



Document title	Proposal on the aggregation procedure in the confidence assessment methodology of the HELCOM Eutrophication Assessment Tool HEAT 3.0
Code	2-1
Category	INF
Agenda Item	2-Eutrophication confidence assessment
Submission date	15.5.2020
Submitted by	Germany
Reference	

Background

This proposal mainly addresses the further development of the confidence assessment in HEAT 3.0 related to the aggregation procedure of different confidence aspects to the overall confidence result. As agreed at HELCOM IN-EUTROPHICATION 16-2020, temporal and spatial confidence should be included in the test implementation in HEAT 3.0 and in this context the aggregation of confidence results on indicator and criteria level to the overall confidence needs to be addressed and decided in the further development.

A brief update of the current status of the confidence assessment and its further development is also included in this document as a basis for further steps in the aggregation procedure of the assessment. The results of the discussions at IN-EUTROPHICATION 16-2020 for agreed or possible new aggregation rules in HEAT 3.0 are also taken into consideration.

Action requested

The Meeting is invited to take note of the information in this document and discuss the proposal for the aggregation of different confidence aspects and indicator values to overall confidence for subsequent test implementation in HEAT 3.0 and agreement of further steps.

Introduction

This document mainly addresses the further development of the confidence assessment in HEAT 3.0 related to the aggregation procedure of different confidence aspects (temporal and spatial, potentially also accuracy aspects) for the indicators used in the assessment to the overall confidence result. At IN-EUTROPHICATION 16-2020 it was agreed that for temporal confidence both proposed aspects of general and specific temporal confidence should be included in the assessment, while for the spatial confidence only the specific aspect based on the percentage of sampled grid cells will be included. By this the confidence assessment will be more differentiated concerning data coverage and representability, but it is also necessary to include accuracy and possibly methodological confidence aspects later on. This would ensure a similar procedure as in the BEAT assessment tool, which also uses some of the eutrophication indicators, and would thus contribute to the harmonisation of the outcomes of the biodiversity and eutrophication status assessments.

Present status of the confidence assessment and further development

The implementation of the confidence proposal for the test assessment as discussed at IN-EUTROPHICATION 16-2020 is currently being prepared. On the basis of the test results, further adjustments are being made with regard to the confidence class boundaries used and other indicator- and area-specific modifications. At present, confidence class boundaries have only been proposed for winter nutrients (DIN, DIP) and chlorophyll. In addition to these indicators, Secchi depth, total nutrients and oxygen will be considered in the further development. For total nutrients and also for Secchi depth, the same class boundaries as for chlorophyll could be used to reflect the same or even longer assessment season as well as the lower natural variability of total nutrients compared to chlorophyll. For oxygen debt, a solution still needs to be found to define the confidence classes according to the calculation procedure of the indicator and the number of data used for this assessment.

The objective is to ensure that confidence assessment with defined class boundaries for the different aspects of confidence is at least possible for all core indicators in offshore areas of the HELCOM sub-basins. In coastal waters it will probably be necessary for Contracting Parties to provide confidence estimates in addition to national data, in particular for those biological indicators which have results in the form of multi-metric indices.

For the test assessment in HEAT 3.0, some changes have been proposed based on the discussions at IN-EUTROPHICATION 16-2020. The spatial confidence assessment will be performed with 30K grid cells in general instead of 20K and 60K grid cells for different areas to reflect that for all assessment areas the same grid cell size should be used as far as possible. Only one exception might be necessary for the small area of The Sound, where 10K grid cells could be used.

It is recommended to include other statistical parameters in HEAT as additional information, such as the standard error and the 95% confidence interval, which can be particularly useful for the accuracy confidence aspect with regard to the probability of classification. The standard error can be used as basis for the calculation of the confidence interval and also for a possible Monte Carlo simulation. A precondition for the inclusion of the standard error is the availability of standard deviation and the number of data. For some indicators like cyanobacteria bloom index or WFD indicators in national coastal waters this information is missing and in these cases confidence estimations should be provided separately or based on expert judgement. The 95% confidence interval including lower and upper limit could be used as additional information, in particular in cases close to the good/moderate boundary to find out where the target values lie within the confidence interval and therefore a clear classification as GES or sub-GES is difficult. These values should be marked to indicate that a more detailed analysis of the data is needed to allow for a final classification. Such cases are not necessarily related to low confidence but may also occur for indicators

with a sufficient number of measurements due to higher standard errors. It should be kept in mind that this procedure of simple confidence interval calculation is only suitable for normally distributed data and a number of measurements > 30 and should therefore be treated with caution when less data is available.

To consider the accuracy confidence aspect it is suggested to include the Monte Carlo simulation similar to the HELCOM BEAT tool as already proposed in previous documents to include a statistical resampling method to determine the uncertainty or probability of the classification results.

Aggregation procedure of the confidence assessment

The different confidence aspects for temporal and spatial and later possibly also accuracy and methodological aspects have to be combined to the indicator confidence. Since we currently have two different temporal but only one spatial aspect, the general temporal (GTC) and the specific temporal confidence (STC) have to be averaged in the first step to result in one total temporal confidence value (TTC). In the second step, the different confidence aspects used are averaged to the indicator confidence as shown in Figure 1.

