frontiers in
MARINE SCIENCE

Marine Ecosystem Ecology

# Tales from a thousand and one ways to integrate marine ecosystem components when assessing the environmental status

Angel Borja, Theo Prins, Nomiki Simboura, Jesper H. Andersen, Torsten Berg, João Carlos Marques, Joao M. Neto, Nadia Papadopoulou, Johnny Reker, Heliana Teixeira and Laura Uusitalo

| | |
|---|---|
| Journal Name: | Frontiers in Marine Science |
| ISSN: | 2296-7745 |
| Article type: | Original Research Article |
| Received on: | 07 May 2014 |
| Accepted on: | 25 Jun 2014 |
| Provisional PDF published on: | 25 Jun 2014 |
| www.frontiersin.org: | www.frontiersin.org |
| Citation: | Borja A, Prins T, Simboura N, Andersen JH, Berg T, Marques JC, Neto JM, Papadopoulou N, Reker J, Teixeira H and Uusitalo L(2014) Tales from a thousand and one ways to integrate marine ecosystem components when assessing the environmental status. *Front. Mar. Sci.* 1:22. doi:10.3389/fmars.2014.00022 |

# Tales from a thousand and one ways to integrate marine ecosystem components when assessing the environmental status

Angel Borja[1*], Theo Prins[2], Nomiki Simboura[3], Jesper H. Andersen[4], Torsten Berg[5], Joao-Carlos Marques[6], Joao M. Neto[6], Nadia Papadopoulou[7], Johnny Reker[8], Heliana Teixeira[9], Laura Uusitalo[10]

(1*) AZTI-Tecnalia, Marine Research Division, Herrera Kaia, Portualdea s/n, 20110 Pasaia, Spain, aborja@azti.es (**corresponding author**)

(2) DELTARES, PO Box 177, 2600 MH Delft, The Netherlands

(3) HCMR - Hellenic Centre for Marine Research, Institute of Oceanography, P.O. Box 712, 19013, Anavissos, Attica, Greece

(4) University of Aarhus, Department of Bioscience, P.O. Box 358, DK-4000 Roskilde, Denmark

(5) MARILIM Aquatic Research GmbH, Heinrich-Wöhlk-Straße 14, 24232 Schönkichen, Germany

(6) IMAR – Institute of Marine Research, Department of Life Sciences, Faculty of Sciences and Technology, University of Coimbra, Largo Marquês de Pombal, 3004-517 Coimbra, Portugal

(7) HCMR - Hellenic Centre for Marine Research, Institute of Marine Biological Resources and Inland Waters, P.O. Box 2214, 71003, Heraklion, Crete, Greece

(8) European Environment Agency, Kongens Nytorv 6, DK-1050 Copenhagen

(9) European Commission, Joint Research Centre, Institute for Environment and Sustainability, 21027 Ispra, Italy

(10) SYKE- Finnish Environment Institute, Marine Research Centre, P.O. Box 140, FIN-00251, Helsinki, Finland

## Abstract

Assessing the environmental status of marine ecosystems is useful when communicating key messages to policymakers or the society, reducing the complex information of the multiple ecosystem and biodiversity components and their important spatial and temporal variability into manageable units. Taking into account the ecosystem components to be addressed (e.g. biological, chemical, physical), the numerous biodiversity elements to be assessed (e.g. from microbes to sea mammals), the different indicators needed to be studied (e.g. in Europe, 56 indicators of status have been selected), and the different assessment scales to be undertaken (e.g. from local to regional sea scale), some criteria to define spatial scales and some guidance on aggregating and integrating information is needed. We have reviewed, from ecological and management perspectives, the approaches for aggregating and integrating currently available for marine status assessment in Europe and other regions of the world. Advantages and shortcomings of the different alternatives are highlighted. We provide some guidance on the steps towards defining rules for aggregation and integration of information at multiple levels of ecosystem organization, providing recommendations on when using specific rules in the assessment. A main conclusion is that any integration principle used should be ecologically-relevant, transparent and well documented, in order to make it comparable across different geographic regions.



**Key words**: ecosystems, marine, indicators, Marine Strategy Framework Directive, descriptors, criteria, assessment, integration

## 1.- Introduction

The requirement to assess the environmental status of marine waters is growing across continents (Borja *et al.*, 2008). It is also one of the challenging tasks to be accomplished in Europe, within the Marine Strategy Framework Directive (MSFD) (European Commission, 2008). The different legislative mandates to asses status coming from the MSFD, Water Framework Directive (WFD) (2000/60/EC) and Habitats Directive (92/43/EEC) and other international initiatives have produced numerous methodologies that can be applied to different ecosystem components, such as various taxonomic or functional groups, habitats, traits, physical features, or to the whole ecosystem (Birk *et al.*, 2012; Halpern *et al.*, 2012). Despite this wealth of methods, determining environmental status and assessing marine ecosystems health in an integrative way is still one of the grand challenges in marine ecosystems ecology research and management (Borja, 2014).

Different attempts to understand, define and assess ecosystem health have been made in recent years (Costanza and Mageau, 1999; Ulanowicz, 2000; Mee *et al.*, 2008; Ojaveer and Eero, 2011; Borja *et al.*, 2013; Tett *et al.*, 2013). The concept of "good environmental status" (GEnS) integrates physical, chemical and biological aspects, together with the services provided by ecosystems, including a sustainable use of the marine resources by society (Borja, 2014). However, synthesizing these aspects into a single value will never appropriately reflect all aspects considered to derive the value (Purvis and Hector, 2000; Derous *et al.*, 2007). Still, this step is useful when communicating key messages to policymakers or the society, reducing the complex information of the multiple ecosystem components and their important spatial and temporal variability into manageable units, which can be used in ecosystem management. Following the recommendation from Mee *et al.* (2008), we use the GEnS acronym because the meaning of 'environmental', within the MSFD, and 'ecological' (good ecological status), within the WFD, is different (see Borja *et al.* (2010), for differences between both concepts), implying a different emphasis between these two major pieces of legislation.

In the case of the MSFD, an appropriate integration process might be even more complex, since the assessment of the status is based upon 11 qualitative descriptors (i.e. D1: biological diversity; D2: non-indigenous species; D3: exploited fish and shellfish; D4: food webs; D5: human-induced eutrophication; D6: seafloor integrity; D7: hydrographical condition; D8: contaminants; D9: contaminants in fish and seafood; D10: litter; and D11: energy and noise), which are further divided into 29 criteria and 56 indicators of health (European Commission, 2010). An overview of MSFD descriptors, criteria and indicators is shown in Table 1.

The aim of this work is to present an overview of the different methods currently available to synthesize the ecosystem complexity, by aggregating and integrating information when assessing the status, focusing mostly on the descriptors related to biodiversity, namely D1, D2, D4, D6 (Cardoso *et al.*, 2010; Prins *et al.*, 2013). This overview would assist managers, through the guidelines provided, in taking decisions for a better management of the marine ecosystems.

## 2.- Ecosystem components combination requirements in assessing the status

There are different methods that can be applied to combine indicators and criteria within descriptors and across descriptors to eventually result in an assessment of GEnS for a specific geographic area. This combination both involves aggregation and integration. The term aggregation is here used for the combination of comparable elements across temporal and spatial scales, indicators and criteria, within a descriptor. The term integration is used for the combination of different elements (e.g. across descriptors). Both combination methods (aggregation and integration) may involve numeric calculations.

In Europe, the MSFD defines environmental status as "the overall state of the environment in marine waters, taking into account the structure, function, and processes of the constituent marine ecosystems together with natural physiographic, geographic, biological, geological and climatic factors, as well as physical, acoustic and chemical conditions, including those resulting from human activities inside or outside the area concerned".

Taking this definition into account, Borja *at al.* (2013) have proposed an operational definition: "GEnS is achieved when physicochemical (including contaminants, litter and noise) and hydrographical conditions are maintained at a level where the structuring components of the ecosystem are present and functioning, enabling the system to be resistant (ability to withstand stress) and resilient (ability to recover after a stressor) to harmful effects of human pressures/activities/impacts, where they maintain and provide the ecosystem services that deliver societal benefits in a sustainable way (i.e. that pressures associated with uses cumulatively do not hinder the ecosystem components in order to retain their natural diversity, productivity and dynamic ecological processes, and where recovery is rapid and sustained if a use ceases)''.

This latter definition includes all MSFD descriptors. Hence, to assess whether or not GEnS has been achieved, some aggregation within and integration across the 11 descriptors is required to move from the evaluation at the level of indicators (the 56 indicators and 29 criteria described in the Commission Decision (European Commission, 2010), see also Table 1) to a global assessment of status, as mentioned also in Cardoso *et al.* (2010). The problem is how to deal with the complex task of combining a high number of indicators and descriptors. To develop a common understanding on this, it is important that Member States are transparent on (i) the process of selecting the indicators to be monitored; (ii) the approaches and combination methods they have used; and (iii) the uncertainties in their indicators and methods.

## 3.- General principles for combination

Based on a literature review, we identified a number of different approaches for combining a number of variables (which could be metrics, indicators, or criteria) into an overall assessment. Some of them have been used within the WFD, others within the RSCs and some others in the MSFD. An overview of the methods is given in Table 2.

When considering the aggregation of indicators, an important factor to be taken into account is the reliability of the individual indicators to be aggregated. With each indicator, it is always possible to make a type I error, i.e. to get a non-GEnS result when the system in fact is in GEnS. The probability of this false positive (FP) signal varies (i) between indicators (Murtaugh, 1996), depending on the natural variability; (ii) with the amount of data used to define the indicator value; and (iii) with the target level compared to the situation in the nature. The risk of getting a FP from each of the individual indicators should affect the aggregation rule as well: if the risk of a FP is a uniform 5% per indicator, on average 1 out of 20 indicators is expected to give a FP; a problem if all indicators should in fact show GEnS. In order to come up with an aggregated assessment in which the risk level is within reasonable bounds, this aspect cannot be overlooked.

*3.1. One-out, all-out (OOAO)*

The OOAO approach is used in the WFD to integrate within and across Biological Quality Elements (BQEs)(CIS, 2003), in order to reach the ecological status of a water body. This approach follows the general concept that the ecological status assigned to a water body depends on the BQE with the lowest status, and consequently, the OOAO approach results in a "worst case".

A prerequisite for the aggregation of various indicators is that they are sensitive to the same pressure (Caroni *et al.*, 2013). In such a case, different aggregation methods can be used to combine parameters (medians, means, etc.). Caroni *et al.* (2013) recommend an OOAO approach when the combination involves parameters/indicators that are sensitive to different pressures. The application of averaging rules may lead to biased results in those cases. The WFD Classification Guidance (CIS, 2003) also advises to use OOAO when combining parameters/indicators that are sensitive to different pressures.

Borja *et al.* (2009a) discussed the challenge of assessing ecological integrity in marine waters, and suggest that simple approaches, such as the 'OOAO' principle of the WFD, may be a useful starting point, but eventually should be avoided. The ecological integrity of an aquatic system should be evaluated using all information available, including as many biological ecosystem elements as is reasonable, and using an ecosystem-based assessment approach. The OOAO rule can be considered a rigorous approach to the precautionary rule, in an ideal world where the status based on each BQE can be measured without error. It results in very conservative assessments (Ojaveer and Eero, 2011). In practice, the inevitable uncertainty associated with monitoring and assessment for each metric and BQE leads to problems of probable underestimation of the true overall status. The OOAO principle has therefore been criticized as it increases the probability of committing a false positive error, leading to an erroneous downgrading of the status of a water body as it has been observed especially within the WFD (Borja and Rodríguez, 2010; Ojaveer and Eero, 2011; Borja *et al.*, 2013; Caroni *et al.*, 2013). In the case of the MSFD, with such large number of descriptors, criteria and indicators, the probability of not achieving good status becomes very high and, probably, unmanageable in practical terms (Borja *et al.*, 2013).

Alternative methods for integrating multiple BQEs in the WFD are currently being considered (Caroni *et al.*, 2013).

*3.2. Averaging approach*

The averaging approach is the most commonly used method to aggregate indicators

(Shin *et al.*, 2012) and consists of simple calculations, using methods such as arithmetic average, hierarchical average, weighted average, median, sum, product or combinations of those rules, to come up with an overall assessment value.

Ojaveer and Eero (2011) showed that in cases where a large number of indicators is available, the choice of e.g. either medians or averages in aggregating indicators did not substantially influence the assessment results. However, this might not necessarily be the case when only a few indicators are available. In such a situation, the result will depend to a larger degree on the distribution of the values involved. A skewed distribution reflecting some major factors and a few ones with very different values will result in very different assessment results for the median compared to assessments based on means. Apart from the mathematical applicability of either method based on the underlying data (e.g. homoscedasticity), the choice of the actual averaging method may be driven by policy decisions focusing on either central trends without much attention to extreme values (median) or focusing on weighting the individual values by their magnitude (arithmetic mean).

