Document title	Proposal on improving the confidence assessment methodology of the HELCOM Eutrophication Assessment Tool HEAT 3.0 – update including assessment examples
Code	4-2
Category	INF
Agenda Item	4.2-Confidence assessment
Submission date	16.1.2020
Submitted by	Germany
Reference	

Background

This document, originally presented at IN-EUTROPHICATION 7-2017 and later also at IN-EUTROPHICATION 10-2018 and 11-E-2018 in an advanced version, proposes adjustment to the HEAT 3.0 assessment tool for assessing confidence based on the development under the BalticBOOST and SPICE projects for improving the confidence assessment within the Biodiversity Assessment Tool (BEAT).

The proposal mainly addresses the temporal and spatial confidence to improve the confidence assessment concerning data coverage and representability. But it includes also suggestions how accuracy and methodological confidence aspects can be supplemented. The intention is to generate differentiated confidence values for all eutrophication indicators, that can also be applied in the BEAT tool for those indicators that are used by HEAT and BEAT, in order to harmonise the outcomes of the biodiversity and eutrophication status assessments. The experience from the ongoing development of the confidence assessment in OSPAR for the revision of the Comprehensive Procedure and different national approaches have also been taken into account for this proposal.

Action requested

The Meeting is invited to take note of the information in this document and discuss the confidence rating proposal for possible test implementation in HEAT 3.0.

Introduction

The HELCOM Eutrophication Assessment Tool (HEAT 3.0) is proposed to be adjusted to develop the confidence methodology, which at present only considers the temporal coverage of monitoring data, to take into account also spatial representability, accuracy and methodological confidence, as proposed for the HELCOM Biodiversity Assessment Tool (BEAT) being developed under the BalticBOOST project.

As these adjustments are already in place in the HELCOM Biodiversity Assessment Tool (BEAT), confidence information needs to be determined for those eutrophication indicators applied also in the biodiversity assessment.

Improving the confidence assessment

At present, HEAT 3.0 assesses confidence as a product of status- and target confidences. Target confidence describes the reliability on the method and/or data used to set the targets, and in terms of the database algorithms, it is pre-defined along with the targets. Status confidence, on the other hand, is based on the temporal coverage of the monitoring data, and can be produced by the database algorithms. The present method is criticized in that it does not take into account how accurate the indicator estimate is in terms of natural variability, or how representable it is spatially, or how confident we are in the monitoring method in general.

The BalticBOOST project, while developing the Biodiversity Assessment Tool (BEAT), defined and developed the confidence assessment methodology further. Firstly, it was proposed that indicator confidence should be based only on the confidence of the monitoring data. The argument was that targets, although scientifically based, are finally a product of political decision-making, and as such not comparable in terms of confidence. The confidence was proposed to be determined through four confidence aspects: a) Temporal coverage, b) Spatial representability, c) Accuracy and d) Methodological confidence. The confidence of each of these is classified either high (=1), moderate (=0.5) or low (=0).

a) Temporal coverage of monitoring data

The aspect of temporal coverage of monitoring data considers the confidence of the indicator to include year-to-year variation in the indicator result. The criteria for evaluating temporal coverage as used for status-confidence in the present HEAT 3.0 tool as ES-score were slightly adapted to consider different indicator data needs of nutrient concentrations in winter and chlorophyll in the growing season (Table 1). In general, the natural variability of chlorophyll in summer is higher than that of nutrient concentrations in winter and the inter-annual variability of chlorophyll is also higher than that of winter nutrients, although the range of variability is area-specific. Therefore, the number of annual observations should be higher for chlorophyll than for nutrients to take this into account. The temporal confidence can be calculated from the eutrophication assessment data, using the database algorithms.

Table 1: Confidence classes for temporal coverage of monitoring data for winter nutrients and chlorophyll in the growing season

Score	Evaluation criteria for winter nutrients (XII-II)	Evaluation criteria for chlorophyll in the growing season (VI-IX)
HIGH	The evaluation is based on more than 15 annual observations during the given assessment period	The evaluation is based on more than 20 annual observations during the given assessment period
MODERATE	The evaluation is based on a minimum of 5 annual observations; or The evaluation is based on more than 15 annual observations during a 5-year period that overlaps with, but is not	The evaluation is based on a minimum of 7 annual observations; or The evaluation is based on more than 20 annual observations during a 5-year period that overlaps with, but is not exactly the same as the given assessment period

