Detect2Protect

Deliverable 3.2

Intra- and interspecific response variability and baseline identification for EBM-based assessment

Author: Elena Gorokhova (Stockholm University)

October 14, 2025

Contents

Foreword

This deliverable was developed within the INSERT project to strengthen the reliability of biological effect-based indicators used for environmental assessment under the Marine Strategy Framework Directive (MSFD). The report focuses on defining robust biomarker baselines and quantifying the natural variability that affects their interpretation. By applying a resampling approach to derive uncertainty-aware Background Assessment Concentrations (BACs), it provides a transparent framework for weighting and comparing biomarker data across species and basins. The outcomes contribute directly to improving the confidence and harmonization of BEI applications in Baltic Sea monitoring.

Summary

Assessing contaminant effects in the Baltic Sea requires reliable baseline levels for biomarkers to distinguish pollution-induced changes from natural biological variation. Yet biomarker responses vary across species, sites, and conditions, which introduces uncertainty in defining consistent Background Assessment Concentrations (BACs). If this variability is not quantified, biological effect-based indicators (BEIs) may become unreliable, either missing early warning signals or overreacting to natural fluctuations.

This deliverable introduces a method to quantify and account for this uncertainty. We identify the main sources of biological and methodological variability and use bootstrap resampling to calculate confidence intervals (CIs) around percentile-based BAC values. The relative CI width describes how precisely a BAC can be estimated, while the coefficient of variation (CV) captures the overall biological variability in the data. Together, these metrics provide a quantitative basis for weighting biomarkers according to their reliability.

Using reference data from the Gulf of Riga (GoR), Bothnian Sea (BS), Western Gotland Basin (WGB), and Bornholm, the approach delivers the first uncertainty-aware BAC estimates for several key biomarkers used in WP3 and the iBEC workflow.

Key Takeaways

- Biomarkers with narrow relative CI width are the most precise and reliable for BAC derivation.
- The coefficient of variation (CV) helps identify biomarkers with naturally high variability but cannot replace CI-based uncertainty analysis.
- Weighting BACs by their statistical reliability allows inclusion of all biomarkers while minimizing bias from uncertain estimates.
- Grouping biomarkers by biological function supports clearer interpretation and more consistent assessment outcomes.

Methods

Definition of BAC values

BAC values were defined for a set of biomarkers (Table 1) as percentile-based thresholds from reference stations free of point sources and with low sediment contamination (Figure 1). Following OSPAR/ICES conventions and iBEC practice:

- **90th percentile** was used for biomarkers where higher values indicate contaminant stress (e.g., oxidative stress, detoxification enzymes).
- **10th percentile** was used where inhibition represents adverse effects (e.g., AChE, RNA/DNA ratio, ORAC/TBARS).
- For some biomarkers, where **both increases and decreases can indicate exposure**, we calculated 10^{th} and 90^{th} percentiles (Table 2; *BAC_Uncertainty_D2P.xlsx*; https://github.com/elenagoro/iBEC-

tool). Use of such thresholds should be agreed upon, and default effect directions should be discussed.

Resampling approach

Using available datasets, bootstrap resampling was applied to quantify uncertainty due to limited sample sizes and sampling variability. For each biomarker:

- **Metadata**: species, basin, sample size, etc.; Tables 1 and 2.
- The selected percentile (10^{th} or 90^{th}) was estimated from the observed data; 1000 bootstrap resamples were drawn, each time recalculating the percentile.
- The distribution of bootstrap estimates was used to derive the mean and the 95% confidence interval (CI).
- **Relative CI width (%)** was calculated as $(CI_{high} - CI_{low})/BAC \times 100$ to provide a unitless measure of uncertainty suitable for comparison across biomarkers.
- **Coefficient of Variation (CV%)**: $(SD_{bootstrap}/mean_{bootstrap}) \times 100$, approximated from CI width when raw bootstrap distributions were not retained.
- Results were visualized with **a traffic-light scheme**: green (<20%) = low uncertainty, yellow (20-40%) = moderate, red (>40%) = high. A summary table can be created showing assigned weights for all biomarker–species–basin combinations. This can be used for weighted BAC calculations or prioritizing monitoring efforts.