The way the indicators are hierarchically arranged influences the assessment results as well, but Ojaveer and Eero (2011) found that these effects were considerably less important than the effects of applying different aggregation rules.

Differential weighting applied to the various indicators can be used when calculating means or medians. An adequate basis for assigning weights is not always available and in such cases an equal weight is recommended by Ojaveer and Eero (2011). Assigning weights often involves expert judgment, and Aubry and Elliott (2006) point out that in some cases, expert opinions on weights can show important divergence.

### *3.3. Conditional rules*

Conditional rules (a specific proportion of the variables have to achieve good status) are an approach where indicators can be combined in different ways for an overall assessment, depending on certain criteria. This provides an opportunity to use expert judgment when combining indicators, in a transparent way. An example of this approach is the application of a conditional rule of at least two out of three indicators (one biotic index and two structural or diversity indices) should pass the threshold in order to achieve GEnS for benthic community condition under D6 in Hellenic waters (Simboura *et al.*, 2012). Tueros *et al.* (2009) present another example of the conditional rule in which when integrating water and sediment variables into an overall assessment of the chemical status and only one sediment or water variable does not meet the objective, while the rest of the variables meet, the final chemical status achieves the objective. This work was also mentioned under the "two out, all out" approach considering the case when two variables do not meet the objective and the final status fails.

Breen *et al.* (2012) used several risk criteria rules and worst-case or integrated approaches when combining evidence before a final assessment. Following Cardoso *et al.* (2010) the integrated approach was applied to Biodiversity, Non-indigenous species, Eutrophication and Seafloor Integrity descriptors, while all other descriptors used a worst case approach following the OOAO principle whereby if one set of evidence suggested that the risk was ‘high’ then ‘high’ was automatically assessed for the entire descriptor.

*3.4. Scoring or rating*

In this method different scores are assigned to a status level (for example, ranging from 1 to 5), for a number of different elements. The scores are summed up to derive a total score which is then rated according to the number of elements taken into account. Different weights can be assigned to the various elements. This method was proposed by Borja *et al*. (2004) to calculate an integrative index of quality and is the basis of many multimetric indices used within the WFD and the MSFD combining different parameters or metrics using the weighted scoring or rating rule into one integrative multimetric index (Birk *et al.*, 2012). It must be recognized here that this approach implies the score values being on a cardinal scale and acting as weighting factors. Otherwise, using an ordinal scale for the scores, summing up the individual elements is mathematically not defined.

Another example is the method developed by Borja *et al*. (2010, 2011b) for a cross-descriptor integration, combining the 11 descriptors of MSFD based on the WFD, HELCOM (2009a, 2009b, 2010) and OSPAR (2010, 2012) experiences. An Ecological Quality Ratio (EQR) was calculated for each indicator of the various MSFD descriptors, with the EQR for the whole descriptor being the average value of the EQR of the indicators. Then, by multiplying the EQR with the percent weight assigned to each descriptor (and summing up to 100), an overall environmental status value was derived.

*3.5. Multimetric indices to combine indicators*

Within the WFD there are many examples of multimetric indices developed for different biological elements, driven by the need to fulfil the detailed requirements of the WFD (see Birk *et al.* (2012) for a complete synthesis).

In addition, within the MSFD, the European Commission established a number of Task Groups consisting of technical experts to help inform the discussions on how to reach a common understanding of the 11 descriptors. Hence, Task Group 6 report on seafloor integrity (Rice *et al.*, 2010) recommends the use of multimetric indices or multivariate techniques for integrating indicators of species composition attributes of this descriptor, such as diversity, distinctness, complementarity/(dis)similarity, or species-area relationships.

There are various other examples of multi-metric indices used to assess the status of the macrobenthos (see Borja *et al.* (2011a) for an overview). Multimetric methods to combine multiple parameters in one assessment may result in more robust indicators, compared to indicators based on single parameters. However, scaling of a multimetric index may be less straightforward, and ideally the various parameters should not be inter-correlated (e.g. the discussion on the TRIX index in Primpas and Karydis (2011)).

*3.6. Multidimensional approaches*

Multivariate methods, such as Discriminant Analysis or Factor Analysis combine parameters in a multi-dimensional space. For assessment purposes, areas need to be classified into groups of GEnS and non-GEnS.

Multivariate methods have the advantage of being more robust and less sensitive to correlation between indicators. However, interpretation is less intuitive than other methods, as information on individual indicators in each ecosystem is lost (Shin *et al.*, 2012) and links to management options are less obvious.

### *3.7. Decision tree*

Decision trees provide the opportunity to apply different, specific, rules to combine individual assessments into an overall assessment. A decision tree allows implementing individual rules at each of its nodes and thus incorporates arbitrary decisions at each step within the decision tree. The decision rules can be quantitative or qualitative as well as based on expert judgement. This gives room for a high degree of flexibility in reaching the final assessment and can thus be used where the other principles fail to represent the intricate interactions, feedback loops and dependencies involved in ecosystem functioning between the ecosystem components.

A simple version of a decision tree involves only having a few conditional rules where a specific proportion or certain individually specified indicators have to achieve good status in order to achieve GEnS. Borja *et al.* (2013) implicitly propose using this kind of decision tree when they take the view that for biodiversity (D1) to be in good status, all other descriptors must be in good status and if one of the pressure descriptors fails, then D1 also fails.

Borja *et al.* (2004, 2009b) describe a methodology that integrates several biological elements (phytoplankton, benthos, algae, phanerogams, and fishes), together with physicochemical elements (including pollutants) into a quality assessment. The proposed methodologies accommodate both WFD and the MSFD. They suggest that the decision tree should give more weight to individual elements taking into account the spatial and temporal variability and the availability of accurate methodologies for some of them (i.e. benthos) and to individual assessment methods which have been used broadly by authors other than the proposers of the method, tested for several different human pressures, and/or intercalibrated with other methods.

### *3.8. Probabilistic approach*

Each of the indicator results are uncertain, due to several factors e.g. natural variation in the sampling sites, random variation in the samples, insufficient scientific understanding about what should be the reference value for good status, etc. Some indicators are bound to include more uncertainty that others, due to differences in the amount of data used, the extent of scientific understanding regarding the issue, and the amplitude of natural variation. If these uncertainties can be approximated, this gives rise to the possibility of taking this information into account when integrating the indicators. The more uncertain indicators will get less weight in the integrated assessment, while the more certain ones will be more reliable and hence get more weight. The calculus of the integrated assessment can be based on Bayesian statistics, giving transparent and coherent rules by which the final score is calculated.

This approach can be combined to one or several of the above-mentioned approaches:

for example, conditional rules can be set in addition to the probabilistic integration rule to include expert judgement; and the principles outlined in the decision tree approach can be applied as well.

Barton *et al.* (2012) demonstrate how to use the probabilistic approach in the DPSIR framework in the case of eutrophication management. There are several other examples in the recent literature about how to evaluate various management measures under uncertainty to optimise one target, such as eutrophication (Barton *et al.*, 2008; Lehikoinen *et al.*, 2013a) and oil spill severity (Lehikoinen *et al.*, 2013b). This approach could be expanded to include several descriptors or indicators.

Probabilistic combination of uncertain indicators would naturally lead to a probability estimate of how likely it is that a marine area is in GEnS; we would, for example, end up with an estimate that the sea area is in GEnS with 70% probability. The managers would then have to decide how much uncertainty they are willing to tolerate; i.e. are they happy if the probability of GEnS is above 50%, or whether they want a higher certainty?

*3.9. High-level integration*

An example of a high-level integration, where assessments for several ecosystem components are merged into a final assessment, is the HELCOM-HOLAS project (HELCOM, 2010). The report presents an indicator-based assessment tool termed HOLAS ('Holistic Assessment of Ecosystem Health Status'). The indicators used in the thematic assessments for eutrophication (HEAT), hazardous substances (CHASE) and biodiversity (BEAT) were integrated into a Holistic Assessment of 'ecosystem health'. The HOLAS tool presented assessment results for three groups: biological indicators, hazardous substances indicators and supporting indicators, and then applied the OOAO principle on the assessment results of those three groups for the final assessment (Figure 1).

This approach, which includes the selection of an agreed reduced set of indicators and agreed weighting rules, could be considered a pragmatic compromise, reducing the risks associated with OOAO while still giving an overall assessment.

An example of such a high level aggregation is the integrative method of Borja *et al.* (2010, 2011b), which includes a weighted scoring or rating method proposed for the MSFD in the southern Bay of Biscay. After aggregating the indicators within each descriptor, each descriptor was weighted according to the human pressure supported by the area. Then the value of each descriptor (i.e. an EQR) was multiplied by the weighting and added to obtain a final value between 0 and 1, being 0 the worst environmental status and 1 the best. This high-level integration was done at spatial and temporal scale. Although these authors combine values across descriptors, leading to a single value of environmental status, it could also be reported as "x out of 11 descriptors" having reached GEnS. In both cases, this allows to take management measures on those human activities impacting more in some of the descriptors or indicators not achieving good status, as shown in Borja *et al.* (2011b).

Halpern *et al.* (2012) developed another method, based more upon human activities and pressures, which presents a high-level integration at country level, using internationally available datasets (Ocean Health Index http://www.oceanhealthindex.org/). Similarly, Micheli *et al.* (2013) looked at cumulative impacts to the marine ecosystems of the Mediterranean and the Black Sea as a whole, while producing impact scores and maps for seven ecoregions and the territorial waters of EU Member states.

A Baltic Sea Health Index (BSHI) will be developed based on: (i) the existing HELCOM toolbox (HEAT, BEAT, CHASE and HOLAS), the MSFD (European Commission, 2008, 2010), and (ii) the Ocean Health Index (Halpern *et al.*, 2012).

Finally, there is a recent high-level integration example in Tett *et al*. (2013), for the North Sea, which includes five steps in the calculation: (i) identify (spatial extent) of ecosystem; (ii) identify spatial granularity and extent of repetitive temporal variability, and decide how to average or integrate over these; (iii) select state variables; (iv) plot trajectory in state space and calculate Euclidian (scalar) distance from (arbitrary) reference condition; and (v) calculate medium-term variability about trend in state space, and use this variability as proxy for (inverse) resilience.

## 4.- Considerations and recommendations when using specific rules

As shown in the previous section, the considerations to be used in combining values and assessing the environmental status are not easily defined. From the lessons learned above, some guidance can be offered:

(i) OOAO is appropriate when:
  - Legal criteria are involved, (e.g. contaminants exceeding legal quality standards, species or habitats failing favourable conservation status under Birds or Habitat Directives, commercial fish stocks failing Maximum Sustainable Yield targets under Common Fisheries Policy).
  - Different pressures are addressed (but in that case other methods can be also used).
  - There is an impact or risk on a future impact.
  - The precautionary principle is applied (e.g. in the case when little information from only a few indicators is available).

(ii) OOAO cannot be used:
  - In cases where indicators show a high level of uncertainty, when various indicators are sensitive to the same pressure, etc. In practice, the uncertainty associated with monitoring and assessment for each indicator/descriptor leads to problems of probable underestimation of the true overall class. Hence, if the error associated to the method used to assess the status of each indicator/descriptor is too high the OOAO approach is not advisable.
  - Note: Often, not all indicators are in the same state of development, or are scientifically sound and fully tested. In some cases P-S-I (Pressure-State-Impact) relations are uncertain. Also, sometimes multiple indicators are used to describe state. While not all of those indicators may be equally important or even comparable, this is done to include indicators that are used as supportive indicators, where P-S-I relations are uncertain. In those cases an aggregation rule such as OOAO should not be applied.

(iii) A ‘two out, all out’ approach can be considered in cases where several methods are combined in one assessment; e.g., when several matrices are used in pollutants to give a broader view of the status (e.g. pollutants in water for an instant picture, pollutants in sediments or biota for a time-integrated result, Tueros *et al.* (2009)).

(iv) Averaging is appropriate when combined variables or indicators are of equal importance or sensitive to the same pressure.

(v) Scoring or decision tree approaches are appropriate when:
- The methods to assess the status of the different indicators/descriptors are in different levels of development. In this case, consider giving more weight to those indicator/assessment methods which have been: (i) used broadly by authors other than the proposers of the method; (ii) tested for several different human pressures; and/or (iii) intercalibrated with other methods.
- It is important to be able to track the different steps involved in the assessment, making the path to the final assessment result transparent.
- Note: Consider different weights for individual indicators/descriptors taking into account the relationship with the pressures within the assessment (sub)region. E.g. if the area is under high fishing pressure the most affected descriptors will be D1, D3, D4, D6 and D11; in turn, D2, D5, D7, D8, D9 and D10 will be less affected.

