



## Baltic Marine Environment Protection Commission

Working Group on the State of the Environment and Nature  
Conservation

STATE & CONSERVATION  
15-2021

Online, 4-8 October 2021

---

<b>Document title</b>	Proposal for ongoing development of confidence assessment for hazardous substances indicators
<b>Code</b>	3J-97
<b>Category</b>	DEC
<b>Agenda Item</b>	3J – Progress of relevant HELCOM expert groups and projects
<b>Submission date</b>	13.9.2021
<b>Submitted by</b>	Baltic Data Flows Project

---

### Background

This document contains a proposal to apply improvements to the evaluation of confidence within the hazardous substances indicators for HOLAS III. Currently the confidence in the majority of hazardous substances indicators is evaluated based on an expert-derived component that is applied to all Scale 4 assessment units for each indicator, irrespective of the underlying data components. The proposal set out below sets out the structure to define a categorical approach towards setting confidence that would address spatial, temporal, methodological and threshold value confidence in an overall evaluation. The evaluation would also be applied at per assessment unit in each indicator.

The work would align the confidence aspect more closely with the approach applied for biodiversity and eutrophication (e.g. under HEAT and BEAT).

The work is being carried out under the Baltic Data Flows project and is expected to be available for application in HOLAS III. Should this not be possible the current approach would be retained for HOLAS III (i.e. that applied under HOLAS II). The changes to the confidence approach is not expected to have any impact on the ability to run the indicator evaluations or the integrated assessment for HOLAS III.

The initial proposals have been reviewed by EN-HZ (EN-HZ 15 and 16-2021), with general support for the improvements and guidance on important factors to consider.

### Action requested

The Meeting is invited to endorse the ongoing development of this improved confidence evaluation towards HOLAS III.

## Improved confidence assessment for HELCOM HZ indicators

This proposal comes as part of the ongoing work under the Baltic Data Flows project Activity 5. In the current HZ indicators a blanket confidence is applied to the whole indicator irrespective of the data (or other factors) entering the evaluation per assessment unit. The current confidence evaluation is a descriptive and expert-based single confidence evaluation that is applied retrospectively to all the assessment units. This same single value is the value that enters the CHASE integrated assessment per indicator to create the resulting integrated confidence in CHASE.

The proposed approach is a way to create a confidence evaluation that addresses the more complex information underlying the indicator evaluations, including considering this information at the assessment unit scale (e.g. per Scale 4 HELCOM Assessment Unit).

It is important to note that the proposed changes do not influence the running of the indicators or integrated assessment (i.e. CHASE) for HOLAS III (i.e. the assessment tool and indicator (i.e. MIME) evaluations remain as defined in other documentation) and that the expanded confidence assessment will only provide additive information on the confidence component. If endorsed this additional confidence information could be applied in HOLAS III to provide a detailed confidence map at the assessment unit scale and those same values also be utilised by CHASE.

The proposed category setting approach and test approaches evaluated here (see Annex 1 excel sheet) are applied for setting the categories for which each assessment period would be assessed against.

### Proposed approach

The proposed approach and structure builds greater harmonisation with the BEAT and HEAT integrated assessment tools applied under the biodiversity and eutrophication assessments, respectively.

The following confidence components are proposed to be included in a categorical way (e.g. classes):

- Temporal
- Spatial
- Methodological
- Threshold value

A categorical approach could be applied by scoring each of these components as either low, moderate or high confidence. The separate components would then be compiled to provide an overall confidence for the given assessment unit (per indicator).

The table below provides an overview of three possible options for combining the separate confidence components per assessment unit (and per indicator): OOA (lowest level) approach, Averaging, or weighting. The Averaging approach was supported by EN-HZ.

**Table 1.** Theoretical example for calculating confidence from categories.

Component	'Score'	Lowest level approach	Averaging*	Weighting		
				Relevance	Weight	Score
Temporal	Moderate		Moderate (3.4) = 3	Moderate relevance	10%	Moderate (3)
Spatial	Low	Low		High relevance	25%	
Accuracy	Moderate			High relevance	25%	
Methodological	High			High relevance	25%	
Threshold value	High			Moderate relevance	15%	

\*Averaging uses the following scoring: *high* = 5, *moderate* = 3, *low* = 1. A rule may also be needed to round up or round down.

### Temporal confidence

The temporal confidence component would provide an evaluation of how well sampled stations within a given assessment unit are. This information can be derived from the data reported to the HELCOM COMBIEN database hosted by ICES (possibly it could also be derived from the indicator evaluation data under MIME, TBC later). The example below uses the data extracted by ICES from COMBINE (see also data summary in excel document Att.1).