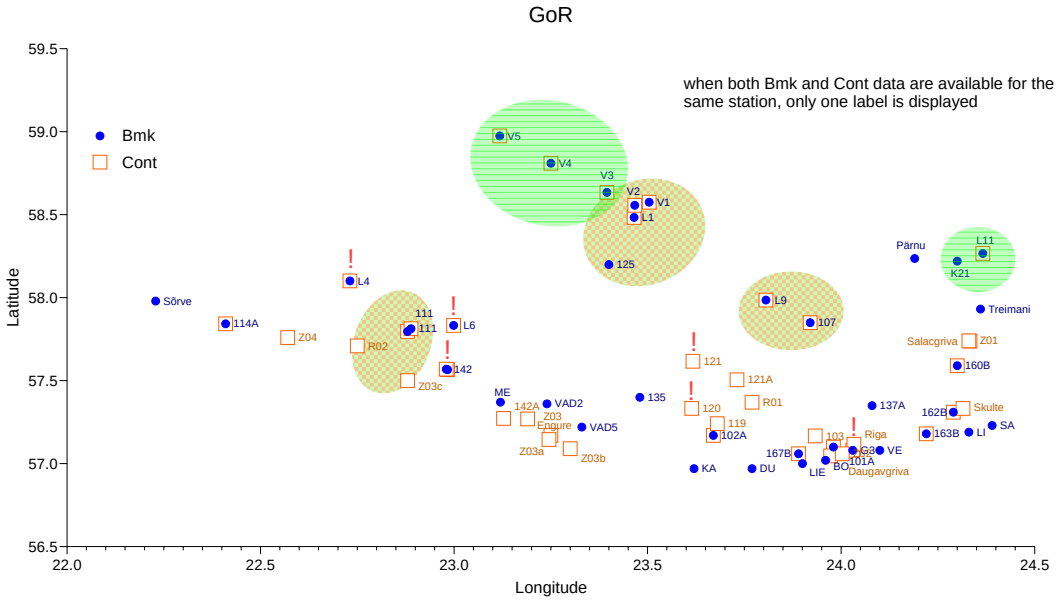

Figure 1. An example of the approach used for setting reference sites (relatively low pollution levels, no point sources) that were defined to calculate background variability of the biomarkers in sentinel species. Geographic distribution of Gulf of Riga monitoring stations where both sediment contaminant concentrations (Cont) and biomarker data (Bmk) were available; generated in BEACON (Interreg) project (2022-2024). Stations shown in colored ellipses indicate reference sites selected for baseline derivation, based on the criteria of (i) contaminant levels not exceeding environmental quality standards (EQS) and (ii) absence of known point sources. Data from these reference stations were used to establish the background variability of the studied biomarkers.

Results and Discussion

Uncertainty in BAC estimates across biomarkers

Bootstrap-derived uncertainty estimates expressed as relative CI width, %, varied markedly among biomarkers and biological functions (Figure 2). Across the full dataset, median uncertainty ranged from below 10% for stable physiological indices to over 100% for some oxidative stress parameters, indicating that the precision of BAC estimates strongly depends on both biomarker type and biological function.

