Service request: "Coherent geographic scales and aggregation rules in assessment and monitoring of Good Environmental Status- Analysis and conceptual phase"

Deltares / AZTI / HCMR / IVM
October 2014

Framework contract No ENV.D2/FRA/2012/0019

Service Request: “Coherent geographic scales and aggregation rules in assessment and monitoring of Good Environmental Status – analysis and conceptual phase”

Delft, 31 October 2014

Consortium - Lead

Deltares
Enabling Delta Life

Sub-contractors

azti tecnalia

hcmr
ΕΛΚΕΘΕ

IVM VU
Institute for Environmental Studies

Coherent geographic scales and aggregation rules for environmental status assessment within the Marine Strategy Framework Directive

Towards a Guidance document

Theo Prins (Deltares)
Angel Borja (AZTI)
Nomiki Simboura (HCMR)
Catherine Tsangaris (HCMR)
Myra van der Meulen (Deltares)
Arjen Boon (Deltares)
Iratxe Menchaca (AZTI)
Alison Gilbert (IVM)

1207879-000

Title
Coherent geographic scales and aggregation rules for environmental status assessment within the Marine Strategy Framework Directive

Client	Project	Reference	Pages
European Commission DG Environment Unit D2	1207879-000	1207879-000-ZKS-0014	53

Keywords
MSFD, monitoring, assessment

Summary
This report deals with the definition of spatial scales and the use of aggregation methods in the assessments of environmental status within the MSFD. Criteria to define spatial scales for assessment areas are discussed. An overview of methods for the aggregation of assessments is presented. The appendix contains a more detailed description of methods and an analysis of the approaches by member states in their Initial Assessments.

References
SFRA0019-SCALES

Version	Date	Author	Initials	Review	Initials	Approval	Initials
1	Feb 2014	Theo Prins					
		Myra van der Meulen					
		Arjen Boon					
		Nomiki Simboura					
		Catherine Tsangaris					
		Angel Borja					
		Iratxe Menchaca					
		Alison Gilbert					
2	April 2014						
3	October 2014						

State
final

Cite as: Prins T.C., Borja A., Simboura N., Tsangaris C., Van der Meulen M.D., Boon A.R., Menchaca I., Gilbert A.J., 2014. Coherent geographic scales and aggregation rules for environmental status assessment within the Marine Strategy Framework Directive. Towards a draft guidance. Deltares/AZTI/HCMR, Report 1207879-000-ZKS-0014 to the European Commission, Delft, 56 pages +annexes.

Contents

Appendices

Executive summary

This report is based on an analysis of national approaches that Member States (MS) have taken in their reporting under Articles 8, 9 and 10 of the Marine Strategy Framework Directive (MSFD), with respect to geographical scaling and aggregation rules, and provides advice for the development of broad EU guidance for coherent geographic scales in assessment and monitoring of GES and for sets of aggregation rules (such as to aggregate data from monitoring for assessments and across criteria for a GES Descriptor).

The objectives of this project were to analyse and compare the national approaches regarding the spatial scales for the environmental assessment of their marine waters, to analyse which aggregation rules have been applied, and to develop guidance for coherent geographic scales and aggregation rules.

An analysis of MS approaches was reported in November 2013 and is included in an Annex to this report. The results were discussed in a WG GES workshop in Brussels in October 2013. This report builds on the results of the workshop and the analysis and presents a next step in the development of guidelines.

Aggregation inevitably causes the loss of information, but information needs can differ, depending on the purpose and may require different levels of aggregation. Environmental assessments address different information needs at different levels and spatial scales, from relatively small spatial scales and low levels of integration to inform on suitable management measures, up to assessments at the level of (sub)regions to follow policy implementation.

Assessment scales should be defined taking into account both ecological considerations such as hydrodynamic and physical-chemical characteristics and biogeography, as well as management perspectives: provide a robust and adequate assessment of environmental state, enable the identification and evaluation of management measures. Spatial assessment scales could be different, depending on the issue, ranging from small scales in the case of local pressures or specific habitats to marine (sub)region or larger scales in the case of wide-spread pressures or species with a large distributional range (e.g. cetaceans).

A method is proposed to develop a system of assessment scales that are nested in a hierarchical way, similar to the approach that has been developed for the Baltic Sea by HELCOM. This could be part of an adaptive management approach where scales can be applied that are suited for the needs of a specific assessment method, allow aggregation to larger scales, while a pragmatic optimization would help to keep the number of assessment areas manageable.

An overview is given of aggregation methods that can be used to combine indicators and criteria within a descriptor. General criteria to decide on the most appropriate aggregation rule are discussed. The 'one-out-all-out' method that is applied in the Water Framework Directive is applicable in some cases, but is not in all cases a suitable approach.

Several methods are discussed that can be used to aggregate assessments across descriptors.

A step-wise approach is proposed that can be used to aggregate assessments at different levels of spatial scales and different levels of integration. The aggregation level will depend on the information that is needed.

This report describes generic approaches and criteria to deal with the spatial scales of assessments and the aggregation of assessments. There are still many open questions and knowledge gaps, and more specific guidance is not yet possible. There is a clear need for further work, which could partly be carried out in pilot projects. An issue that requires further development is the aggregation of biodiversity related indicators (under Descriptors 1, 4, 6), that encompass many different features, and methods to combine those in a meaningful way in assessments have not been developed yet. The combination of spatial assessment scales with time scales for assessments is another issue that needs further development. The nested approach towards spatial scale has been developed for the Baltic Sea for some topics, but has not been applied for all elements of the MSFD and has not been applied in other regional seas. The effects of uncertainty in data for assessment results and the risks of misclassification should be considered when more specific aggregation methods are developed. Finally, several options for aggregation across descriptors are discussed in the report. There are methods available to combine descriptors in integrated assessments and appropriate approaches and to present this in a meaningful way, but further developmental work is needed. The upcoming review of the Commission Decision on criteria and methodological standards can be helpful in this respect.

Most importantly, aggregation rules should be transparent and agreed, at least at a regional level, if not EU level, to allow for comparative assessments and communication of results (as required by MSFD).

1 Introduction

1.1 General background

The 2012 reporting for Marine Strategy Framework Directive (MSFD, 2008/56/EC) Articles 8, 9 and 10 constitutes three important steps in the first six-year management cycle of the MSFD. With the reporting on the initial assessment of the marine waters (Art. 8), the determination of Good Environmental Status (GES, Art. 9) and the identification of environmental targets and associated indicators (Art. 10) the Member States (MS) should have identified all relevant issues concerning drivers, pressures, state and impacts in the marine environment.

Article 3(5) of the MSFD requires that good environmental status is determined at the level of the marine region or subregion as referred to in Article 4, on the basis of the qualitative Descriptors in Annex I to the Directive. However, for the assessment of whether GES has been achieved or not, finer scales can be used. In the Water Framework Directive (WFD, 2000/60/EC) assessments of ecological and chemical status are done at the level of estuarine and coastal water bodies. The geographical scale to be used for assessments under the MSFD is not well defined. Consequently, in the first reporting of implementation in 2012 the geographical scales adopted for the assessment of environmental status varied considerably between descriptors, and differed widely among MS. Member State determinations of GES need to be consistent across the marine region or subregion.

Assessments of the status of the marine environment have a spatial characteristic (for example, what area are we dealing with, what ecosystem components and what pressures are important in this area), which may work out differently for the various descriptors or criteria and indicators within a descriptor due to differences in spatial distribution of ecosystem components and human activities. Therefore, the first question that needs to be addressed is:

- *What is the appropriate spatial scale for the assessment of each Descriptor including its criteria and indicators?*

When assessment scales have been defined, the question of scaling up from the individual, specific or sectorial assessments to an assessment for the whole (sub-)region needs to be considered:

- *How to scale up from assessment areas to larger geographic scales?*

A third question deals with the aggregation of the various assessments at different levels:

- *How to aggregate indicators within a criterion, or criteria within a descriptor, or all the descriptors to come to a comprehensive and balanced judgement of the status of marine waters through GES?*

In January 2013, the European Commission put out a Service Request, asking for an analysis of national approaches that Member States have taken in their reporting under Articles 8, 9 and 10 of the MSFD, with respect to geographic scaling and aggregation rules, and for the development of broad EU guidance for coherent geographic scales in assessment of GES and for sets of aggregation rules.