Score	Evaluation criteria for winter nutrients (XII-II)	Evaluation criteria for chlorophyll in the growing season (VI-IX)
	exactly the same as the given assessment period	
LOW	The evaluation is based on less than 5 annual observations; or The evaluation is based on more than 5 annual observations during a 5-year period that has temporal overlaps with, but is not exactly the same as the given assessment period; or The evaluation is based on a period that does not have temporal overlap with the given assessment period	The evaluation is based on less than 7 annual observations; or The evaluation is based on more than 7 annual observations during a 5-year period that has temporal overlaps with, but is not exactly the same as the given assessment period; or The evaluation is based on a period that does not have temporal overlap with the given assessment period

The proposed class boundaries for temporal confidence estimation apply primarily to in-situ data, but not neccessarily to earth observation and ferry box data. According to the HELCOM HEAT 3.0 manual for eutrophication assessment, the classification for the ES score with class boundaries of <5, 5-15 and >15 was used in the HOLAS II assessment for chlorophyll even when using satellite data. Considering the large amount of data, the class boundaries of the confidence assessment should be adjusted for satellite and ferry box data in the future.

This general temporal approach can be supplemented by a detailed assessment of the indicator-specific assessment periods, e.g. the distribution of observations in the individual months of the growing season for chlorophyll, or in the winter period for some of the nutrients. The number of missing months in the respective assessment periods of the different indicators will then determine the class boundaries. For the specific temporal confidence, the requirements for regular observations should also be higher for chlorophyll than for winter nutrients to take into account the different natural variabilities of the indicators.

The proposed class boundaries for the specific confidence for winter nutrients and chlorophyll in the growing season are summarised in Table 2. If the number of missing months is less than 3 in the whole assessment period of 5 years for example, the confidence will be rated as high. Moderate confidence will be assigned when observations are missing in 3 up to 5 months of the assessed period. If more than 5 months are missing, the confidence will be rated as low. Although the assessment season for chlorophyll is one month longer than for winter nutrients, the same class boundaries should be applied to reflect the higher data requirements for chlorophyll, because in relation to the total number, fewer months may be missing for chlorophyll to result in the same confidence class. This is to ensure regular sampling in the confidence assessment and to take into account the higher natural variability.

For indicators that are evaluated throughout the year like total nitrogen and total phosphorus, a less strict adjustment could be made. For indicator and area-specific adaptations the standard deviation of indicator values as percentage or the min-max variation could be included for areas with steeper gradients.

Table 2: Confidence classes for specific temporal coverage for winter nutrients and chlorophyll in growing season related to the number of missing observation months

Score	Evaluation criteria for winter nutrients (XII-II) and chl in growing season (VI-IX)
HIGH	< 3 missing months
MODERATE	> 3 - ≤ 5 missing months
LOW	> 5 missing months

In a first approach, the general and the specific temporal confidence have been combined to the total temporal confidence by simply calculating the arithmetic mean value. Of course, the application of a weighted approach would also be possible, for example, using a ratio of two-thirds to one-third, and should be investigated further. In order to classify the results of the integrated confidence, the same value ranges have been used as in the biodiversity assessment of BEAT:

> 0.75 as high, 0.5-0.75 as moderate and < 0.5 as low.

For the confidence test assessment, four different HELCOM sub-basins have been selected based on different data coverage and area size: Kattegat, Eastern Gotland Basin, Gulf of Riga and the Gulf of Finland. In all areas the indicators winter DIP and summer in-situ chlorophyll have been included to investigate differences in the confidence assessment. For the Gulf of Finland, a comparison between different assessment periods was additionally carried out, which is therefore presented in a separate table (Table 4).

Table 3 shows basic data information and results of temporal confidence test assessments in the selected HELCOM sub-basins.

Table 3: Temporal confidence assessment for DIP and chlorophyll indicator related to total number of observations and the number of missing observations in the assessment period 2011-2016

Area	Indicator	No. of observations	Minimum no. of annual observations	General temporal confidence class	No. of missing month	Specific temporal confidence class	Total temporal confidence
SEA-001 Kattegat	DIP (XII-II)	271	36	High (1.0)	0	High (1.0)	High (1.0)
	Chl-a (VI-IX)	300	43	High (1.0)	0	High (1.0)	High (1.0)
SEA-009 Eastern Gotland Basin	DIP (XII-II)	135	13	Moderate (0.5)	1	High (1.0)	Moderate (0.75)
	Chl-a (VI-IX)	227	30	High (1.0)	0	High (1.0)	High (1.0)
SEA-011 Gulf of Riga	DIP (XII-II)	25	2	Low (0.0)	14	Low (0.0)	Low (0.0)
	Chl-a (VI-IX)	70	7	Low (0.0)	14	Low (0.0)	Low (0.0)

The results of the temporal confidence assessment showed substantial differences between the different sub-basins. High confidence was determined in the Kattegat for both indicators and both temporal confidence aspects and also for chlorophyll in the Eastern Gotland Basin. The Gulf of Riga was classified as low based on the relatively low number of observations and a higher number of missing months in the respective assessment periods of winter and the growing season. The Gulf of Finland was also classified as low in total (Table 4), although the general temporal confidence was moderate due to the significantly higher number of observations compared to the Gulf of Riga.