To reflect the differences in sampling matrix the categories could also be adjusted, for example water sampling requires a higher frequency of sampling than sediment or biota monitoring, the latter being more end points for contaminants and less transient or susceptible to seasonal etc changes. This confidence aspect would also be applied at the assessment unit level irrespective of the 'full' (data series of 3 or more consecutive years) or 'initial' (data series of less than 3 consecutive years) data categorisation used for presenting assessment symbols, those being derived based on the statistical evaluation carried out in the MIME assessment tool.

The following parameters with relevance are possible to summarise or calculate from COMBINE at the HELCOM Scale 4 assessment unit level: number of stations within that assessment unit, number of sampling visits within that unit, the start and end years of the time series, the number of year in the time series, the length of the time series (i.e. years between start and end years), the average number of sampling visits per year, and the average spatial area per monitored sampling station in that assessment unit.

In the examples provided in the excel attachment (Att.1) the proposal is to use the 'average number of sampling visits per year' component to derive the temporal confidence component. This value is derived based on the number of visits and the length of the time series for that assessment unit.

Sampling matrix type and coastal or open sea will require different categorical settings, as presented in the excel attachment (Att.1).

An example is shown in the table below.

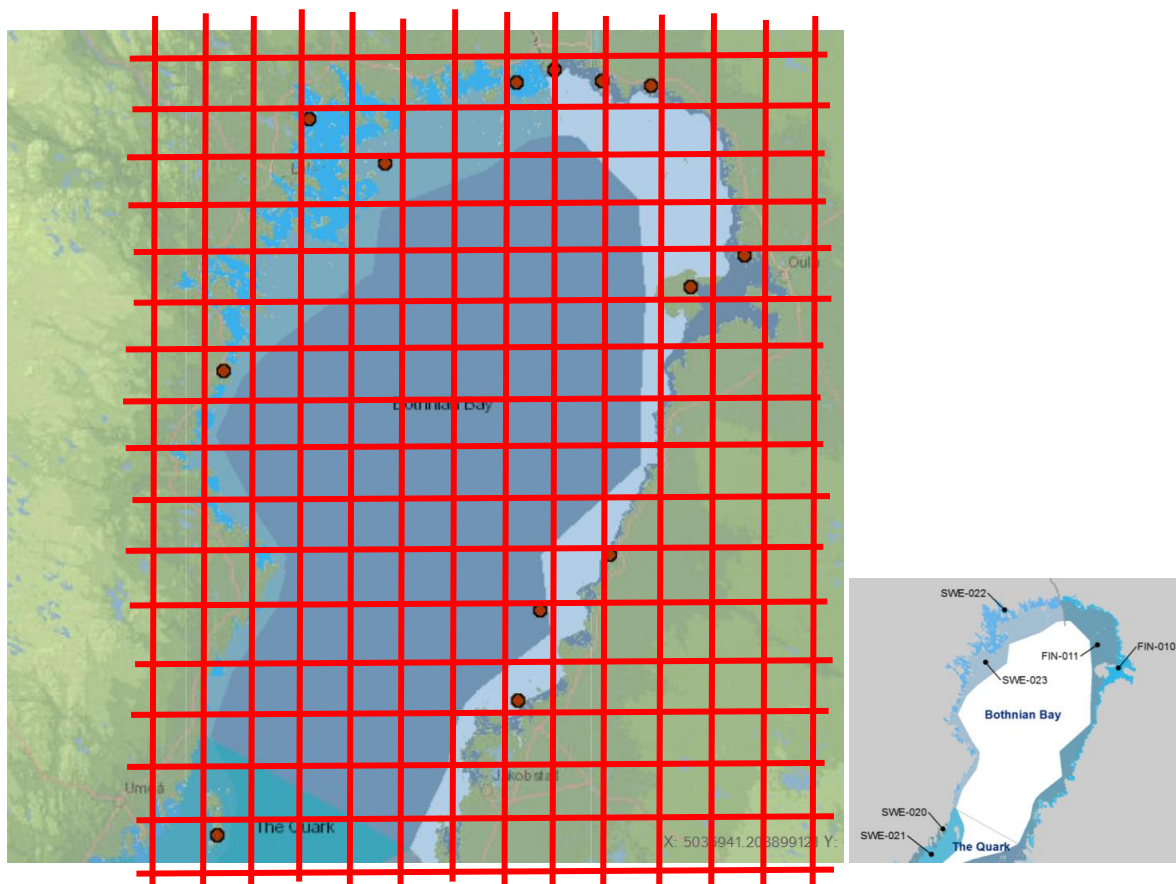
**Table 2.** Example of categorical temporal confidence setting. Note this may differ between substance or sampling matrix in the final version and this is an example of how it would be applied. The example here is for PAHs in biota for coastal waters.