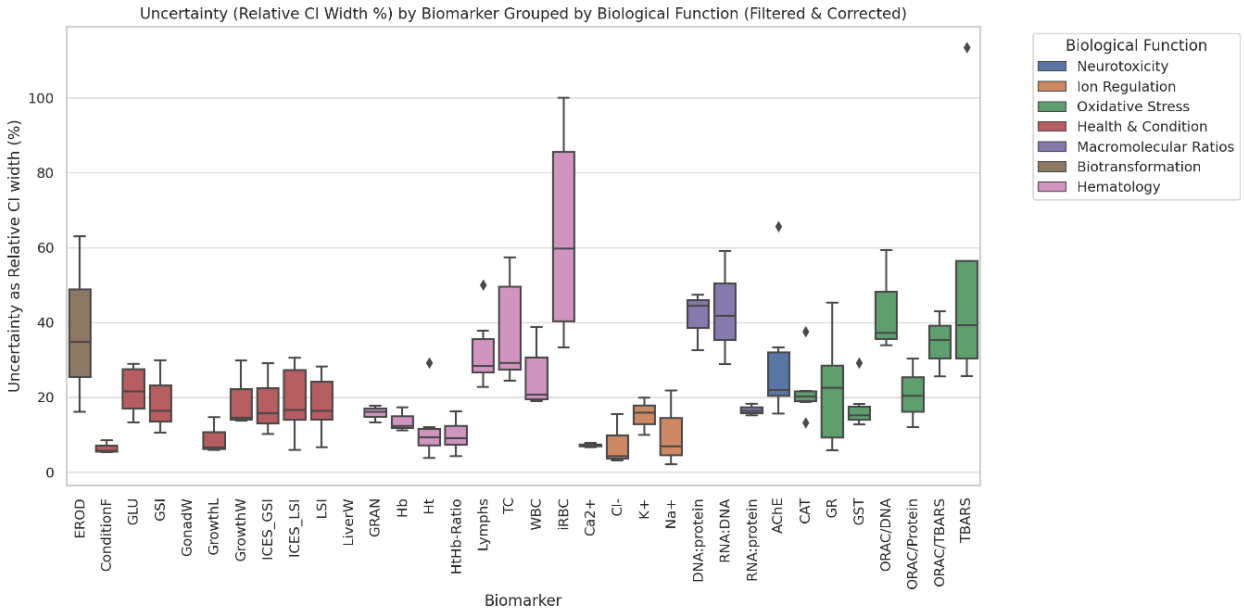

Figure 2. Uncertainty of percentile-based BAC estimates expressed as relative confidence interval (CI) width (%) for individual biomarkers, grouped by biological function. The plot illustrates how the precision of BACs varies across functional categories and taxa. Biomarkers related to physiological condition and ion regulation show the lowest uncertainty, reflecting stable baseline distributions, whereas oxidative stress and hematological markers display wider CIs, indicating greater natural variability and sensitivity to environmental or physiological factors. Data are from Table 2; *BAC_Uncertainty_D2P.xlsx*; https://github.com/elenagoro/iBEC-tool.

For fish (perch) biomarkers, uncertainty levels were generally moderate but variable across functional groups:

- *Biotransformation* (EROD, GST) showed variable uncertainty (16–63%), consistent with moderate biological variability and limited sample availability at reference sites (≤30/site).
- *Health and condition indices* (GSI, LSI, condition factor, growth metrics) exhibited generally low uncertainty (<20%), reflecting stable distributions in reference populations. Glucose BAC had generally low to intermediate uncertainty (<20%).
- *Hematological parameters* (hematocrit, hemoglobin, lymphocytes, granulocytes) displayed heterogeneous uncertainty: while Hb, Ht and their ratio had low CI widths (<<20%), lymphocyte

and total cell counts exceeded 50%, indicating high inter-individual variation and sensitivity to short-term physiological changes.

- *Ion regulation markers* (Na^+, K^+, Cl^-, Ca^{2+}) showed consistently low uncertainty (<15%), suggesting both low between-individual variability and compact tails of the distribution, which is an advantage for integration into BEI assessments. Note that this does not imply that a biomarker itself is effective in predicting exposure.
- *Oxidative stress biomarkers* (GR, GST) have variable uncertainty. GR shows higher uncertainty (35–60%), while GST is more stable (15–30%).

For invertebrate biomarkers, the pattern was comparable:

- *Neurotoxicity* (AChE) displayed low uncertainty (<15%), supporting its robustness as a diagnostic marker.
- *Oxidative stress biomarkers* (CAT, GST, GR, ORAC, TBARS, ORAC/TBARS) had variable uncertainties (Table 3). CAT, GST, and ORAC/Protein show the lowest uncertainty (≈12-22%), making them the most stable oxidative stress indicators across basins. In contrast, TBARS and ORAC/DNA display the highest variability (up to 113%), reflecting large biological and site-related differences that limit their reliability for BAC derivation and application.
- *Macromolecular ratio indicators* (RNA/DNA, RNA/protein, DNA/protein) were assessed only in amphipods. RNA/Protein shows the lowest uncertainty (≈10–18%), making it the most reliable condition biomarker, while RNA/DNA and DNA/Protein display moderate variability (≈22-40%) across basins.