Table 4: Comparison of temporal confidence in the Gulf of Finland between the assessment periods 2007-2011 and 2011-2016 for DIP and chlorophyll

Area	Indicator	No. of observations	Minimum no. of annual observations	General temporal confidence class	No. of missing month	Specific temporal confidence class	Total temporal confidence
SEA-013 Gulf of Finland 2007-2011	DIP (XII-II)	88	8	Moderate (0.5)	5	Moderate (0.5)	Moderate (0.5)
	Chl-a (VI-IX)	67	4	Low (0.0)	9	Low (0.0)	Low (0.0)
SEA-013 Gulf of Finland 2011-2016	DIP (XII-II)	100	12	Moderate (0.5)	8	Low (0.0)	Low (0.25)

Area	Indicator	No. of observations	Minimum no. of annual observations	General temporal confidence class	No. of missing month	Specific temporal confidence class	Total temporal confidence
	Chl-a (VI-IX)	156	14	Moderate (0.5)	8	Low (0.0)	Low (0.25)

The comparison of the assessment periods from 2007-2011 and 2011-2016 in the Gulf of Finland showed differences in particular in the number of observations, which led to an improved general temporal confidence assessment for chlorophyll, but not for DIP. However, since the specific temporal confidence was still rated as low due to the relatively high number of missing months in the period 2011-2016, no class improvements were observed in the overall temporal confidence. The same comparison was made for the spatial confidence in the Gulf of Finland and is summarised in Table 8.

At present, the general temporal confidence classification is based on the condition of a minimum number of annual observations. For the temporal confidence assessment of chlorophyll in the Gulf of Finland in the period 2011-2016, this resulted in a moderate classification based on one year with only 14 samples, while the average number of observations of all other years was 24, indicating a higher confidence. However, to ensure a sufficient number of regular observations over the entire assessment period, the strict condition of the annual minimum number of observations should be maintained in line with the precautionary principle.

b) Spatial representability of monitoring data

The spatial representability of monitoring data assesses how well the indicator covers the spatial variation within the assessment unit. However, the measure should be relatively simple for potential implementation in HEAT 3.0, describing for example how well the data is evenly distributed within the area. As a first approach to cover the spatial representability in the confidence test assessment, the number of observations in the assessment period was related to a fixed grid cell size of 20K in the different assessment areas. Since the same grid cell size is already used for the calculation and assessment of chlorophyll satellite data, this will facilitate the implementation of the confidence assessment approach in the workspace. For practical reasons 60K grid cells can be used for some of the larger HELCOM sub-basins, e.g. the Eastern Gotland Basin or the Bothnian Sea.

In the BEAT tool, spatial representability is considered high if the data represents reliably at least 90% of the relevant habitat types occurring in the area, or in cases with a clear spatial gradient or patchiness in the parameter value, the monitoring set to cover at least 90% of this variation. When the representability or variation (in case of gradients) is covered by 70-89% moderate confidence is assigned to the indicator. Confidence is considered to be bad if less than 70% of relevant habitats or less than 60% of the variation in gradients are covered. This is usually based on expert judgement and can not be calculated within the HEAT tool.

The number of observations was summed up for the assessment period and related to the number of grid cells per assessment area in the different selected HELCOM sub-basin example areas. This resulted in a number of observations per grid cell within the assessment unit without considering the evenness of the distribution of observations within the assessment area. For the three scoring categories, ‘high’, ‘moderate’, and ‘low’, general numbers of observations per grid cell were defined, differentiated for nutrients and chlorophyll to meet the different data requirements (Table 5). Adaptations might be useful in future, in particular for coastal areas depending on the steepness of gradients and other regional specifications.