1.2 Objectives of this report

The objective of the Service request is to develop guidance on the application of geographic scales and aggregation rules in the assessment of the status of the marine environment under the MSFD.

The objectives stated in the Service Request are to:

1. assess the electronic and text reporting undertaken by Member States (MS) under Articles 8, 9 and 10 of the MSFD with the aim to analyse and compare the national approaches taken per descriptor regarding the scales for the assessment of the environmental state of their marine waters, determining GES and setting environmental targets.
2. analyse which aggregation rules have been applied, if any, by MS in their reports. Based on the results of these analyses and further comparison with regional approaches and methods applied in research projects, identify issues that require further consultation by MS, Regional Sea Conventions (RSC) and the European Commission.
3. develop broad EU guidance for coherent geographic scales in assessment and monitoring of GES and for sets of aggregation rules and organize a debate with MS on this.

In the framework of this service, a report was made with an analysis of MS approaches towards geographical scaling and aggregation in the Initial assessments. The report gives an overview of existing methods applied by RSCs and in other assessments. This analytical report addresses the above mentioned questions 1 and 2. The analytical report is included in Annex I, and was presented and discussed in a WG GES workshop in Brussels on 23rd October 2013.

This report builds on the results of the analysis and the discussions in the WG GES workshop. It aims to provide elements to develop guidance for a coherent approach to defining geographic areas for the assessment of environmental status and for the aggregation of these assessments from smaller areas up to the (sub)region scale as well as for the aggregation of criteria/indicators for each Descriptor within these assessment areas

The report is part of the Service Contract SFRA0019 - SCALES under the agreement of the ‘Framework contract for services related to development of methodological standards in relation to good environmental status of the seas under MSFD (ENV.D.2/FRA/2012/0019)’ between the European Commission/DG Environment and Deltares, as lead partner of a consortium with AZTI, HCMR, IVM and SYKE.

1.3 Report outline

Chapter 2 sketches the scope of this report. The question how to define an appropriate spatial scale for assessments of the marine environment is treated in Chapter 3. This chapter provides guidelines and criteria for the definition of assessment scales. Chapter 4 deals with the aggregation of assessments across criteria and indicators, as well as across descriptors, and aggregation from assessment areas to larger geographic scales. Chapter 5 discusses knowledge gaps and suggestions for further work.

2 Scope of the report

The final objective of this guidance is to enable the definition of a set of assessment areas within each (sub)region and to outline possible aggregation rules between criteria/indicators for each Descriptor; this will support improved coherence in the implementation of the MSFD and increase the comparability of the assessments of environmental status, with respect to the use of geographic scales and aggregation in those assessments.

The term scale is used in many different ways and may connote different aspects of space and time. Scale refers to the spatial or temporal dimension and is characterized by both grain and extent. Grain refers to the finest level of spatial or temporal resolution. Extent refers to the size of the area or the duration of time under consideration (Turner *et al.* 1993).

Time scales and spatial scales are closely related, and similar questions apply to both spatial and temporal scales: what are the scales for monitoring and assessments, what are the scales for reporting. However, in this report we will only elaborate on the issue of spatial scales as most questions seemed to revolve around geographical scales (Clausen *et al.* 2011; OSPAR 2011).

According to Article 3.5 of the MSFD, good environmental status of MS marine waters shall be determined at the level of the marine region or subregion, on the basis of the eleven qualitative descriptors from Annex I. GES is therefore a topic with a large spatial extent covering all EU marine waters, where MS have to collaborate at (sub)regional scale to define GES.

However, GES can be assessed at different scales (‘grain’). Assessments need to be done at spatial scales that are ecologically relevant, to provide information on the environmental status which is relevant to ecosystem-based management. The assessments have to support management of the human activities and pressures in the marine environment, in order to achieve GES in line with the ecosystem-based approach (see EEA 2014 for definition). The choice of an appropriate scale is critical and scales have to be selected in relation to the question that needs to be answered (Turner *et al.* 2001). As GES encompasses many different aspects of the marine environment, spatial scales that are relevant from an ecological and management perspective, may differ between the various descriptors. This raises questions with respect to the spatial scales that should be applied in assessments, and with respect to ways to aggregate assessment results.

There are basically three issues with respect to scaling and aggregation. The first question is the definition of the scales that are appropriate to assess the environmental status of the marine waters. This scale may be different, depending on the environmental issue. This issue is discussed in Chapter 3.

The second issue is the integration of a number of assessments, each dealing with a specific topic. To assess whether good environmental status has been achieved, 29 criteria and 56 associated indicators have been developed in relation to each of the eleven descriptors from Annex I of the MSFD (EC 2010). It is foreseen that in some cases several assessments for each indicator may have to be developed as different ecosystem components have to be considered (Clausen *et al.* 2011). Consequently, the number of operational indicators could even be higher than 56.

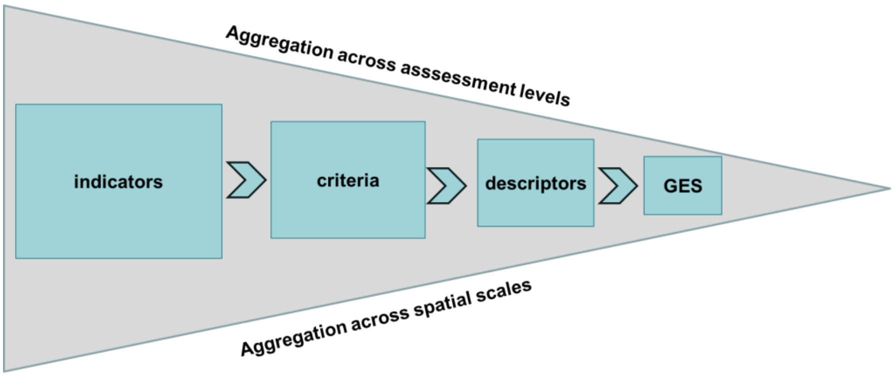


Figure 2.1. Schematic representation of the aggregation across assessment levels and spatial scales.

For the assessment of environmental status the assessments at the level of the more than 50 indicators have to be aggregated to the higher levels in order to assess the status of an ecosystem component (e.g. species or habitat) or a pressure from a human activity (Figure 2.1). Thirdly, an aggregation may be useful to go from assessments at the scale of individual assessment areas to assessments or presentations at the scale of marine regions and subregions

Aggregation inevitably causes the loss of information (Van Beurden and Douven 1999, Vermaat *et al.* 2005), but information needs can differ, depending on the purpose and may require different levels of aggregation.

Assessments are carried out to evaluate the (change of) environmental status, the impact of human activities and the effect of policy measures. The purpose of this evaluation of environmental status is to identify the main risks for the marine environment, inform managers and policy-makers about the environmental impacts of human activities, the need for measures, and the progress towards achieving GES. The assessments address different information needs at different levels and different spatial scales. Member States have a primary interest in assessing the status in the marine waters under their jurisdiction, to identify main risks and the need for measures. This requires information at least at the level of their marine waters within a (sub)region but often at finer scales and often a low level of integration. For example, indicators and small assessment areas are in many cases more suitable to link pressures to environmental impacts, and to inform on suitable management measures. At a European level, assessments of the environmental status to follow progress towards GES require approaches at a larger spatial scale which may also require higher integration levels (Figure 2.2).