Confidence evaluation	Temporal component categories
High	More than 2 visits/samples per year
Moderate	Over 1 visit per year
Low	1 visit or less per year

### Spatial confidence

Spatial confidence would provide an evaluation of how dispersed or well sampled an assessment unit is (though the spatial spread of that sampling within the assessment unit itself would not be directly addressed and is an issue require significant resources to evaluate, as well as having direct implications for monitoring programmes).

The example provided in the figure below aims to illustrate how such an approach can be applied and converted to a categorical evaluation of confidence. From such information it can be derived/visualised how many stations are sampled for a given substance within each HELCOM Scale 4 assessment unit. In addition the spatial are a of each of those assessment unit scan be collated.



**Figure 1.** The Bothnian Bay containing the assessment points for PCBs, dioxins and furans assessment from HOLAS II. This is an example to show options for spatial confidence setting.

The following parameters with relevance are possible to summarise or calculate from COMBINE at the HELCOM Scale 4 assessment unit level: number of stations within that assessment unit, number of sampling visits within that unit, the start and end years of the time series, the number of year in the time series, the length of the time series (i.e. years between start and end years), the average number of sampling visits per year, and the average spatial area per monitored sampling station in that assessment unit.

In the examples provided in the excel attachment (Att.1) the proposal is to use the 'area per station' component to derive the spatial confidence component. This component is derived by taking the area of the given assessment unit and dividing it by the number of stations sampled within that assessment unit to give an average area of the assessment unit addressed per station. These values are then categorised. Sampling matrix type and coastal or open sea will require different categorical settings, as presented in the excel attachment (Att.1). This reflects for example the more homogeneous nature of open sea assessment areas compared to coastal areas for hazardous substances.

An example is shown in the table below.

**Table 2.** Example of categorical temporal confidence setting. Note this may differ between substance or sampling matrix in the final version and this is an example of how it would be applied. The example here is for PAHs in biota for coastal waters.

Confidence evaluation	Temporal component categories
High	Less than 1000 km <sup>2</sup> per station
Moderate	Between 1000 and 3000km <sup>2</sup>
Low	Over 3000km <sup>2</sup>

### Methodological confidence

The methodological confidence would address the details underlying the assessment itself. This would be an expert evaluation and reliant on the following parts, see table below.

**Table 3.** Proposed categorical approach for assessing methodological confidence.

Method component	Confidence score
Regionally agreed MIME tool (monitoring and assessment guideline, or equivalent, in place)	High
Regionally agreed MIME tool (no monitoring and assessment guideline in place)	Moderate
Internationally agreed methodology (monitoring and assessment guideline, or equivalent, in place)	High
Internationally agreed methodology (no monitoring and assessment guideline in place)	Moderate

Other regionally agreed methodology (monitoring and assessment guideline, or equivalent, in place)	High
Other regionally agreed methodology (no monitoring and assessment guideline in place)	Moderate

NOTE: all methodologies above would need to be regionally approved under HELCOM, this being the minimum for application, thus moderate confidence in the method (i.e. allowing some room for improvement) is anticipated to be the lowest required category.

### Threshold value confidence

This can be derived from the extraction table and indicator reports.

**Table 4.** Evaluation of threshold values for confidence.

Threshold value type	Confidence scoring
EQS value	High
QS derived from an EQS	High
BAC from other region	Low
BAC from HELCOM region	Moderate
Scientific publications	Moderate
Scientific publications where values applied internationally (e.g. in other regional sea commissions for example)	High

Any known scientific issues or processes where threshold values were under review (e.g. PBDE or Diclofenac) could have a penalty applied, lowering their confidence rating by one level.

### Implications for MIME and CHASE

There are no direct implications for the running of either MIME or CHASE for HOLAS III, i.e. they will produce an indicator evaluation and integrated assessment irrespective of these proposed changes. The major changes, if implemented, would be that when presenting an indicator status evaluation (e.g. a map) that provides a status evaluation per indicator and per assessment unit (at scale 4) then there would also be an evaluation of confidence applied and presented at the same assessment scale (i.e. not as a per indicator descriptive overview as was applied in HOLAS II). For CHASE the previous confidence evaluation (e.g. high for indicator x) was applied per assessment unit assessed also but as the same score for all assessment units in any given indicator. This new approach would enable the more detailed confidence (i.e. representative of each assessed assessment unit) to be taken into CHASE, thus it could be expected that the confidence evaluations in CHASE may also become more varied as they rely on the indicator confidence assessments (though the CHASE integration of the status part would not differ due to this).

## Annex 1. Examples and testing of category setting to be applied.



4-1 Att.1 Possible  
confidence setting app