Journal of Environmental Management 344 (2023) 118625

Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: www.elsevier.com/locate/jenvman

Research article

Mixing apples and oranges: Assessing ecological status and its confidence from multiple and diverse indicators

Jacob Carstensen [a,*], Ciaran Murray [b], Mats Lindegarth [c,d]

[a] Aarhus University, Department of Ecoscience, Frederiksborgvej 399, DK-4000, Roskilde, Denmark
[b] NIVA-Denmark, Njalsgade 76, 2300 København S, Denmark
[c] Gothenburg University, Department of Marine Sciences, Tjärnö Marine Laboratory, S-45296, Strömstad, Sweden
[d] Swedish Institute for the Marine Environment, Box 260, S-40530, Göteborg, Sweden

ARTICLE INFO

Handling editor: Jason Michael Evans

Keywords:
Cumulative impacts
Integrated assessment
Monitoring data
Online tool
Pressure diagnosis
Uncertainty assessment
Water framework directive

ABSTRACT

Ecosystem responses to increasing human pressures are complex and diverse, affecting organisms across all trophic levels. This has prompted the development of methods that integrate information across many indicators for environmental management. Legislative frameworks such as the European Water Framework Directive (WFD), specifically prescribe that integrated assessme nt (IA) of ecological status must consider indicators representing various biological and supporting quality elements. We present a general approach for an IA system based on a piece-wise linear transformation of indicator distributions to a standardized scale, allowing for integrating information from multiple and diverse indicators through a policy-dependent aggregation scheme. Uncertainties associated with monitoring data used for calculating indicators and their propagation throughout the integration scheme allow for confidence assessment at all levels of the hierarchical integration. Specific pressures leading to ecological impact can be identified through the most impaired indicators in the hierarchical and transparent aggregation scheme. The IA and its confidence are facilitated though the development of an online tool that accesses information from monitoring databases and presents the outcome at all levels of the assessment, ensuring consistency and transparency in the calculations for all potential stakeholders. We demonstrate the versality and applicability of the approach using indicators and aggregation principles from the Swedish national guidelines for assessing ecological status of rivers, lakes and coastal waters according to the WFD. Although the approach and the tool were developed specifically for the WFD ecological status assessment in Sweden, the generality of the approach implies that it can easily be adapted to the WFD assessment methods of other countries as well as other policies, where an integrated assessment is required.

1. Novelty and relevance statement

Methods for integrated assessment of ecological status are becoming increasingly important for environmental management with growing pressures on ecosystems. We present a novel approach and an online tool for integrating information from multiple and diverse indicators to provide a holistic assessment of ecological status for water bodies following the guidelines of the European Water Framework Directive (WFD). The hierarchical integration of indicators through their statistical distributions provides transparency in the assessment and allows confidence assessment at all levels of integration. Moreover, it allows identification of indicator(s) compromising good ecological status and the likely pressure(s) responsible. The approach and tool, developed according to the WFD guidelines in Sweden, can easily be adapted to WFD assessment methods in other countries as well as other policies.

2. Introduction

Climate change, increasing human activities and exploitation of resources have led to growing pressures and impacts on aquatic ecosystems, prompting management responses to ensure the integrity of these ecosystems and the services they provide (Halpern et al., 2012; Keeler et al., 2012). The cumulative impact of multiple stressors affects organisms at all trophic levels of the ecosystem, either directly by targeting certain species (e.g. fishing, anti-fouling) or indirectly by changing the environmental state (e.g. reduced light conditions, acidification,

* Corresponding author.
E-mail address: jac@ecos.au.dk (J. Carstensen).

https://doi.org/10.1016/j.jenvman.2023.118625
Received 9 May 2023; Received in revised form 19 June 2023; Accepted 11 July 2023
Available online 17 July 2023

hypoxia) and thereby the habitats harboring these organisms. Moreover, the organization of the food web and species' interactions are affected by the uneven distribution of impacts on organisms, essentially changing the functioning of aquatic ecosystems.

Legislative frameworks have been adopted in many countries and regions to safeguard aquatic ecosystems by regulating human activities in a holistic ecosystem-based approach (Kirkfeldt, 2019). These frameworks are inherently adaptive, operating in cycles involving decision-making, implementing, monitoring, evaluating and adjusting (Allen et al., 2011). The complexity of aquatic ecosystems has driven the development of integrative assessments that combine a broad array of indicators of state changes and ecological impacts (Borja et al., 2008). These integrative assessments are typically based on an appropriate scaling and weighting of several indicators representing key features of the aquatic ecosystem. For example, the HELCOM Eutrophication Assessment Tool (HEAT) combined indicators of nutrients, chlorophyll *a*, Secchi depth and benthic invertebrates by scaling these against their reference conditions and averaging these scaled indicator values (Andersen et al., 2011). Similarly, observed concentrations of hazardous substances were scaled against critical thresholds and averaged in the Chemical Status Assessment Tool (CHASE) for the Baltic Sea and North Sea (Andersen et al., 2016). However, there are many approaches to integrate indicators, but essential features are that any integration principle should be ecologically relevant, transparent and documented (Borja et al., 2014).

The European Water Framework Directive (WFD) employs an integrated assessment (IA) of Biological Quality Elements (BQEs) using the one-out-all-out (OOAO) principle, i.e. the ecological status of a water body (WFD term for aquatic assessment unit) depends on the BQE with the lowest status (CIS, 2005). BQEs include phytoplankton, phytobenthos, macrophytes, macroinvertebrates and fish, and BQE indicators should capture relevant information on abundance, biomass, community composition and diversity. The EU member states have generally developed their own indicators for each BQE (Birk et al., 2012), either as one composite indicator (aggregation contained in the indicator formulation) or by aggregating several indicators with an appropriate, yet undefined, principle (CIS, 2005). The ecological status of a BQE is quantified on an Ecological Quality Ratio (EQR) scale from 0 to 1 that is divided into five classes representing bad, poor, moderate, good, and high status. Furthermore, if the BQEs suggest high or good status and the status of supporting quality elements (SQE) (hydromorphological and physico-chemical quality elements) is lower, the ecological status can be downgraded by one class (CIS, 2005). However, the principles used for aggregating indicators and for combining ecological and supporting elements have not been laid out in the guidelines since the WFD is a framework directive and consequently, countries can employ quite diverse principles in their implementation. The choice of aggregation method can considerably affect the overall assessment outcome (Langhals et al., 2014), challenging the principles of comparability and transparency of IA systems across Europe (Hering et al., 2010). These challenges are not specific to the WFD but also inherent to the implementation of other environmental directives such as the Marine Strategy Framework Directive (MSFD) (Probst and Lynam, 2016). Thus, there is a strong need to harmonize aggregation principles, acknowledging that it is not possible to harmonize indicators due to differences in monitoring data and aquatic ecosystems across Europe.

The WFD requires that all water bodies should achieve at least good ecological status, by 2027 at the latest, and river-basin management plans should be implemented for reducing relevant pressures if this is not the case. Any IA is inherently associated with uncertainty and it is imperative to quantify the confidence of the status classification as the program of measures can indeed be costly. Particularly, the use of the OOAO-principle results in higher uncertainty of the IA compared to other combination rules (Moe et al., 2015). Misclassifications can either lead to unnecessary implementation costs or compromising ecosystem services. The uncertainty of an assessment can be reduced by more and better monitoring data as well as improving the indicator calculation routines to account for variability ascribed to other factors than the human pressures such as seasonality, winds, etc. (Carstensen, 2007). Despite the obvious benefits of quantifying the confidence of an integrated assessment, most IA studies do not consider this aspect at all (van Beest et al., 2021). Andersen et al. (2010) developed a 'confidence rating' system based on heuristic aggregation principles, albeit not rooted in a theoretical probabilistic framework. Confidence assessment remains a challenge, requiring the estimation of variance components for different sources of uncertainty, which is often only possible through analyses of large datasets (Carstensen and Lindegarth, 2016), or large surveys dedicated specifically to the purpose (Dromph et al., 2013; Dudley et al., 2013). Nevertheless, it is important that confidence assessment becomes an integral part of IA systems (van Beest et al., 2021).

In a Swedish context, it is absolutely necessary to develop a systematic and coherent approach to status assessment, with supporting tools to integrate indicators and estimate confidence in classification etc. This is particularly important considering the task of assessing approximately 650 coastal waterbodies, more than 7500 lakes and 15,500 streams included in WFD regulations. Therefore, one of the main aims of the Swedish research program WATERS (2010–2016), was to develop a common framework for classification and uncertainty assessment. Thus, the objective of this paper is to develop a general and transparent framework for integrating indicators for assessing ecological status of water bodies according to the WFD and to assess the confidence of the assessment. The aim is also to demonstrate the applicability of this framework through developing an internet-based open-source tool that enables stakeholders to carry out WFD integrated assessments. Finally, we propose that the framework can easily be adapted to other legislative frameworks that use different aggregation rules.

3. From monitoring data to integrated assessment

Integrated assessment of ecological status is primarily a measurement-driven process, although it can also be supported by model predictions (Schuwirth et al., 2019). EU member states are required to establish monitoring programs for the WFD ecological assessment, involving both biological and physical-chemical measurements. Whereas the physical-chemical measurements are generally carried out according to international standards, biological measurements are typically done according to national standards that are often not coordinated with other countries. Even within countries, regional differences in biological monitoring methods exist. Consequently, methods for aggregating the raw measurement are quite diverse and mostly specific to each country.

