Baltic Marine Environment Protection Commission

Second HELCOM BalticBOOST Workshop on the HOLAS II Biodiversity assessment tool

Copenhagen, Denmark, 14-15 September 2016

Co-funded by the European Union

Document title	Document 1 - Testing the HELCOM Biodiversity Assessment Tool (BEAT 3.0)
Category	CMNT
Submission date	30.8.2016
Submitted by	BalticBOOST

Background

The First HELCOM BalticBOOST workshop on the HOLAS II biodiversity assessment tool (HELCOM BalticBOOST Biodiv WS 1-2016) was held in February 2016 in Copenhagen. The workshop recommended that the integrated biodiversity assessment in HOLAS II should be based on concepts used in the BEAT tool (HELCOM 2010) and its subsequent developments, and advised the BalticBOOST project to develop the tool further, including to test alternative structures and spatial scales, develop solutions for entering core indicators to the tool, and develop approaches to include confidence assessment to the tool (Outcome of the workshop).

The fifth Meeting of the Project for the development of the second holistic assessment of the Baltic Sea (HELCOM HOLAS II 5-2016) supported the recommendations of the HELCOM BalticBOOST Biodiv WS 1-2016 and further invited the project to test the recent proposals from the EC workshop on guidance for implementation of assessments under MSFD Article 8 to use the so called ‘species approach’ vs the ‘criteria approach’ when integrating indicators for birds, fish and mammals (paragraph 4.14 of the Outcome).

This document describes the structure of the further developed tool, proposed to be named BEAT 3.0, and presents results of the various test scenarios as requested by the previous HELCOM BalticBOOST workshop and the HOLAS II 5-2016 meeting. The test scenarios resulted in 14 alternative assessment results in the selected pilot areas, which were evaluated against a set of evaluation criteria.

Action required

The workshop is invited to:

- evaluate the test results, including:
  - proposed approach to include trend-based indicators
  - the alternative structures (species vs criteria approach)
  - weighted averaging vs OOAO approach
  - spatial representation
  - potential inclusion of additional indicators
  - use of the same indicator under several criteria
- recommend the most feasible tool structure
- advise on the finalization of the tool for use in the HOLAS II assessment.

Testing the HELCOM Biodiversity Assessment Tool (BEAT 3.0)

1. Introduction

The first integrated assessment of the Baltic Sea marine biodiversity was carried out in 2009 (HELCOM 2009) and the assessment was updated for the Initial Holistic Assessment in 2010 (HELCOM 2010). Both the assessments were made using the HELCOM Biodiversity Assessment Tool (BEAT 1.0, Andersen et al. 2014). Integrated biodiversity assessments were further developed and improved in the Baltic Sea region in the HARMONY project, MARMONI project and the DEVOTES project. All the projects have given valuable improvements to the assessment methodology which were finally evaluated by the First HELCOM BalticBOOST workshop on the HOLAS II biodiversity assessment tool (HELCOM BalticBOOST Biodiv WS 1-2016) which was held in February 2016 in Copenhagen. The BalticBOOST project has the task to develop and test the tool for operational use in the Second HELCOM holistic assessment of the Baltic Sea.

The workshop recommended that the HELCOM HOLAS II biodiversity assessment tool should be based primarily on the similar concept as in the first integrated assessment but improvements from the later projects should be included to the development of the tool. The workshop further advised the BalticBOOST project to develop the tool, test alternative structures, integration methods and spatial scales, develop solutions for entering core indicators to the tool and develop approaches to include confidence assessment to the tool (Outcome of the workshop). The HELCOM HOLAS II biodiversity assessment tool is proposed to be named BEAT 3.0.