This schematic general approach was supplemented by using a more specific approach to take into account the distribution of observations within the areas. This has been done by counting the number of sampled and not sampled grid cells within the defined area and then calculating the percentage of sampled grid cells in relation to the total number of grid cells in this area. The proposed class boundaries are listed below and have been differentiated for winter nutrients and chlorophyll in the growing season to take into account the different natural variabilities of these indicators with a higher data requirement for chlorophyll than for nutrients (Table 5). This general approach is based on a similar procedure used in the latest national assessment according to the OSPAR Comprehensive Procedure in German coastal and marine waters in a simplified way. The numbers of observations per grid cell have been extracted and modified from the German approach for the spatial confidence assessment.

Table 5: Proposed class boundaries for spatial confidence rating for the number of observations per grid (20K) and the percentage of sampled grid cells per area

Score	Evaluation criteria for general spatial confidence – n/grid (20K)		Evaluation criteria for specific spatial confidence - % of sampled grid cells	
Parameter	**Winter nutrients**	**Chlorophyll**	**Winter nutrients**	**Chlorophyll**
HIGH	> 4	> 5	> 70 %	> 80 %
MODERATE	2-4	3-5	50 – 70 %	60 – 80 %
LOW	< 2	< 3	< 50 %	< 60 %

Following the scheme given in Table 5, a high confidence score for the general spatial confidence will require a total of more than 4 samplings of winter nutrients or more than 5 samplings of chlorophyll per 20K-grid-cell. In addition to this, a high confidence score for the specific spatial confidence will be defined as reaching a sampling coverage of more than 70 % of the area-related grid cells for winter nutrients and more than 80 % for chlorophyll. The latter will take into account the higher natural variability of chlorophyll and the resulting higher requirements for the data density.

Table 6 lists the confidence class boundaries for the number of observations per 60K grid cells, which can be used for larger sub-basins. The number of observations per grid cell was not simply translated by a factor of nine from 20k to the larger grid cells of 60k, but it was taken into account that the two HELCOM sub-basins proposed for the use of 60K grids are very large and central areas which for practical reasons cannot be sampled to the same extent as smaller and near-coastal areas and where lower gradients are assumed. Therefore, the class boundaries have been set less strict. For the specific spatial confidence of 60K grids, the same percentage of sampled grid cells should be applied as for 20K grids.

Table 6: Proposed class boundaries for spatial confidence rating for the number of observations per grid for 60K grid cell size

Score	Evaluation criteria for general spatial confidence – n/grid (60K)	
Parameter	**Winter nutrients**	**Chlorophyll**
HIGH	> 16	> 20
MODERATE	8-16	10-20
LOW	< 8	< 10

Both spatial confidence assessments have been carried out for the same HELCOM example areas as for the temporal confidence rating for the indicators DIP and chlorophyll. The number of samplings (n) was related to the number of 20K-grids/area (20K grid = 400 km^2) in the sub-basins of Kattegat, Gulf of Riga and Gulf of Finland, while for the Eastern Gotland Basin 60K-grid-cells were used with the proposed class boundaries for the resulting confidence estimation as listed in Table 6. The reason for this was that large areas like the Eastern Gotland Basin or the Bothnian Sea in the central and northern part of the HELCOM region could not be sampled in the same way as smaller and near-coastal areas. Based on the assumed lower variability in these areas, a lower number of observations should be sufficient in these exceptional cases. It will be recommended to use 60K-grid-cells only for those areas with a size of more than 100 20K-grid-cells. Furthermore, it shall be made clear at this point that it shall not be a free choice whether either 20K- or

60K-grid-cells are used for the confidence assessment in the different areas, but the decision on that will rather be determined by area size and location.

In the OSPAR Common Indicator assessment, so-called quadrants (defined squares of around 3275 km^2) have been used for the estimation of data coverage in the region of the Greater North Sea. The confidence rating for the upcoming assessment of the OSPAR Quality Status Report in 2023 is currently being developed based on a new assessment area division and it is not yet decided whether a similar approach to the one proposed for HELCOM or another approach more focussed on the accuracy of the confidence rating will be used in future.

As it had been done before for the temporal confidence rating, the general and specific spatial confidence values have been averaged in order to calculate the total spatial confidence. Depending on the importance of the different spatial confidence apsects, a weighted approach would be also possible.

The following Table 7 gives an overview of the spatial confidence assessment results for DIP and chlorophyll in the different sub-basins.