(vi) Probabilistic approach:
- Consider carefully the uncertainties related to all of the various parts of the problem; be sure not to overestimate the well-known uncertainties (e.g. natural variance and sampling bias) and underestimate the poorly known uncertainties (e.g. insufficient knowledge or competing hypotheses about ecological interactions; combined effects of various pressures that may be strengthen or weaken each other, etc.).
- Consider using expert knowledge in evaluating the various uncertainties.
- If using expert judgement to weigh the different indicators in addition to the uncertainty estimate, make sure that the weighing is based on the relative importance of the indicators, not on the perceived uncertainty; otherwise you will end up double counting the effect of uncertainty in the final evaluation.

(vii) Multimetric and multivariate methods are appropriate when:
- Integrating several indicators of species composition or several indicators of eutrophication or seafloor integrity (e.g. in D1, D5, D6)
- It is advisable to verify that stakeholders and managers can understand the interpretation of the results, and results must be presented in a clear way.

(viii) For any of the described methods take into account that:
- Using as many ecosystem components/indicators/criteria as reasonable and available will make the analysis more robust.
- Integrate across state descriptors (D1, D3, D4, D6) differently than across pressure descriptors (D2, D5, D7, D8, D9, D10, D11), giving higher weight to state-based descriptors.

## 5.- Application of combination rules in assessments

As shown above, the WFD focuses on the structure of the ecosystem using a limited number of biodiversity components (the BQEs), that are combined through the precautionary OOAO approach (Borja *et al.*, 2010). In contrast, the MSFD can be considered to follow a 'holistic functional approach', as it takes into account not only structure (biodiversity components, habitats), but also function (e.g. food webs, seafloor integrity) and processes (e.g. biogeochemical cycles) of the marine ecosystems. The MSFD also uses descriptors that not only relate to biological and physicochemical state indicators but also to pressure indicators (Borja *et al.*, 2010, 2013). The MSFD requires the determination of GEnS on the basis of the qualitative descriptors in Annex I, but does not specifically require one single GEnS assessment, in contrast to the WFD.

There are many methodological challenges and uncertainties involved in establishing a holistic ecosystem assessment, when it is based on the large number of descriptors, associated criteria and indicators defined under the MSFD. The choice of indicator aggregation rules is essential, as the final outcome of the assessment may be very sensitive to those indicator aggregation rules (Ojaveer and Eero, 2011; Borja *et al.*, 2013; Caroni *et al.*, 2013). As shown in the previous section, different methodologies can be applied for aggregating indicators, which vary, amongst others, in the way the outliers influence the aggregate value.

When aggregating indicators most researchers agree that multiple accounting should be avoided. For example, phytoplankton indicators under D1 should be indicative of biodiversity state while under D5 it should be an estimator of the level of eutrophication. Similarly, macroinvertebrates under D1 should represent biodiversity state and under D6 also the state change from pressures on the seafloor. In these cases, although the datasets used could be the same, the main characteristics of the indicators to be used within each descriptor should be different, e.g. the value of macroinvertebrates indicators under D1 (rarity of species, endangered species, engineer species presence, etc.) and the condition of benthic community under D6 (ratio of opportunistic/sensitive, multimetric methods to assess the status, etc.). Of course, for aggregating indicators within the same criterion it is important that all indicators have the same level of maturity and that sufficient data are available.

There are at least four levels of combination required to move from evaluation of the individual metrics or indicators identified by the Task Groups to an assessment of GEnS (Cardoso *et al.*, 2010). As an example, using D6 (Seafloor integrity), Figure 2 shows: (i) aggregation of metrics/indices within indicators (see names of indicators in Table 1); (ii) aggregation of indicators within the criteria of a descriptor (for complex descriptors), e.g. criteria 6.1 (physical damage) and 6.2 (condition of benthic community); (iii) status across all the criteria of a descriptor; and (iv) integration of status across all descriptors.

As one moves up the scale from metric/indicator level to overall GEnS, the diversity of features that have to be combined increases rapidly (Figure 2). This poses several challenges arising from the diversity of metrics, scales, performance features (sensitivity, specificity, etc.) and inherent nature (state indicators, pressure indicators, impact indicators) of the metrics that must be integrated.

### *5.1. Aggregation of indicators and criteria (combination within a descriptor)*

Cardoso *et al.* (2010) summarize the methods for an integration within a MSFD descriptor, categorizing them into two wider categories: (i) integrative assessments combining indicators and/or attributes appropriate to local conditions; and (ii) assessment by worst case (in this context, ‘worst case’ means that GEnS will be set at the environmental status of the indicator and/or attribute assessed at the worst state for the area of concern).

Table 3 summarizes the approaches to aggregate attributes within each descriptor. In some cases the MSFD Task Groups propose deconstructing the ecosystem into ‘descriptor indicators’ and then recombining them again to give a pass/fail for the GEnS, using (in four cases) the OOAO principle (Table 3). Borja *et al.* (2013) emphasize that such a ‘deconstructive structural approach’ makes large assumptions about the functioning of the system and does not consider the weighting of the different indicators and descriptors. It implies that recombining a set of structural attributes gives an accurate representation of the ecosystem functioning.

An example of this accurate representation is shown by Tett *et al*. (2013), who assess the ecosystem health of the North Sea, using different attributes and components of the ecosystem. These components include structure or organization, vigour, resilience, hierarchy and trajectory in state space. All the information from the different components are combined and synthesized for a holistic approach to assess the ecosystem health.

Other approaches have been used in aggregating indicators within each descriptor. For example, Borja *et al.* (2011b) use the biodiversity valuation approach, in assessing biodiversity within the MSFD, integrating several biodiversity components (zooplankton, macroalgae, macroinvertebrates, fishes, cetaceans and seabirds). Biodiversity valuation maps aim at the compilation of all available biological and ecological information for a selected study area and allocate an integrated intrinsic biological value to the subzones (Derous *et al.*, 2007). Details on valuation methodology can be consulted in Pascual *et al*. (2011) (see Figure 4 in that paper). This methodology provides information for each of the components and their integrative valuation, together with the reliability of the result, taking into account spatial and temporal data availability (Derous *et al.*, 2007). The advantage of this method is that the current information used to valuate biodiversity can be adapted to the requirements of the MSFD indicators. Moreover, this method can avoid duplication of indicators in two descriptors (e.g. D1 and D6), since the metrics used could be different. This information can be converted into environmental status values, as shown in Borja *et al*. (2011b).

*5.2. Integration of descriptors (combination across descriptors)*

Discussion on how to integrate the results of each descriptor into an overall assessment of GEnS for regions or subregions was not part of the Terms of Reference for the Task Groups. However, work within Task Group 6 (Sea floor integrity) identified a method for integration and assessment that might also be appropriate, if applied across all descriptors, at a regional scale (Cardoso *et al.*, 2010). As these authors pointed out, cross-descriptor integration at the scale of (sub)regional seas runs the risk of blending and obscuring the information that is necessary to follow progress towards GEnS and to inform decision-makers about the effects and the efficiency of policies and management. It may lead to masking of problems within specific descriptors.

Borja *et al.* (2013) describe at least 8 options to determine GEnS in a regional sea context (Table 4). These authors detail the concept behind these options, and propose the decision rule more adequate for the assessment method to be used, depending on the circumstances i.e. data availability, lack of monitoring, etc. In addition, these authors consider what type and amount of data are required, and then discuss the pros and cons of the different options. The implementation of a complex directive, such as the MSFD, requires a high amount of data to assess the environmental status in a robust way. Hence, the options from 1 to 8 proposed in Table 4 are sequentially less demanding of new data, and the degree of detailed environmental assessment is also decreasing.

As such, Option 1, which is most similar to the WFD approach, deconstructs GEnS into the 11 descriptors and then into the component indicators, assessing each components for each area before attempting to produce an overall assessment (Table 4). However, having a complete dataset covering all descriptors and indicators for the assessment is difficult, if not impossible to achieve in practical terms. The use of pressure maps as an estimator of the environmental status and possible impacts to marine ecosystems could be considered instead (see Table 4). This would, however, build on the substantial assumption that the level of pressure is adequately representing the current state on all different levels of ecosystem components. Option 7, in contrast, only uses published data for the activities, and then infers a static relationship between activity, pressures, state changes and impacts both on the natural and the human system. Here, the number of underlying assumptions is even larger than using pressure maps, since the method relies on predefined and static DPSIR relations. Between these extremes, there are several intermediate options to integrate and present information, each with its own requirements, pros and cons (Table 4).

### 5.2.1. One-out, all-out (OOAO)

Although the MSFD describes the GEnS individually for each of the 11 descriptors, this does not necessarily imply the ability to have GEnS at the level of all the descriptors, nor does it mean that each descriptor should necessarily be graded individually in a binary way (i.e. good or not good environmental status) (Borja *et al*., 2013).

It could be argued that the 11 descriptors together summarize the way in which the ecosystem functions in terms of the MSFD view. As Member States have to consider each of the descriptors to determine good environmental status, this could be interpreted as a requirement to achieve GEnS for each of these descriptors. In that case, applying OOAO is the only integration method that can be applied to arrive at an overall assessment of GEnS, leading to a high probability of not achieving GEnS.

This assumes that the 11 descriptors, and the associated indicators, can be considered a coherent and consistent framework that adequately reflects the environmental status. In that situation, state descriptors not achieving GEnS would be accompanied by pressure descriptors not achieving GEnS, if the reaction of the ecosystem components is immediate, acting on the same time scale as the pressures. If this is not the case, for example if a pressure descriptor (e.g. D5 or D8) indicates that the level of the pressure is too high to achieve GEnS, while state descriptors (e.g. D1 or D4) do not reflect this, there is clearly an inconsistency in the assumed MSFD assessment framework, indicating that it does not capture delayed responses of state indicators to changing pressure indicators. That could be interpreted as a need for further research on the nature of P-S-I relations and the consistency in environmental targets for the descriptors involved, since our current state of knowledge on quantitative causal relations between pressures, state changes and impacts is limited. In addition, nearly all ecosystem components are subject to the true cumulative effects of many simultaneous pressures related to a range of human activities (Crain *et al.*, 2008; Stelzenmüller *et al.*, 2010; Knights *et al.*, 2013). This means that, for some descriptors at least, there is a large scientific uncertainty associated with the definition of environmental targets and GEnS. Uncertainties in target setting, in the performance of an action (e.g. ecosystem state post-management) or in the contribution of individual driver(s) causing state change can undermine decision making when implementing environmental policy and can limit our ability to identify what should be managed, and what the impact of management might be (Knights *et al.*, 2014). Consequently, developing a consistent assessment framework for all descriptors and indicators is an extremely challenging task, and using the OOAO approach is not appropriate.

### 5.2.2. Alternative approaches

The usefulness of integrating descriptors to one single value (overall GEnS assessment based on combination of the 11 descriptors) is under discussion by the Member States and the European Commission groups for the implementation of the MSFD. An argument against integration across descriptors is that it may not be informative any more since it results in loss of information at a crucial level where different elements are combined that cannot be integrated without major concessions.

The abovementioned groups have suggested that an integration across the biodiversity-related descriptors (D1, D2, D4, D6) might be an option, splitting those descriptors into various groups (e.g. functional or species groups). If a species or species group is assessed under more than one descriptor different aspects should be considered (e.g. chlorophyll a under D5 and phytoplankton species composition under D1).

However, if an integration across all descriptors is decided, Borja *et al*. (2010) suggest that the 11 descriptors are hierarchical and do not have an equal weighting when assessing the overall GEnS. Hence, Borja *et al*. (2013) suggest that for biodiversity (D1) to be fulfilled requires all others to be met and similarly if one of the stressor or pressure-related descriptors (e.g. D11, energy including noise) fails then by definition the biodiversity will be adversely affected at some point. This approach addresses the conceptual drawback of the OOAO principle and allows to have delayed responses to changing pressure regimes without drawing false conclusions and still being precautionary.

In addition to the problem of combining indicators (seen in the previous section) and descriptors the MSFD requires Member States to integrate and geographically scale-up the assessments at the level of a region or subregion (Borja *et al.*, 2010). This differs strongly from the approach under the WFD, which is restricted to quality assessments at the scale of a water body (Hering *et al.*, 2010). This means that the GEnS assessments of the different Member States within a regional sea need to be comparable and should avoid anomalies at the borders of Member States in order to enable synthesising of the assessments into a region-wide assessment (Borja *et al*., 2013). This requires both comparable methods and associated combination rules to ensure minimum standards for GEnS reporting across Member States. As such, we advocate a set of common principles (expanded from Claussen *et al.,* 2011, as shown in Borja *et al*., 2013):

- The combination across levels of different complexity should accommodate different alternatives, i.e., aggregation below descriptor level (across indicators within criteria, and criteria within descriptors, as shown in the previous section) and can certainly differ from descriptor level integration;

- Integration across state descriptors (D1, D3, D4, D6) should be done differently than across pressure descriptors (D2, D5, D7, D8, D9, D10, D11), but avoiding double counting of indicators in different descriptors (e.g. phytoplankton under D1 and D5, macroinvertebrates under D1 and D6).

