystem status assessment.

## 1.2 Data availability

The simplest DNA-based data are coming from metabarcoding, which is a process with many steps including the isolation, amplification and reading of specific sequences, representatives of taxa, most frequently identified at the species level. This process has been standardized long ago and a global aggregator of such data is already at operational level: iBOL, the international barcoding of life. Accordingly, the metabarcoding data have been selected for the calculation of the proposed DNA-based indices. They can be used either as raw numbers or slightly transformed through simple calculations whose meaning has already been tested by other types of data, such as species occurrence, abundance, biomass or productivity. Metabarcoding data are being selected in Europe from a range of observation networks, such as EMOBON.

# 1. Approach

The approach followed for the identification of the proposed DNA-based indices is explained in the following paragraphs.

## 1.1. Review

The ANERIS Consortium went through a detailed review of the relevant literature and resources available on the internet (e.g. observation networks, databases, global aggregators, existing software, web services and workflows). This review gave substantial information that guided the selection of the proposed DNA-based indices so that they would be operational. The review process was tremendously facilitated by the results of two EU funded projects: DEVOTES (2) and GES4Seas (4).

## 1.2. Selection

The indicator selection process included the review of the ecological status indices provided by the project DEVOTES (DEVelopment Of innovative Tools for understanding marine biodiversity and assessing good Environmental Status) and especially by the software developed Nested Environmental status Assessment Tool) (NEAT) (1). By following the primary criteria mentioned above, the following DNA-based indices are proposed:

1. DNA-based trends for the arrival of new non-indigenous taxa.

This index is applied to metabarcoding occurrence (and when available on relative abundance) data. There is already a standardized and operational workflow that identifies the new non-indigenous taxa in a pelagic sample, hence in the pelagical habitat.

2. DNA-based trends in the distributional range of invasive taxa.

This index is applied to metabarcoding occurrence data. There are already standardized and operational web services that can calculate expansions on distributional ranges of taxa in pelagic habitats, which can be transformed into proper workflows available online.

3. DNA-based taxonomic relatedness.

This index is applied to metabarcoding occurrence data. At the same time, it uses the higher classification of the taxa identified in a pelagic sample, a process that can provide expected ranges in the values of the index and/or departures from those expectations. There're already standardized and operational web services that can calculate expansions on distributional ranges of taxa in pelagic habitats, which can be transformed in proper workflows available online.

## 1.3. Metadata

The metadata to be used in order for the above indices to be officially described and published will follow the standard introduced by the project DEVOTES and is available in the NEAT software. An example of that metadata is provided below for the M-AMBI index:

**Original name**: Multivariate AZTI Marine Biotic Index (M-AMBI) (5; 6)

**Data Required:**

- Species Abundance Data (N_i): Number of individuals of species i in a sample.
- Species Sensitivity/Tolerance Classifications:

  *Sensitive species: EG_I*

  *Tolerant species: EG_III, EG_IV*

  *Opportunistic species: EG_V*

- Environmental Variables: Physical-chemical parameters like salinity, depth, and sediment grain size.

**Metrics and Formulas:**

### 1. Species Richness (S):

$S$ = Total number of species in a sample.

### 2. Shannon-Wiener Diversity Index (H'):

$H' = -\Sigma(p_i * ln(p_i))$

Where:

- $p_i = N_i / N$ is the proportion of individuals of species i.

- $N = \Sigma(N_i)$ is the total number of individuals.

### 3. AZTI Marine Biotic Index (AMBI):

$AMBI = (\Sigma(j * EG_j)) / (\Sigma(EG_j))$

Where:

- $j$ = ecological group number ($EG_1$ for sensitive species, $EG_5$ for opportunistic species).
- $EG_j$ = abundance of species in group j.

### M-AMBI Calculation Process:

1. Standardization of Metrics:

Each metric (S, H', AMBI) is standardized using the formula:

$x' = (x - min(x)) / (max(x) - min(x))$

Where:

- $x$ is the value of the metric.
- $min(x)$ and $max(x)$ are the observed minimum and maximum values across all samples.

2. Factor Analysis:

Apply PCA or similar techniques to integrate standardized metrics (S', H', AMBI') into a single composite index.