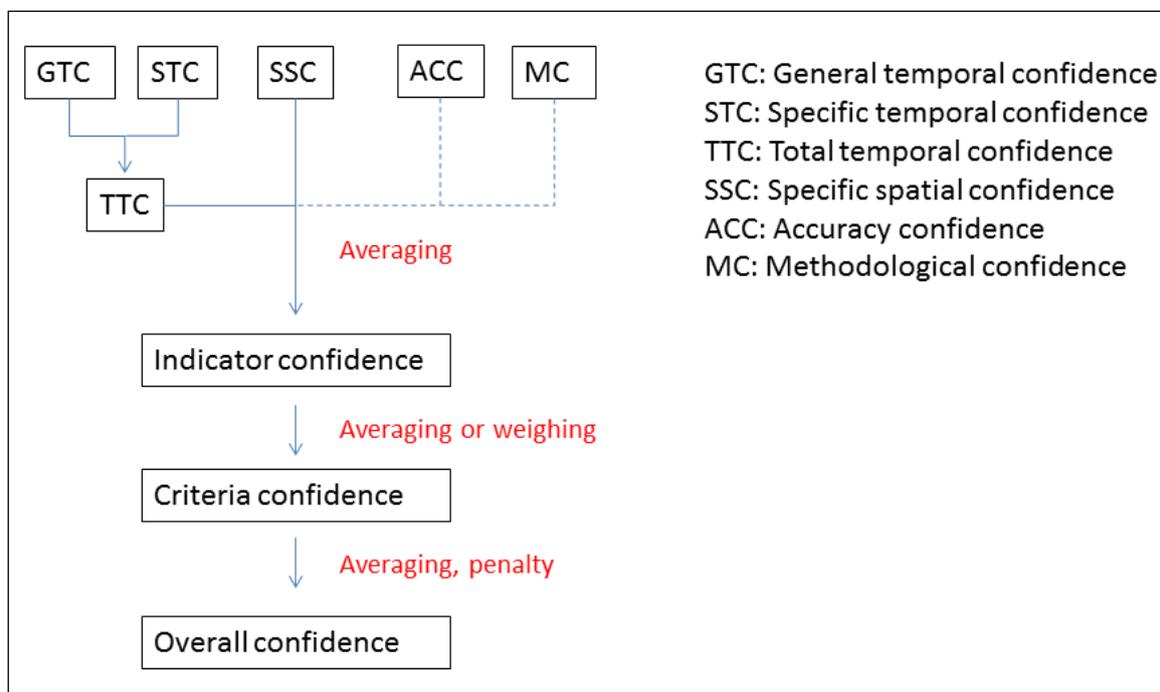


Figure 1: Aggregation of different confidence aspects for resulting indicator confidence

This means that currently temporal (TTC) and spatial (SSC) confidence is averaged for the indicator result. Further confidence results for accuracy and methodological aspects (ACC, MC) may be added later.

The different indicator confidence results have to be aggregated to result in the overall confidence of the assessment (Figure 2). The same procedure should be used for this as for the status assessment of the indicators in HEAT 3.0. Indicators of the same criteria group should be averaged or weighted according to the area-specific definitions, e.g. a ratio of 33:67 for DIN:DIP in the Bothnian Bay, while for almost all other HELCOM sub-basins the indicator confidence values can be simply averaged (ratio of 50:50) for criterion 1 to obtain a confidence result at the criteria level.

Based on the outcome of IN-EUTROPHICATION 16-2020 this would mean that Secchi depth will be averaged with oxygen debt indicator and/or zoobenthos indicator (BQI) for criterion 3 (indirect effects) and no longer together with chlorophyll in criterion 2 (direct effects) if it is applied in all areas. Pending a final decision, Secchi depth could be placed either in criterion 2 or 3 (as shown in Figure 2), depending on the area, which would affect both status and confidence assessments.