The chain of data processing from monitoring data to integrated assessment is long and involves many steps (Fig. 1). Biological monitoring data typically consist of measurements of certain taxonomical resolution, e.g. species-specific abundances or biomass. These raw observations are subsequently aggregated into an index that expresses specific features of the biological community, e.g. the phytoplankton functional trait index (FTI) for lakes (Solheim et al., 2013) or the benthic quality index (BQI) for marine macrofauna (Rosenberg et al., 2004). Similarly, physical-chemical measurements at discrete depths can be aggregated to represent the entire or parts of the water column (e.g. surface chlorophyll *a*), or extrapolated to estimate the spatial extent of hypoxia. In some cases, the index and observation are the same, e.g. Secchi depth. An index represents discrete information in time and space. However, an assessment should typically represent an entire period and spatial entity (e.g. water body for the WFD) and therefore, indices are aggregated to form indicators that aim at describing certain characteristics covering the entire ecosystem and assessment period (e.g. mean summer chlorophyll *a* in a water body over a 6-year period). When several indicators are employed to characterize a BQE or SQE, they need to be aggregated by means of a decision rule, which in most

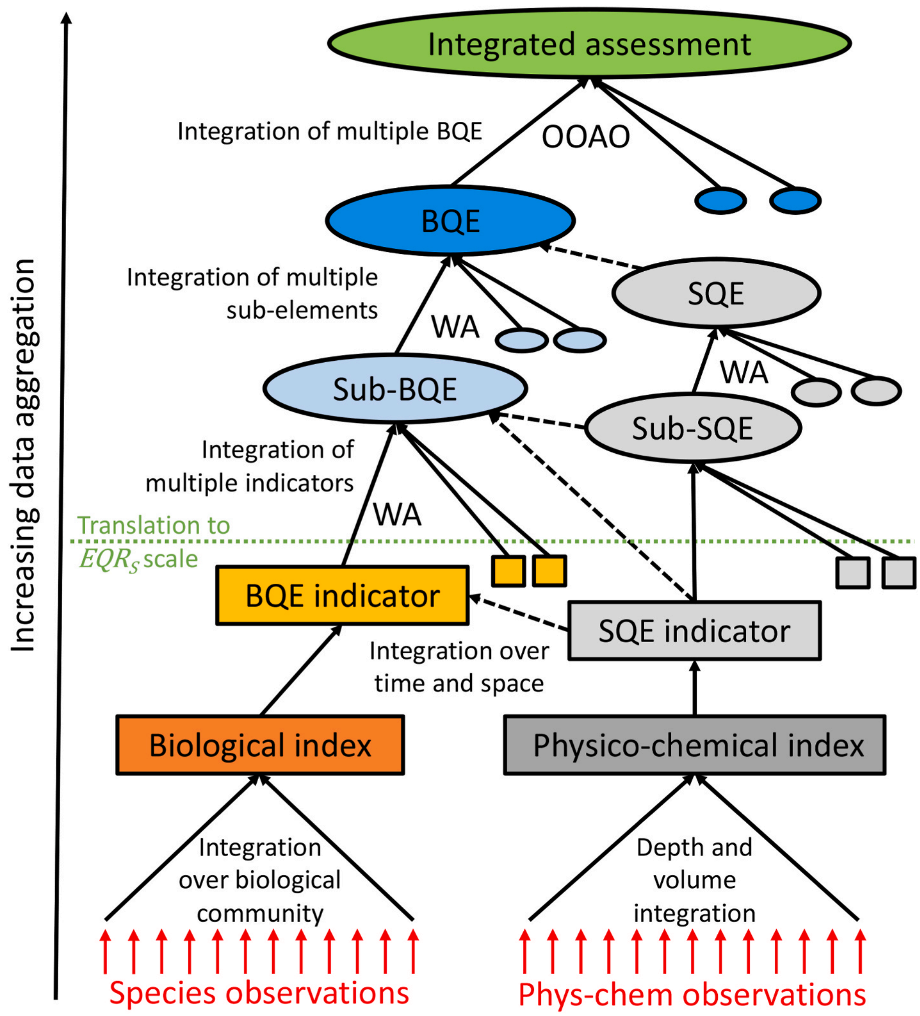

Fig. 1. The data flow from observations to integrated assessment through different levels of aggregation. The green dotted line marks the transition from indicators to EQR_S-scale. Method of aggregation: WA=Weighted Average, OOAO=One-out-all-out.

cases is averaging (with or without weighting indicators). In some cases, the aggregation is carried out in two steps through intermediate sub-elements (e.g. macroalgae and angiosperms for the BQE benthic vegetation and nitrogen and phosphorus for the SQE nutrient concentrations). This means that indicators are first aggregated to sub-elements that are subsequently aggregated to BQE or SQE. At this aggregation level, the BQE status can be downgraded if there is a discrepancy with the SQE status (CIS, 2005). However, it is also possible for SQE sub-elements or SQE indicators to be used for downgrading BQE sub-elements and BQE indicators, respectively. Finally, the BQEs are aggregated to the IA following the OOAO-principle. The actual integrated assessment method is decided by member states, including diverse decision rules as will be shown for the Swedish IA method below.

3.1. *Indicator standardization*

Ecological and biodiversity indicators are highly diverse in their computation and use different numerical scales (Teixeira et al., 2016), which means that they cannot be aggregated without standardization through some transformation. Moreover, the same indicator value can express quite different status depending on which water body and period it represents. For example, a chlorophyll *a* mean of 5 µg L^{-1} may be considered excessively high for winter conditions but relatively low for summer conditions in a river-dominated temperate estuary, and at the same time excessively high at any time of the year for an oligotrophic coastal system. This is acknowledged in the WFD and other legislative frameworks by relating the actual indicator value (*Status*) to a reference point. The WFD defines the reference condition (*RefCond*) as characterizing an ecosystem state with minimal disturbance from human activities and uses relative deviations (Ecological Quality Ratio (*EQR*)) from the reference condition to characterize the degree of disturbance, i.e.

$$EQR = \begin{cases} \dfrac{RefCond}{Status} & \textit{for indicators responding positively to pressures} \\ \quad or & \\ \dfrac{Status}{RefCond} & \textit{for indicators responding negatively to pressures} \end{cases} \tag{1}$$

Assuming the indicator only takes positive values, this EQR-transformation will map indicators onto the interval between 0 and 1, with 1 representing the *RefCond*. However, this transformation will not yield a consistent representation of the entire EQR-scale for many indicators. For instance, a halving of the mean Secchi depth, even for the most polluted water bodies, would be considered an extreme deterioration of water quality and hence, the EQR-transformation (Eq. (1)) would, in practice, utilize only the range from 0.5 to 1. The simplicity of the EQR-transformation imply that the statuses of indicators are projected onto different intervals. Furthermore, the response of the indicator to pressures may not be linear or inverse (Eq. (1)), resulting in class boundaries that are unevenly distributed over the EQR-scale. Thus, although the EQR-transformation, in principle, restricts the outcome space within the interval between 0 and 1, the transformation does not produce compatible EQR-metrices. EQR-values for different indicators are not comparable and even EQR-values for the same indicator are not comparable, if their *RefCond* values are different. In other words, the EQR-transformation (Eq. (1)) produces "apples and oranges" that cannot be directly compared with each other. Therefore, class boundaries are specific to the given indicator and the group of water bodies sharing the same *RefCond* value and pressure-response relationship (within same type according to the WFD). This specificity of class boundaries creates a lack of transparency and problems with communicating the results of the IA in a consistent manner. For example, an indicator EQR-value of 0.75 can represent different status classes among water bodies.

To accommodate these deficiencies with the simple EQR-transformation (Eq. (1)), we propose a piecewise linear transformation that maps the indicator value onto a standardized EQR-scale (EQR_S), where each of the five status classes are represented by an equal interval of 0.2 (Fig. 2). Hence, the boundaries Poor/Bad (B_{PB}), Moderate/Poor (B_{MP}), Good/Moderate (B_{GM}) and High/Good (B_{HG}) for the estimated indicator ($\widehat{I}$) always correspond to 0.2, 0.4, 0.6 and 0.8 on the EQR_S-scale. Additionally, the estimated indicator is assumed bounded by the reference condition at one end (*RefCond*) and an extremely poor value (B_{min}) at the other end of the scale. Essentially, the piecewise linear transformation is a discrete representation of a reversed pressure-response relationship with the EQR_S-scale representing different levels of pressure. For an indicator responding negatively to increasing pressure (Fig. 2A) the EQR_S is calculated as

$$EQR_S = \begin{cases} 0 & if\ \widehat{I} \leq B_{min} \\ \dfrac{\widehat{I} - B_{min}}{B_{PB} - B_{min}} \bullet 0.2 & if\ B_{min} < \widehat{I} \leq B_{PB} \\ \dfrac{\widehat{I} - B_{PB}}{B_{MP} - B_{PB}} \bullet 0.2 + 0.2 & if\ B_{PB} < \widehat{I} \leq B_{MP} \\ \dfrac{\widehat{I} - B_{MP}}{B_{GM} - B_{MP}} \bullet 0.2 + 0.4 & if\ B_{MP} < \widehat{I} \leq B_{GM} \\ \dfrac{\widehat{I} - B_{GM}}{B_{HG} - B_{GM}} \bullet 0.2 + 0.6 & if\ B_{GM} < \widehat{I} \leq B_{HG} \\ \dfrac{\widehat{I} - B_{HG}}{RC - B_{HG}} \bullet 0.2 + 0.8 & if\ B_{HG} < \widehat{I} \leq RefCond \\ 1 & if\ \widehat{I} > RefCond \end{cases} \tag{2}$$

For an indicator responding positively to increasing pressure (Fig. 2B), a similar transformation can be found by reversing the indicator scale. Although both *EQR* (Eq. (1)) and EQR_S (Eq. (2)) require boundaries to be supplied for the status assessment, we submit that specifying boundaries at the indicator scale is more informative and transparent than specifying boundaries at the EQR-scale. Importantly, however, the advantage of using EQR_S is the translation to a common scale, enabling indicator comparison and integration (see below).