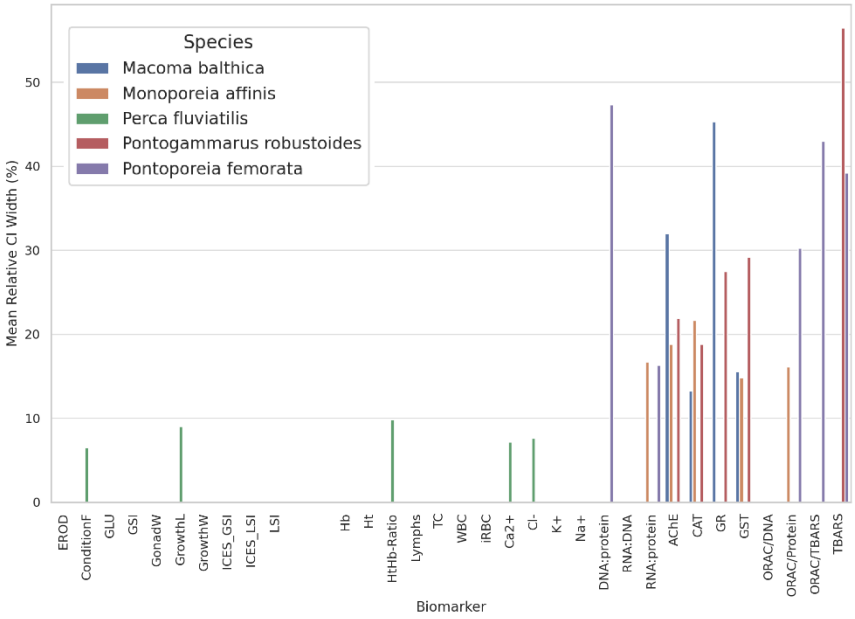

Figure 3. The five most reliable BAC estimates per species, represented by biomarkers with the lowest mean relative confidence interval (CI) width (%). Each bar shows the mean uncertainty of the percentile-based BAC value derived from reference data. Lower CI widths indicate greater precision in the baseline definition, reflecting stable background variability rather than the biological performance of the biomarker itself. BACs for physiological and condition-related endpoints in *Perca fluviatilis* (e.g., GSI, ConditionF, ionic balance) were the most reliable, whereas oxidative stress and macromolecular ratio markers showed wider confidence intervals, particularly in invertebrates.

Overall, biomarkers reflecting stable physiological traits (ion balance, growth, general condition) yielded the most precise BACs, while stress-responsive or molecular-level markers (oxidative stress, hematology, macromolecular ratios) were characterized by wider confidence intervals (Figure 3). Across invertebrate species, CAT, GST, ORAC/Protein, AChE, and the RNA/Protein ratio show the lowest uncertainty (≈10–25%), making them the most reliable for BAC derivation. In contrast, GR, TBARS, and ORAC/DNA exhibit high variability (up to 60–114%), reducing the precision of BAC estimates. These differences suggest the importance of function-specific interpretation when establishing BEI baselines.

Uncertainty in BAC estimates across species

The analysis identified the most statistically reliable BAC estimates, those with the lowest mean uncertainty, across species (Figure 3). In *Perca fluviatilis*, BACs for condition, ionic and energy-balance indices (e.g., GSI, condition factor, Na^+, Ca^{2+}) showed narrow confidence intervals (<10%), indicating that these baseline values are well constrained by the available reference data despite a relatively low sample size ($n \leq 30$; Table 2; *BAC_Uncertainty_D2P.xlsx*). In contrast, BACs for oxidative stress biomarkers displayed greater uncertainty (>25%), reflecting both natural variability and relatively low sample sizes.

For *Macoma balthica*, the biomarkers with the lowest uncertainty are CAT, GST, and GR. These enzymatic indicators provide consistent baseline values across basins and are well suited for deriving precise BACs. Including AChE as a complementary neurotoxicity marker can further strengthen assessments due to its stable response in this species.

For amphipods (*Monoporeia affinis, Pontoporeia femorata,* and *Pontogammarus robustoides*), the most reliable biomarkers are GST, CAT, ORAC/Protein, and RNA/Protein, each with uncertainty typically below ~20-25%. These indicators combine metabolic and antioxidant information with low variability, making them strong candidates for standardized BAC setting and inter-basin comparison.

How variability and confidence interact in BEI assessment

In most cases, the bootstrap mean BAC closely matched the empirical percentile estimate, confirming that BAC values derived from reference data are statistically stable (Figure 4). However, the CI width indicates that markers with higher biological or environmental variability (e.g., oxidative stress indicators) require either larger reference datasets or temporal replication to improve precision.