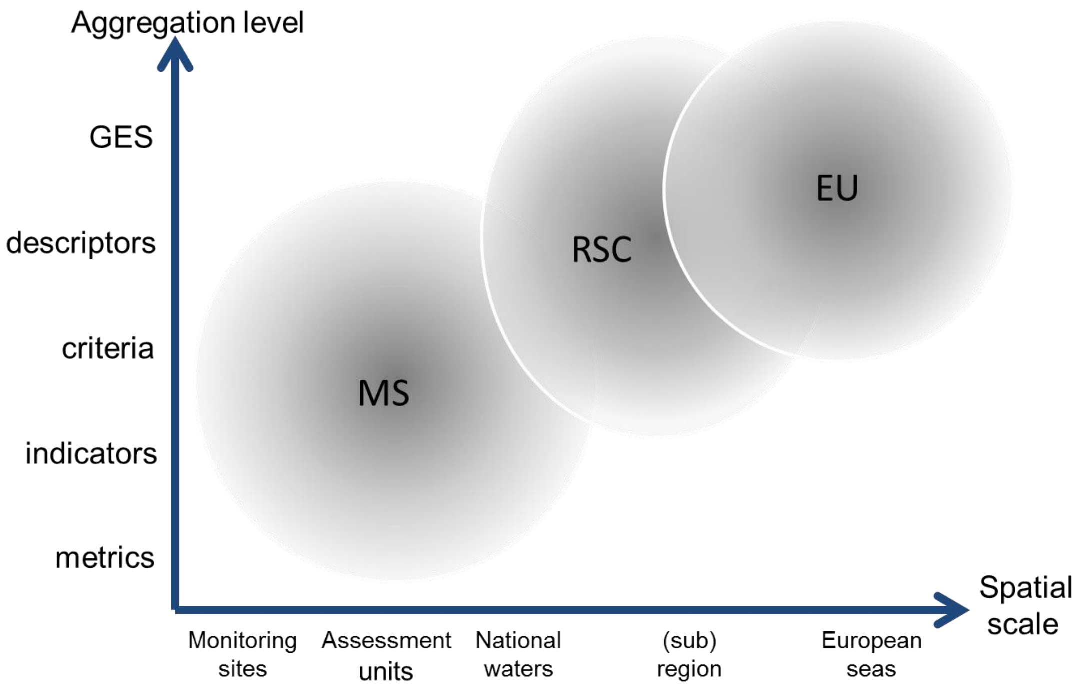


Figure 2.2. Differences in information needs and associated spatial scales and aggregation levels.

This report provides generic approaches that are applicable at different spatial scales and levels of aggregation. This could go from the spatial scale of monitoring sites and the level of metrics to the scale of marine regions and subregions and assessments of GES. In this report, the emphasis lies on the steps that are needed to go from assessments at the scale of one assessment unit, to assessments that go beyond the spatial scale of the marine waters under jurisdiction of a MS and beyond the aggregation level of indicators (Figure 2.2).

One of the first steps when assessing marine status is to obtain a comprehensive view of the relations between the human activities (at sea or on land), the pressures they exert on the marine environment, and the change that is caused in the state of the environment, leading to impacts on ecosystem services. To obtain such a view, an often used method is the DPSIR (Driver, Pressure, State, Impact, Response) approach (OECD 1993; EEA 1999). The DPSIR conceptual framework provides an overall mechanism for analysing environmental problems, with regards to sustainable development. In EEA’s definition (Gabrielsen and Bosch 2013), ‘Driving Forces’ are social, demographic and economic developments in societies and the corresponding changes in lifestyles, overall levels of consumption and production patterns. These Driving Forces lead to a range to human activities which aim to meet these 'societal demands' and these in turn lead to a number of ‘Pressures’ on the natural environment, through release of substances (emissions), physical and biological changes and the use of resources by human activities. These pressures degrade the ‘State’ of the environment, expressed as quantity and quality of physical, biological and chemical phenomena. These changes then have ‘Impacts’ upon human health, ecosystems and materials. The impacts lead to a ‘Response’ by society to prevent, compensate, ameliorate or adapt to changes in the state of the environment, by policy measures, such as regulations, information and taxes; these can be directed at any other part of the system (Figure 2.3).

The DPSIR approach is helpful to structure indicators and focus on causal relations regarding environmental problems. However, it has been criticized as being inappropriate as an analytical tool, because it ignores the complexity of environmental and socio-economic issues and definitions of the DPSIR are ambiguous (Maxim *et al.* 2009; Spangenberg *et al.* 2009). For the scope of this report, the DPSIR provides a useful framework to structure the indicators, criteria and descriptors from the Commission Decision (EC 2010), which are a mixture of Pressure, State and Impact descriptors according to the definition given above. But this does not imply that they can be placed in a simple, linear and deterministic description of the marine environment. Similarly, geographic scales for Pressure and State are potentially different and will not always match.

The approach in this report focusses on an approach to spatial scales related to the natural system. There may be a mismatch with the scale of the socio-economic system (Cumming *et al.* 2006).

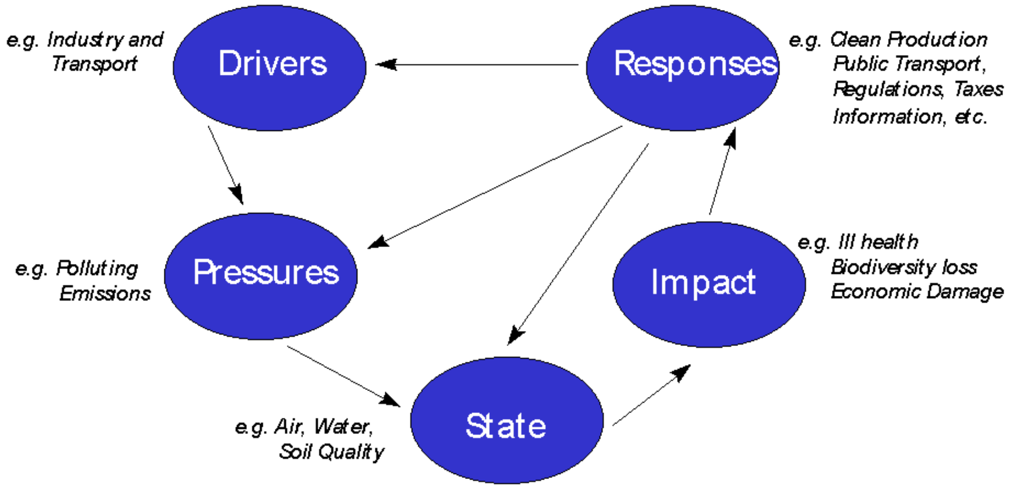


Figure 2.3. DPSIR framework for reporting on environmental issues (Source: EEA).

3 Definition of spatial assessment scales

This chapter deals with the question how to define the spatial scale for assessment areas that leads to meaningful information to support the management of European marine waters.

The general principles and considerations for the application of rules for geographic scaling have been discussed in a number of documents relating to the MSFD implementation and assessments of the marine environment, like the Task group reports drafted in 2010 (Cardoso *et al.* 2010, Cochrane *et al.* 2010, Ferreira *et al.* 2010, Galgani *et al.* 2010, Law *et al.* 2010, Olenin *et al.* 2010, Piet *et al.* 2010, Rice *et al.* 2010, Rogers *et al.* 2010, Swartenbroux *et al.* 2010, Tasker *et al.* 2010) and other documents (OSPAR 2011). The analytical report (see Appendix B) provides a detailed overview of existing methods of spatial scaling. This chapter discusses general principles, criteria for scaling and proposes steps for the definition of assessment areas.

3.1 MSFD requirements

The MSFD requires that good environmental status is determined at the level of the marine region or subregion (Art. 3.5), on the basis of the qualitative descriptors in Annex I of the MSFD. For the Baltic Sea and Black Sea GES should be determined at the level of the marine region. The North-East Atlantic Ocean and the Mediterranean Sea have each been divided into 4 subregions where GES should be determined:

a) the Baltic Sea
b) the North-east Atlantic Ocean
    - Macaronesia
    - Bay of Biscay and the Iberian coast
    - Celtic Seas
    - Greater North Sea
c) the Mediterranean Sea
    - Western Mediterranean Sea
    - Adriatic Sea
    - Ionian Sea and the Central Mediterranean Sea
    - Aegean-Levantine Sea
d) the Black Sea

3.2 General principles for the definition of assessment areas

The definition of assessment areas needs to address spatial scales at different levels.

The highest level is the level of the marine (sub)region. In some cases (some biodiversity issues, commercial fish stocks) the geographic assessment scale may exceed the scale of the (sub)region. However, in many cases, the scale of the regions and subregions is too large for meaningful assessments, as too large assessment areas will mask the more local pressures and their impacts and will not provide the information that is necessary to decide on management measures.