3.2. Indicator uncertainty assessment

For assessing the confidence of the IA, the uncertainties at the base of the aggregation scheme and their propagation through the various data processing steps need to be quantified (Fig. 1). Monitoring data and derived indices are typically sampled heterogeneously in time and space, and consequently the error structure is complex, involving variance contributions from spatial and temporal sources of variation as well as their interactions (Carstensen and Lindegarth, 2016). Different people/institutions carrying out sampling and sample analyses add further variation to the measurement data. Carstensen and Lindegarth (2016) presented a framework that listed a broad range of temporal, spatial and methodological variations affecting typical monitoring data, and demonstrated that quantifying the different variance components and using these in the indicator calculations gave more correct estimates of the indicator confidence. They assumed that the indicator was calculated from a linear combination of parameter estimates (i.e. typically mean and standardization parameters). However, even though this approach has statistical and computational advantages, many indicators involve more complex algebra and decision rules, and consequently do not fulfil this linear assumption. Given that indicator calculations can be quite complex, explicit formulas for calculating indicator uncertainty often cannot be derived. Therefore, we propose a more general methodology for quantifying the uncertainty of an indicator based on Monte Carlo simulations (Fig. 3).

An indicator (I) is calculated from a number of indices (y_i, i = 1, …,N or $\boldsymbol{y}$ in vector notation) through an indicator function $f(\ )$ that aggregates $\boldsymbol{y}$ into a single number (an indicator value, i.e. $\widehat{I} = f(\boldsymbol{y})$). These indices are assumed normal distributed, either directly or through an appropriately chosen link function $g(\ )$. Examples of the latter are $g(\ ) = \log(\ )$ for lognormal distributed indices (e.g. chlorophyll *a*) and $g(\ ) = \text{logit}(\ )$ for proportions (e.g. proportion of cyanobacteria). If $\boldsymbol{y}$ is normally distributed, $g(\ )$ becomes the identify function. Thus, $g(\boldsymbol{y})$ is normally distributed with an estimated mean ($\widehat{\mu}$) and a covariance matrix that depends on several variance components for spatial, temporal and methodological uncertainties as described in Carstensen and Lindegarth (2016). Consequently, the distribution of the indicator can be calculated by Monte Carlo simulations of errors from these variance components through the inverse link function $g^{-1}(\ )$ and the indicator function $f(\ )$ (Fig. 3). Following the recommendations from Carstensen and Lindegarth (2016), the variance components for the different indices should

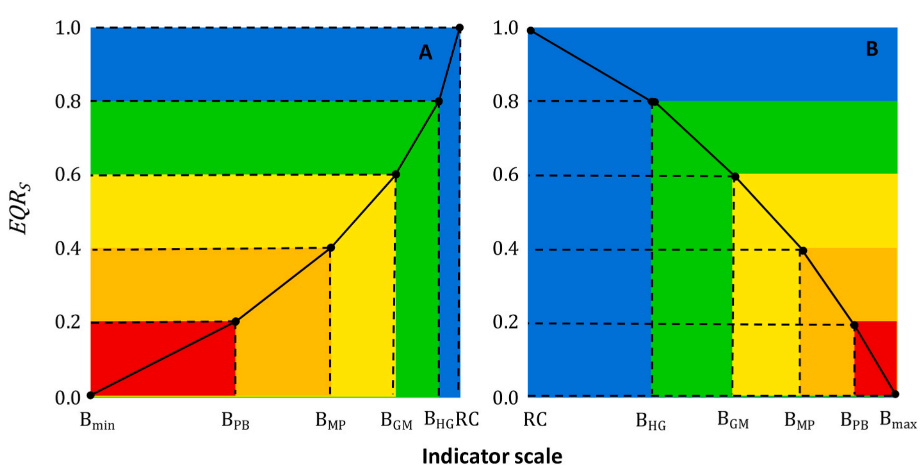

Fig. 2. The piecewise linear EQR-transformation (see Eq. (2)) for an indicator with responding negatively (A) or positively (B) to a pressure. RC = *RefCond*.

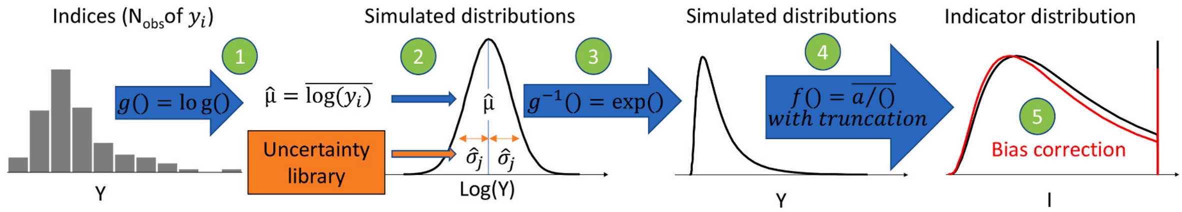

Fig. 3. Illustration of indicator uncertainty assessment for a lognormal distributed variable and an indicator function that averages the inverse of observations with an imposed truncation point. The different steps are indicated by green circles. Step 1: Observations/indices are log-transformed, and the mean of the log-transformed indices ($\hat{\mu}$) is calculated. Step 2: Monte Carlo realizations of N_{obs} log-transformed observations are simulated using the estimated mean $\hat{\mu}$ and variances from the uncertainty library to match the sampling properties of the indices. Step 3: The Monte Carlo realizations are back-transformed to lognormal distributions. Step 4: An indicator value is calculated for each realization using the indicator function (e.g. an average of inverse transformed simulations with truncation), and the indicator calculations of all Monte Carlo realizations produce an indicator distribution. Step 5: The indicator distribution is adjusted to have a mean equal to the indicator function applied to the indices.

be estimated from large data sets as a separate analysis and parameterized in the indicator calculation. Such variance estimates ($\hat{\sigma}_j^2$) are part of the IA tool configuration, referred to as the 'uncertainty library'. Thus, the Monte Carlo simulations are based on a mean estimated from the observed indices and parameterized variances.

The median of the Monte Carlo simulations will approach the indicator value for the observed indices ($\hat{I} = f(\mathbf{y})$) with increasing number of realizations, but the distribution of the simulations will only be normal (mean and median are equal) if both $f(\ )$ and $g(\ )$ are linear transformations. Any other functional form will produce a skewed distribution with mean and median differing. It is desirable for the simulated indicator distribution to have a mean equal to $f(\mathbf{y})$ and therefore a bias correction is introduced that adjusts the simulated distributions to ensure this requirement (Fig. 3). The bias correction, calculated as the difference between mean of the simulations and $f(\mathbf{y})$ after applying the link function $g(\ )$, is employed to adjust $\hat{\boldsymbol{\mu}}$ before applying the inverse link function $g^{-1}(\ )$ and the indicator function $f(\ )$ to ensure that constraints imposed by these transformations are maintained. Finally, the indicator translation is applied to the bias-corrected indicator distribution, resulting in an indicator distribution on the standardized scale (EQR_S).

3.3. Integrated assessment of indicators

The indicator information can be aggregated in many ways, potentially leading to diverse outcomes in the final IA (Borja et al., 2014). One of the first approaches for an IA was introduced by Vollenweider et al. (1998) with the TRIX indicator that scaled indices for productivity (chlorophyll *a* and oxygen concentration), nutrients (nitrogen and phosphorus concentrations), and supplementary water quality factors (Secchi depth) within their validity range followed by averaging after log-transforming indices as well as their upper and lower boundaries. The TRIX indicator has been applied, in original or slightly modified form, to many different ecosystems throughout the world, e.g. the Black Sea (Moncheva et al., 2001), the Persian Gulf (Taebi et al., 2005), the Mar Menor Lagoon (Spain) and the Montego estuary (Portugal) (Salas et al., 2008), the Southeast Mexico (Herrera-Silveira and Morales-Ojeda, 2009), and the Adriatic Sea (Andricevic et al., 2021). In contrast to the bounded intervals used for standardization in TRIX, the HELCOM HEAT (Andersen et al., 2011) and CHASE (Andersen et al., 2016) assessment tools for the Baltic Sea use only a single reference value for standardization (i.e. computing a ratio), before aggregating indicators by weighted averaging. The aggregation principle of these tools implicitly assumes that the standardized indicators have similar distributions that express comparable ecological status. However, this assumption can be questioned, as described above for chlorophyll *a* and Secchi depth, and importantly, the standardization is sensitive to the chosen upper/lower boundary or reference value. In fact, these values are often difficult to estimate as they represent conditions that are unlikely observed. The standardization with the piecewise linear model (Eq. (2)) also uses a bounded interval for the transformation, but it is less sensitive to the determination of the bounding range, since it uses several intermediate boundaries (B_{HG}, B_{GM}, B_{MP} and B_{PB}) to "steer the transformation". These waypoints of the transformation are more well determined as they can be better associated with observations. Importantly, the piecewise linear transformation can be seen as a discretized version of a more complex pressure-response curve, allowing for larger degree of freedom in translating an indicator to the standardized EQR-scale.