Understanding both CV% (overall variability) and relative CI width (uncertainty of the baseline) helps determine how confidently a biomarker can be used to detect pollution effects (Figure 5):

- **High CV% and wide CI** mean the biomarker naturally varies a lot, making it difficult to distinguish pollution-related changes from normal biological fluctuations. Such biomarkers should be used with caution or supported by additional indicators.
- **Low CV% and narrow CI** indicate a stable biomarker with well-defined baseline conditions—ideal for early warning and trend detection.
- **A high CV but narrow CI** suggests that although the data vary, the percentile-based BAC (baseline threshold) remains statistically robust.

- **A low CV but wide CI** points to uncertainty caused by small sample sizes or uneven data distribution, meaning more reference data are needed.
- **Mismatches between CV and CI** reveal potential data or sampling limitations:

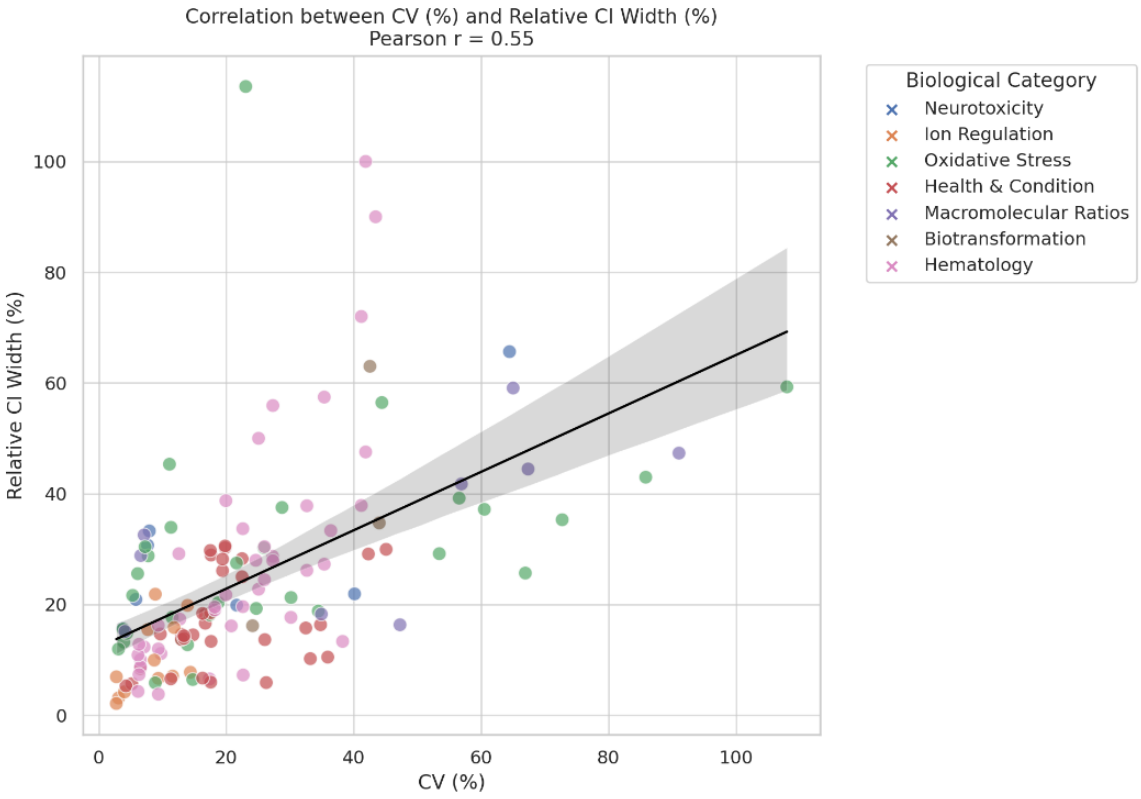

Figure 4. Relationship between coefficient of variation (CV%) and relative confidence interval (CI) width (%) across biomarkers and biological categories. The positive correlation (Pearson r = 0.55) indicates that biomarkers with higher intrinsic variability tend to yield less precise percentile-based BAC estimates. However, deviations from this trend (cases where relative CI width is large despite a moderate CV, or vice versa) highlight that sampling design and data distribution (e.g. skewness, sample size, presence of outliers) also influence BAC uncertainty. A high CV with a narrow CI typically reflects consistent percentile estimates despite large overall spread, whereas a low CV but wide CI suggests insufficient data density near the target percentile. Together, these two metrics provide complementary insight: CV reflects total biological variability, while relative CI width quantifies the uncertainty in defining a robust reference threshold.