3. Normalization:

Normalize the composite score to a scale of 0 to 1:

*M-AMBI = (Composite Score - min(Composite Score)) / (max(Composite Score) - min(Composite Score))*

**DPSIR Framework Application** (3)**:**

- Driver: Human activities like industrial discharge, urban runoff, and agriculture.
- Pressure: Pollution, eutrophication, sediment disturbance.
- State: Changes in species composition, diversity, and ecological balance.
- Impact: Biodiversity loss, habitat degradation, and reduced ecosystem services.
- Response: Mitigation measures such as pollution reduction and habitat restoration.

**Ecological Status Classification:**

The M-AMBI value is used to classify ecological status into five categories:

- High: $\text{M-AMBI} \geq 0.8$
- Good: $0.6 \leq \text{M-AMBI} < 0.8$
- Moderate: $0.4 \leq \text{M-AMBI} < 0.6$
- Poor: $0.2 \leq \text{M-AMBI} < 0.4$
- Bad: $\text{M-AMBI} < 0.2$

**Applications:**

- Assessment of marine and coastal water bodies' ecological quality.
- Compliance with the European Water Framework Directive (WFD).
- Monitoring the impact of anthropogenic pressures on benthic ecosystems.

# 2. Next steps

The next steps towards the description, testing and standardized publication of the proposed DNA-based indices are:

## 2.1. Datasets

Relevant datasets for the application of the proposed DNA-based indices will be downloaded from databases or global aggregators in order to test the proposed indicators. The data will need to comply with some standards and if possible to be FAIR compliant. The datasets with which the DNA-based indices will be tested will make part of their official proposal through publication.

## 2.2. Workflows

Existing pieces of software and types of code that can execute the calculation of the values of the indices will be first converted into web services and, where necessary, will be orchestrated in workflows, so after the loading of the data and the specification of certain parametres the calculation of the values will be executed with a single click.

## 2.3. Testing

Depending on the availability of the data, exhaustive testing will be attempted to find the capabilities and limits of the DNA-based indices in reflecting the ecological status of the pelagic habitat. The testing will be made exclusively in the standardized virtual research environment and through the workflows and services hosted there in MyLifeWatch platform (https://my.lifewatch.eu).

## 2.4. Documentation

The entire process will be finally summarized in the Deliverable 2.3: Development of new DNA-based indicators for pelagic habitats environmental status monitoring.

# Acknowledgements

The authors acknowledge the financial support provided by the European Union’s Horizon Europe research and innovation programme under Grant Agreement No. [101094924], which has enabled the successful implementation of the ANERIS project. This funding has been essential for advancing our research and development activities in marine biodiversity assessment.

## References

1. *Berg, Torsten & Murray, Ciarán & Carstensen, Jacob & Andersen, Jesper. (2016). NEAT – Nested Environmental status Assessment Tool Manual – Version 1 .1. 10.13140/RG.2.1.4144.6161*
2. *DEVOTES, 2024. [http://www.devotes-project.eu/](http://www.devotes-project.eu/)*
3. *Gari, Sirak Robele; Newton, Alice; Icely, John D. (2014). A review of the application and evolution of the DPSIR framework with an emphasis on coastal social-ecological systems. Zenodo. [https://doi.org/10.1016/j.ocecoaman.2014.11.013](https://doi.org/10.1016/j.ocecoaman.2014.11.013)*
4. *GES4SEAS, 2024. [https://www.ges4seas.eu/](https://www.ges4seas.eu/)*
5. *Sigovini, M., Keppel, E., & Tagliapietra, D. (2013). M-AMBI revisited: Looking inside a widely-used benthic index. Hydrobiologia, 717(1), 41–50.*
6. *Teixeira, Heliana, Berg, Torsten, Uusitalo, Laura, Fürhaupter, Karin, Heiskanen, Anna-Stiina, Mazik, Krysia, Lynam, Christopher P., Neville, Suzanna, Rodriguez, J. German, Papadopoulou, Nadia et al. (2016). A Catalogue of Marine Biodiversity Indicators. Frontiers in Marine Science, 3, 207. DOI: 10.3389/fmars.2016.00207*