Table 7: Spatial confidence assessment for DIP and chlorophyll indicator in different assessment areas related to the number of observations per grid and the percentage of sampled grid cells per area

Area	Indicator	No. of observations	No. of 20K-/60K-grids/area	n/grid	General spatial confidence class	% of sampled grids	Specific spatial confidence class	Total spatial confidence
SEA-001 Kattegat (20K)	DIP (XII-II)	271	38.1	7.1	High (1.0)	49.9	Low (0.0)	Moderate (0.50)
	Chl-a (VI-IX)	300	38.1	7.9	High (1.0)	28.9	Low (0.0)	Moderate (0.50)
SEA-009 Eastern Gotland Basin (60K)	DIP (XII-II)	135	19.7	6.9	Low (0.0)	86.3	High (1.0)	Moderate (0.50)
	Chl-a (VI-IX)	227	19.7	11.5	Moderate (0.50)	96.4	High (1.0)	Moderate (0.75)
SEA-011 Gulf of Riga (20K)	DIP (XII-II)	25	21.7	1.2	Low (0.0)	41.5	Low (0.0)	Low (0.0)
	Chl-a (VI-IX)	70	21.7	3.2	Moderate (0.5)	55.3	Low (0.0)	Low (0.25)

The total spatial confidence was rated ’moderate’ in the Kattegat and the Eastern Gotland Basin, while the assessment result in the Gulf of Riga was ’low’ for both indicators. The individual spatial confidence aspects showed quite different results based on the varying number of observations and different area sizes. The major difference in the classification of the percentage of sampled grid cells between the 20K and 60K areas indicates a possible need for adaptation of confidence class boundaries to harmonise assessment results based on different grid cell sizes.

The following Table 8 lists the underlying data for the spatial confidence test assessment in the Gulf of Finland for the assessment periods 2007-2011 and 2011-2016 in order to enable a direct comparison and resulting differences. The number of samplings has increased slightly for DIP, but significantly for chlorophyll. Therefore, the number of samplings per 20K-grid is in the same range for DIP, while it has more than doubled for chlorophyll associated with a change in the classification from low to moderate. However, the increase in the number of samples has not resulted in a significantly higher number of sampled grid cells, so that the percentage increased only slightly and did not lead to any change in the classification of the specific spatial confidence. Although the percentage of grid cells sampled in the period 2011-2016 was higher for chlorophyll than for DIP, the resulting classification was better for DIP than for chlorophyll, based on the different confidence class boundaries for these two indicators reflecting the higher data requirement for chlorophyll.

Table 8: Comparison of spatial confidence in the Gulf of Finland between the assessment periods 2007-2011 and 2011-2016 for DIP and chlorophyll

Area	Indicator	No. of observations	No. of 20K-/60K-grids/area	n/grid	General spatial confidence class	% of sampled grids	Specific spatial confidence class	Total spatial confidence
SEA-013 Gulf of Finland 2007-2011	DIP (XII-II)	88	41.5	2.1	Moderate (0.5)	48.2	Low (0.0)	Low (0.25)
	Chl-a (VI-IX)	67	41.5	1.6	Low (0.0)	48.2	Low (0.0)	Low (0.0)
SEA-013 Gulf of Finland 2011-2016	DIP (XII-II)	100	41.5	2.4	Moderate (0.5)	50.6	Moderate (0.5)	Moderate (0.5)
	Chl-a (VI-IX)	156	41.5	3.8	Moderate (0.5)	55.4	Low (0.0)	Low (0.25)

The comparison in the Gulf of Finland showed improvements between the two assessment periods, but also the need of further monitoring efforts.

The following Figure 1 gives an overview of the distribution of chlorophyll and DIP samplings in the Eastern Gotland Basin, the Gulf of Finland and the Gulf of Riga for the estimation of data coverage in the assessment period 2011-2016.