The development towards the second holistic assessment of the Baltic Sea is coordinated through the HELCOM HOLAS II project which is guided by the HOLAS II core team. In its fifth meeting (April 2016), the team supported the recommendations of the previous workshop but also further advised the project to:

- in case of trend-based core indicators, to explore the use of an interim approach, where experts will be asked to estimate the distance to GES based on categories (paragraph 4.12 of the Outcome);
- take into consideration the evolving drafts of the European Commission Decision on GES criteria in the further testing of the tool, including to:
    - test the proposal to end the integration at the level of the ecosystem elements (birds, fish, mammals, pelagic and benthic habitats) vs to make an overall biodiversity assessment
    - test the recent proposals from the EC workshop on MSFD Article 8 assessments to use the so-called species approach vs the criteria approach when integrating indicators for birds, fish and mammals (paragraph 4.14 of the Outcome).

The tests of BEAT 3.0 were carried out in four case study areas: Gulf of Finland, Gulf of Riga, Bornholm Basin and Kattegatt. In addition, an integrated test, based on the case study areas, was done for Baltic Sea. The tests were carried out using available indicator data in the HELCOM biodiversity indicator reports and also utilizing the eutrophication indicators from the EUTRO-OPER project. The project was also requested to test the use of 'additional indicators' in the tool. Additional indicators have been proposed by Contracting Parties as a supplement to the HELCOM core indicators. This document presents the updated BEAT 3.0 tool and its test results according to the alternative scenarios, requested by the previous workshop and the HOLAS II core team. Detailed descriptions of the tool structure and results from the test scenarios are given in Attachments 1 and 2 (separate excel files) and summarized descriptions and results are given in the document text. In order to compare the test results, the document also includes evaluation criteria against which the results are compared and recommendations are given.

2. Basic structure of the BEAT 3.0

The BEAT 3.0 tool is an indicator-based assessment tool. The tool requires stand-alone indicators which have quantitative (or numeric/ semi-quantitative) result values, a GES boundary value and minimum and maximum values. Thus, indicators with a trend GES target cannot be used as such in the assessment.

HELCOM BalticBOOST Biodiv WS 1-2016 suggested alternatives for how indicators with trend GES targets could be included: a) set an artificial threshold for defining interim GES value "if the desired trends was met, what would the value be in 2020?", b) treat the slope as it was a single value and use as a threshold, and c) develop the assessment tool to allow use of trend-based GES (paragraph 15 of the Outcome). BalticBOOST evaluated these alternatives and concluded that it cannot be recommended to use trends in a status assessment. A proposed approach from BalticBOOST to overcome this problem, supported by HELCOM HOLAS II 5-2016, proposes how to include indicators with a trend GES target in BEAT 3.0 (Figure 1). In this approach the indicator experts are asked to judge if the indicator is in GES or not and to evaluate the distance to GES, resulting in a categorical approach with four classes: sub-GES, far from GES; sub-GES close to GES; just above GES; and clearly in GES. The workshop is invited to consider the proposed approach.

The indicators are arranged to pre-defined groups (='structure') where the integration will take place within and among the groups. BEAT 3.0 produces a single assessment result for the biodiversity assessment per assessment unit but lower-level status results can also be extracted from the tool (e.g. a status of a certain GES criterion or marine element). The status result is given at a scale between 0 and 1, where 0.6 defines the GES boundary, and allows evaluating the distance to the GES boundary; both below and above the boundary. In the output, this can be shown as different shades of GES and sub-GES colors. The assessment units follow the HELCOM Monitoring and Assessment Strategy.

The BEAT 3.0 includes an assessment of confidence which is based on indicator confidence, as presented further in document 2 of this workshop. In the BEAT 3.0 structure, the overall confidence is calculated similarly through the tool structure as the indicator values. The confidence assessment is not included in this document, but results will be presented at the workshop.

The BEAT 3.0 has been coded in R and the test scenarios were run with the R-coded version of the tool. The work continues to embed the BEAT 3.0 into a workspace where it can be run without knowing the R language. The workspace will also include a graphical component to allow production of visual outputs from the tool.