TRIX, HEAT and CHASE produce results on continuous scales, whereas other IA approaches employ decision trees, eventually combined with a discrete scoring system. The ASSETS methodology for eutrophication assessment considers a number of primary and secondary symptoms evaluated in a decision-tree structure (Bricker et al., 2003) and was used in the National Estuarine Eutrophication Assessment (NEEA) in the US (Bricker et al., 2007) as well as in China (Xiao et al., 2007) and Spain (Garmendia et al., 2012). Borja et al. (2009) developed a decision tree for integrating water, sediment and biomonitors for chemical status in the Basque Country in Spain. Taxon-specific scoring systems have also been developed for specific BQE such as the ASPT index (Armitage et al., 1983) and AMBI index (Borja et al. 2000). A scoring system, whether employed at the species level or at an aggregated level, essentially translates continuous information onto a discrete representation that involves a loss of information. Whereas the information loss may be less important when aggregating taxon-specific information across a rich and diverse community, the discretization of relatively few indicators, through scoring or decision tree structures, in an IA could represent a considerable information loss. Essentially, the objective of scoring systems and tree structures is similar to that of IA systems using a continuous scale, i.e. translating information onto a common scale, but the artefacts of the discretization are typically overlooked. Hence, such discretization approaches should be avoided for IA, unless the thresholds of the discretization represent tipping points for abrupt ecosystem change.

Borja et al., 2016 argued that any integration principle should be ecologically relevant, transparent and well documented. Whereas previous IA systems have generally used a relatively flat integration structure using weighted averages of indicators, we propose a hierarchical approach for aggregation to achieve a more equal representation of BQE, SQE and their sub-elements (Fig. 1; see also Swedish aggregation principles below). The hierarchical structure for aggregating indicator information is also implemented in the NEAT tool used for MSFD assessments of environmental status (Uusitalo et al., 2016). A hierarchical aggregation scheme also implies a weighted average of indicators, but instead of explicitly supplying weights for aggregation, these are implicitly supplied through the hierarchy of information aggregation. However, it is also possible to include weighting of indicators and sub-elements in the hierarchical aggregation. As an example, if there are three indicators for the macroalgae sub-element, these can have equal weights (w = 0.33 for all three) or alternatively, one indicator can be evaluated as twice as important as the other two indicators (w = 0.5, 0.25 and 0.25, respectively). The hierarchical structure also provides a more flexible aggregation in the presence of occasionally missing indicators, which would require adjusting all weights with a flat aggregation structure. Moreover, the hierarchical aggregation represents a modular structure that easily allows for introducing new indicators or removing obsolete indicators. In the hierarchical aggregation structure, sub-elements are aggregated with equal weight unless they have different importance for the ecosystem. For instance, the two sub-elements macroalgae and angiosperms could be weighted differently, according to whether the water body is dominated by hard substrate (rocky shores) or soft bottoms (sandy shores).

The hierarchical structure also offers greater transparency as opposed to composite indicators like TRIX. Since the IA is carried out in a stepwise manner, the status assessments of all indicators and their aggregations in the hierarchical structure are performed and can be evaluated. Consequently, the IA does not only inform whether or not good ecological status is achieved, it also specifically identifies the indicators responsible for failure to achieve good ecological status by going backwards through the transparent hierarchy. From the overall assessment, it is clear which BQEs do not reach good ecological status and from that one can trace the sub-elements and indicators contributing to this less favorable status. In contrast, it is more difficult to pinpoint the 'problematic' indicators with approaches like TRIX, HEAT and CHASE. Thus, a hierarchical structure for aggregation also direct management responses towards focusing on measures to improve the status of those indicators or BQE that show less favorable conditions.

Finally, the uncertainty of indicators and their propagation are also calculated at each step of the aggregation in the hierarchical structure. This allows for assessing the confidence of the status assessments from the indicator level to the overall IA. Indicators, sub-elements or BQEs that contribute with large uncertainty or that are critical to assessing whether good ecological status has been achieved (i.e. close to the good-moderate boundary) can be identified. Such information is important for revising and updating the monitoring program with the aim of improving the confidence of the assessment. In fact, we recommend to continuously assess the ecological status during the 6-year assessment period and improve the monitoring of those BQEs that are at risk of not achieving good ecological status, i.e. implementing an adaptive monitoring program that focuses on obtaining the most critical information for the IA.

4. Toolbox development

In Sweden, an online tool (https://waters.p.niva.no) has been developed for the ecological assessment of water bodies according to the WFD based on the indicator standardization, uncertainty assessment and integration presented above. The indicator calculations including uncertainty assessment were programmed in R, and the user interface was developed using "Shiny", a package for building web applications using R. Whereas earlier tools were developed as downloadable applications, where the user had to supply indicator values and boundaries (e.g. HEAT and NEAT), we acknowledged the need for providing validated routines for calculating indicator values and particularly, the uncertainty distribution associated with these. This online tool receives index values (e.g. biological indices, integrated values of discrete water samples) from the national monitoring databases hosted by the Swedish Agricultural University (SLU) for freshwater data and the Swedish Meteorological and Hydrological Institute (SMHI) for marine data. These indices are stored in an internal database of the tool and used for dynamically calculating indicator values using different data selection criteria. The indicator calculation follows the Swedish guidelines (Swedish Agency for Marine and Water Management, 2019), but it is possible to deselect criteria for data minimum requirements. All indicators, for which the underlying indices are available, are calculated along with their distributions, but it is also possible to deselect indicators in the aggregation schemes in case that indicator is deemed less relevant or highly uncertain. Additionally, the tool allows extrapolation of the status of a specific BQE from other water bodies sharing the same type, in cases where there is no local monitoring data to support the BQE status assessment. For example, if there are no fish indices for a given lake, the status of the fish BQE can be optionally chosen as the average of EQR_S distributions for the fish BQE in other lakes with similar physical characteristics. Importantly, all indicator aggregations are carried out on EQR_S-transformed indicator value.

Ecological status of Swedish rivers is based on eight biological indicators for the BQE and five indicators for the SQE that are aggregated through a hierarchical scheme, where SQE indicators can downgrade ecological status by one class (Fig. 4). The SQE indicators are combined with the BQE indicators that are consider sensitive to the given SQE indicator. For example, the IPS indicator for benthic diatoms is considered a sentinel for eutrophication (Kwandrans et al., 1998) and therefore, it is combined with the nutrient indicator for Total Phosphorus, such that TP can lower the status of the IPS indicator when the TP indicator suggests a poorer status than the IPS indicator. Another indicator for benthic diatoms, the ACID indicator, describe changes in the community that are related to acidification (Andrén and Jarlman, 2008) and therefore, it is combined with an indicator for changes in pH that can potentially lower the status of the ACID indicator. Since these two indicators respond to two different pressures, they are combined with the OOAO-principle for a combined status assessment of the BQE 'Macrophytes and phytobenthos'. The BQE 'Benthic invertebrates' is assessed using two biological indicators, where ASPT is a general benthic quality indicator (Armitage et al., 1983) and DJ is an indicator that is particularly sensitive to low oxygen concentrations (Dahl et al., 2004)) and therefore combined with a supporting indicator for oxygen conditions. These two indicators for benthic invertebrates are aggregated by averaging. The BQE 'Fish' is assessed by a main biological indicator (VIX) and modifications of this to indicate responses to acidification, morphological and hydrological changes (Beier et al., 2007). The fish status is calculated by aggregating the main VIX indicator with the relevant pressure-specific VIX indicators. The ecological status of rivers is found from the status of the three BQEs using the OOAO-principle.

Ecological status of Swedish lakes is based on 12 biological indicators for the BQE and five indicators for the SQE in a hierarchical scheme like that for rivers (Fig. 5). The BQE 'Phytoplankton' has two sub-elements (biomass and composition), where the status for phytoplankton biomass is found by averaging the EQR_S distributions for chlorophyll *a* and actual phytoplankton biomass before combination with the PTI indicator (Phillips et al., 2013). The nutrient indicator (TP) can potentially lower the status of the phytoplankton BQE. Only one biological indicator is used for the BQE 'Macrophytes' (TMI) that responds to eutrophication as pressure, which is assessed using Secchi depth as SQE that potentially can lower the status. The status of benthic diatoms is assessed as for rivers using the same two biological indicators supported by two chemical indicators (Fig. 4) that according to the Swedish national guidelines are combined using the OOAO-principle. The BQE 'Benthic invertebrates' is assessed using the general indicator ASPT (Armitage et al., 1983), the eutrophication-sensitive BQI indicator (Wiederholm, 1980) combined with oxygen as SQE indicator, and the acidification-sensitive MILA indicator (Johnson et al., 2007) combined with pH change as SQE indicator. The status of the BQE 'Fish' is assessed with the general biological indicator EQR8, the eutrophication-specific indicator EIndexW3 and the acidification-specific indicator AIndexW5 (Holmgren et al., 2007), where the two latter indicators can be downgraded with the TP and pH indicators, respectively. The ecological status of lakes is found from the status of the five BQEs using the OOAO-principle.