Thus in most cases, assessment and reporting need to be done at smaller scales. As stated in the Commission Decision (EC 2010), when the assessment needs to start at a relatively small spatial scale to be ecologically meaningful (for instance because pressures are localised), it could be necessary to scale up assessments to broader scales, such as at the levels of subdivisions, subregions and regions. The criteria to define smaller spatial scales are based on the specificities of a particular area or topic, which can be related to two perspectives:

- Management perspective

The assessment scales have to be chosen in such a way that assessments provide the right information to the process of policy implementation and management of marine areas. For this purpose, it is crucial that assessment areas are defined that provide a robust and adequate assessment of environmental state, and that enable the identification of management measures, and the evaluation of their effectiveness.

From a management perspective, subsidiarity between different policies, and the level of enforcement in different areas (e.g. difference between territorial waters and EEZ) must be considered as well. Already existing assessment areas for reporting under other EU Directives and for the RSCs may be useful for MSFD purposes as well. Examples of these are assessment areas that have been used in eutrophication assessments by HELCOM and OSPAR (see Annex §2.4).

Assessment areas must be designed in relation to risks for the marine environment, caused by the main drivers (D) and human activities, as mentioned previously. The impacts (I) of pressures (P) are generally greater near the source (either land-based or sea-based) and decrease with distance from the source. For static pressures like land-based sources of pollution, this means that there is a gradient of decreasing pressures and impacts from the coast to offshore areas. For mobile activities (e.g. shipping, trawling), pressures and impacts may be high at a small, local scale (‘grain’, e.g. direct impact of bottom trawling) while the human activity causing the pressure is found over a large extent. The density and intensity of human activities is generally higher near the coast as well. A finer spatial resolution of assessment areas may be required in coastal areas than in offshore areas where less human activities take place. However, it must be realized that offshore activities are increasing.

Assessments should make it possible to inform managers and policymakers on the environmental impacts of human activities, and link these impacts to pressures and activities. Through this link between pressures, state (S) and impacts, management responses (R) can be identified. Consequently, the spatial scale of assessments must, as much as possible, reflect those D-P-S-I-R relationships previously mentioned.

Too large areas can mask local pressures and their impacts, and are therefore not suitable for management purposes. On the other hand, too small areas result in a high monitoring burden, and may lead to inadequate assessments when the spatial distribution of ecosystem components is not sufficiently covered, hampering an evaluation of the wider effects or the cumulative impacts of local pressures.

A risk-based approach (Fig. 3.1) helps to prioritize areas and indicators for monitoring and assessment, by identifying areas that are vulnerable due to a combination of high sensitivity and high pressures. RSCs have used a risk-based approach already, with a higher density of monitoring stations and a smaller spatial scale of assessment areas in the coastal zone (for example HELCOM 2009a; OSPAR 2008). Transboundary effects of pressures have to be taken into account in a risk-based approach.

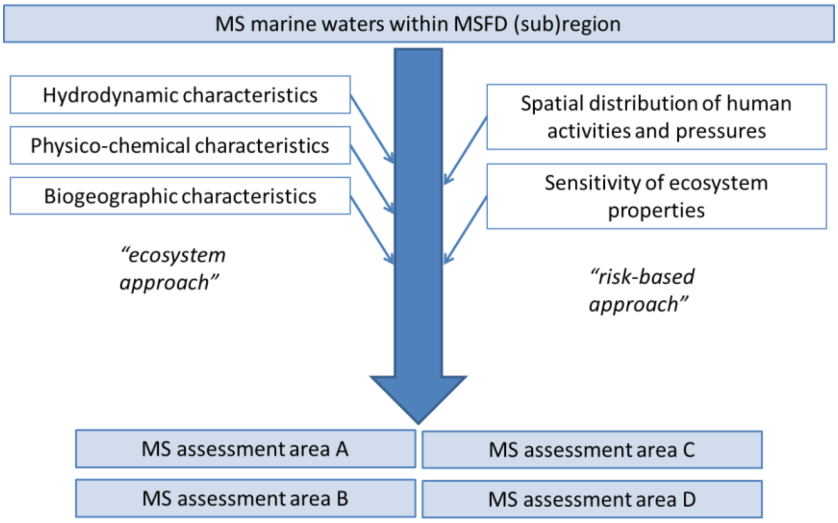


Figure 3.1 Schematic picture of the definition of assessment areas within a (sub)region. (MS: Member State). The two approaches in the scheme are not mutually exclusive

- Ecological considerations

At a spatial scale smaller than the marine region and subregion, relatively distinct ecological units can be delimited on the basis of their physical, chemical and biological characteristics (Figure 3.1). The further development of a marine ecosystem typology (Maes *et al.* 2013) may be helpful in the definition of these areas. For environmental assessments, a definition of smaller assessment areas at the level of metrics, indicators, criteria or descriptors may be necessary.

The MSFD indicates that hydrological, oceanographic and biogeographic features should be taken into account in defining the (sub)regions (Art. 3.2). Assessment areas within those (sub)regions can be further specified, first of all by using hydrological and oceanographic characteristics, in particular seawater temperature, salinity, mixing characteristics, frontal systems, turbidity (but also depth, currents, wave action and nutrient characteristics where appropriate) to define water masses of similar overall character within each (sub)region. In addition, biogeographic distribution patterns, related to benthic or pelagic habitats or marine populations are important and need to be taken into consideration when defining these assessment areas. The boundaries between such areas should wherever possible be based on marked changes in hydrodynamic, physic-chemical and biogeographic characteristics, but where changes are more gradual, more pragmatic factors such as the physiographic shape of the coastline and administrative boundaries may also be used, provided that the set of areas within a (sub)region overall are ecologically-based. The identification of a set of ecological assessment areas within a (sub)region provides the basis for assessment of ecosystem characteristics and habitats occurring within the area, as it provides a specific geographical area in which to determine the extent of impacts and whether GES and associated targets have been met (OSPAR 2012).

The features to define assessment areas are not all equally important for all descriptors, criteria and indicators. For descriptors like D5 (Eutrophication), D8 (Contaminants) and D9

(Contaminants in seafood) with (often) clearly localized sources of pollution (e.g. rivers or point sources), hydrodynamic characteristics play an important role. For descriptors like D1 (Biodiversity), D3 (Commercial fish and shellfish), D4 (Food webs) and D6 (Seafloor integrity) habitat patterns and biogeographic characteristics are often more important. For ecologically relevant scales of the latter descriptors, the assessment should cover the entire range of the species or of discrete populations (e.g. for large/mobile species). For habitats/communities it is most appropriate to assess the status within biogeographic zones, as functionally similar habitats can have wider distributions (Cochrane *et al.* 2010).

Activities may result in different types of pressures, e.g. both localised pressures and pressures operating at a larger spatial scale. For example, pressures and impacts arising from fisheries operate both at the larger scale of stocks of commercial species and at smaller, patchy scales in relation to physical impacts on the marine environment, like in the case of bottom trawling.

Concluding, ecological assessment areas must be defined in such a way that they adequately reflect both the ecological scales exhibited in each (sub)region and the links to areas which are effective for management measures. Ecosystem-based management is an integrated approach to management that consider the entire ecosystem including humans; It is 1) a spatial approach that builds around 2) acknowledging connections, 3) cumulative impacts and 4) multiple objectives (EEA 2014). This indicates that the risk-based approach is integrated in the ecosystem approach (Figure 3.1).

Size of assessment areas may vary from small areas for a specific biological feature to large areas relating to highly mobile species, homogenous habitats or large-scale food webs. On the basis of ecological considerations solely, assessment areas could be different between the various indicators and descriptors. Obviously, from a management and reporting perspective a large number of different scales will be less desirable, and a balance is needed to support the assessment process. The number of assessment areas should in principle be kept to a minimum to prevent an overly complicated assessment process (Cochrane *et al.* 2010).

3.3 Criteria for spatial assessment scales

As mentioned before, for the assessment of environmental status 29 criteria and 56 associated indicators have been developed in relation to the eleven descriptors from Annex I of the MSFD (EC 2010). The number of operational indicators may be even higher than 56 as in some cases several assessments for each indicator may have to be developed (Clausen *et al.* 2011). Assessment areas need to be defined in such a way that the assessments provide useful and adequate information but also keep the monitoring and reporting effort within reasonable limits. This paragraph provides a number of criteria for the definition of adequate spatial scales. The following paragraph will propose steps to keep the number of assessment areas manageable from the perspective of monitoring and reporting.