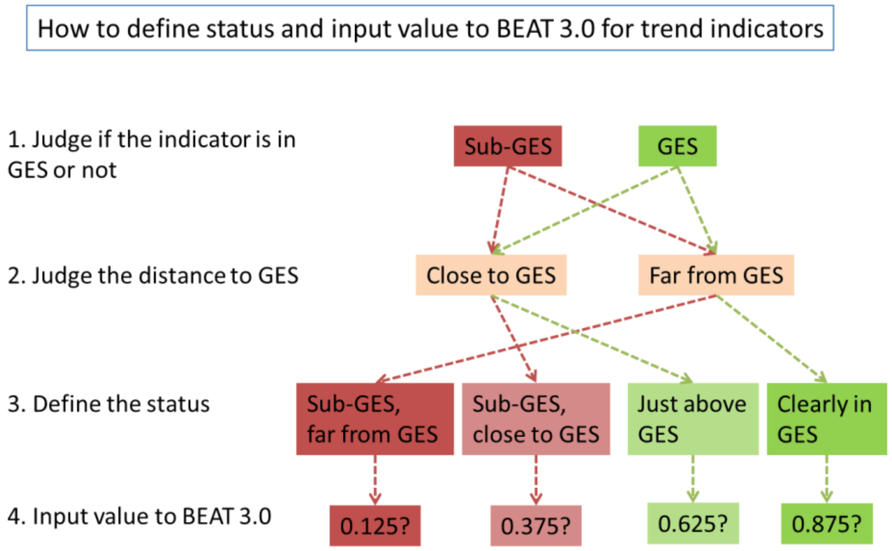
How to define status and input value to BEAT 3.0 for trend indicators
1. Judge if the indicator is in GES or not
2. Judge the distance to GES
3. Define the status
4. Input value to BEAT 3.0
Sub-GES
GES
Close to GES
Far from GES
Sub-GES, far from GES
Sub-GES, close to GES
Just above GES
Clearly in GES
0.125?
0.375?
0.625?
0.875?

Figure 1. Proposed method on how to include indicators with a trend based GES target.

3. Testing scenarios for alternative structures, integration rules and spatial aggregations

The HELCOM BalticBOOST Biodiv 1/2016 workshop and the fifth meeting of the HOLAS II core team asked the BalticBOOST project to test the developed tool with different scenarios to give advice on how the biodiversity assessment in HOLAS II best fulfills its purpose. The tested scenarios are divided in the following themes:

3.1 Alternative structures

Two different structures were tested: the ‘criteria approach’ and the ‘species approach’ (see Figure 2a and 2b). In the *criteria approach*, indicators are integrated to criteria and further to species group. In the *species approach*, indicators are evaluated per species before integrating to the species groups. These alternative structures were only tested for marine mammals, as indicators from other species groups do not at this point support the species approach, i.e. indicator results are not available at species level.