- Consideration of a different contribution of the two types of descriptors for the overall GEnS evaluation – giving state descriptors a higher weight, as receptors of the impacts caused by pressures. The rationale for this, as recognized by Claussen *et al.* (2011), is that "in principle, where GEnS for state-based descriptors (D1, 3, 4, 6) is achieved it follows that GEnS for pressure-based descriptors should also be met". This principle makes the assumption that the state eventually will reflect ceasing pressures. When the state descriptors finally reach a satisfactory level then the pressures must be having a limited (or mitigated) impact.

5.2.3. Visualizing and communicating the status

The outlined alternative approach also shows that concerns on integration across descriptors do not necessarily have to be a problem. There are some methods which have demonstrated that integrating the information into single values (Borja *et al*., 2011b), maps (HELCOM, 2010) or radar schemes (Halpern *et al.*, 2012) is still helpful and informative for ecosystem management, despite the involved loss of information that is inherent to a single number. Information can be retained when always presenting that single number together with the main underlying data, ideally visualizing the different levels of aggregation, allowing the lookup of the status at any level and relating the status with the actual pressures that lead to the synthesized value.

As an example, the Ocean Health Index (Halpern *et al*., 2012) provides weighted index scores for environmental health, both a global area-weighted average and scores by country (Figure 3). The outer ring of the radar scheme is the maximum possible score for each goal, and a goal's score and weight (relative contribution) are represented by the petal's length and width, respectively. This way of visualizing the integration could be adapted for the MSFD, integrating at the level of region or subregion, but also showing the values within each descriptor. This would still allow managers to extract relevant information and take actions at different levels: small (or local) scale, large (regional) scale, integrative (whole ecosystem status), or for each descriptor.

Another example, applied specifically for the MSFD, using all descriptors and most of the indicators, can be consulted in Borja *et al*. (2011b). These authors studied a system in which the main driver for the whole area is fishing, whilst at local level some pressures such as waste discharges are important. Although the overall environmental status of the area was considered good, after the integration of all indicators and descriptors, two of the descriptors (fishing and food webs) were not in good status

(Table 5). Interestingly, biodiversity was close to the boundary to good status (Table 5), suggesting that the system could be unbalanced by fishing, but affecting various biological descriptors to different degrees. This means that the pressure must be managed to avoid problems in the future, especially because the descriptors already in less than good status showed a negative trend (Table 5).

Hence, from the examples above and the given reasoning, both main choices are still useful: either integrate or not integrate information across descriptors. Irrespectively of which combination proposal(s) is adopted and at which level, the precautionary principle should always be followed in absence of more robust knowledge (Borja et al., 2013).As a summary, the pros and cons of each decision are shown in Table 6.

## 6.- Proposed steps for combination

As a possible approach for the combination of assessments we propose the following steps (Figure 4):

- Assessments start at a low level, viz. the level of indicators and spatial scales that were defined for each specific indicator. This would result in assessment results for each indicator and each assessment area incorporating the levels of spatial assessment that was described as a nested approach (Step 1 - spatial scales).

- Within one descriptor, this could result in a number of assessments for the different indicators, that all use the same scales for their assessment areas. This could be the case for descriptors like D5 and D8. In those cases, the assessments at indicator level can be aggregated to assessments at descriptor level for each assessment area, using suitable aggregation rules (Step 2 - aggregation within a descriptor). These steps are already commonly used procedures in OSPAR and HELCOM assessments for eutrophication and contaminants.

- For other descriptors, the spatial scales for indicators may not be the same for all indicators. This could be the case for biodiversity, where a different spatial scale may be used depending on the species or habitat. Although integration of different biodiversity components and functional groups is required, methods need further development, and a number of EU projects are focussing on this issue.

Aggregation up to this level gives a detailed assessment result that suits the information needs for identifying environmental problems and needs for measures. The result of those steps at European level would be a very high number of assessment results, for each descriptor and assessment area (comparable to presenting the WFD assessments at water body level).

The following steps could provide information at a higher level of integration presenting the required overview of the current status of the overall environmental state and the progress towards GEnS:

- Within a descriptor, the assessment results of all assessment areas within a subregion can be presented in a more integrated way (Step 3 - spatial aggregation).

  - Generally, use of OOAO (if one assessment area fails GEnS, the whole subregion fails) is not useful, as it gives a very conservative result and is not informative. Also, if the pressure is highly localized this approach is not adequate, since the whole subregion could fail GEnS due to a single location (which, of course, will need specific management measures).
  - In some cases, for example if a pressure is more or less homogeneous across a whole subregion (fishing, shipping), it could be useful to apply OOAO

- Percentage of surface area achieving GEnS: This could be a more useful approach, if the extent and intensity of a pressure can be quantified. For example, if the pressure is present in 45% of the surface area of a subregion, but the surface area not achieving GEnS is only 2%, it could be concluded that the subregion does not achieve GEnS in 2% of its area, where management measures are needed.

- Other metrics

For some descriptors, surface area may be a good measure to express status at a subregional level: for example, D5, D8, and D10. For other descriptors, surface area is not suitable but other metrics should be considered, e.g. D1: numbers of species/habitats failing to achieve favourable conservation status; D3: number of stocks failing to meet "Maximum Sustainable Yield".

The end result of Step 3 could present the level at which GEnS is achieved at subregional scale as a pie chart. The aggregation results of Step 3 could be integrated across descriptors in a final presentation per subregion, using methods such as radar plots, or methods similar to the Ocean Health Index (Step 4 - aggregation across descriptors). In this step, weighted approaches as suggested in previous sections would be considered.

## 7.- Concluding remarks

From the information provided in this overview, some conclusions can be highlighted:

- Some kind of integration across indicators, criteria and descriptors is required to arrive at assessment of GEnS or 'ecosystem health'.
- Integration principles should be ecologically-relevant, transparent and documented.
- Integrated assessment should not only present a classification result (primary assessment) but also address uncertainties and assess confidence of the classification result (as a secondary assessment). When carrying out an assessment at a specific scale, the decisions made in regard to integration principles/rules should be available as a sort of third assessment or backlog.
- Assessments should be planned around the question(s) to be addressed and the tool(s) to be used. Monitoring should subsequently be designed to meet the requirements of the planned assessments.
- This study provides information on combining methods to integrate ecosystem components to assess status and guidelines for scientists and managers on the steps

to be followed, when deciding on assessment scales and combination approaches. Integration of taxonomic, functional and key or keystone biodiversity components into an overall biodiversity assessment able to link to GEnS and to ecosystem service provision and the sustainable management of detrimental human activities is the next challenge.


## Acknowledgements

This work has been done under Framework contract No ENV.D2/FRA/2012/0019 (Coherent geographic scales and aggregation rules in assessment and monitoring of Good Environmental Status – analysis and conceptual phase), of the European Directorate General of Environment; and DEVOTES project (DEVelopment Of innovative Tools for understanding marine biodiversity and assessing good Environmental Status) funded by the European Union under the 7th Framework Program 'The Ocean of Tomorrow' Theme (grant agreement no. 308392) (www.devotes-project.eu). JHA was supported by the WATERS project (Waterbody Assessment Tools for Ecological Reference conditions and status in Sweden). María C. Uyarra (AZTI-Tecnalia) and Mike Elliott (University of Hull) provided constructive comments to the first version of the manuscript. This is contribution number XXX from the Marine Research Division (AZTI-Tecnalia).


## References


Aubry, A., M. Elliott, 2006. The use of environmental integrative indicators to assess seabed disturbance in estuaries and coasts: Application to the Humber Estuary, UK. *Marine Pollution Bulletin*, **53**: 175-185.

Barton, D. N., T. Saloranta, S. J. Moe, H. O. Eggestad, S. Kuikka, 2008. Bayesian belief networks as a meta-modelling tool in integrated river basin management - Pros and cons in evaluating nutrient abatement decisions under uncertainty in a Norwegian river basin. *Ecological Economics*, **66**: 91-104.

Barton, D. N., S. Kuikka, O. Varis, L. Uusitalo, H. J. Henriksen, M. Borsuk, A. de la Hera, R. Farmani, S. Johnson, J. D. Linnell, 2012. Bayesian networks in environmental and resource management. *Integrated environmental assessment and management*, **8**: 418-429.

Birk, S., W. Bonne, A. Borja, S. Brucet, A. Courrat, S. Poikane, A. Solimini, W. van de Bund, N. Zampoukas, D. Hering, 2012. Three hundred ways to assess Europe's surface waters: An almost complete overview of biological methods to implement the Water Framework Directive. *Ecological Indicators*, **18**: 31-41.

Borja, A., 2014. Grand Challenges in Marine Ecosystems Ecology. *Frontiers in Marine Science*, **1**: 1-6 DOI: 10.3389/fmars.2014.00001.

Borja, Á., J. G. Rodríguez, 2010. Problems associated with the 'one-out, all-out' principle, when using multiple ecosystem components in assessing the ecological status of marine waters. *Marine Pollution Bulletin*, **60**: 1143-1146.

Borja, A., V. Valencia, J. Franco, I. Muxika, J. Bald, M. J. Belzunce, O. Solaun, 2004. The water framework directive: water alone, or in association with sediment and biota, in determining quality standards? *Marine Pollution Bulletin*, **49**: 8-11.

Borja, A., S. B. Bricker, D. M. Dauer, N. T. Demetriades, J. G. Ferreira, A. T. Forbes, P. Hutchings, X. Jia, R. Kenchington, J. C. Marques, C. Zhu, 2008. Overview of integrative tools and methods in assessing ecological integrity in estuarine and coastal systems worldwide. Marine Pollution Bulletin, 56: 1519-1537.

Borja, A., A. Ranasinghe, S. B. Weisberg, 2009a. Assessing ecological integrity in marine waters, using multiple indices and ecosystem components: Challenges for the future. *Marine Pollution Bulletin,* **59**: 1-4.

Borja, A., J. Bald, J. Franco, J. Larreta, I. Muxika, M. Revilla, J. G. Rodríguez, O. Solaun, A. Uriarte, V. Valencia, 2009b. Using multiple ecosystem components, in assessing ecological status in Spanish (Basque Country) Atlantic marine waters. *Marine Pollution Bulletin,* **59**: 54-64.

Borja, A., M. Elliott, J. Carstensen, A. S. Heiskanen, W. van de Bund, 2010. Marine management--towards an integrated implementation of the European Marine Strategy Framework and the Water Framework Directives. *Marine Pollution Bulletin,* **60**: 2175-2186.

Borja, A., E. Barbone, A. Basset, G. Borgersen, M. Brkljacic, M. Elliott, J. M. Garmendia, J. C. Marques, K. Mazik, I. Muxika, J. M. Neto, K. Norling, J. G. Rodríguez, I. Rosati, B. Rygg, H. Teixeira, A. Trayanova, 2011a. Response of single benthic metrics and multi-metric methods to anthropogenic pressure gradients, in five distinct European coastal and transitional ecosystems. *Marine Pollution Bulletin,* **62**: 499-513.

Borja, A., I. Galparsoro, X. Irigoien, A. Iriondo, I. Menchaca, I. Muxika, M. Pascual, I. Quincoces, M. Revilla, J. German Rodriguez, M. Santurtun, O. Solaun, A. Uriarte, V. Valencia, I. Zorita, 2011b. Implementation of the European Marine Strategy Framework Directive: a methodological approach for the assessment of environmental status, from the Basque Country (Bay of Biscay). *Marine Pollution Bulletin,* **62**: 889-904.

Borja, A., M. Elliott, J. H. Andersen, A. C. Cardoso, J. Carstensen, J. G. Ferreira, A.-S. Heiskanen, J. C. Marques, J. M. Neto, H. Teixeira, L. Uusitalo, M. C. Uyarra, N. Zampoukas, 2013. Good Environmental Status of marine ecosystems: What is it and how do we know when we have attained it? *Marine Pollution Bulletin,* **76**: 16-27.

Breen, P., Robinson, L.A., Rogers, S.I., Knights, A.M., Piet, G., Churilova, T., Margonski, P., Papadopoulou, N., Akoglu, E., Eriksson, A., Fineko, Z., Fleming-Lehtinen, V., Galil, B., Goodsir, F., Goren, M., Kryvenko, O., Leppanen, J.M., Markantonatou, V., Moncheva, S., Oguz, T., Paltriguera, L., Stefanova, K., Timofte, F., and Thomsen, F. 2012. An environmental assessment of risk in achieving good environmental status to support regional prioritisation of management in Europe. Marine Policy 36(5):1033-1043.

Bricker, S. B., J. G. Ferreira, T. Simas, 2003. An integrated methodology for assessment of estuarine trophic status. Ecological Modelling, 169: 39-60.

Cardoso, A. C., S. Cochrane, H. Doemer, J. G. Ferreira, F. Galgani, C. Hagebro, G. Hanke, N. Hoepffner, P. D. Keizer, R. Law, S. Olenin, G. J. Piet, J. Rice, S. I. Rogers, F. Swartenbroux, M. L. Tasker, W. van de Bund, 2010. Scientific support to the European Commission on the Marine Strategy Framework Directive. Management Group Report. *EUR 24336 EN – Joint Research Centre, Luxembourg: Office for Official Publications of the European Communities*: 57 pp.