When looking at spatial scales, the characteristics of indicators, criteria and descriptors has to be taken into account. An often used approach is the distinction between pressure-related and state-related indicators and descriptors (e.g. EC 2011), although for many descriptors it should be realized that they contain a mixture of pressure, state and impact indicators. Nevertheless, the approach to spatial scales for pressures can be distinguished from the approach for state, and steps to define spatial scales could be different.

A number of criteria for defining scales for indicators, criteria and descriptors are proposed, based on the management perspective and the ecological considerations mentioned in the previous paragraph. In all cases there are two basic questions behind the definition of an assessment scale, 1) what feature is being assessed? and 2) what is the logical scale for the assessment?

From a management perspective, the definition of spatial scales can be linked to the risk-based approach which should assess the link between P-S-I criteria/indicators. In this perspective, issues like the spatial scale of pressures and impacts, the impacts of one single pressure on various indicators/descriptors, the cumulative impacts of pressures, trans-boundary problems and time scales of impacts should be considered. Some activities may result in both localised pressures and in pressures operating at a larger spatial scale. For example, pressures and impacts arising from fishing activity operate both at the larger scale of stocks of commercial species and at smaller, patchy scales in relation to physical impacts on the marine environment, like in the case of bottom trawling. This risk-based approach may be particularly relevant for Pressure-related criteria and indicators.

Criteria to be considered are:

- The intensity and the extent of the pressures, for example along the coastal zone in relation to hydrodynamic characteristics
  - Hydrodynamic characteristics (currents, transport patterns, mixing) in conjunction with the morphology of the coastal area may control the impacts of a pressure. This can be the case, for example, for eutrophication phenomena (D5) as well as the dispersal and concentration level of contaminants in water (D8) and biota (D9).
  - Assessments could then be focused on the areas with the highest pressure and highest risk, based on oceanographic characteristics. Examples of such an approach are the Physically Sensitive Area (PSA) (Ferreira *et al.* 2010), the EUTRISK indices developed by the JRC (Druon *et al.* 2004), and the assessment areas used for eutrophication assessments by HELCOM and OSPAR (HELCOM 2009a; OSPAR 2008). HELCOM recommends that assessment of eutrophication indicators may be most relevant at the sub-basin scale in the open sea combined with water body or water type level in the coastal zone (compatibility with WFD scales). OSPAR uses an area specific approach in eutrophication assessments, which takes into account hydrodynamic characteristics and the proximity to nutrient sources.
- The vulnerability of the ecosystem components to a pressure
  - The occurrence of sensitive ecosystem components in combination with the presence of high pressures identifies a risk. Examples of such components are particularly sensitive habitats like seagrass meadows, biogenic reefs or coralligenous habitats, or endangered or protected populations like sea turtles or marine mammals
- Cumulative impacts
  - Ecosystem components may be exposed to a range of pressures that have additive and/or cumulative impacts. Assessment of environmental status needs a defined scale relevant for that particular component. Within that scale the cumulative impact of the pressures must be assessed. There are several examples of tools that have been developed to identify and assess cumulative impacts at large scale within the MSFD (e.g. Andersen *et al.* 2013; Knights *et al.* 2013, Korpinen *et al.* 2013).
- Transboundary effects
  - If a water mass defined by hydrological and oceanographic characteristics covers an area that falls under the jurisdiction of several MS but is exposed to a similar pressure, the spatial scale should take into account the trans-boundary effects of this

  pressure. This may be particularly relevant for descriptors D5, D8, D10. There are many examples where, due to transport patterns, discharges of nutrients and contaminants from one source (e.g. a large river) may cause impacts at some distance from the source (the marine waters of neighbouring MS). Similarly, species distributions often cover the marine waters of various MS, where pressures on this species may occur at a local scale but have impacts on a much larger scale.

- Ecological and biogeographic characteristics
  - Some pressures may occur at a small spatial scale but still have the risk of large-scale impacts. An example is the assessment of the impacts of invasive alien species which generally should begin at the local scale, such as “hot-spots” and “stepping stone areas” for species introductions. Criteria like dispersal, vectors of introduction, pathways etc., are important factors that need to be assessed at a local or regional scale depending on the species biogeography (Zenetos *et al.* 2012).

These features are not all equally important for all descriptors, criteria and indicators. For descriptors like D5, D8 and D9 with (often) clearly localized sources of pollution (e.g. rivers or other point sources), hydrodynamic characteristics play an important role. But in other cases pressures may be widespread (e.g. noise related to ship traffic).

From an ecological perspective, the assessment scale may vary from small areas for a specific biological feature to large areas relating to highly mobile species, homogenous habitats or large-scale food webs. Consequently, assessment areas may need to encompass marine waters of several MS in some cases. For ecologically-relevant scales for the biodiversity descriptors, the assessment should cover the entire range of the species or of discrete populations (e.g. for large/mobile species). For habitats, biogeographic characteristics and patterns are important. For habitats/communities it is most appropriate to assess those within biogeographic zones, as functionally similar habitats can have wider distributions. In some cases, biodiversity-related components may exceed the scale of a (sub)region; this may be the case for groups like migratory birds, marine mammals and some (commercial) fish stocks (D3). Further development of methods for the assessment of biodiversity issues at regional scale is necessary (HELCOM 2009b).

The ecological perspective is, in particular, important for State-related criteria and indicators.

Criteria for defining scales could be related to the ecological considerations mentioned earlier, like:

- hydrological and oceanographic criteria
- biogeographical criteria

For biodiversity-related descriptors (like D1, D4) a suitable set of ecological assessment areas must be defined. The assessment scales should adequately reflect both the ecological scales of the biodiversity components (species, habitats, ecosystems) in each region/sub-region and the link to areas which are effective for management measures.

The outcomes of a status assessment are highly dependent on the geographical scale at which they are undertaken. Policies are often applied at specific geographic scales related to the scope of the policy or to national jurisdiction. The choice of an assessment scale should not lead to differences in status classifications for a species or habitat between different policy frameworks.

An example of how to take into account conservation priorities, biogeography and managerial issues in defining spatial scales for the mapping of three key Mediterranean habitats related to Descriptors D1 and D6 (i.e. seagrass *Posidonia oceanica* meadows, coralligenous formations, and marine caves) is the work of Giakoumi et al. (2013). Different scenarios were determined through a systematic planning approach dealing with large-scale heterogeneity, among which the scale of the whole Mediterranean basin and the ecoregion scale. Ecoregions are areas of relatively homogeneous species composition, clearly distinct from adjacent systems (Spalding *et al.* 2007). In the Mediterranean, eight ecoregions are defined, which are smaller units within some of the subregions (Alboran Sea, Algero-Provenzal Basin, Tyrrhenian Sea, Tunisian Plateau/Gulf of Sidra, Adriatic Sea, Ionian Sea, Aegean Sea including the Sea of Marmara,and Levantine Sea). The authors suggested that planning at the ecoregional level ensures better representativeness of the selected conservation features and adequate protection of species, functional, and genetic diversity across the basin.

In the approach towards spatial scales for the MSFD, environmental assessments for other EU legislation, like the WFD or the Bird and Habitat Directives should be considered. These Directives operate at different scales and the assessments under these Directives only apply to certain areas (for example, only coastal waters under the WFD), and additionally the Directives cover only some elements of GES. MSFD assessments should complement the other assessments for an efficient assessment of all Directives.

3.3.1 Matching pressures to state and impact.

The FP7 project ODEMM has developed a framework to link drivers, pressures, state and impacts in the marine environment (see http://www.odemm.com).

It provides an example of methods to establish links between P-S-I, which should be considered when defining spatial scales.

3.4 Steps towards defining spatial scale

As discussed above, there are many criteria to take into consideration when deciding on spatial assessment scales. In addition to the management perspective and ecological considerations discussed in §3.2, this also includes the question of the final objective of the assessment, i.e. what information is needed and who will use the information?