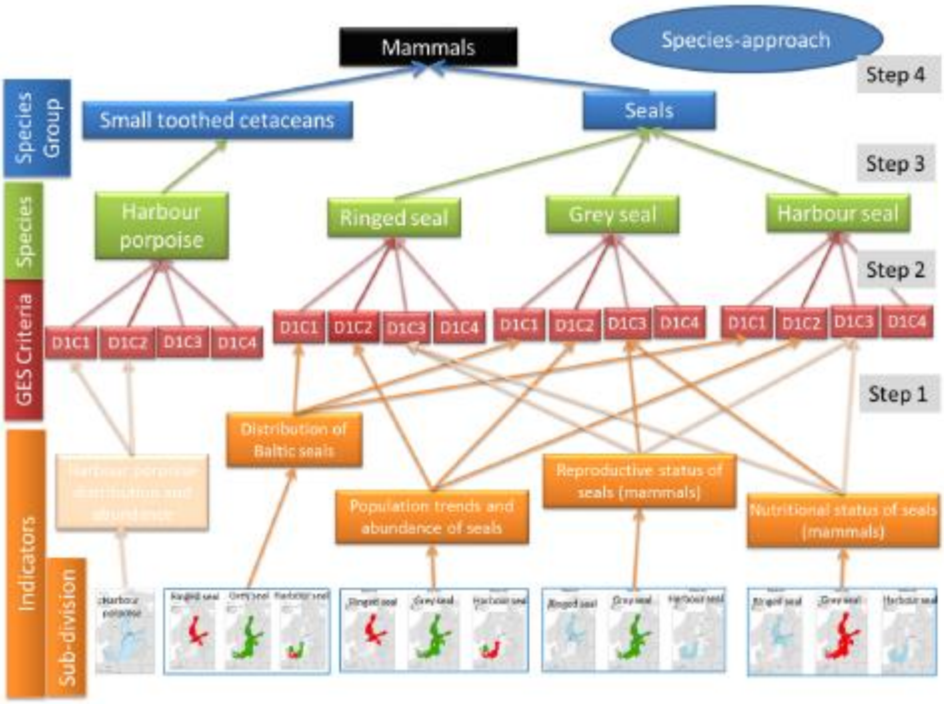

Figure 2a. Schematic representation of species approaches (Outcome of HOLAS II 5-2016 Annex 3).

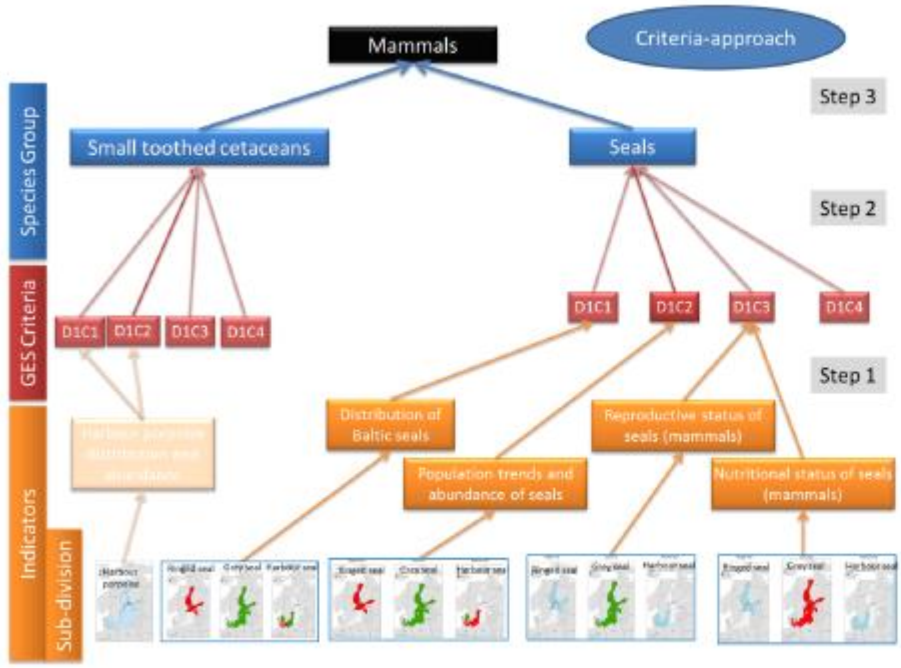

Figure 2b. Schematic representation of the criteria approach (Outcome of HOLAS II 5-2016 Annex 3).

3.2 Integration approaches

Different integration approaches were tested. Weighted averaging at all steps, or the one-out-all-out (OOAO) principle was used at different levels (steps in Figure 2a and b) in the integration process. The aspect to end the assessment at the ecosystem component level and not have an overall assessment of biodiversity is also included in this theme.

3.3 Spatial representation

Three approaches were tested:

1) using the spatial scales defined in the HELCOM indicator reports,
2) downscaling the indicators to HELCOM spatial assessment unit 4, and
3) same as in the second approach, but down-weighting indicators in areas where indicator experts considered the use of the indicator inappropriate, i.e. the species/habitat is not present in the sub-unit, and the indicator should thus not be assessed at this level. This information was collected from the indicator experts in the request for additional data.