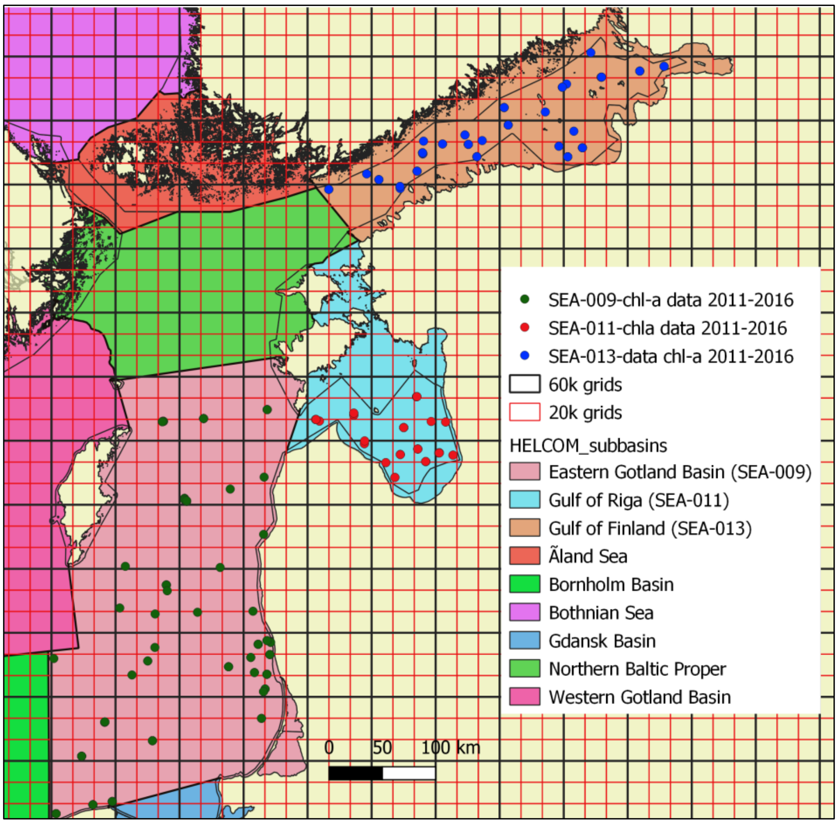

Figure 1: Distribution of chlorophyll samples in the assessment period 2011-2016 in the selected example areas of the Gulf of Finland, the Gulf of Riga and Eastern Gotland Basin

Figure 2 shows the distribution of chlorophyll and DIP samplings in the Gulf of Finland between the two assessment periods in more detail, in order to count the number of sampled and not-sampled (empty) grid cells for assessing the specific spatial confidence. Of a total of 41.5 20K-grids for the Gulf of Finland, 20 grids had been sampled for DIP and 23 grids for chlorophyll in the period 2011-2016, corresponding to 50.6 % and 55.4 % of the total grid cells. According to the proposed class boundaries for the specific spatial confidence the percentage for chlorophyll was rated as 'low' in contrast to the 'moderate' classification for DIP with a lower percentage of sampled grid cells.