Caroni, R., W. Bund, R. T. Clarke, R. K. Johnson, 2013. Combination of multiple biological quality elements into waterbody assessment of surface waters. *Hydrobiologia*, **704**: 437-451.

CIS, 2003. Overall Approach to the Classification of Ecological Status and Ecological Potential. *Water Framework Directive Common Implementation Strategy Working Group 2 A Ecological Status (ECOSTAT), Rome,* 53 pp.

Claussen, U., D. Connor, L. de Vrees, J. M. Leppänen, J. Percelay, M. Kapari, O. Mihail, G. Ejdung, J. Rendell, 2011. Common Understanding of (Initial) Assessment, Determination of Good Environmental Status (GES) & Establishment of Environmental Targets (Articles 8, 9 & 10 MSFD). European Commission: 73 p.

Costanza, R., M. Mageau, 1999. What is a healthy ecosystem? Aquatic Ecology, 33: 105-115.

Crain, C. M., K. Kroeker, B. S. Halpern, 2008. Interactive and cumulative effects of multiple human stressors in marine systems. Ecology Letters, 11: 1304-1315.

Derous, S., T. Agardy, H. Hillewaert, K. Hostens, G. Jamieson, L. Lieberknecht, J. Mees, I. Moulaert, S. Olenin, D. Paelinckx, M. Rabaut, E. Rachor, J. Roff, E. W. M. Stienen, J. T. van der Wal, V. van Lancker, E. Verfaillie, M. Vincx, J. M. Weslawski, S. Degraer, 2007. A concept for biological valuation in the marine environment. *Oceanologia*, **49**: 99-128.

European Commission, 2008. Directive 2008/56/EC of the European Parliament and of the Council establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive). Official Journal of the European Union, L164: 19-40.

European Commission, 2010. Commission Decision of 1 September 2010 on criteria and methodological standards on good environmental status of marine waters (notified under document C(2010) 5956)(2010/477/EU). *Official Journal of the European Union*, **L232**: 12-24.

Feary, D.A., A.M. Fowler, T.J. Ward, 2014. Developing a rapid method for undertaking the World Ocean Assessment in data-poor regions – A case study using the South China Sea Large Marine Ecosystem. *Ocean & Coastal Management*, 95: 129-137.

Ferreira, J. G., J. H. Andersen, A. Borja, S. B. Bricker, J. Camp, M. Cardoso da Silva, E. Garcés, A.-S. Heiskanen, C. Humborg, L. Ignatiades, C. Lancelot, A. Menesguen, P. Tett, N. Hoepffner, U. Claussen, 2011. Overview of eutrophication indicators to assess environmental status within the European Marine Strategy Framework Directive. Estuarine, Coastal and Shelf Science, 93: 117-131.

Halpern, B. S., S. Walbridge, K. A. Selkoe, C. V. Kappel, F. Micheli, C. D'Agrosa, J. F. Bruno, K. S. Casey, C. Ebert, H. E. Fox, R. Fujita, D. Heinemann, H. S. Lenihan, E. M. P. Madin, M. T. Perry, E. R. Selig, M. Spalding, R. Steneck, R. Watson, 2008. A Global Map of Human Impact on Marine Ecosystems. Science, 319: 948-952.

Halpern, B. S., C. Longo, D. Hardy, K. L. McLeod, J. F. Samhouri, S. K. Katona, K. Kleisner, S. E. Lester, J. O'Leary, M. Ranelletti, A. A. Rosenberg, C. Scarborough, E. R. Selig, B. D. Best, D. R. Brumbaugh, F. S. Chapin, L. B. Crowder, K. L. Daly, S. C. Doney, C. Elfes, M. J. Fogarty, S. D. Gaines, K. I. Jacobsen, L. B. Karrer, H. M. Leslie, E. Neeley, D. Pauly, S. Polasky, B. Ris, K. St Martin, G. S. Stone, U. R. Sumaila, D. Zeller, 2012. An index to assess the health and benefits of the global ocean. *Nature*, **488**: 615-620.

HELCOM, 2009a. Eutrophication in the Baltic Sea – An integrated thematic assessment of the effects of nutrient enrichment and eutrophication in the Baltic Sea region. *Baltic Sea Environmental Proceedings*, **115B**: 152 pp.

HELCOM, 2009b. Biodiversity in the Baltic Sea – An integrated thematic assessment on biodiversity and nature conservation in the Baltic Sea. *Baltic Sea Environmental Proceedings*, **116B**: 192 pp.

HELCOM, 2010. Ecosystem Health of the Baltic Sea 2003–2007: HELCOM Initial Holistic Assessment. *Baltic Sea Environmental Proceedings*, **122**: 68 pp.

HELCOM, 2013. Draft HELCOM Monitoring and assessment strategy. 30 pp. http://www.helcom.fi/Documents/Action%20areas/Monitoring%20and%20assessment/Monitoring%20and%20assessment%20strategy/Monitoring%20and%20assessment%20strategy.pdf

Hering, D., A. Borja, J. Carstensen, L. Carvalho, M. Elliott, C. K. Feld, A. S. Heiskanen, R. K. Johnson, J. Moe, D. Pont, A. L. Solheim, W. de Bund, 2010. The European Water Framework Directive at the age of 10: a critical review of the achievements with recommendations for the future. *The Science of the Total Environment*, **408**: 4007-4019.

Knights, A. M., R. S. Koss, L. A. Robinson, 2013. Identifying common pressure pathways from a complex network of human activities to support ecosystem-based management. *Ecological Applications*, **23**: 755-765.

Knights, A.M., F. Culhane, S.S. Hussain, K.N. Papadopoulou, G.J. Piet, J. Raakær, S.I. Rogers, L.A. Robinson,. 2014. A step-wise process of decision-making under uncertainty when implementing environmental policy. Environmental Science & Policy 39, 56-64.

Korpinen, S., L. Meski, J. H. Andersen, M. Laamanen, 2012. Human pressures and their potential impact on the Baltic Sea ecosystem. Ecological Indicators, 15: 105-114.

Lehikoinen, A., I. Helle, E. Klemola, S. Mäntyniemi, S. Kuikka, H. Pitkänen, 2013a. Evaluating the impact of nutrient abatement measures on the ecological status of coastal waters: a Bayesian network for decision analysis. *International Journal of Multicriteria Decision Making*, **In press.**

Lehikoinen, A., E. Luoma, S. Mantyniemi, S. Kuikka, 2013b. Optimizing the recovery efficiency of Finnish oil combating vessels in the Gulf of Finland using Bayesian Networks. *Environmental Science & Technology*, **47**: 1792-1799.

Mee, L. D., R. L. Jefferson, D. d. A. Laffoley, M. Elliott, 2008. How good is good? Human values and Europe's proposed Marine Strategy Directive. Marine Pollution Bulletin, 56: 187-204.

Micheli F, B.S. Halpern, S. Walbridge, S. Ciriaco F. Ferretti, S. Fraschetti, R. Lewison, L. Nykjaer, A. A. Rosenberg (2013) Cumulative Human Impacts on Mediterranean and Black Sea Marine Ecosystems: Assessing Current Pressures and Opportunities. PLoS ONE 8(12): e79889. doi:10.1371/journal.pone.0079889

Murtaugh, P.A., 1996. The statistical evaluation of ecological indicators. Ecological Applications 6(1): 132-139.

Ojaveer, H., M. Eero, 2011. Methodological challenges in assessing the environmental status of a marine ecosystem: case study of the Baltic Sea. *PLoS One*, **6**: e19231.

OSPAR, 2009. Eutrophication Status of the OSPAR Maritime Area. Second OSPAR Integrated Report. Eutrophication Series. OSPAR Commission, London, 107 pp. http://qsr2010.ospar.org/media/assessments/p00372_Second_integrated_report.pdf

OSPAR, 2010. OSPAR Quality Status Report 2010. OSPAR Commission.

OSPAR, 2011. Starting point for developing guidance on geographic scales. ICG-MSFD(4)11/3/4-E, 11 pp.

OSPAR, 2012. MSFD Advice Manual and Background Document on Biodiversity. London, Publication Number: 581/2012, 141 pp.

Pascual, M., A. Borja, S. V. Eede, K. Deneudt, M. Vincx, I. Galparsoro, I. Legorburu,

2011. Marine biological valuation mapping of the Basque continental shelf (Bay of Biscay), within the context of marine spatial planning. *Estuarine, Coastal and Shelf Science*, **95**: 186-198.

Primpas, I., M. Karydis, 2011. Scaling the trophic index (TRIX) in oligotrophic marine environments. *Environmental Monitoring and Assessment*, **178**: 257-269.

Prins, T., M. van der Meulen, M., A. Boon, N. Simboura, C. Tsangaris, A. Borja., I. Menchaca, 2013. "Coherent geographic scales and aggregation rules in assessment and monitoring of Good Environmental Status" Analysis and conceptual phase" Deltares – The Netherlands November 2013 Analytical report under Framework contract No NV.D2/FRA/2012/0019. http://ec.europa.eu/environment/marine/publications/

Purvis, A., A. Hector, 2000. Getting the measure of biodiversity. Nature, 405: 212-219.

Rice, J., C. Arvanitidis, A. Borja, C. Frid, J. Hiddink, J. Krause, P. Lorance, S. A. Ragnarsson, M. Skold, B. Trabucco, 2010. Marine Strategy Framework Directive – Task Group 6 Report Seafloor integrity. *EUR 24334 EN – Joint Research Centre, Luxembourg: Office for Official Publications of the European Communities*: 73 pp.

Shin, Y.-J., A. Bundy, L. J. Shannon, J. L. Blanchard, R. Chuenpagdee, M. Coll, B. Knight, C. Lynam, G. Piet, A. J. Richardson, 2012. Global in scope and regionally rich: an IndiSeas workshop helps shape the future of marine ecosystem indicators. *Reviews in Fish Biology and Fisheries*, **22**: 835-845.

Simboura, N., A. Zenetos, M. A. Pancucci-Papadopoulou, S. Reizopoulou, N. Streftaris, 2012. Indicators for the Sea-floor Integrity of the Hellenic Seas under the European Marine Strategy Framework Directive: establishing the thresholds and standards for Good Environmental Status. Mediterranean Marine Science, 13: 140-152.

Solheim, A. L., K. Austnes, P. Kristensen, M. Peterlin, V. Kodes, R. Collins, S. Semeradova, A. Künitzer, R. Filippi, H. Prchalová, C. Spiteri, T. Prins, 2012. Ecological and chemical status and pressures in European waters. Thematic assessment for EEA Water 2012 Report. ETC/ICM Technical Report, 1: 146 pp.

Stelzenmüller, V., J. Lee, A. South, S. I. Rogers, 2010. Quantifying cumulative impacts of human pressures on the marine environment: a geospatial modelling framework. Marine Ecology Progress Series, 398: 19-32.

Tett, P., R. J. Gowen, S. J. Painting, M. Elliott, R. Forster, D. K. Mills, E. Bresnan, E. Capuzzo, T. F. Fernandes, J. Foden, R. J. Geider, L. C. Gilpin, M. Huxham, A. L. McQuatters-Gollop, S. J. Malcolm, S. Saux-Picart, T. Platt, M. F. Racault, S. Sathyendranath, J. van der Molen, M. Wilkinson, 2013. Framework for understanding marine ecosystem health. *Marine Ecology Progress Series*, **494**: 1-27.

Tueros, I., Á. Borja, J. Larreta, J. G. Rodríguez, V. Valencia, E. Millán, 2009. Integrating long-term water and sediment pollution data, in assessing chemical status within the European Water Framework Directive. Marine Pollution Bulletin, 58: 1389-1400.

Ulanowicz, R. E., 2000. Toward the Measurement of Ecological Integrity. In: Pimentel, D., L. Westra, R. F. Noss (eds) Ecological Integrity: Integrating Environment, Conservation, and Health. Island Press, Washington DC, 99-113.

## Figure captions

Figure 1. (a) Example of an integrated assessment of ecosystem health in the Baltic Sea 2003-2007 based on the HOLAS tool. (b) Screenshot to illustrate how the HOLAS classification tool for the Gulf of Finland works. See HELCOM (2010) for details. Courtesy by Helsinki Commission.