Ecological status of Swedish coastal waters is based on four biological indicators for the BQE and eight indicators for the SQE in a hierarchical scheme similar to that for rivers and lakes (Fig. 6). Only biomass is considered for the BQE 'Phytoplankton' and the two indicators for

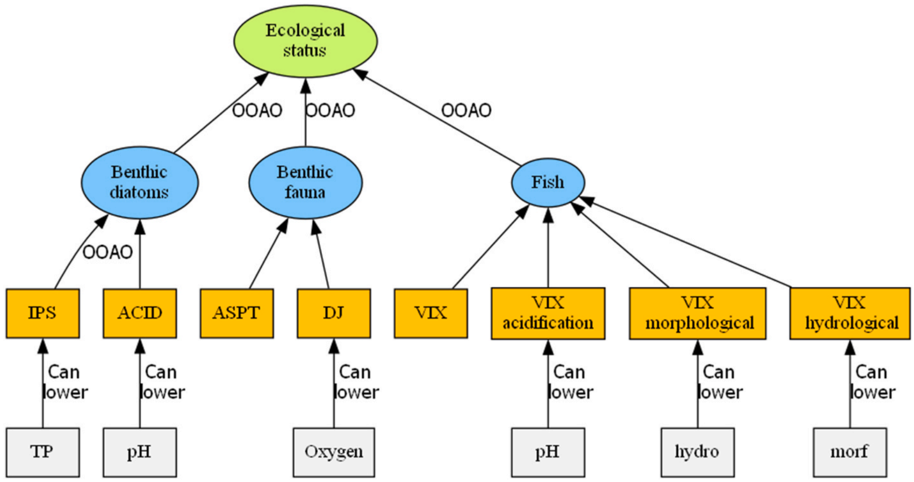

Fig. 4. Integrated assessment for Swedish rivers. The aggregation scheme is based on BQE indicators (orange) and SQE indicators (gray). SQE indicators can potentially downgrade BQE indicators in high or good status by one class level. Ecological status is found by applying the OOAO-principle to the BQEs (blue).

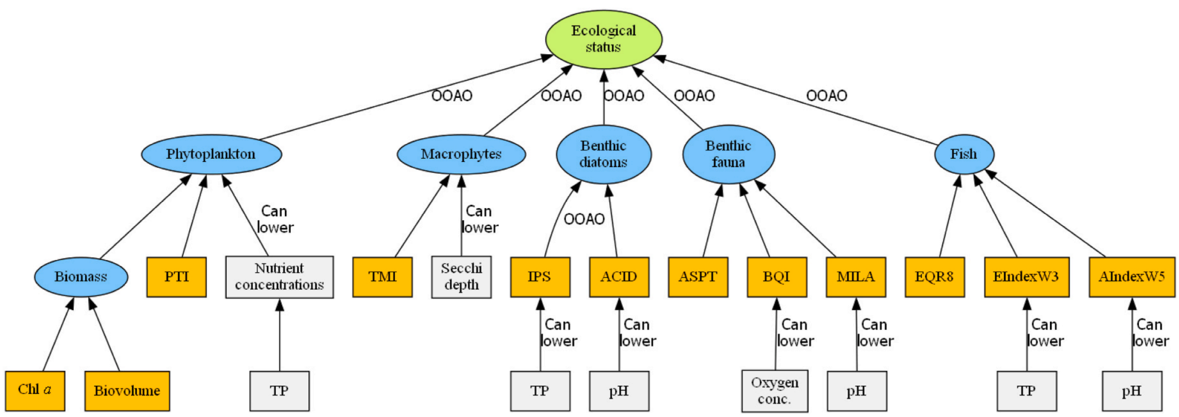

Fig. 5. Integrated assessment for Swedish lakes. Same information as in Fig. 4.

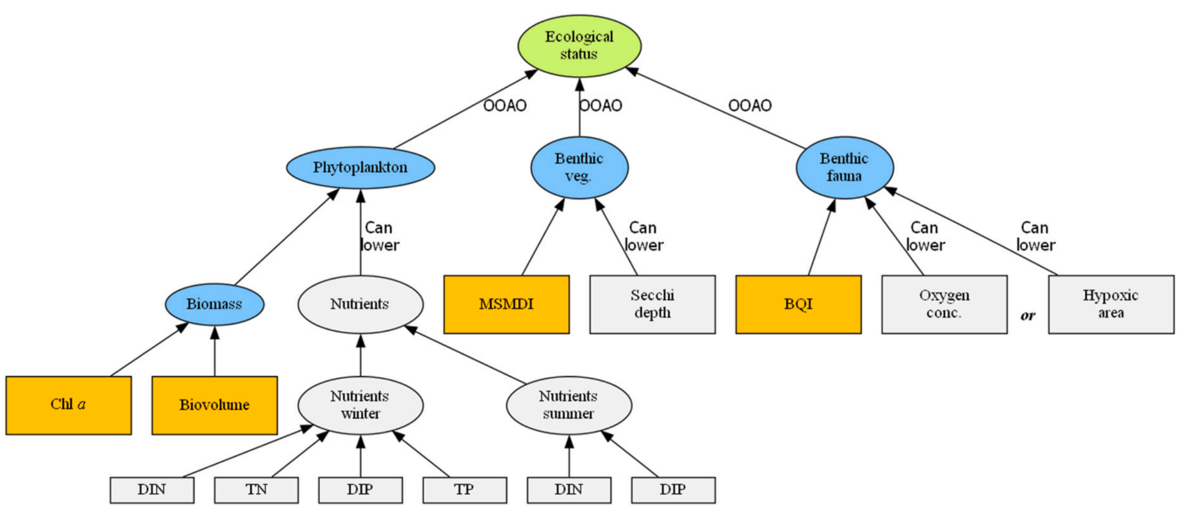

Fig. 6. Integrated assessment for Swedish coastal waters. Same information as in Fig. 4.

chlorophyll *a* and phytoplankton biomass are aggregated to the sub-element 'biomass' and combined with an overall status for nutrient conditions that is found by first aggregating inorganic and total nutrients for winter and aggregating inorganic nutrients for summer, and then aggregating winter and summer nutrient status into a combined nutrient status. In Swedish coastal waters, both nitrogen and phosphorus are considered important for controlling eutrophication. The BQE 'Benthic vegetation' is assessed using a multi-species maximum depth index (MSMDI), which is supported by Secchi depth as SQE indicator. The BQE 'Benthic fauna' is assessed by the BQI indicator (Rosenberg et al., 2004) and since the benthic fauna is sensitive to low oxygen conditions, this indicator is supported by a SQE indicator for low oxygen conditions. The ecological status of coastal waters is found from the status of the three BQEs using the OOAO-principle.

The aggregation scheme uses a generic approach to assess indicator values and uncertainties, and the tool only needs to have specified the indicator function ($f(\ )$) and the link function for the observations ($g(\ )$) and its inverse ($g^{-1}(\ )$) (Fig. 3) as well as the reference and boundary values for the subsequent transformation to the EQR_S-scale (Fig. 2). These functions are supplemented as a source file in R that is linked with the main source. For some indicators, reference and boundary values are not constant but given by functions of hydromorphological features or other variables and in such cases, these functions also need to be provided as a source file and configured through parsing a parameter vector to the functions. The variance parameters for assessing the indicator uncertainty are provided through a configuration file.

Similarly, the aggregation schemes, here exemplified with the current Swedish approach, can easily be adapted to other national guidelines, using different sets of indicators and aggregation principles. Such adaptation would require access to indices calculated from the databases hosting the national monitoring data, a source file with the indicator functions, link functions for normalization of the different indices and a parameter file with values for *RefCond* and boundaries as well as variances for different sources of uncertainty for confidence assessment (Carstensen and Lindegarth, 2016). Finally, the aggregation scheme for BQE and SQE is simple to modify for other national approaches, not only for WFD assessments but also for other directives such as the MSFD and EU Bird and Habitat Directive (BHD). The significance of the proposed method is the general method of combining indicator distributions through the EQR_S-transformation. Importantly, the integrated assessment (IA) is based on aggregating distributions rather than values such that confidence assessment can be made at all levels of aggregation.

5. Case Study: Gullmar Fjord

The Gullmar Fjord on the west coast of Sweden is a true threshold fjord with a sill depth around 40 m and maximum depth of 120 m. It receives freshwater inputs from a number of smaller rivers mixing with the surface water that is dominated by brackish water outflow from the Baltic Sea, whilst the bottom water is episodically replenished by denser water from the Skagerrak. This creates an almost permanently stratified system with infrequent ventilation of the bottom waters that are oxygen deficient most of the time. The coastal ecosystem has been well monitored for many years and therefore ideally suited for demonstrating the IA tool. In this example, the most recent 6-year assessment period (2013–2018) was chosen. The assessment included chlorophyll *a* and phytoplankton biovolume as indicators for phytoplankton and BQI as an indicator for benthic fauna. Inorganic and total nitrogen and phosphorus (winter and summer) were included as indicators of nutrient status, an SQE having the potential to downgrade phytoplankton status. Average oxygen concentration in bottom waters and the areal extent of hypoxia (defined as <2 ml L^{-1}) were indicators of the SQE for oxygen conditions, potentially downgrading benthic fauna status (Fig. 7). The MSMDI indicator for benthic vegetation was not included. It should be stressed that this case exemplifies the integrated assessment tool and does not represent an official status classification of the Swedish coastal water body. Moreover, the status is assessed by the median of the distributions, employing the face-value approach as described in the WFD Common Implementation Strategy (CIS, 2003).

For the 6-year assessment period, 38 observations of phytoplankton biovolume, 46 observations of chlorophyll *a*, and 23 BQI indices were used to characterize phytoplankton and benthic fauna BQE (Fig. 7a). These biological data were supplemented by 48 and 62 observations of summer and winter nutrient concentrations, respectively, as well as 22 observations of bottom oxygen concentrations and extent of hypoxia (data not shown). Distributions for biovolume and chlorophyll *a* were right skewed, suggesting lognormal distribution and hence, used $\log(\ )$ as link function, whereas BQI was approximately normally distributed and therefore, the identity link function was employed (i.e. no transformation).