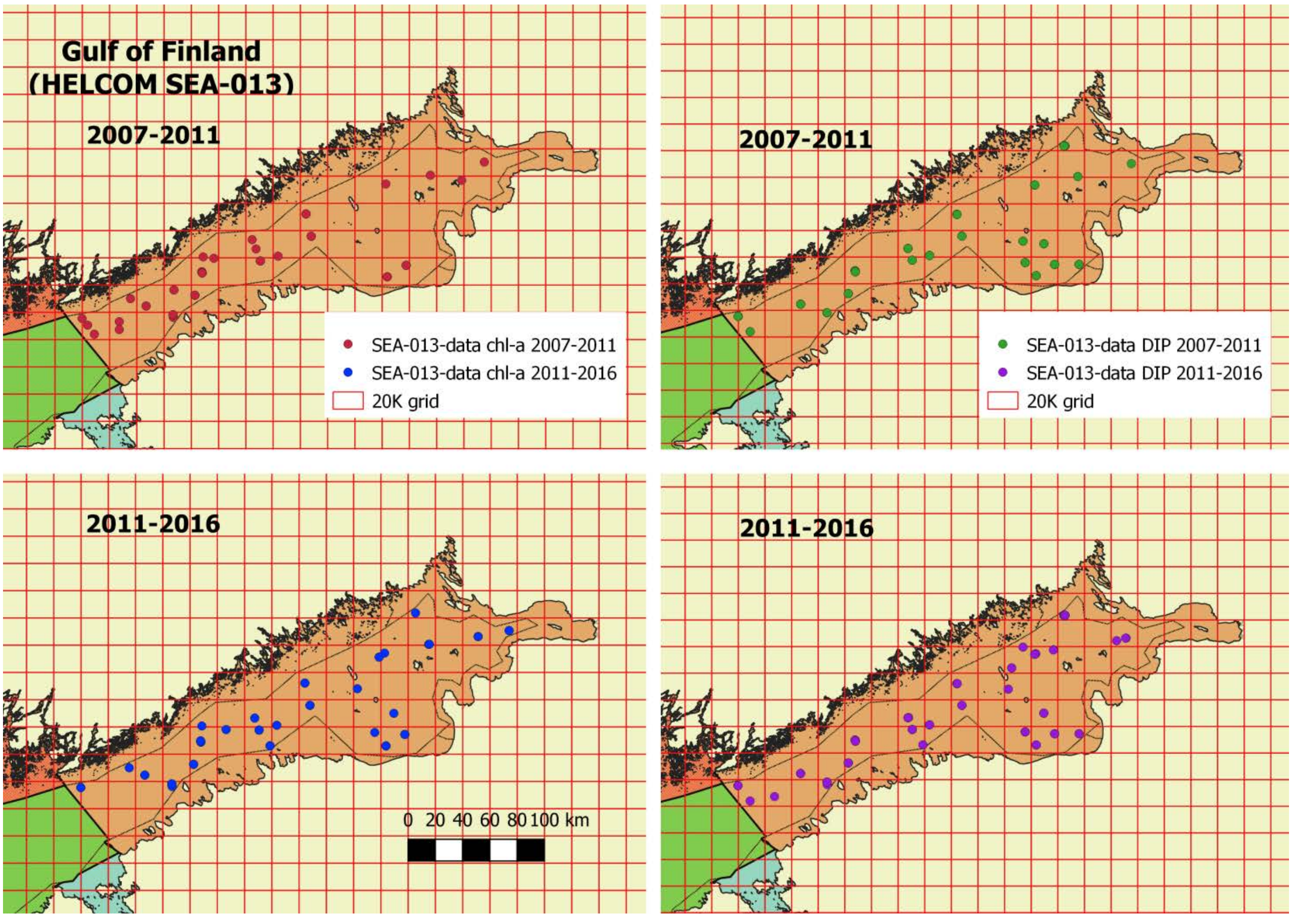

Figure 2: Distribution of chlorophyll samples in the assessment period 2011-2016 in the Gulf of Finland

In the current approach of this proposal, the final overall confidence has been calculated as the average of temporal and spatial confidence. Within the next steps, further refinements of the aggregation procedure such as weighting can also be tested, once the importance of the various confidence aspects has been agreed on. In the HELCOM BEAT tool, an integrated confidence assessment is carried out in parallel to the status assessment, which means that the basic integrated confidence is calculated following the same assessment structure and integration rules as for the corresponding status assessment. This procedure could serve as guidance also for setting up the assessment in HEAT 3.0.

The following Table 9 shows the overall, i.e. temporal plus spatial, confidence assessment in different HELCOM sub-basins.

Table 9: Final overall confidence assessment for DIP and chlorophyll indicator in different assessment areas

Area	Indicator	Total temporal confidence	Total spatial confidence	Final overall confidence
SEA-001 Kattegat (20K)	DIP (XII-II)	High (1.0)	Moderate (0.50)	Moderate (0.75)
	Chl-a (VI-IX)	High (1.0)	Moderate (0.50)	Moderate (0.75)

Area	Indicator	Total temporal confidence	Total spatial confidence	Final overall confidence
SEA-009 Eastern Gotland Basin (60K)	DIP (XII-II)	Moderate (0.75)	Moderate (0.50)	Moderate (0.625)
	Chl-a (VI-IX)	High (1.0)	Moderate (0.75)	High (0.875)
SEA-011 Gulf of Riga (20K)	DIP (XII-II)	Low (0.0)	Low (0.0)	Low (0.0)
	Chl-a (VI-IX)	Low (0.0)	Low (0.25)	Low (0.125)

The sub-basins selected for the confidence test showed different assessment results. The final overall confidence for the Kattegat resulted in ’moderate’ confidence for DIP and chlorophyll, whereas in the Gulf of Riga both indicators were assessed as ’low’. In the Eastern Gotland Basin, the confidence result for DIP was ‘moderate’, while for chlorophyll even ’high’ confidence was determined. The different indicator results are mainly due to the different number of observations for DIP and chlorophyll, which also influenced the spatial confidence, but continuous sampling during the assessment periods and evenness of spatial distribution contributed to the assessment result as well. The different data requirements for the two indicators have finally determined the resulting classification.

The comparison between the assessment periods 2007-2011 and 2011-2016 in the Gulf of Finland revealed only minor differences in the overall confidence assessment (Table 10). Both indicators were finally classified as ’low’ in both assessment periods, although the calculated number showed some differences within the ’low’ category. However, due to changes in the individual aspects of confidence, improvements and deteriorations could be observed between the two periods.

Table 10: Comparison of overall confidence in the Gulf of Finland between the assessment periods 2007-2011 and 2011-2016 for DIP and chlorophyll

Area	Indicator	Total temporal confidence	Total spatial confidence	Final overall confidence
SEA-013 Gulf of Finland 2007-2011	DIP (XII-II)	Moderate (0.5)	Low (0.25)	Low (0.375)
	Chl-a (VI-IX)	Low (0.0)	Low (0.0)	Low (0.0)
SEA-013 Gulf of Finland 2011-2016	DIP (XII-II)	Low (0.25)	Moderate (0.5)	Low (0.375)
	Chl-a (VI-IX)	Low (0.25)	Low (0.25)	Low (0.25)

The different temporal and spatial confidence aspects allowed a more detailed and accurate assessment of the monitoring data. Furthermore, gaps in temporal or spatial coverage can be clearly identified and be used as valuable hints to possible improvements of the sampling strategy.