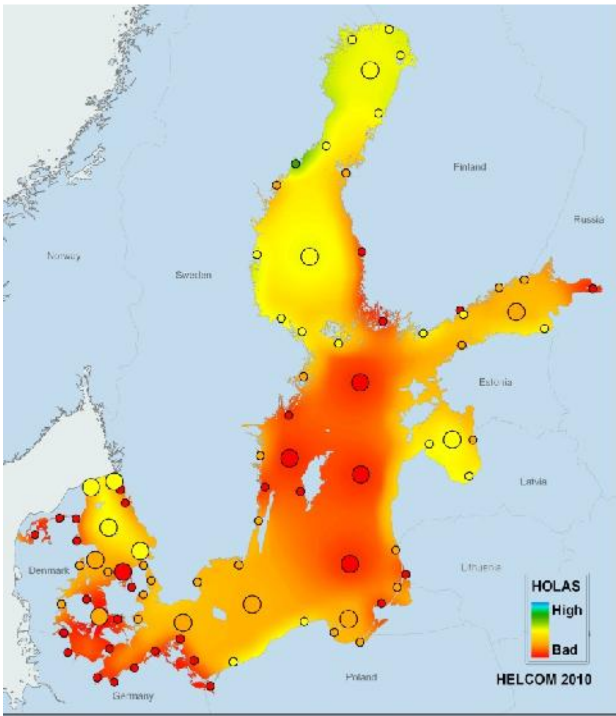
Figure 2. Diagram of a possible approach for aggregation of indicators and criteria and integration of descriptors (D), using D6 as an example. For indicators and criteria description, see Table 1.

Figure 3. Ocean Health Index scores (inside circle) and individual goal scores (coloured petals) for global area-weighted average of all studied countries (modified from Halpern *et al.*, 2012).

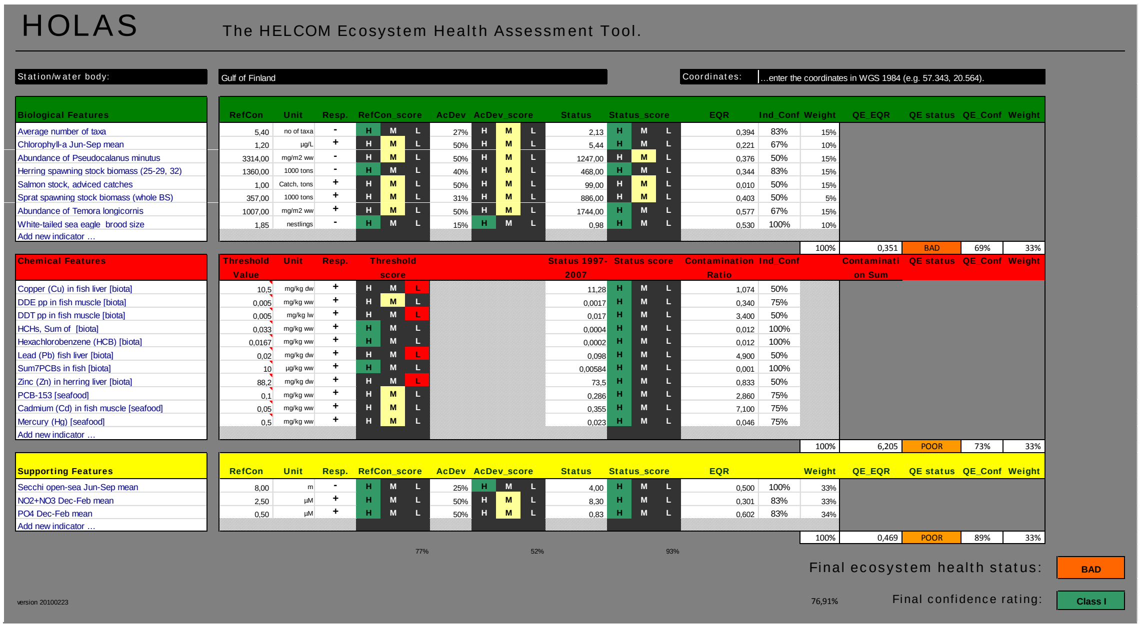
Figure 4. Schematic view of steps for combination towards an assessment at subregional level. GEnS: Good Environmental Status.

Fig 1

(a)




(b)



HOLAS — The HELCOM Ecosystem Health Assessment Tool.

Station/water body: Gulf of Finland — Coordinates: ...enter the coordinates in WGS 1984 (e.g. 57.343, 20.564).

| Biological Features | RefCon | Unit | Resp. | RefCon_score | AcDev | AcDev_score | Status | Status_score | EQR | Ind_Conf | Weight | QE_EQR | QE status | QE_Conf | Weight |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Average number of taxa | 5,40 | no of taxa | - | H M L | 27% | H M L | 2,13 | H M L | 0,394 | 83% | 15% | | | | |
| Chlorophyll-a Jun-Sep mean | 1,20 | µg/L | + | H M L | 50% | H M L | 5,44 | H M L | 0,221 | 67% | 10% | | | | |
| Abundance of Pseudocalanus minutus | 3314,00 | mg/m2 ww | - | H M L | 50% | H M L | 1247,00 | H M L | 0,376 | 50% | 15% | | | | |
| Herring spawning stock biomass (25-29, 32) | 1360,00 | 1000 tons | - | H M L | 40% | H M L | 468,00 | H M L | 0,344 | 83% | 15% | | | | |
| Salmon stock, adviced catches | 1,00 | Catch, tons | + | H M L | 50% | H M L | 99,00 | H M L | 0,010 | 50% | 15% | | | | |
| Sprat spawning stock biomass (whole BS) | 357,00 | 1000 tons | + | H M L | 31% | H M L | 886,00 | H M L | 0,403 | 50% | 5% | | | | |
| Abundance of Temora longicornis | 1007,00 | mg/m2 ww | + | H M L | 50% | H M L | 1744,00 | H M L | 0,577 | 67% | 15% | | | | |
| White-tailed sea eagle brood size | 1,85 | nestlings | - | H M L | 15% | H M L | 0,98 | H M L | 0,530 | 100% | 10% | | | | |
| Add new indicator ... | | | | | | | | | | | | | | | |
| | | | | | | | | | | | 100% | 0,351 | BAD | 69% | 33% |

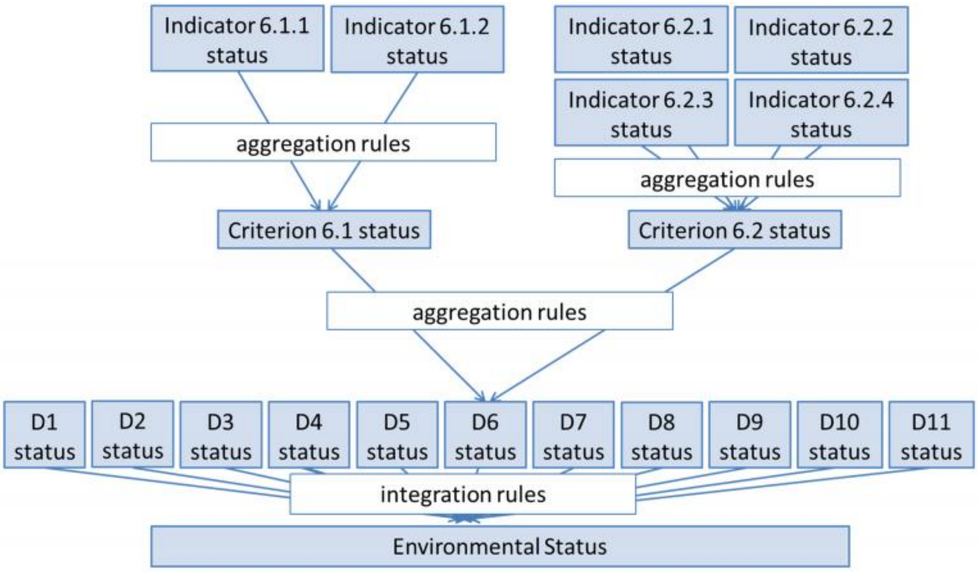
| Chemical Features | Threshold Value | Unit | Resp. | Threshold score | Status 1997-2007 | Status score | Contamination Ratio | Ind_Conf | Contamination Sum | QE status | QE_Conf | Weight |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Copper (Cu) in fish liver [biota] | 10,5 | mg/kg dw | + | H M L | 11,28 | H M L | 1,074 | 50% | | | | |
| DDE pp in fish muscle [biota] | 0,005 | mg/kg ww | + | H M L | 0,0017 | H M L | 0,340 | 75% | | | | |
| DDT pp in fish muscle [biota] | 0,005 | mg/kg lw | + | H M L | 0,017 | H M L | 3,400 | 50% | | | | |
| HCHs, Sum of [biota] | 0,033 | mg/kg ww | + | H M L | 0,0004 | H M L | 0,012 | 100% | | | | |
| Hexachlorobenzene (HCB) [biota] | 0,0167 | mg/kg ww | + | H M L | 0,0002 | H M L | 0,012 | 100% | | | | |
| Lead (Pb) fish liver [biota] | 0,02 | mg/kg dw | + | H M L | 0,098 | H M L | 4,900 | 50% | | | | |
| Sum7PCBs in fish [biota] | 10 | µg/kg ww | + | H M L | 0,00584 | H M L | 0,001 | 100% | | | | |
| Zinc (Zn) in herring liver [biota] | 88,2 | mg/kg dw | + | H M L | 73,5 | H M L | 0,833 | 50% | | | | |
| PCB-153 [seafood] | 0,1 | mg/kg ww | + | H M L | 0,286 | H M L | 2,860 | 75% | | | | |
| Cadmium (Cd) in fish muscle [seafood] | 0,05 | mg/kg ww | + | H M L | 0,355 | H M L | 7,100 | 75% | | | | |
| Mercury (Hg) [seafood] | 0,5 | mg/kg ww | + | H M L | 0,023 | H M L | 0,046 | 75% | | | | |
| Add new indicator ... | | | | | | | | | | | | |
| | | | | | | | | 100% | 6,205 | POOR | 73% | 33% |

| Supporting Features | RefCon | Unit | Resp. | RefCon_score | AcDev | AcDev_score | Status | Status_score | EQR | | Weight | QE_EQR | QE status | QE_Conf | Weight |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Secchi open-sea Jun-Sep mean | 8,00 | m | - | H M L | 25% | H M L | 4,00 | H M L | 0,500 | 100% | 33% | | | | |
| NO2+NO3 Dec-Feb mean | 2,50 | µM | + | H M L | 50% | H M L | 8,30 | H M L | 0,301 | 83% | 33% | | | | |
| PO4 Dec-Feb mean | 0,50 | µM | + | H M L | 50% | H M L | 0,83 | H M L | 0,602 | 83% | 34% | | | | |
| Add new indicator ... | | | | | | | | | | | | | | | |
| | | | | | | | | | | | 100% | 0,469 | POOR | 89% | 33% |
| | | | | 77% | | 52% | | 93% | | | | | | | |

Final ecosystem health status: BAD

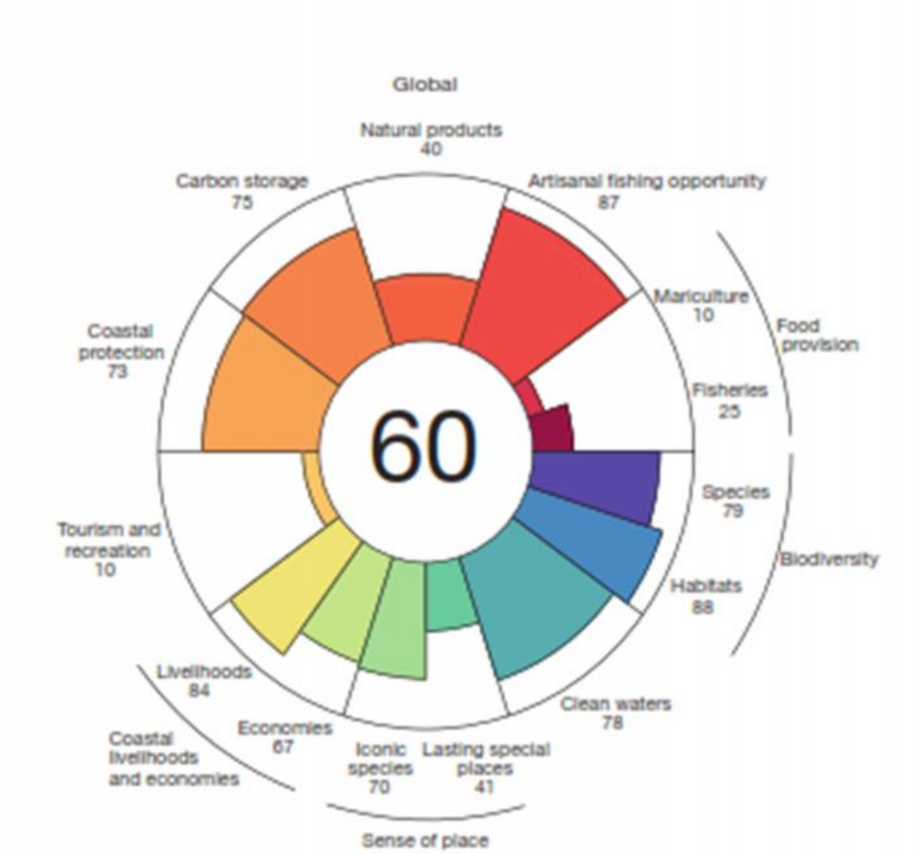
version 20100223 — 76,91% — Final confidence rating: Class I