Estimated means (−0.985 for log(biovolume), 0.266 for log(chlorophyll *a*) and 11.84 for BQI) combined with parameterized variances (i.e. from the uncertainty library) were used to simulate the sampling distributions with the link function (Fig. 7b), followed by back-transformation using the inverse link function (Fig. 7c). According to the Swedish Agency for Marine and Water Management (2019), individual observations of biovolume and chlorophyll *a* (Chla) should be EQR-transformed according to Eq. (1) with truncation of values exceeding 1. Applying this EQR-transformation yielded distributions close to 1 since most observations were close to or even below *RefCond* (Fig. 7d). For BQI, no transformation was employed, and the distributions of these steps were identical (Fig. 7b–d).

The simulated sample distributions were aggregated to indicator values by a 3-step aggregation over 1) each year and station, 2) each year, and 3) period according to the guidelines (Swedish Agency for Marine and Water Management, 2019). Performing this aggregation on each of the 10,000 Monte Carlo simulations yielded the indicator distributions (Fig. 7e), with means equal 0.960, 0.909 and 11.84 for EQR (biovolume), EQR(Chla) and BQI, respectively. After bias correction, indicator means were 0.944, 0.914 and 11.84 for EQR(biovolume), EQR (Chla) and BQI, respectively, yielding mean values closer, although not identical, to the phytoplankton indicator values calculated from the observations/indices (0.919 for EQR(biovolume) and 0.935 for EQR (Chla)). Thus, the bias corrections only partly adjust for the difference between the indicator calculation based on the observations/indices and simulations, due to the non-linear transformations involved in the phytoplankton indicator calculations. For BQI, the bias correction was almost zero (only small numerical variability), since all data processing for the indicator calculation was linear. The distributions of EQR(biovolume) and EQR(Chla) were mainly above the boundary between high and good status, whereas the BQI indicator distribution was almost equally divided between good and moderate status (Fig. 7e).

The WFD class boundaries are not equidistantly distributed over the indicator scale (Fig. 7e), in contrast to applying the EQR_S-transformation (Fig. 7f). It should be stressed that indicator distributions on the EQR_S scale are continuous even though they are displayed as discrete. In fact, the EQR_S-transformation does not change the distribution among status classes, i.e. probabilities of the different classes are the same before and after the transformation (cf. Fig. 7e versus 7f). Thus, for Gullmar Fjord the status of phytoplankton biovolume was high with 100% confidence, and the status of chlorophyll *a* was high with 73% probability, good with 23% probability, moderate with 3% probability and poor with 1% probability. Similarly, BQI was in good or moderate status with 47% and 53% probabilities, respectively.

Combining the two phytoplankton indicators (i.e. their EQR_S distributions) for the phytoplankton BQE resulted in a status assessment of high with 94% probability and good with 6% probability (Fig. 7g). This distribution, found by averaging EQR_S-values for biovolume and chlorophyll *a* for each of the 10,000 simulations, represents a compromise between the two indicators. Since BQI constitutes a composite index and thus the only indicator, no aggregation was done for benthic fauna. According to the assessment procedure, the phytoplankton status should be compared with the nutrient status, which is primarily good with 80% confidence although also with 18% probability of high and 2% probability of moderate status (Fig. S1). The status for phytoplankton is downgraded by one status class from high or good, if the nutrient status is lower (Fig. 6). Similarly, the status of benthic fauna can be downgraded using the EQR_S distribution for oxygen (Fig. S2). Employing the decision rule to the EQR_S distributions for phytoplankton and nutrients as well as for EQR_S distributions for benthic fauna and oxygen, i.e. for each of the 10,000 simulations, resulted in a relatively large shift for

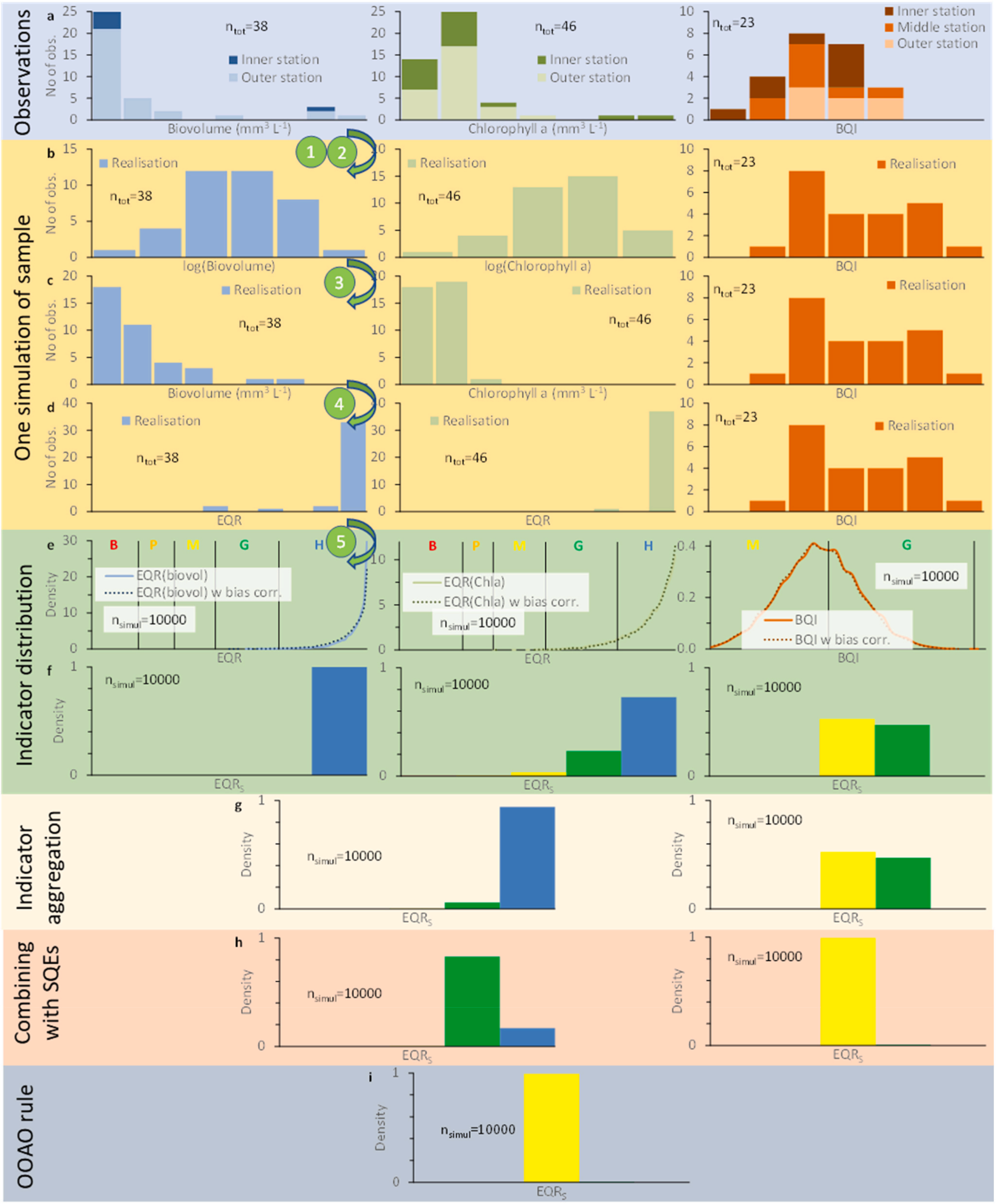

Fig. 7. Exemplification of the different steps of calculating indicator distributions (numbers in green circles refer to the steps in Fig. 3), standardization to EQR_S-scale (Eq. (2)), indicator aggregation, combination with SQE and application of OOAO-principle for IA for the coastal water body Gullmar Fjord in Sweden (cf. Fig. 6). For illustrating the different steps, only indicator calculations for phytoplankton and benthic fauna BQE are shown. a) Monitoring observations/indices of biovolume (left), chlorophyll a (middle) and BQI (right). b-d) Examples of one Monte Carlo simulation with the same sampling properties as the observations/indices. Simulations are shown with (b) the link transformation $g(\,)$, (c) after back-transformation $g^{-1}(\,)$ and (d) applying EQR-transformation associated with the indicator function for biovolume and chlorophyll a. For BQI these three distributions are identical since the link function is the identity function. e) The distribution of 10,000 simulations of the sample distribution after applying the indicator function $f(\,)$ without and with the bias correction. Vertical lines mark class boundaries for the five status classes. f) Translation of the indicator to the EQR_S-scale (cf. Fig. 2). g) Aggregation of indicator EQR_S distributions to BQE distribution. h) Combining BQE distributions with SQE distributions (SI) for potential downgrading status. i) Application of the OOAO-principle to the two BQE distributions.

phytoplankton (from high status with 94% confidence to good status with 83% confidence) and a narrow distribution for benthic fauna with 99% confidence of moderate status (Fig. 7h). Combining these two distributions with the OOAO-principle yielded an overall moderate status with 99% confidence (Fig. 7i). The status of benthic fauna assessed through BQI was responsible for the failure to achieve good ecological status, which points to unfavorable oxygen conditions in the water column and surface sediments due to eutrophication as the main cause. This example demonstrates how indicators and indicator aggregations are calculated based on a probabilistic framework by combining distributions from the bottom of the hierarchy to the overall assessment, and how this integrated assessment can be used to identify the most relevant cause and pressure by going backwards through the hierarchy of aggregation rules.