In addition to the “content-driven” approach, there is also the need to develop a system of assessment areas that is coherent, consistent and manageable.

And finally, there may be reasons to adapt the spatial scales for assessments over time. Autonomous ecological changes or ecological changes in response to management may occur that require an adaptation of the spatial scale (Cumming *et al.* 2006); similarly, changes in pressures (magnitude, extent) may result in a need to re-evaluate assessment scales. Hence, the choice of assessment scales needs to be part of adaptive management.

In theory, the criteria to define assessment areas could be applied to all 56 MSFD indicators mentioned in EC (2010) separately. This could result in 56 (or even more) different configurations of assessment areas, each of which suiting the exact needs of a specific indicator, and ranging in scale from small-sized assessment areas (like WFD water bodies) to assessment areas at the scale of the (sub)region.

It is clear that this would result in a high monitoring and management burden. A solution to keep this manageable is a nested hierarchical approach as the one developed by HELCOM.

In such an approach, different levels are nested within each other. At the largest scale, the delimitation must be compatible with the marine (sub)regions (MSFD Art. 4.2). An example is shown in Figure 3.2. Small-sized assessment areas (at the lower aggregation level) fit within larger-sized assessment areas (at the higher level). This approach shows that it is possible to aggregate the results of assessments at a small scale to an assessment at a larger scale and enables the definition of different scales depending on the nature of the component to be assessed.

In the approach developed by HELCOM (HELCOM 2013) five hierarchical spatial scales are defined:

1) the whole Baltic Sea,
2) a division of the Baltic Sea into 19 sub-basins defined by ecological criteria,
3) a further division of sub-basins into coastal and offshore areas,
4) a division of all areas by national boundaries,
5) a further division of the coastal areas to WFD water bodies/types

In HELCOM's view, the various hierarchical division levels can be used depending on the needs. The levels need further specification according to the particular state and pressure components to be assessed. For example, monitoring and assessment of mobile marine mammals such as grey seals may require the whole Baltic Sea scale while assessment of eutrophication indicators may be most relevant at the sub-basin scale in the open sea combined with water body or type level in the coastal zone. HELCOM recommends that the scale to be used should be chosen from the five possible scales (HELCOM 2013).

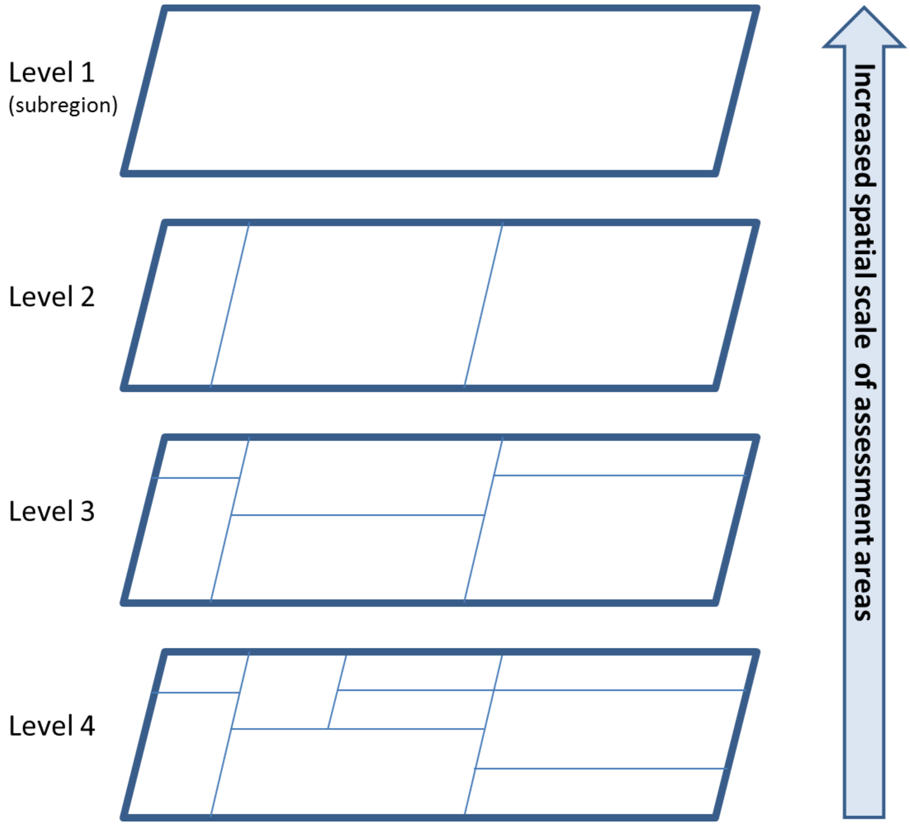


Figure 3.2 Schematic picture of the nested hierarchical definition of assessment areas

The approach by HELCOM is still under development and has not been applied yet to all MSFD indicators. As already discussed, depending on the character of the state and pressure descriptors/indicators a specific scale may be required, and this could easily result in a high number of different “configurations” of scales, which seems less desirable. A nested design of assessment scales in combination with a pragmatic optimization as part of an adaptive management approach to scaling would help to keep the number of assessment areas at a manageable level, using the following steps (Figure 3.3):

1 Define scales for ecosystem components related to state indicators and descriptors, using ecosystem characteristics as a basis, and taking into account the pressures on those state indicators (hydrological, oceanographic, biogeographic features). This can result in different choices for scales, for different indicators or descriptors.
2 Define scales for ecosystem components related to pressure indicators and descriptors (where necessary at smaller scales for local pressures). Again, this can result in different choices for scales, for different indicators or descriptors.
3 Consider assessment scales used in the framework of other policies (e.g. WFD, BD, HD, CFP).
4 Optimize the number of assessment areas to limit the monitoring and reporting burden, by combining assessment areas into one, nested, system consisting of a number of different levels of spatial scales.

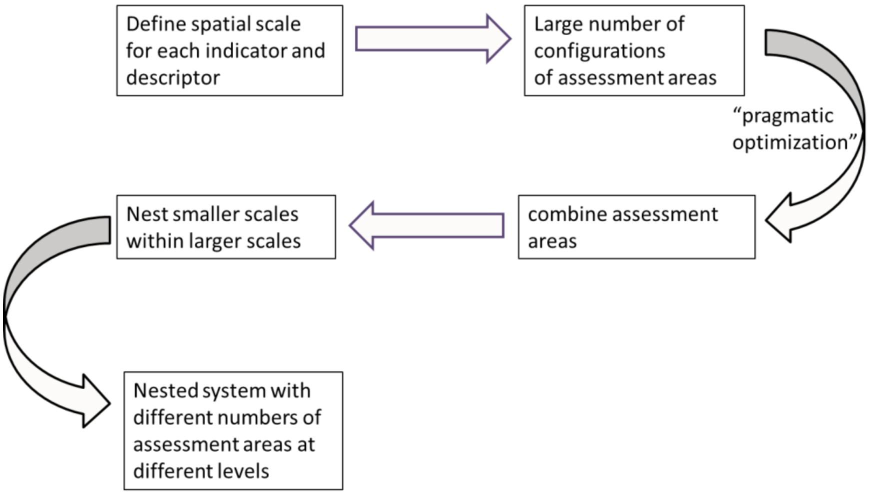


Figure 3.3 Schematic picture of a stepwise approach for the definition of assessment scales

3.5 Recommendations

- A decision is needed for each marine region and subregion on the relevant spatial assessment scale of indicators and descriptors, in particular in those cases where assessment scales go beyond the scale of national boundaries.
- We recommend to coordinate the choice of relevant spatial scales within each marine (sub)region or at EU level.
- A system of nested spatial scales can help to find a balance between a large number of assessment scales fit for the purpose of each specific assessment and an acceptable monitoring and reporting effort. We recommend to develop such a system for each marine region or subregion or at EU level.
- Guidance on assessment scales should be introduced into the GES Decision on criteria and methodological standards, should it be decided to revise the 2010 Decision.

4 Aggregation of assessments

This chapter discusses the different methods that can be applied to aggregate criteria and indicators within descriptors and across descriptors, and methods to aggregate assessments across assessment areas to eventually come to an assessment of GES for a geographic area. The analytical report (see Appendix B) provides a detailed overview of existing methods for aggregation. This chapter discusses general principles and criteria, and proposes steps for aggregation of assessments.