In practice, **combining both metrics** allows identifying which BEIs are reliable for routine assessment, which need further data support, and which are too variable to provide clear signals of contaminant stress. For example, when evaluating a biomarker dataset, the joint examination of CV and relative CI width can help decide whether additional data are needed or whether the existing baseline is sufficiently precise. A biomarker showing both low CV and narrow CI can be confidently used for assessment and trend detection, while one with a high CV but narrow CI suggests that variability is well captured statistically and may not compromise baseline reliability. Conversely, a low CV but wide CI would indicate insufficient reference data density near the percentile used for BAC calculation, calling for expanded sampling rather than

discarding the marker (Figure 5). This combined interpretation makes it possible to distinguish true biological variability from statistical uncertainty, improving the robustness of BEI datasets before integration into broader assessments.

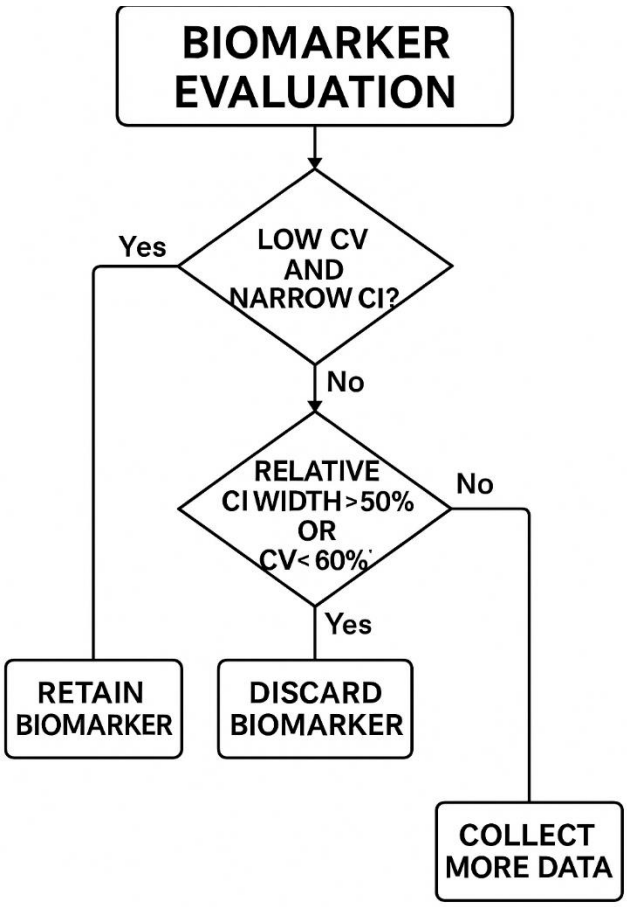

Figure 5. Decision framework for evaluating biomarker suitability based on biological and statistical variability. The flowchart illustrates how the coefficient of variation (CV) and the relative confidence interval (CI) width jointly guide data interpretation. Biomarkers with both low CV and narrow CI are considered reliable for baseline definition and trend detection. Those with high CV but narrow CI can be retained with caution, as their variability is well represented by the dataset. When the CI is wide but CV is low, additional data are needed to improve statistical precision. Biomarkers showing both high CV and wide CI are too uncertain for operational use and should be discarded or treated as exploratory indicators.

Conclusions

This work establishes a transparent method to quantify and integrate the uncertainty of BACs into the evaluation of biological effects. Bootstrap-derived confidence intervals provide a direct measure of baseline precision, allowing each BAC to be classified by reliability. Translating relative CI width into color-coded uncertainty classes (🟢 🟡 🔴) enables consistent communication and application in assessment.

Within the iBEC-tool and weight of evidence (WoE) framework, this weighting can be operationalized by:

- Assigning quantitative weights to biomarkers according to their uncertainty (e.g., 1.0 for 🟢, 0.7 for 🟡, 0.4 for 🔴; Table 4).
- Using these weights to adjust the contribution of individual biomarkers when aggregating BEI results across sites or species.
- Prioritizing high-confidence BACs for quantitative evaluation while retaining less certain ones as contextual or supporting evidence.

From a management standpoint, this information helps evaluate the confidence level of BAC thresholds used in BEI assessment. A precise BAC (narrow CI) ensures that deviations observed in monitoring data can be interpreted with higher confidence as genuine biological responses to contaminant exposure, whereas imprecise BACs (wide CI) increase the risk of false positives or negatives.