It can be assumed that it will be possible to implement the confidence procedures of this proposal in HEAT 3.0 including the new spatial and updated temporal confidence aspects. Based on the HELCOM assessment units in combination with the 20k-grid-layer, the exact spatial estimation of sampled grid cells per assessment area should be possible. The fact that the same 20K-grid-layer is also used for the confidence estimation of the ES-score for chlorophyll earth observation and ferry box data in the HELCOM HEAT assessment could be helpful for the implementation. In the current HELCOM approach, the number of 20k-grid-cells containing data are counted and multiplied with the number of observation days during the year. The ES-score is calculated separately for the different data types in the HELCOM confidence assessment, but currently the same criteria based on the number of observations are used for the confidence classification. It should be investigated in detail whether this procedure will still be suitable in future or confidence class boundaries should be adapted for earth observation and ferry box data in difference to in-situ data for the next assessment.

After implementation of the confidence approach, the assessment results of the different confidence aspects in the eutrophication test assessment will provide more information on temporal and spatial data coverage of in-situ observations in the different assessment areas and can be used to evaluate the proposed method.

c) Accuracy of indicator result

The accuracy of the indicator result describes how certain the indicator estimate is, in relation to the variability of the data. It is primarily assigned as the standard error of the monitoring data.

The standard error can be achieved from all the present CORE eutrophication indicators, and the eutrophication database algorithms can be developed to calculate it when updating the indicators form the monitoring data. For earth observation data like chlorophyll a or cyanobacterial blooms, the standard error of daily means could also be calculated and provided by ICES algorithms. For biological indicators like the Benthic Quality Index, expert evaluations for the accuracy confidence from the BEAT tool can be used instead of the standard error.

If potential new indicators lack a value for standard error, an optional approach would be needed. For biodiversity indicators, as proposed by BalticBOOST, a categorical approach is carried out if the standard error is not available for new indicators. This is a compliance check by expert judgement of the probability that the indicator signal clearly reflects that GES is achieved/not achieved. High confidence is assigned if GES has ‘most likely’ been / has not been achieved (by at least 90% probability). Moderate confidence is judged if the probability is ‘likely’ (70-89% probability) and low confidence is judged if the probability of correctly indicating the status evaluation of the indicator is ‘unsure’ (less than 70% probability). This type of accuracy could not be calculated by the database algorithms.

Another possibility could be to include a statistical resampling method like the Monte Carlo simulation as used in the NEAT assessment tool of the DEVOTES project or the bootstrapping method applied for biological indicators in coastal areas to determine the uncertainty or probability of the classification results.

d) Methodological confidence

The aspect of methodological confidence considers the quality of the monitoring methodology. High confidence is assigned if the monitoring has been conducted according to HELCOM guidelines (for parameters where these are available) and the data is quality assured according to HELCOM or other internationally accepted guidelines. Moderate confidence is assigned if the monitoring has been conducted only partly according to HELCOM guidelines and/or the data originates from mixed sources, and is partly quality assured according to HELCOM or other international standards and/or the data is quality assured, but according to local standards. If monitoring has not been conducted according to HELCOM guidelines or the data has not been quality assured, the methodological confidence is considered low.

Methodological confidence is not calculated from the data, but is predefined for each parameter, based on information available in the HELCOM Monitoring Manual.

To differentiate methodological confidence, it might also be useful to distinguish between different data sources by defining different class boundaries for methodological reliability of in-situ, satellite, ferry box or modelled data.

If the meeting agreed on the proposed confidence amendments aiming at specifying temporal coverage and spatial representability for future eutrophication assessments, the different confidence aspects can be implemented in HEAT 3.0 in the framework of the eutrophication test assessment project with ICES. After evaluation of the test assessment results, further steps for modifications in the HELCOM workspace could be planned.

References

Brockmann U., Topcu D., Schütt M., Leujak W. (2017): Third assessment of the eutrophication status of German coastal and marine waters 2006-2014 in the North Sea according to the OSPAR Comprehensive Procedure, 108 pp.

OSPAR (2013): Common Procedure for the Identification of the Eutrophication Status of the OSPAR Maritime Area. OSPAR, reference no. 2013-8, 66 pp.