6. Conclusions

IA tools are needed for assessing compliance with ecological targets and such tools should deliver the best possible support for decision makers. We propose an approach for integrating information from multiple and diverse indicators through an aggregation scheme that can be customized to relevant policies and demonstrate its applicability for assessing ecological status according to the WFD in Sweden. These IA tools should be transparent, allowing to identify the specific indicator(s) failing to achieve ecological objectives and the most important pressure (s) responsible. We argue that such decision support tools should be developed as online applications with direct access to monitoring databases to ensure consistency in the calculations, allowing all stakeholders to derive identical results in a transparent manner. Furthermore, we emphasize the ability to assess the confidence of such IA through a probabilistic framework based on indicator distributions rather than indicator values. However, given that the confidence of the IA tool relies on monitoring data from complex ecosystems, any ecological status assessment should be supervised ensuring that input data are reliable and chosen indicators are relevant. Consequently, such an IA tool must be flexible and deliver input to the decision – not the decision itself.

Author contribution

Jacob Carstensen: Conceptualization, Methodology, Software, Writing – Original Draft **Ciaran Murray:** Methodology, Software, Writing – Review & Editing **Mats Lindegarth:** Conceptualization, Writing – Review & Editing, Funding Acquisition.

Funding sources

This study was primarily funded by a contract from the Swedish Environmental Protection Agency and the Swedish Agency for Water and Marine Management (Contract no 802-0179-10). JC and CJM further received support from the GES4SEAS project, funded by the European Union under the Horizon Europe program (grant agreement no. 101059877).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data are available through the link in the paper

Acknowledgements

We thank the many persons at the Swedish County Administrative Board (Länsstyrelsen) for testing the integrated assessment tool and providing useful recommendations for improvement. Niklas Hansson and Jonas Svensson from the Swedish Water and Management Agency are acknowledged for valuable inputs and discussions. This study was supported by the Swedish Water and Management Agency (WATERS project) and the GES4SEAS project funded by the European Union under the Horizon Europe program (grant agreement no. 101059877).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jenvman.2023.118625.

References

Allen, C.R., Fontaine, J.J., Pope, K.L., Garmestani, A.S., 2011. Adaptive management for a turbulent future. J. Environ. Manag. 92, 1339–1345. https://doi.org/10.1016/j.jenvman.2010.11.019.

Andersen, J.H., Murray, C.J., Kaartokallio, H., Axe, P., Molvær, J., 2010. A simple method for confidence rating of eutrophication status classifications. Mar. Pollut. Bull. 60, 919–924. https://doi.org/10.1016/j.marpolbul.2010.03.020.

Andersen, J.H., Axe, P., Backer, H., Carstensen, J., Claussen, U., Fleming-Lehtinen, V., Järvinen, M., Kaartokallio, H., Knuuttila, S., Korpinen, S., Laamanen, M., Lysiak-Pastuszak, E., Martin, G., Møhlenberg, F., Murray, C., Nausch, G., Norkko, A., Villnäs, A., 2011. Getting the measure of eutrophication in the Baltic Sea: towards improved assessment principles and methods. Biogeochem. 106, 137–156. https://doi.org/10.1007/s10533-010-9508-4.

Andersen, J.H., Murray, C., Larsen, M.M., Green, N., Høgåsen, T., Dahlgren, E., Garnaga-Budré, G., Gustavson, K., Haarich, M., Kallenbach, E.M.F., Mannio, J., Strand, J., Korpinen, S., 2016. Development and testing of a prototype tool for integrated assessment of chemical status in marine environments. Environ. Monit. Assess. 188, 115. https://doi.org/10.1007/s10661-016-5121-x.

Andrén, C., Jarlman, A., 2008. Benthic diatoms as indicators of acidity in streams. Fundam. Appl. Limnol. 173 (3), 237–253. https://doi.org/10.1127/1863-9135/2008/0173-0237.

Andricevic, R., Kekez, T., Vojkovic, M., 2021. Trophic status assessment of Central Eastern Adriatic Sa using water quality variables and loading capacity concept for estuaries. Mar. Pollut. Bull. 173, 113126 https://doi.org/10.1016/j.marpolbul.2021.113126.

Armitage, P.D., Moss, D., Wright, J.F., Furse, M.T., 1983. The performance of a new biological water quality score system based on macroinvertebrates over a wide range of unpolluted running-waters. Water Res. 17, 333–347. https://doi.org/10.1016/0043-1354(83)90188-4.

Beier, U., Degerman, E., Sers, B., Bergquist, B., Dahlberg, M., 2007. Bedömningsgrunder för fiskfaunans status i rinnande vatten – utveckling och tillämpning av VIX. Environmental quality criteria to determine the status of fish in running waters – development and application of VIX, 2007 Fiskeriverket Informerar 5, 59. p (In Swedish with English summary). https://www.slu.se/globalassets/ew/org/inst/aqua/externwebb/databaser/elprovfiskedatabasen/bedomningsgrunder-for-fiskfaunans-status-i-rinnande-vatten—utveckling-och-tillampning-av-vix_finfo-2007_5.pdf.

Birk, S., Bonne, W., Borja, A., Courrat, A., Poikane, S., Solimini, A., van de Bund, W., Zampoukas, N., Hering, D., 2012. Three hundred ways to assess Europe's surface waters: an almost complete overview of biological methods to implement the Water Framework Directive. Ecol. Indicat. 18, 31–41. https://doi.org/10.1016/j.ecolind.2011.10.009.

Borja, A., Perez, J.F., 2000. A marine biotic index to establish ecological quality of soft-bottom benthos within European estuarine and coastal environments. Mar. Pollut. Bull. 40, 1100–1114. https://doi.org/10.1016/S0025-326X(00)00061-8.

Borja, A., Bald, J., Franco, J., Larreta, J., Muxika, I., Revilla, M., Rodríguez, J.G., Solaun, O., Uriarte, A., Valencia, V., 2009. Using multiple ecosystem components, in assessing ecological status in Spanish (Basque Country) Atlantic marine waters. Mar. Poll. Bullet. 59, 54–64. https://doi.org/10.1016/j.marpolbul.2008.11.010.

Borja, A., Bricker, S.B., Dauer, D.M., Demetriades, N.T., Ferreira, J.G., Forbes, A.T., Hutchings, P., Jia, X., Kenchington, R., Marques, J.C., Zhu, C., 2008. Overview of integrative tools and methods in assessing ecological integrity in estuarine and coastal systems worldwide. Mar. Pollut. Bull. 56, 1519–1537. https://doi.org/10.1016/j.marpolbul.2008.07.005.

Borja, A., Prins, T., Simboura, N., Andersen, J.H., Berg, T., Marques, J.C., Neto, J.M., Papadopoulou, N., Reker, J., Teixeira, H., Uusitalo, L., 2014. Tales from a thousand and one ways to integrate marine ecosystem components when assessing the environmental status. Front. Mar. Sci. 1 https://doi.org/10.3389/fmars.2014.00022 article 72.

Borja, A., Elliott, M., Andersen, J.H., Berg, T., Carstensen, J., Halpern, B.S., Heiskanen, A.-S., Korpinen, S., Lowndes, J.S.S., Martin, G., Rodriguez-Ezpeleta, N., 2016. Overview of integrative assessment of marine systems: the Ecosystem Approach in practice. Front. Mar. Sci. 3 https://doi.org/10.3389/fmars.2016.00020 article 20.

Bricker, S.B., Ferreira, J.G., Simas, T., 2003. An integrated methodology for assessment of estuarine trophic status. Ecol. Model. 169, 39–60. https://doi.org/10.1016/S0304-3800(03)00199-6.