To improve BAC precision across species: (*i*) prioritize enzymatic and ratio-based biomarkers with low uncertainty; (*ii*) normalize variable biomarkers by protein content or physiological indices; (*iii*) standardize reference site selection to better capture background variability; and (*iv*) increase replicate numbers for highly variable biomarkers. These measures will enhance the reliability and cross-species comparability of biomarker-based thresholds in invertebrate assessments.

This approach strengthens the interpretative power of BEIs by linking statistical confidence with biological meaning. It supports a tiered, evidence-based assessment, where robust BACs form the core of the quantitative evaluation and variable biomarkers inform diagnostic interpretation and monitoring design. The methodology aligns with the MSFD Descriptor 8 objective of ensuring transparent, reproducible, and uncertainty-aware integration of biological effects into marine environmental status assessment.

Table 1. Overview of the biomarkers included in the BAC uncertainty analysis, grouped by biological function*, species, and geographic basin.

Biological function	Species	Basins	Biomarkers
Biotransformation / detoxification	*Perca fluviatilis*	BS, Bornholm, WGB	EROD, GST
Cellular damage / macromolecular ratios	*Monoporeia affinis, Pontoporeia femorata*	BS, WGB, Bornholm	DNA/protein, RNA/protein, RNA/DNA
Fish health & condition	*Perca fluviatilis*	BS, Bornholm, WGB	ConditionF, GSI, LSI, GrowthL, GrowthW, GLU
Hematology / immune response	*Perca fluviatilis*	BS, Bornholm, WGB	TC, GRAN, Hb, Ht, HtHb-Ratio, Lymphs, WBC, iRBC
Ion regulation / osmoregulation	*Perca fluviatilis*	BS, Bornholm, WGB	Ca2+, Cl-, K+, Na+
Neurotoxicity	*Macoma balthica, Monoporeia affinis, Pontogammarus robustoides, Pontoporeia femorata*	BS, GoR, WGB	AChE
Oxidative stress / antioxidant defense	*Macoma balthica, Monoporeia affinis, Perca fluviatilis, Pontogammarus robustoides*	BS, Bornholm, GoR, WGB	GR, CAT, GST, TBARS, ORAC/TBARS, ORAC/DNA, ORAC/Protein

*Functional interpretation. **Condition and energy allocation**: Condition factor (ConditionF) - composite index of body condition (Fulton's K); decreases with malnutrition, chronic exposure, or metabolic stress. Gonadosomatic index (GSI) – Ratio of gonad to body weight; reduced GSI indicates reproductive suppression or delayed gametogenesis; Liver somatic index (LSI) – Ratio of liver to body weight; may decrease under metabolic depletion or increase due to detoxification enzyme induction; Growth (length, weight; GrowthL/W) – Slower growth reflects chronic toxicity or nutrient limitation; **Hematology and immune status:** Hematocrit (Ht) – Fraction of blood volume occupied by red cells; low Ht signals anemia or hemodilution; Hemoglobin (Hb) – Oxygen transport capacity; decreases with anemia or red cell damage; Ht/Hb ratio – Reflects erythrocyte quality and oxygen-carrying efficiency; altered in anemia or hemoconcentration; Lymphocytes (Lymphs) – Decline during chronic exposure or immune suppression; may transiently increase in early stress; Granulocytes (GRAN) – Increase under acute stress, infection, or inflammation; White blood cell count (WBC) and Total cell count (TC) – Indicate overall immune activity; low values imply immune depression. Immature red blood cells (iRBC) – Rise during recovery from anemia or drop when hematopoiesis is inhibited. **Ion regulation and osmoregulation**: Sodium (Na^+) – Decreases with gill or renal dysfunction; Potassium (K^+) – Increases with membrane damage or cell leakage; Chloride (Cl^-) – Typically decreases under gill impairment; Calcium (Ca^{2+}) – Declines with impaired gill ion exchange and reduced osmoregulatory efficiency. **Biotransformation and detoxification:** EROD (ethoxyresorufin-O-deethylase) – Phase I enzyme activity linked to CYP1A induction; increases upon exposure to planar organic contaminants such as PAHs and PCBs; **Oxidative stress and antioxidant defense**: Catalase (CAT) – Converts hydrogen peroxide to water; elevated under oxidative stress; Glutathione S-transferase (GST) – Conjugates lectrophilic compounds; induced by a wide range of xenobiotics. Glutathione reductase (GR) – Maintains cellular redox balance; increases in oxidative challenge; Metabolic and stress markers: Glucose (GLU) – Short-term stress indicator; rises in acute stress, may decline in chronic or severe exposure when energy reserves are depleted.