3.4 Number of indicators

The test assessment was run for different sets of data using different numbers of indicators: using only HELCOM biodiversity core indicators, adding also relevant eutrophication core indicators (Secchi depth, oxygen, zoobenthos indices, macrophyte and phytoplankton indicators) and WFD indicators used nationally, as well as adding additional indicators suggested by the Contracting Parties[1]. The number of indicators is presented in Table 1.

[1] HOLAS II 5-2016 agreed that Contracting Parties will propose national/regional indicators to be used in HOLAS II to complement the HELCOM core indicators. The proposals will be discussed at HOLAS II 6-2016.

Table 1. Number of indicators in the BalticBOOST Theme 1.1 case study areas. For the pre-core indicators the spatial coverage is not clear.

Case study area	Core indicators	Pre-core indicators	WFD + Eutro Core indicators	Additional indicators
Gulf of Finland	21	14	13	1
Gulf of Riga	21	14	12	-
Bornholm Basin	17	14	9	(2)
Kattegat	10	14	6	-

3.5 Indicators that could be used under more than one criterion

The assessment result when using same indicators under several criteria was tested. The options were:

1) indicators used only once,
2) same indicator used only once per descriptor,
3) indicator used for all criteria it contributed to.

For testing the different scenarios a set of default choices (see table 2) were made to reduce the amount of total combinations. Thus, we tested in total 14 different scenarios. In the tests, we used the criteria as described in the original MSFD Commission Decision (2010/477/EU). BEAT 3.0 can easily be modified to use the suggested criteria in the revision of the Commission Decision, by updating the indicator-criteria links.

Table 2. The tested themes and alternative test scenarios. The marked scenarios were used as the default option when testing the other themes.

	Alternative test scenarios		
Themes for testing	1	2	3
Assessment structure	criteria based approach	species based approach	
Integration approaches	OOAO	weighted averaging	weighted averaging, OOAO at high level
Spatial representation	using indicator defined scales	down-scaling indicators to finest level	down-scaling indicators using weights
Number of indicators	BD core indicators	BD core + WFD and Eutro core indicators	all available indicators (including national indicators)
Indicators with multiple criteria	indicator used only once	indicator used once per relevant descriptor	indicator used for all relevant criteria

4. Results and evaluation of the BEAT 3.0 test scenarios

All the 14 test scenarios were successfully analyzed by the BEAT 3.0. Below the scenarios are analyzed against each of the evaluation criteria and finally a synthesis and recommendations are given to support decisions of the HELCOM Contracting Parties on the assessment structure.

4.1 Deviation of the integrated status result

Tables with results from the tests are presented in detail in Attachment 2. Here, a brief summary of the main points is presented.

Alternative structures

The two alternative scenarios of the integration structure, tested for mammals, did not differ in their result if the integration was made using a weighted averaging approach, as the weight of the indicators follow throughout the assessment. Thus, the way how to get to the ecosystem component through the structures, did not matter. However, if applying OOAO, the assessment result changes using both the species and the criteria approach. Using the OOAO rule, the assessment outcome was showing poorer status when using the species based approach in Gulf of Finland and Gulf of Riga, whereas the criteria based approach showed poorer status in the Bornholm Basin and Kattegatt.

Integration approaches

The different integration methods affected the assessment outcome. On Baltic Sea scale, the assessment results varied from 0.52 with weighted averaging, to 0.24 with OOAO at ecosystem component level, and further to 0.12 with OOAO at species group level. Integrating from criteria, averaging gave the result 0.60 and OOAO 0.45. Results from the case study areas are presented in Attachment 2. It is important to note that the integration method should not be chosen based on the result, but this is a fundamental principle of the assessment to be decided on.

Spatial representation

In the spatial representation of indicators only small differences (at highest 0.05 difference from the mean) in the assessment result was observed between the scenarios. However, when using indicators at their defined scales birds were not represented in any of the selected test areas, only at Baltic Sea level. If stopping the integration of indicators at the ecosystem component level 2 (i.e. not producing an overall biodiversity assessment), the spatial representation can be dealt with within the ecosystem components. The spatial representation of indicators within an ecosystem component is fairly similar, e.g. bird indicators area assessed at the same spatial scale. However, if results are displayed at a finer spatial scale than the indicator is assessed on, downscaling the result will be needed.