Bricker, S., Longstaff, B., Dennison, W., Jones, A., Boicourt, K., Wicks, C., Woerner, J., 2007. Effects of nutrient enrichment in the nation's estuaries: a decade of change, national estuarine eutrophication assessment update. In: NOAA Coastal Ocean Program Decision Analysis Series No. 26. National Centers for Coastal Ocean Science, Silver Spring, MD, p. 322.
Carstensen, J., 2007. Statistical principles for ecological status classification of Water Framework Directive monitoring data. Mar. Pollut. Bull. 55, 3–15. https://doi.org/10.1016/j.marpolbul.2006.08.016.
Carstensen, J., Lindegarth, M., 2016. Confidence in ecological indicators: a framework for quantifying uncertainty components from monitoring data. Ecol. Indicat. 67, 306–317. https://doi.org/10.1016/j.ecolind.2016.03.002.
CIS, 2003. Monitoring under the water framework directive. Guidance document no. 7. In: ECOSTAT Working Group 2.7. European Commission, Luxembourg, p. 160. http://forum.europa.eu.int/.
CIS, 2005. Overall Approach to the Classification of Ecological Status and Ecological Potential. Guidance Document No. 13. ECOSTAT Working Group 2A. European Commission, Luxembourg. 53 pp. Available from: http://forum.europa.eu.int/.
Dahl, J., Johnson, R.K., Sandin, L., 2004. Detection of organic pollution of streams in southern Sweden using benthic macroinvertebrates. Hydrobiologia 516, 161–172. https://doi.org/10.1023/B:HYDR.0000025264.35531.cb.
Dromph, K.M., Agusti, S., Basset, A., Franco, J., Henriksen, P., Icely, J., Lehtinen, S., Moncheva, S., Revilla, M., Roselli, L., Sørensen, K., 2013. Sources of uncer-tainty in assessment of marine phytoplankton communities. Hydrobiologia 704, 253–264. https://doi.org/10.1007/s10750-012-1353-0.
Dudley, B., Dunbar, M., Penning, E., Kolada, A., Hellsten, S., Oggioni, A., Bertrin, V., Ecke, F., Søndergaard, M., 2013. Measurements of uncertainty in macrophytemetrics used to assess European lake water quality. Hydrobiologia 704, 179–191. https://doi.org/10.1007/s10750-012-1338-z.
Garmendia, M., Bricker, S., Revilla, M., Borja, A., Franco, J., Bald, J., Valencia, V., 2012. Eutrophication assessment in Basque estuaries: comparing a North American and a European method. Estuar. Coast 35, 991–1006. https://doi.org/10.1007/s12237-012-9489-8.
Halpern, B.S., Longo, C., Hardy, D., McLeod, K.L., Samhouri, J.F., et al., 2012. An index to assess the health and benefits of the global ocean. Nature 488, 615–620. https://doi.org/10.1038/nature11397.
Hering, D., Borja, A., Carstensen, J., Carvalho, L., Elliott, M., Feld, C.K., Heiskanen, A.-S., Johnson, R.K., Moe, J., Pont, D., Solheim, A.L., van de Bund, W., 2010. The European Water Framework Directive at the age of 10: a critical review of the achievements with recommendations for the future. Sci. Total Environ. 408, 4007–4019. https://doi.org/10.1016/j.scitotenv.2010.05.031.
Herrera-Silveira, J.A., Morales-Ojeda, S.M., 2009. Evaluation of the health status of a coastal ecosystem in southeast Mexico: assessment of water quality, phytoplankton and submerged aquatic vegetation. Mar. Pollut. Bull. 59, 72–86. https://doi.org/10.1016/j.marpolbul.2008.11.017.
Holmgren, K., Kinnerbäck, A., Pakkasmaa, S., Bergquist, B., Beier, U., 2007. Bedömningsgrunder för fiskfaunans status i sjöar, 2007 Fiskeriverket Informerar 3, 54. p (In Swedish with English summary). https://www.slu.se/globalassets/ew/org/inst/aqua/externwebb/radgivning/bedomning-av-ekologisk-status/finfo_2007_3.pdf.
Johnson, R.K., Goedkoop, W., Fölster, J., Wilander, A., 2007. Relationships between macroinvertebrate assemblages of stony littoral habitats and water chemistry variables indicative of acid-stress. Water, Air, Soil Pollut. 7, 323–330. https://doi.org/10.1007/s11267-006-9072-z.
Keeler, B.L., Polasky, S., Brauman, K.A., Dalzell, B., 2012. Linking water quality and well-being for improved assessment and valuation of ecosystem services. Proc. Natl. Acad. Sci. USA 109, 18619–18624. https://doi.org/10.1073/pnas.1215991109.
Kirkfeldt, T.S., 2019. An ocean of concepts: why choosing between ecosystem-based management, ecosystem-based approach and ecosystem approach makes a difference. Mar. Pol. 106, 103541 https://doi.org/10.1016/j.marpol.2019.103541.
Kwandrans, J., Eloranta, P., Kawecka, B., Wojtan, K., 1998. Use of benthic diatom communities to evaluate water quality in rivers of southern Poland. J. Appl. Phycol. 10, 193–201. https://doi.org/10.1023/A:1008087114256.
Langhals, S.D., Reichert, P., Schuwirth, N., 2014. The method matters: a guide for indicator aggregation in ecological assessments. Ecol. Indicat. 45, 494–507. https://doi.org/10.1016/j.ecolind.2014.05.014.
Moe, S.J., Solheim, A.L., Soszka, H., Gołub, M., Huturowicz, A., Kolada, A., Picinska-Faltynowicz, J., 2015. Integrated assessment of ecological status and misclassification of lakes: the role of uncertainty and index combination rules. Ecol. Indicat. 48, 605–615. https://doi.org/10.1016/j.ecolind.2014.08.018.
Moncheva, S., Gotsis-Skretas, O., Pagou, K., Krastev, A., 2001. Phytoplankton blooms in Black Sea and mediterranean coastal ecosystems subjected to anthropogenic eutrophication: similarities and differences. Estuar. Coast Shelf Sci. 53, 281–295. https://doi.org/10.1006/ecss.2001.0767.
Phillips, G., Lyche-Solheim, A., Skjelbred, B., Mischke, U., Drakare, S., Free, G., Järvinen, M., de Hoyos, C., Morabito, G., Poikane, S., Carvalho, L., 2013. A phytoplankton trophic index to assess the status of lakes for the Water Framework Directive. Hydrobiologia 704, 75–95. https://doi.org/10.1007/s10750-012-1390-8.
Probst, W.N., Lynam, C., 2016. Integrated assessment results depend on aggregation method and framework structure – a case study within the European Marine Strategy Framework Directive. Ecol. Indicat. 61, 871–881. https://doi.org/10.1016/j.ecolind.2015.10.040.
Rosenberg, R., Blomqvist, M., Nilsson, H.C., Cederwall, H., Dimming, A., 2004. Marine quality assessment by use of benthic species-abundance distributions: a proposed new protocol within the European Water Framework Directive. Mar. Pollut. Bull. 49, 728–739. https://doi.org/10.1016/j.marpolbul.2004.05.013.
Salas, F., Teixeira, H., Marcos, C., Marques, J.C., Pérez-Ruzafa, A., 2008. Applicability of the trophic index TRIX in two transitional ecosystems: the Mar Menor lagoon (Spain) and the Mondego estuary (Portugal). ICES J. Mar. Sci. 65, 1442–1448. https://doi.org/10.1093/icesjms/fsn123.
Schuwirth, N., Borgwardt, F., Domisch, S., Friedrichs, M., Kattwinkel, M., Kneis, D., Kuermmerlen, M., Langhans, S.D., Martínez-López, J., Vermeiren, P., 2019. How to make ecological models useful for environmental management. Ecol. Model. 411, 108784 https://doi.org/10.1016/j.ecolmodel.2019.108784.
Solheim, A.L., Feld, C.K., Birk, S., Phillips, G., Carvalho, L., Morabito, G., Mischke, U., Willby, N., Søndergaard, M., Hellsten, S., Kolada, A., Mjelde, M., Höhmer, J., Miler, O., Argillier, C., Jeppesen, E., Lauridsen, T.L., Poikane, S., 2013. Ecological status assessment of European lakes: a comparison of metrics for phytoplankton, macrophytes, benthic invertebrates and fish. Hydrobiologia 704, 57–74. https://doi.org/10.1007/s10750-012-1436-y.
Swedish Agency for Marine and Water Management, 2019. Klassificering Och Miljökvalitetsnormer Avseende Ytvatten (HVMFS 2019:25). https://www.havochvatten.se/vagledning-foreskrifter-och-lagar/foreskrifter/register-vattenforvaltning/klassificering-och-miljokvalitetsnormer-avseende-ytvatten-hvmfs-201925.html.
Taebi, S., Etemad–Shahidi, A., Golamreza, A.F., 2005. Examination of three eutrophication indices to characterize water quality in the North East of Persian Gulf. J. Coast. Res., SI 42, 405–411. https://www.jstor.org/stable/25737009.
Teixeira, H., Berg, T., Uusitalo, L., Fürhaupter, K., Heiskanen, A.-S., Mazik, K., Lynam, C. P., Neville, S., Rodriguez, J.G., Papadopoulou, N., Moncheva, S., Chulilova, T., Kryvenko, O., Krause-Jensen, D., Zaiko, A., Verrisimo, H., Pantazi, M., Carvalho, S., Patricio, J., Uyarra, M.C., Borja, A., 2016. A catalogue of marine biodiversity indicators. Front. Mar. Sci. 3 https://doi.org/10.3389/fmars.2016.00207 article 207.
Uusitalo, L., Blanchet, H., Andersen, J.H., Beauchard, O., Berg, T., Bianchelli, S., Cantafaro, A., Carstensen, J., Carugati, L., Cochrane, S., Danovaro, R., Heiskanen, A.-S., Karvinen, V., Moncheva, S., Murray, C., Neto, J.M., Nygård, H., Pantazi, M., Papadopoulou, N., Simboura, N., Srėbalienė, G., Uyarra, M.C., Borja, A., 2016. Indicator-based assessment of marine biological diversity–lessons from 10 case studies across the European seas. Front. Mar. Sci. 3, 159. https://doi.org/10.3389/fmars.2016.00159.
Van Beest, F., Nygård, H., Fliming, V., Carstensen, J., 2021. On the uncertainty and confidence in decision support tools (DSTs) with insights from the Baltic Sea ecosystem. Ambio 50, 393–399. https://doi.org/10.1007/s13280-020-01385-x.
Vollenweider, R.A., Giovanardi, F., Montanari, G., Rinaldi, A., 1998. Characterization of the trophic conditions of marine coastal waters with special reference to the NW Adriatic Sea: proposal for a trophic scale, turbidity and generalized water quality index. Environmentrics 9, 329–357. https://doi.org/10.1002/(SICI)1099-095X(199805/06)9:3%3C329::AID-ENV308%3E3.0.CO;2-9.
Wiederholm, T., 1980. Use of benthos in lake monitoring. J. Water Pollut. Control Fed. 537–547. https://www.jstor.org/stable/25040750.
Xiao, Y., Ferreira, J.G., Bricker, S.B., Nunes, J.P., Zhu, M., Zhang, X., 2007. Trophic assessment in Chinese coastal systems-review of methods and application to the changjiang (yangtze) estuary and jiaozhou bay. Estuar. Coast 30, 901–918. https://doi.org/10.1007/BF02841384.