Table 2. Summary of biomarker-specific Background Assessment Concentrations (BACs) and their associated uncertainty metrics (bootstrap 95% CI, relative CI width, and coefficient of variation) calculated from reference stations across basins. The table provides a quantitative basis for evaluating biomarker reliability and variability in deriving BAC thresholds for MSFD Descriptor 8 assessments.

See *BAC_Uncertainty_D2P.xlsx*; https://github.com/elenagoro/iBEC-tool.

Table 3. Uncertainty ranges (Relative CI width, %) for oxidative stress biomarkers in Baltic Sea invertebrates, showing variability across species and basins. Enzymatic antioxidants (CAT, GST) and protein-normalized ORAC indices exhibit low uncertainty, whereas lipid peroxidation markers (TBARS) show the widest and least consistent ranges.

Biomarker	Species observed	Range (% Rel. CI width)	Biomarker description	Notes (concise)
CAT	*Monoporeia affinis*, *Macoma balthica*, *Pontogammarus robustoides*	13–22	Antioxidant enzyme (catalase)	Low uncertainty; stable across basins.
GR	*Monoporeia affinis*, *Macoma balthica*, *Pontogammarus robustoides*	28–45	Antioxidant enzyme (glutathione reductase)	Moderate uncertainty; higher in *Macoma*.
GST	*Monoporeia affinis*, *Macoma balthica*, *Pontogammarus robustoides*	15–29	Detoxification enzyme (glutathione S-transferase)	Low uncertainty; consistent across basins.
ORAC/DNA	*Monoporeia affinis*, *Pontoporeia femorata*	34–59	Antioxidant capacity normalized to DNA	Moderate–high uncertainty; variable across species.
ORAC/Protein	*Monoporeia affinis*, *Pontoporeia femorata*	12–30	Antioxidant capacity normalized to protein	Generally low uncertainty; most stable ratio.
ORAC/TBARS	*Monoporeia affinis*, *Pontoporeia femorata*	26–43	Oxidative balance ratio (antioxidant vs. lipid peroxidation)	Mid-range variability across basins.
TBARS	*Monoporeia affinis*, *Pontoporeia femorata*, *Pontogammarus robustoides*	26–113	Lipid peroxidation marker	Highest uncertainty; large spread within *Monoporeia*.

Table 4. Uncertainty-based weighting of BAC values used in BEI assessment*

Color-coded weighting scheme showing how uncertainty in percentile-based BAC estimates is translated into assessment weights. Each uncertainty class (low, moderate, high) includes representative biomarkers with similar relative CI widths. The color bullets indicate the visual codes used in data tables and BEI dashboards. See Table 2 for the full set of biomarkers and the color-coded weighting.

Uncertainty class	Relative CI width (%)	Color code	Weight (W)	Example biomarkers (species)	Interpretation
Low	< 25	🟢	1.0	AChE (*Macoma balthica, Perca fluviatilis*); GST (*M. balthica, P. fluviatilis*); CAT (*Monoporeia affinis*); Na^+, Ca^{2+}, GSI, ConditionF (*P. fluviatilis*)	Well-constrained BACs with narrow confidence intervals; reference baselines considered robust for integration in BEI assessments.
Moderate	25–45	🟡	0.7	EROD (*P. fluviatilis*); RNA/DNA (*Pontoporeia femorata*); GR (*M. balthica*)	BAC estimates show moderate precision; acceptable for assessment when supported by replication or complementary indicators.
High	> 45	🔴	0.4	TBARS (*Monoporeia affinis*); ORAC/DNA (*Pontoporeia femorata*)	Highly uncertain BACs; wide confidence intervals indicating strong natural or methodological variability. Use only for exploratory analyses.

*Assessment relevance:

– The classification reflects uncertainty in BAC estimates, not biomarker responsiveness to exposure.

– Lower uncertainty (low CI width) indicates more statistically reliable reference values, whereas higher uncertainty highlights limited precision of the BAC rather than variability in biological response.

– The adjusted weights (1.0, 0.7, 0.4) provide a balanced scaling for uncertainty integration in composite indices.