Number of indicators

When testing number of indicators (see Table1), the overall biodiversity assessment score did surprisingly not vary much. The coverage of ecosystem components and criteria increased using more indicators and thus including the confidence assessment would be necessary to evaluate the differences between the scenarios properly. Highest deviation from the mean (0.08) was observed in the Bornholm Basin. Within the ecosystem components the benthic habitat, and also pelagic habitat in Bornholm Basin and Kattegatt, showed large variation between the scenarios due to the poor representability of indicators for these habitats in the biodiversity core indicators.

Use of indicators under several criteria

Using the same indicator under several criteria is not affecting the assessment result for the ecosystem components or overall biodiversity status, but changed the representability of indicators for the criteria. Using the same indicator once under several descriptors changed the descriptor results only slightly (<0.01

difference in assessment score), but included some more criteria and more ecosystem components per descriptor. Using same indicator several time under the same descriptor changed the descriptor results a bit (0.07), but the differences within criteria and ecosystem component changed remarkably (up to 0.27).

4.2 Confidence of the status result

The confidence assessment will be updated before the workshop.

4.3 Number of indicators per criteria and ecosystem component

Using only HELCOM core indicators for biodiversity, there was a lack of indicators in many of the proposed MSFD GES criteria (Table 3). Also the benthic habitat is unrepresented, when using only the agreed core indicators. Adding HELCOM core indicators for eutrophication and national WFD indicators improved the number of indicators for the pelagic and benthic habitats. When using same indicators under several descriptors, additional criteria could be covered.

Table 3. Number of core indicators per criteria and ecosystem component. Bold numbers refer to agree core indicators, numbers in brackets are core and pre-core indicators that are not yet operational. Numbers in italics refer to the scenario where the same indicator was used under several descriptors and grey numbers refer to indicators used under the WFD that are relevant to use for the assessment of biodiversity.

	Mammals	Birds	Fish	Benthic	Pelagic
D1C1	**3**				
D1C2	**3** (1)	**2** (1)	**3**		
D1C3	**2**		(1)	(1)	(2)
D1C4					
D1C5				(1)	
D1C6			**2**	(2) 7	(2) *2* 4
D1C7					(1)
D3C2			*3*		
D4C1			(1)		
D4C2	*2*		*(2)*		
D4C3	*5*	*2*	*2*		**2**
D6C1				*(2)*	
D6C2				*(2)*	

4.4 Evaluation of the test result

The tested scenarios are here evaluated based on four evaluation criteria:

- deviation in result; a criterion to identify scenarios that affect the outcome of the assessment.
- differences between the case study areas; a criterion that assesses if the tested scenarios behaved the same way in all case study areas.
- missing indicators; a criterion to identify how ecosystem components and criteria are covered
- number of indicators per aggregation; a criterion assessing how well the indicators are represented at different aggregation scales.

Scenarios differing from the mean assessment result were highlighted using red color in the traffic light evaluations, not necessarily reflecting if it is considered good or bad.

Testing the alternative integration rules in the assessment, highlighted the difference in the output depending on the chosen approach. OOAO differed most from the tested scenarios, and is thus given red in the traffic light evaluation (Table 4), although it best fulfills the precautionary principle emphasizing the element in poorest condition. The result showed a similar pattern in all tested areas. The evaluation criteria on missing indicators and number of indicators per aggregation were given a red light in the OOAO scenario as the poorest indicator will decide the outcome.

Table 4. Traffic-light evaluation of the different integration approaches tested.