Fig 2



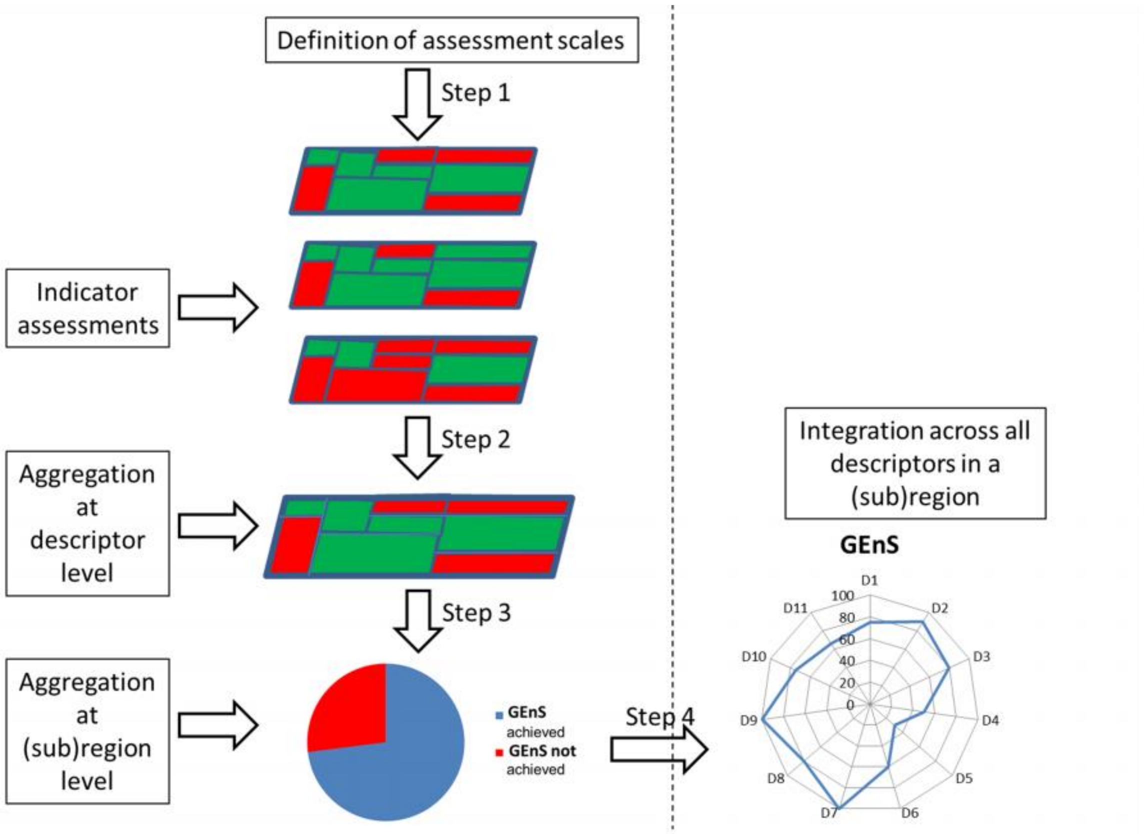
Indicator 6.1.1 status
Indicator 6.1.2 status
Indicator 6.2.1 status
Indicator 6.2.2 status
Indicator 6.2.3 status
Indicator 6.2.4 status
aggregation rules
aggregation rules
Criterion 6.1 status
Criterion 6.2 status
aggregation rules
D1 status
D2 status
D3 status
D4 status
D5 status
D6 status
D7 status
D8 status
D9 status
D10 status
D11 status
integration rules
Environmental Status

Fig 3



Global
Natural products
40
Artisanal fishing opportunity
87
Carbon storage
75
Mariculture
10
Food provision
Coastal protection
73
Fisheries
25
60
Species
79
Tourism and recreation
10
Biodiversity
Habitats
88
Livelihoods
84
Clean waters
78
Economies
67
Coastal livelihoods and economies
Iconic species
70
Lasting special places
41
Sense of place

Fig 4



Definition of assessment scales
Step 1
Indicator assessments
Step 2
Aggregation at descriptor level
Step 3
Aggregation at (sub)region level
GEnS achieved
GEnS not achieved
Step 4
Integration across all descriptors in a (sub)region
GEnS
D1
D2
D3
D4
D5
D6
D7
D8
D9
D10
D11
100
80
60
40
20
0

Table 1. Descriptors, criteria and indicators selected by the European Commission (2010), for ecosystem-based assessment and management of European seas, within the Marine Strategy Framework Directive.

| Descriptors | Criteria | Indicators |
|---|---|---|
| 1. **BIOLOGICAL DIVERSITY** is maintained. The quality and occurrence of habitats and the distribution and abundance of species are in line with prevailing physiographic, geographic and climatic conditions. | 1.1. Species distribution | 1.1.1. Distributional range<br>1.1.2. Distributional pattern within the latter, where appropriate<br>1.1.3. Area covered by the species (for sessile/benthic species) |
| | 1.2 Population size | 1.2.1. Population abundance and/or biomass, as appropriate |
| | 1.3 Population condition | 1.3.1. Population demographic characteristics (e.g. body size or age class structure, sex ratio, fecundity rates, survival/ mortality rates)<br>1.3.2. Population genetic structure, where appropriate |
| | 1.4. Habitat distribution | 1.4.1. Distributional range<br>1.4.2. Distributional pattern |
| | 1.5. Habitat extent | 1.5.1. Habitat area<br>1.5.2. Habitat volume, where relevant |
| | 1.6. Habitat condition | 1.6.1. Condition of the typical species and communities<br>1.6.2. Relative abundance and/or biomass, as appropriate<br>1.6.3. Physical, hydrological and chemical conditions |
| | 1.7. Ecosystem structure | 1.7.1. Composition and relative proportions of ecosystem components (habitats and species) |
| 2. **NON-INDIGENOUS SPECIES** introduced by human activities are at levels that do not adversely alter the ecosystems. | 2.1. Abundance and state characterisation of non-indigenous species, in particular invasive species | 2.1.1. Trends in abundance, temporal occurrence and spatial distribution in the wild of non-indigenous species, particularly invasive non-indigenous species, notably in risk areas, in relation to the main vectors and pathways of spreading of such species |
| | 2.2. Environmental impact of invasive non-indigenous species | 2.2.1. Ratio between invasive non-indigenous species and native species in some well-studied taxonomic groups (e.g. fish, macroalgae, molluscs) that may provide a measure of change in species composition (e.g. further to the displacement of native species)<br>2.2.2 Impacts of non-indigenous invasive species at the level of species, habitats and ecosystem, where feasible |
| 3. Populations of all **COMMERCIALLY EXPLOITED FISH AND SHELLFISH** are within safe biological limits, exhibiting a population age and size distribution that is indicative of a healthy stock. | 3.1. Level of pressure of the fishing activity | 3.1.1. Fishing mortality (F)<br>3.1.2. Ratio between catch and biomass index (hereinafter 'catch/biomass ratio') (if analytical assessments yielding values for F are not available) |
| | 3.2. Reproductive capacity of the stock | 3.2.1. Spawning Stock Biomass (SSB)<br>3.2.2. Biomass indices (if analytical assessments yielding values for SSB are not available) |
| | 3.3. Population age and size distribution | 3.3.1. Proportion of fish larger than the mean size of first sexual maturation<br>3.3.2. Mean maximum length across all species found in research vessel surveys (3.3.2)<br>3.3.3. 95 % percentile of the fish length distribution observed in research vessel surveys<br>3.3.4. Size at first sexual maturation, which may reflect the extent of undesirable genetic effects of exploitation (secondary indicator) |
| 4. All elements of the marine **FOOD WEBS**, to the extent that they are known, occur at normal abundance and diversity and levels capable of ensuring the long-term abundance of | 4.1. Productivity of key species or trophic groups | 4.1.1. Performance of key predator species using their production per unit biomass (productivity) |
| | 4.2. Proportion of selected species at the top of food webs | 4.2.1. Large fish (by weight) |
| | 4.3. Abundance/distribution of key trophic | 4.3.1. Abundance trends of functionally important selected groups/species. Detailed indicators need to be further specified, taking account of their |

| | | |
|---|---|---|
| the species and the retention of their full reproductive capacity. | groups/species | importance to the food webs, on the basis of suitable groups/species in a region, sub-region or subdivision, including where appropriate: (i) groups with fast turnover rates (e.g. phytoplankton, zooplankton, jellyfish, bivalve molluscs, short-living pelagic fish) that will respond quickly to ecosystem change and are useful as early warning indicators, (ii) groups/species that are targeted by human activities or that are indirectly affected by them (in particular, by-catch and discards), (iii) habitat-defining groups/species, (iv) groups/species at the top of the food web, (v) long-distance anadromous and catadromous migrating species, and (vi) groups/species that are tightly linked to specific groups/species at another trophic level |
| 5. Human-induced **EUTROPHICATION** is minimised, especially adverse effects thereof, such as losses in biodiversity, ecosystem degradation, harmful algae blooms and oxygen deficiency in bottom waters. | 5.1. Nutrient levels | 5.1.1. Nutrients concentration in the water column<br>5.1.2. Nutrient ratios (silica, nitrogen and phosphorus), where appropriate |
| | 5.2. Direct effects of nutrient enrichment | 5.2.1. Chlorophyll concentration in the water column<br>5.2.2. Water transparency related to increase in suspended algae, where relevant<br>5.2.3. Abundance of opportunistic macroalgae<br>5.2.4. Species shift in floristic composition such as diatom to flagellate ratio, benthic to pelagic shifts, as well as bloom events of nuisance/toxic algal blooms (e.g. cyanobacteria) caused by human activities |
| | 5.3. Indirect effects of nutrient enrichment | 5.3.1. Abundance of perennial seaweeds and seagrasses (e.g. fucoids, eelgrass and Neptune grass) adversely impacted by decrease in water transparency<br>5.3.2. Dissolved oxygen, i.e. changes due to increased organic matter decomposition and size of the area concerned |
| 6. **SEA-FLOOR INTEGRITY** is at a level that ensures that the structure and functions of the ecosystems are safeguarded and benthic ecosystems, in particular, are not adversely affected. | 6.1. Physical damage, having regard to substrate characteristics | 6.1.1. Type, abundance, biomass and areal extent of relevant biogenic substrate<br>6.1.2. Extent of the seabed significantly affected by human activities for the different substrate types |
| | 6.2. Condition of benthic community | 6.2.1. Presence of particularly sensitive and/or tolerant species<br>6.2.2. Multi-metric indexes assessing benthic community condition and functionality, such as species diversity and richness, proportion of opportunistic to sensitive species<br>6.2.3. Proportion of biomass or number of individuals in the macrobenthos above some specified length/size<br>6.2.4. Parameters describing the characteristics (shape, slope and intercept) of the size spectrum of the benthic community |
| 7. **PERMANENT ALTERATION OF HYDROGRAPHICAL CONDITIONS** does not adversely affect marine ecosystems. | 7.1. Spatial characterisation of permanent alterations | 7.1.1. Extent of area affected by permanent alterations |
| | 7.2. Impact of permanent hydrographical changes | 7.2.1. Spatial extent of habitats affected by the permanent alteration<br>7.2.2. Changes in habitats, in particular the functions provided (e.g. spawning, breeding and feeding areas and migration routes of fish, birds and mammals), due to altered hydrographical conditions |
| 8. Concentrations of **CONTAMINANTS** are at levels not giving rise to pollution effects. | 8.1. Concentration of contaminants | 8.1.1. Concentration of the contaminants mentioned above, measured in the relevant matrix (such as biota, sediment and water) in a way that ensures comparability with the assessments under Directive 2000/60/EC |
| | 8.2. Effects of contaminants | 8.2.1. Levels of pollution effects on the ecosystem components concerned, having regard to the selected biological processes and taxonomic groups where a cause/effect relationship has been established and needs to be monitored<br>8.2.2. Occurrence, origin (where possible), extent of significant acute pollution events (e.g. slicks from oil and oil products) and their impact on biota physically affected by this pollution |
| 9. **CONTAMINANTS** | 9.1. Levels, number and | 9.1.1. Actual levels of contaminants that have been detected and number of |

| | | |
|---|---|---|
| **IN FISH AND OTHER SEAFOOD FOR HUMAN CONSUMPTION** do not exceed levels established by Community legislation or other relevant standards. | frequency of contaminants | contaminants which have exceeded maximum regulatory levels<br>9.1.2. Frequency of regulatory levels being exceeded |
| 10. Properties and quantities of **MARINE LITTER** do not cause harm to the coastal and marine environment. | 10.1. Characteristics of litter in the marine and coastal environments | 10.1.1. Trends in the amount of litter washed ashore and/or deposited on coastlines, including analysis of its composition, spatial distribution and, where possible, source<br>10.1.2. Trends in the amount of litter in the water column (including floating at the surface) and deposited on the sea- floor, including analysis of its composition, spatial distribution and, where possible, source<br>10.1.3. Trends in the amount, distribution and, where possible, composition of micro-particles (in particular micro- plastics) |
| | 10.2. Impacts of litter on marine life | 10.2.1. Trends in the amount and composition of litter ingested by marine animals (e.g. stomach analysis) |
| 11. Introduction of energy, including **UNDERWATER NOISE**, is at levels that do not adversely affect the marine environment. | 11.1. Distribution in time and place of loud, low and mid frequency impulsive sounds | 11.1.1. Proportion of days and their distribution within a calendar year over areas of a determined surface, as well as their spatial distribution, in which anthropogenic sound sources exceed levels that are likely to entail significant impact on marine animals measured as Sound Exposure Level (in dB re 1μPa 2.s) or as peak sound pressure level (in dB re $1\mu Pa_{peak}$) at one metre, measured over the frequency band 10 Hz to 10 kHz |
| | 11.2. Continuous low frequency sound | 11.2.1. Trends in the ambient noise level within the 1/3 octave bands 63 and 125 Hz (centre frequency) (re 1μPa RMS; average noise level in these octave bands over a year) measured by observation stations and/or with the use of models if appropriate |