4.1 MSFD requirements

Article 3(4) of 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*".

Article 3(5) of the MSFD defines good environmental status,that "*shall be determined at the level of the marine region or subregion as referred to in Article 4, on the basis of the qualitative descriptors in Annex I"*.

To assess whether or not GES has been achieved, some aggregation may be needed within and across the 11 Descriptors to move from assessment at the level of indicators (the 56 indicators described in the Commission decision (EC 2010) to a more overall assessment of status, as mentioned also in Cardoso *et al.* (2010). This is also to ensure the assessments are coherent and comparable across the Member States of a (sub)region and at EU level.

4.2 Level of aggregation

As indicated by Cardoso *et al.* (2010) there are various levels at which assessment results can be aggregated. Based on the Management group report (Cardoso *et al.* 2010) and the criteria and indicators from the Commission Decision (EC 2010) at least four levels of aggregation or integration can be identified:

(i) Aggregation of metrics/indices within indicators;
(ii) Aggregation of indicators within the criteria of a Descriptor (for complex Descriptors);
(iii) Status across all the criteria of a Descriptor; and
(iv) Status across all Descriptors.

For each aggregation step it is essential to consider the purpose of the assessment (what is the question that needs to be answered) and the need for aggregation (does it supply the right answer). The level of aggregation depends on the type of information that is needed, as discussed in Chapter 2. The information need and the assessment that provides this information determine the level of aggregation, and should be considered also when deciding on the methods of aggregation. For some questions, aggregation up to the first and second level mentioned above could be enough, for example to inform on measures targeting the main pressures and impacts. The third and particularly the fourth aggregation level are more directed at information needs at the marine (sub)region and EU level. The choice for a level of aggregation also depends on the characteristics of the descriptor.

As one moves up the scale from metric/indicator level to overall GES, the diversity of features that have to be integrated increases rapidly. 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.

A relatively simple case of aggregation can be illustrated by an example given here for Descriptor 5 (Figure 4.1). This descriptor has two pressure indicators for criterion 5.1 ‘nutrient levels’, four state/impact indicators for criterion 5.2 ‘direct effects’ and two impact indicators for criterion 5.3 ‘indirect effects’. The first integration step mentioned above depends on the definition of each of the indicators. The second step, aggregating indicators within a criterion, and the third step, aggregating criteria within a descriptor, both require a decision on whether it is useful to aggregate to a higher level and a choice of the most appropriate aggregation rule. In the case of eutrophication, there are already existing assessment tools that apply those aggregation steps and can serve as an example (OSPAR 2008, HELCOM 2009a).

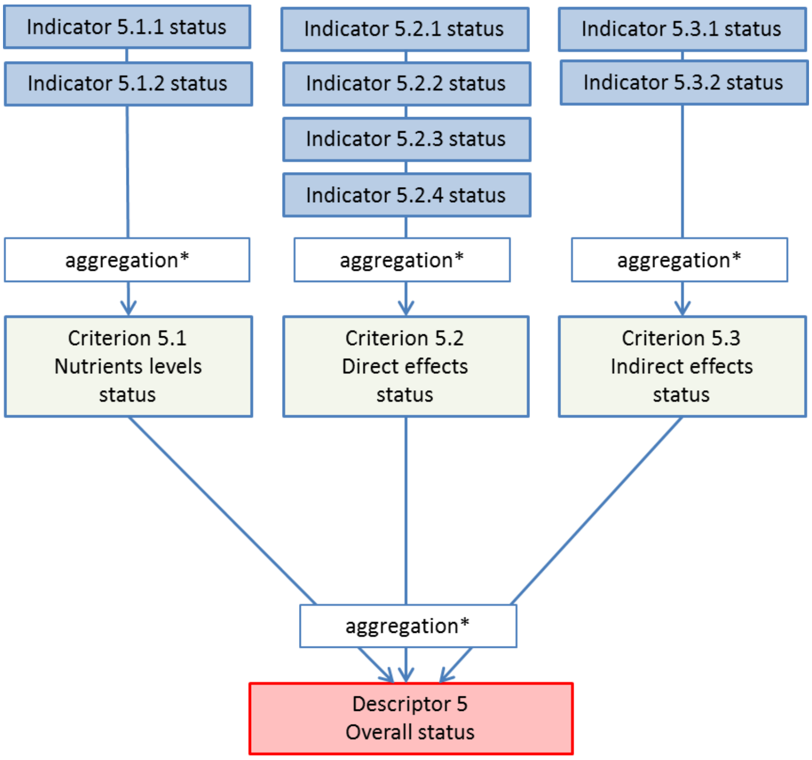


*Figure 4.1. Example of a potential approach for aggregation of indicators and criteria for D5. *All aggregation steps involve a decision on whether and how to aggregate.*

A much more complicated case of aggregation can be illustrated by an example for Descriptor 1. In the case of Descriptor 1, assessments are required at several ecological levels, viz. at species, functional/species group, habitat and ecosystem level. For the species and habitat levels, three criteria are mentioned in EC (2010), all but one consisting of several indicators. For the ecosystem level one criterion and one indicator is mentioned.

For the species level this means that, for each of the criteria, the second integration step consists of aggregation of the indicators within the criterion. As the indicators refer to species-specific characteristics, this aggregation can only be done for a selected species. Obviously, the third integration step, aggregation across the criteria can only be done for the selected species as well. This implies that, in the case of assessments for a number of species, an additional aggregation step can be identified, that consists of aggregating the results of the assessments for a number of species (Figure 4.2). There could also be an intermediate step consisting of aggregation of species within species groups or functional groups. For example, by using groups for highly mobile or widely dispersed species (birds, mammals, reptiles, fish, cephalopods) (EC 2012a) and other groups (phytoplankton, zooplankton, angiosperms, macroalgae and invertebrate bottom fauna) from MSFD Annex III Table 1. For all the aggregation steps, a decision is needed on the usefulness of aggregation and on the rules for aggregation.

The habitat level represents a similar case where various indicators and criteria apply to a specific habitat, and aggregation can be done for a specific habitat but also across habitats (Figure 4.3). Like for species, an intermediate step could consist of aggregating habitats within predominant habitat type or special habitat (Habitat Directive Annex I habitat (see EC 2012a for habitat types).

Finally, aggregation of all criteria within descriptor 1 would require combination of assessments at species, habitat and ecosystem level (Figure 4.4). If additional aggregation steps are used for species (within species/functional groups; not shown in Fig. 4.2) and/or habitats (within habitat types; not shown in Fig. 4.3), the figure will become more complex.

The examples presented here for D1 and D5 should merely be considered as illustrations of the range of options for aggregation within and across indicators and criteria and within descriptors. As stated before, challenges arise from the diversity and inherent nature of the metrics that are integrated.

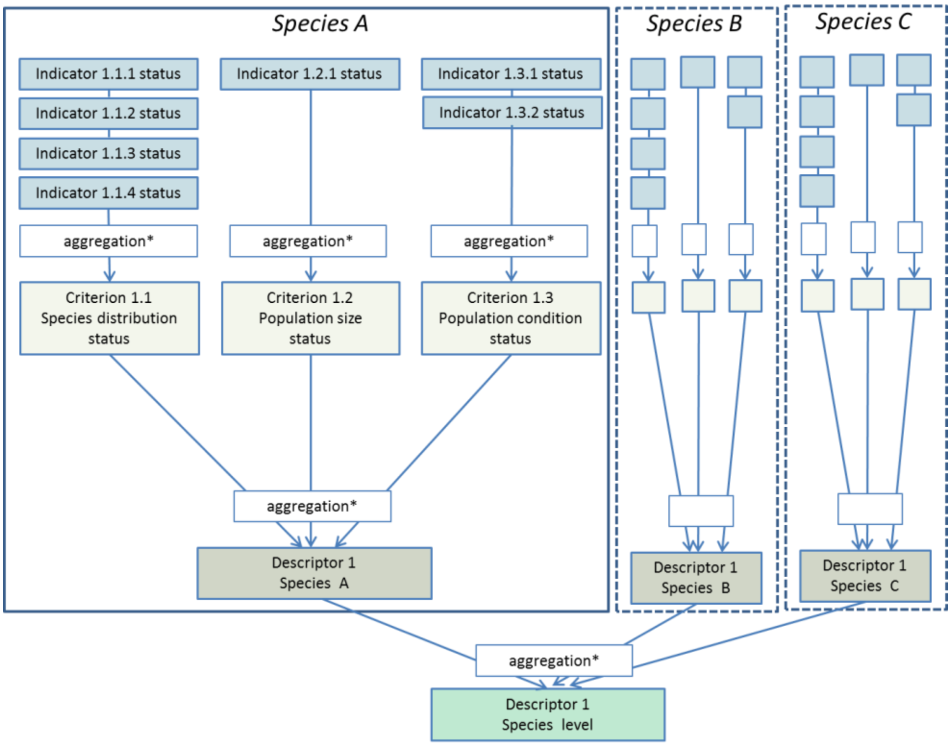


Figure 4.2. Example of a potential approach for aggregation of indicators and criteria for the species level of D1. *All aggregation steps involve a decision on whether and how to aggregate.An intermediate step, aggregating species within species or functional groups before aggregation at Descriptor level, could be added.