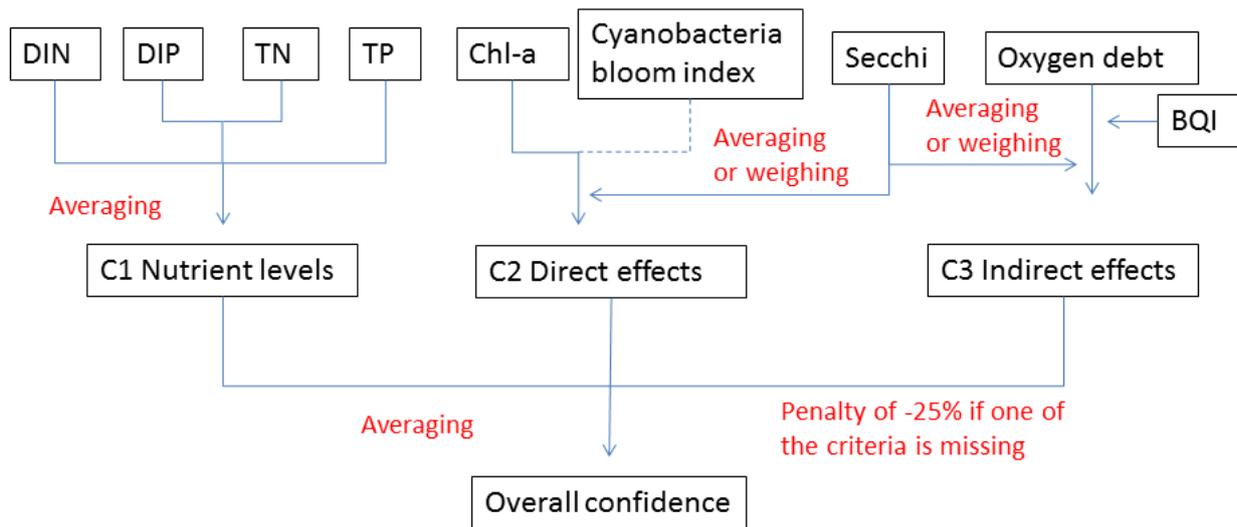


Figure 2: Aggregation of confidence on indicator and criteria level to result in the overall confidence of the eutrophication assessment

If the shallow water oxygen indicator will be added to criterion 3 in future, a decision will be necessary on averaging or weighing these indicators, at least for those sub-basins where both oxygen indicators will be used together with Secchi depth. This will also influence the confidence assessment accordingly.

The different indicator confidence results are combined to criteria results according to the aggregation principle of the status assessment:

1. DIN, DIP, TN and TP are averaged in C1 (Nutrient levels), with the exception of two sub-basins (Gulf of Riga and Bothnian Bay) where indicator weights have to be used
2. Chlorophyll-a and cyanobacterial bloom index (if indicator used and confidence of this indicator will be provided as stated in the Eutrophication Manual: 'estimated by data submitter') are averaged in C2 (direct effects); in case Secchi depth is also used in selected sub-basins in C2, the area-specific indicator weights related to chlorophyll and Secchi depth should be used as laid down in the Eutrophication Manual
3. Oxygen, Secchi depth and possibly benthic fauna (BQI) are averaged or weighted in C3 (indirect effects). When shifting Secchi depth from the second to the third criterion, it has to be decided whether oxygen and Secchi are to be treated equally or with a different weighing or percentage distribution in the status assessment. This decision will also determine the way in which these indicators are averaged for the confidence assessment.

The three criteria confidence results are averaged to the overall confidence and no one-out-all-out principle will be applied in the final step following the procedure for the confidence assessment currently used in HEAT 3.0. In order to ensure that a robust assessment is carried out taking into account all criteria and as many indicators as possible, a penalty may be imposed if one of the three criteria cannot be included because the relevant indicators are missing. This applies in particular to criteria C2 and C3 for the determination of eutrophication effects. In the HEAT approach, elevated nutrient concentrations alone without measurable effects in criteria groups C2 and C3 can also determine the classification in the non-good status (in contrast to the OSPAR assessment procedure), but whether such effects of nutrient enrichment are detectable is shown in particular by the indicators of criteria C2 and C3. Direct and indirect effects should therefore be the decisive criteria. The regulation of a penalty could be specified in such a way that the overall confidence is only reduced by -25% if C2 or C3 are missing, and thus the evaluation of the effects is only possible to a limited extent.

The second rule for imposing a penalty in the case a criterion is only represented by one indicator should be carefully evaluated in order not to worsen the confidence result more than necessary. Considering the discussion at IN-EUTROPHICATION 16-2020 on proposed new aggregation rules and the shifting of Secchi depth, possible single nutrient criteria for nitrogen and phosphorus, or the establishment of a fourth criteria group for ecosystem functioning in future may lead to situations where only one indicator represents a criteria group. However, a fundamental downgrading of the confidence assessment associated with this is not beneficial if, through changes in the aggregation rules, a more differentiated and ecologically meaningful eutrophication assessment is aimed for. Therefore, it is recommended not to use a general penalty of -25% if a criterion is only represented by one indicator. This can be demonstrated by a specific sub-basin example for the Kattegat. Due to the shift of Secchi depth to the third criteria group and the fact that cyanobacteria bloom index is not applicable in this sub-basin, this would mean that only chlorophyll will be assessed for criterion 2. On the basis of a high number of observations and also an adequate spatial distribution, it would not be appropriate to impose a penalty and end up with the same confidence result as other chlorophyll indicator results based on fewer observations and a worse distribution. This would be sending out the wrong signal in terms of observation efforts and monitoring strategy planning. In principle, the objective of carrying out the eutrophication assessment on the basis of the largest possible number of meaningful indicators should continue to be achieved in order to provide reliable results.

The aggregation of the confidence results should follow the same procedure than the status assessment as long as individual confidence results on indicator level will be provided (in HEAT 3.0 and later also in assessment reports) and effects of the application of the aggregation rules can be followed in the subsequent confidence assessment procedure. A completely different approach, e.g. a simple averaging of the confidence results of all indicators without taking into account the allocations to criteria and weightings, is not considered useful and cannot be recommended, since it would be contrary to the aim of harmonising BEAT and HEAT tools.