1162

Table 2. Approaches for combining different metrics, indicators or criteria to assess the status, including the advantages and disadvantages of each approach, as considered by the authors. Key: GEnS: Good environmental status.

| General | Details of | Advantages | Disadvantages |
|---|---|---|---|
| One-out all-out (OOAO) principle<br><br>CIS (2003), Borja et al. (2009a), Borja and Rodríguez (2010), Ojaveer and Eero (2011), Caroni et al. (2013) | All variables have to achieve good status | Most comprehensive approach. Follows the precautionary principle | Trends in quality are hard to measure. Does not consider weighting of different indicators and descriptors. Chance of failing to achieve good status very high. |
| | As a variation, Tueros et al. (2009) proposed the Two-out all-out: if two variables do not meet the required standard, good status is not achieved | More robust compared to OOAO approach | See above |
| Averaging approach<br><br>Ojaveer and Eero (2011), Shin et al. (2012) | Non-weighted: Variable values are combined, using the arithmetic average or median | Indicator values can be calculated at each level of aggregation. Recommended when combined parameters are sensitive to a single pressure | Assumes all variables are of equal importance |
| | Weighted: Like the previous method, with different weights assigned to the various variables | Reflects the links between descriptors and avoids double counting | High data requirements Problem of agreeing on weights |
| | Hierarchical: With variables defined at different hierarchical levels | Reflects the hierarchy among descriptors and avoids double counting Different calculation rules can be applied at different levels | Problem of agreeing on hierarchy |
| Conditional rules<br><br>Tueros et al. (2009), Simboura et al. (2012), Breen et al. (2012) | A specific proportion of the variables have to achieve good status | Focuses on the key aspects (i.e. biodiversity descriptors) | Assumes that GEnS is well represented by a selection of variables |
| Scoring or rating<br><br>Borja et al. (2004, 2010, 2011b), Birk et al. (2012) | Sum of weighted scores | Different weights can be assigned to the various elements | Problem of agreeing on weights. Metrics may not be sensitive to the same pressures |
| Multimetric approaches<br><br>Rice et al. (2010), Borja et al. (2011a), Birk et al. (2012) | Multi-metric indices | Integrates multiple indicators into one value. May result in more robust indicators, compared to indicators based on single parameters | Correlations between parameters can be an issue. Results are hard to communicate to managers. Metrics may not be sensitive to the same pressures |
| Multi-dimensional approaches<br><br>Shin et al. (2012) | Multivariate analyses | No need to set rigid target values, since values are represented within a domain | Results are hard to communicate to managers |

| General | Details of | Advantages | Disadvantages |
|---|---|---|---|
| Decision tree<br><br>Borja et al. (2004, 2009b, 2013) | Integrating elements into a quality assessment using specific decision rules | Possible to combine different types of elements, flexible approach | Only quantitative up to a certain level |
| Probabilistic<br><br>Barton et al. (2008, 2012), Lehikoinen et al. (2013a, 2013b) | Bayesian statistics | Produces a probability estimate of how likely the area is in GEnS; managers can decide the acceptable undertainty | Difficult to calculate |
| High-level integration<br><br>HELCOM (2010), Borja et al. (2010, 2011b), Halpern et al. (2012), Tett et al. (2013) | Assessment results for three groups: biological indicators, hazardous substances indicators and supporting indicators, each applying OOAO | Reduces the risks associated with OOAO while still giving an overall assessment | Technical details |

Table 3. Summary of Task Group approaches to aggregate attributes within a Descriptor (Cardoso *et al*., 2010).

| **Aggregation of attributes** | **Descriptor** |
|---|---|
| Integrative assessments (combining attributes appropriate to local conditions) | D1 Biodiversity<br>D2 Non-indigenous species<br>D5 Eutrofication<br>D6 Seafloor integrity |
| Assessment by worst case (Descriptor not in good status if any attribute is not OK) | D3 Commercial fish (3 attributes)<br>D4 Food webs (2 attributes)<br>D8 Contaminants (3 attributes)<br>D9 Contaminants in fish (1 attribute)<br>D10 Litter (3 attributes)<br>D11 Energy and noise (3 attributes) |

Table 4. Options for determining if an area/regional sea is in Good Environmental Status (GEnS) (modified from Borja *et al.*, 2013). Key: OOAO: 'one out, all out' principle.

| Option | Decision rule | Data requirements | Pros | Cons | Examples in place |
|---|---|---|---|---|---|
| *Either:*<br><br>1. fulfilling all the indicators in all the descriptors | All indicators are met irrespective of weighting (OOAO) | Data needed for all aspects on regional seas scale | Most comprehensive approach | Unreasonable data requirements; all areas will fail on at least one indicator; may include double-counting | None |
| *Or:*<br><br>2.fulfilling the indicators in all descriptors but as a weighted list according to the hierarchy of the descriptors | Agreeing the weighting | Data needed for all aspects on regional seas scale | Reflects the interlinked nature of the descriptors and avoids double counting | Unreasonable data requirements; problem of agreeing the weighting | HELCOM (2010); Borja *et al.* (2011b); Aubry and Elliott (2006) |
| *Or:*<br><br>3.fulfilling the indicators just for the biodiversity descriptor and making sure these encompass all other quality changes | All biodiversity indicators are met irrespective of weighting | Data needed for all components of biodiversity | Focuses on the main aspect | Assumes that the biodiversity descriptor really does encompass all others | Feary *et al.* (2014) |
| *Or:*<br><br>4.create a synthesis indicator which takes the view that 'GEnS is the ability of an area to support ecosystem services, produce societal benefits and still maintain and protect the conservation features' | Integration of the information from different descriptors and indicators, and evaluation of the overall benefits | Data needed for the indicators included in that synthesis indicator, valuation of the ecosystem services and benefits | Fulfils the main aim of marine management (see text) | Requires a new indicator and an agreement in the way of integrate the information; trade-offs between ecosystem services and their beneficiaries require either economic, ethical or political evaluation and decision, and cannot be based only on ecological knowledge | Borja *et al.* (2011b) |
| *Or:* | then if an area has | An expert judgement | It may reflect the | It may be too subjective (i.e. | Bricker *et al.* |

| | | | | | |
|---|---|---|---|---|---|
| 5.have a check-list (ticking boxes) of all the aspects needed | e.g. more than 60% of the boxes ticked then it is in GEnS | approach, based on 'probability of evidence' | state of the science; if done rigorously then it may be the easiest to implement | based on soft intelligence) | (2003); Ferreira *et al.* (2011) |
| *Or:*<br><br>6.have a summary diagram such as a spiders-web diagram showing the 'shape of GEnS according to several headline indicators' | The shape of the diagram | | Easy to understand and show to managers | The decision on when GEnS is achieved | Halpern *et al.* (2012) |
| *Or:*<br><br>7.not reporting the environmental status but only the list of pressures (i.e. on the premise that if an area has no obvious pressures then any changes in the area must be due to natural changes which are outside the control of management) | No pressures in an area sufficient to cause adverse effects | Quantitative maps of pressures | Can be derived by national databases, mapping, pressure lists | Relates to 'cause' rather than 'effect', difficult to set boundaries between pressure status classes: is it sufficient to base the assessment on the list of pressures, while those can have very different spatial extent and strength? | Aubry and Elliott (2006); Halpern *et al.* (2008); Korpinen *et al.* (2012); Solheim *et al.* (2012) |
| *Or:*<br><br>8.a combination of all/some of these when there are insufficient data in some areas or for some descriptors or indicators | | Combination of pressures and descriptors data | Information available from Member States reports | Either requires too much information (hence unreasonable) or too little (hence inaccurate) | None |

Table 5. Example of an assessment of the environmental status, within the Marine Strategy Framework Directive, in the Basque Country offshore waters (Bay of Biscay) (modified from Borja *et al*., 2011b). EQS: Environmental Quality Standards; EQR: Ecological Quality Ratio, both based upon the Water Framework Directive (WFD); NA: not available. Trends: red colour, negative; green colour, positive (in both cases can be increasing/decreasing, depending on the indicator).

| Qualitative Descriptors | Explanation of the indicators used | Reference conditions/EQS | Recent trend | Reliability (%) | Weight (%) | EQR | Final Environmental Status | Final Confidence ratio |
|---|---|---|---|---|---|---|---|---|
| 1.- Biological diversity | integrated biological value | | NA | 69 | 15 | 0.51 | 0.08 | 10.35 |
| 2.- Non-indigenous species | ratio non-indigenous sp. | OSPAR | ▲ | 80 | 10 | 0.98 | 0.10 | 8 |
| 3.- Exploited fish and shellfish | | | ▼ | 100 | 15 | $\boxed{0.48}$ | 0.07 | 15 |
| | fishing mortality <reference | | | 100 | | 0.18 | | |
| | Spawning stock <reference | | | 100 | | 0.67 | | |
| | % large fish | | | 100 | | 0.59 | | |
| 4.- Marine food webs | | | ▼ | 70 | 10 | $\boxed{0.40}$ | 0.04 | 7 |
| 5.- Human-induced eutrophication | | WFD | ▼ | 94 | 10 | 0.96 | 0.10 | 9.4 |
| | Nutrients in good status | | | 100 | | 0.80 | | |
| | Chlorophyll in high status | | | 100 | | 1.00 | | |
| | Optical properties in high status | | | 100 | | 1.00 | | |
| | Bloom frequency in high status | | | 70 | | 1.00 | | |
| | Oxygen in high status | | | 100 | | 1.00 | | |
| 6.- Seafloor integrity | | WFD | ► | 100 | 10 | 0.89 | 0.09 | 10 |
| | Area not affected | | | 100 | | 0.87 | | |
| | % presence sensitive sp. | | | 100 | | 0.98 | | |
| | Mean M-AMBI value | | | 100 | | 0.83 | | |
| 7.- Alteration of hydrographical conditions | | | ► | 100 | 2 | 1.00 | 0.02 | 2 |
| 8.- Concentrations of contaminants | High % of samples <EQS | WFD | ▼ | 100 | 9 | 0.80 | 0.07 | 9 |
| 9.- Contaminants in fish and other seafood | Values are 30% of the most affected in the NEA | WFD | ▼ | 30 | 9 | 0.60 | 0.05 | 2.7 |
| 10.- Marine litter | Values are 50% of the most affected in Europe | OSPAR | ▲ | 30 | 5 | 0.57 | 0.03 | 1.5 |
| 11.- Energy & underwater noise | Moderate ship activity | OSPAR | NA | 10 | 5 | 0.70 | 0.04 | 0.5 |
| **Final assessment** | | | | | **100** | | **0.68 Good** | **75.5 High** |

Table 6. Pros and cons of the decision of integrating the information across descriptors

| Procedure | Pros | Cons |
|---|---|---|
| No integration | - Direct detection of problems (management needs) for each descriptor<br>- Useful for local managers (close to specific or local pressures)<br>- Reduces multiple accounting<br>- Easiest to implement | - Does not fulfil the main aim of marine management in an integrative way<br>- Does not fully reflect the ecosystem-based approach<br>- Difficult to compare across Member States and regions |
| Integration (all descriptors or a subset) | - Progress towards GEnS relevant at regional scale (comparable across regional seas and countries)<br>- Environmental status defined in an integrative way, as health of the ecosystem (full ecosystem-based approach)<br>- Most comprehensive approach<br>- Reflect the interlinked nature of the descriptors<br>- Easy to communicate in policy and societal domains | - Loss of information on specific issues, obscuring the progress towards GEnS<br>- Can mask problems from specific descriptors/pressures<br>- May include multiple accounting<br>- May be too subjective, as it typically involves expert judgment |