	Integration approaches		
Evaluation criteria	OOAO	weighted averaging	weighted averaging, OOAO at high level
Deviation in result	🟥	🟨	🟩
Differences between areas	🟩	🟩	🟩
Missing indicators	🟥	🟨	🟨
Number of indicators per aggregation	🟥	🟩	🟩

Using the indicator defined spatial scales will overall give a poor representation of indicators at the most detailed spatial scale, as few indicators are assessed at this scale (Table 5). As the option to down-weight indicators in areas where they were not considered representative was only scarcely used, the difference between the down-scaling scenarios were small.

Table 5. Traffic-light evaluation of the different spatial representations of the indicators.

	Spatial representation		
Evaluation criteria	using indicator defined scales	down-scaling indicators to finest level	down-scaling indicators using weights
Deviation in result	🟥	🟩	🟩
Differences between areas	🟨	🟨	🟨
Missing indicators	🟥	🟨	🟨
Number of indicators per aggregation	🟥	🟩	🟩

Testing different sets of indicators showed interestingly that the overall result did not vary much. However, when it comes to representability in the ecosystem components and MSFD criteria the biodiversity core indicators alone were covering less ecosystem components and criteria than if adding WFD and eutrophication indicators as indicators for the benthic and pelagic habitats (Table 6).

Table 6. Traffic-light evaluation of the number of indicators used.

	Number of indicators		
Evaluation criteria	BD core indicators	BD core + WFD and Eutro core indicators	all available indicators (including national indicators)
Deviation in result	🟩	🟩	🟩
Differences between areas	🟨	🟨	🟨
Missing indicators	🟥	🟨	🟨
Number of indicators per aggregation	🟥	🟨	🟨

In the test where indicators were used under several criteria, the assessment result was quite similar if the criteria were under different descriptors (Table 7). Using all criteria the indicators have been evaluated to correspond to, the assessment results deviated from the other options. Although the coverage of criteria was highest in this option, the duplication of indicators under same descriptor should be avoided.

Table 7. Traffic-light evaluation of assigning indicators to several criteria.

	Indicators with multiple criteria		
Evaluation criteria	indicator used only once	indicator used once per relevant descriptor	indicator used for all relevant criteria
Deviation in result	🟩	🟩	🟥
Differences between areas	🟩	🟩	🟩
Missing indicators	🟥	🟨	🟨
Number of indicators per aggregation	🟥	🟨	🟩

4.5 Cross-criteria analysis and recommendations for the structure of the BEAT 3.0

Alternative structures (species vs criteria approach): The alternative structures gave different outcomes in the different case study areas when using the OOAO rule. In our tests, the species based approach better reflected the status of species, whereas specific aspects of the ecosystem component, e.g. abundance, were better reflected in the criteria based approach. In a biodiversity assessment the species based approach gives a more easily communicable result.

Weighted averaging vs OOAO approach: The test results showed that using OOAO for integration of indicator results will give high weight to single indicators. Weighted averaging can, however, fade out signals of concern. Using weighted averaging with OOAO at a high level will consider the indicators evenly, but still include the precaution in the final result. Stopping the assessment at the ecosystem component level is more informative than a single value for overall biodiversity.

Spatial representation: Based on the indicators used in the test scenarios it is recommended that indicators are downscaled in order to secure representability of all ecosystem components in sub-areas. This follows the recommendation of HELCOM BalticBOOST Biodiv WS 1-2016 that results should be presented at the most detailed spatial level feasible.

Potential inclusion of additional indicators: The tests also showed that using only HELCOM core indicators for biodiversity will reduce the robustness of the assessment as the number of indicators per criterion and ecosystem component would be quite low. Including relevant eutrophication and WFD indicators (those that contribute to describing habitat condition) can improve the robustness, although all MSFD criteria cannot be assigned.

Use of the same indicator under several criteria: Using the same indicator under several descriptors can be recommended, but not to use the same indicator under several criteria within a descriptor.

It should be noted, that the choices on integration and aggregation methods as well as use of indicators should not be based on the status assessment result. These are fundamental questions on how to perform the assessment, and the scenario test results should only be used as illustration how the choices influence the assessment.