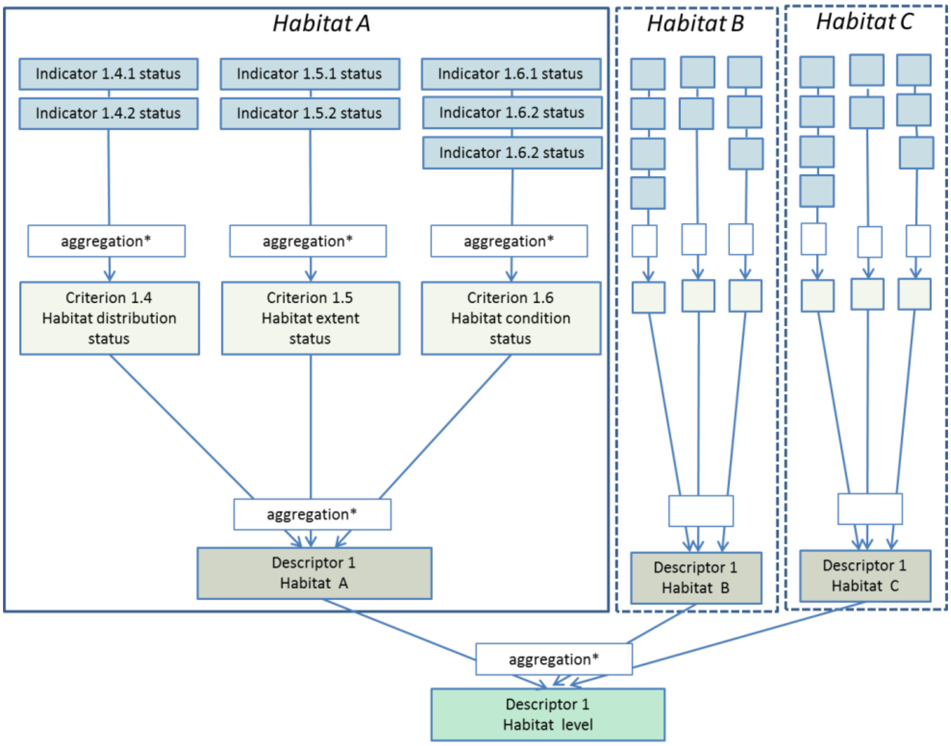


Figure 4.3. Example of a potential approach for aggregation of indicators and criteria for the habitat level of D1. *All aggregation steps involve a decision on whether and how to aggregate. An intermediate step, aggregating habitats within habitat types before aggregation at Descriptor level, could be added.

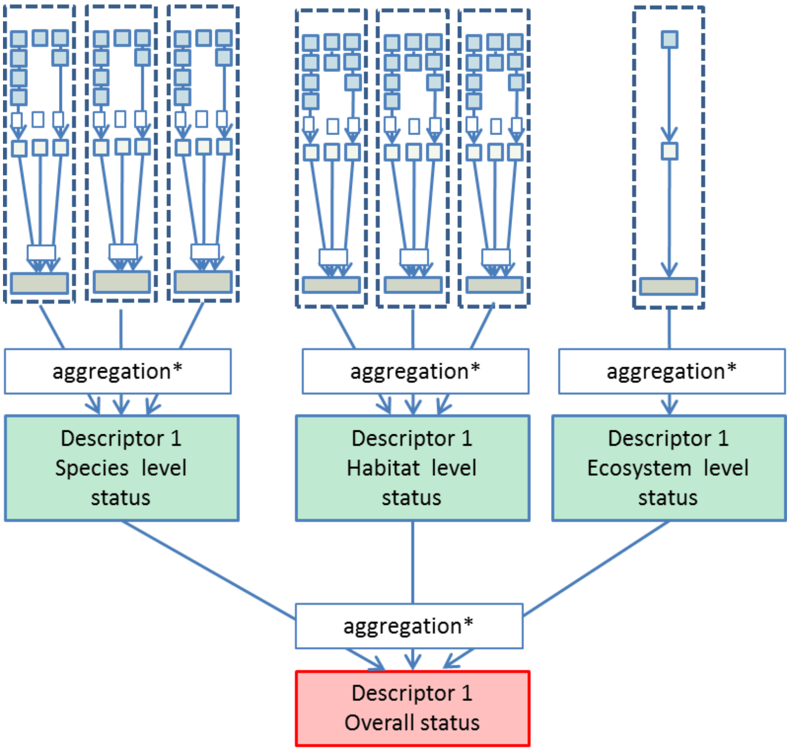


*Figure 4.4. Example of a potential approach for aggregation of indicators and criteria for D1. *All aggregation steps involve a decision on whether and how to aggregate. Intermediate steps for species and habitat groups could be added.*

4.3 General principles for aggregation

Based on a literature review, we identified a number of different approaches for aggregation rules that combine variables (which could be metrics, indicators, or criteria) into an overall assessment. Some of them have been used within the WFD, others within the Regional Sea Conventions and some others in the MSFD.

An overview of the methods is given in Table 4.1. A more detailed explanation of the methods can be found in the Analytical report (see Appendix B).

Table 4.1. Approaches for aggregation of different metrics, indicators or criteria to assess the status, including the advantages and disadvantages of each approach (see §4.3 in Appendix B for references).

General approach	Details of method	Advantages	Disadvantages
One-out all-out (OOAO) principle	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. May include double-counting.
	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
Conditional rules	A specific proportion of the variables have to achieve good status	Can help to focus on the key aspects	Assumes that GES is well represented by a selection of variables
Averaging approach	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. May hide hot spots of impacts
	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
Scoring or rating	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	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	Multivariate analyses	No need to set rigid target values, since values are represented within a domain	Results are hard to communicate to managers
Decision tree	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
High-level integration	Assessment results for some pre-defined groups (for example, biological indicators, hazardous substances indicators, supporting indicators, each applying OOAO).	Reduces the risks associated with OOAO while still giving an overall assessment	Technical details

4.3.1 One-out, all-out (OOAO)

The OOAO approach is used in the WFD to integrate within and across Biological Quality Elements (BQEs) (EC 2005). This approach follows the general concept that a particular status assigned to a water body depends on the quality element with the lowest status. The objective is to ensure "that the negative impact of the most dominant pressure on the most sensitive quality element is not averaged out and obscured by minor impacts of less severe pressures or by less sensitive quality elements responding to the same pressure." (EC 2012b)

A prerequisite for the combination of various parameters is that they are sensitive to the same pressure (Caroni *et al.* 2013). In that case, different methods can be used to combine parameters (medians, averages, etc.). Caroni *et al.* (2013) recommend an OOAO approach when aggregation 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 (EC 2005) also advises to use OOAO when combining parameters/indicators that are sensitive to different pressures.

Several criteria are suggested for cases where OOAO should be applied:
(i) when different pressures are addressed,
(ii) when there is an impact or risk of a future impact, and
(iii) when legal standards are involved (e.g. contaminants exceeding legal quality standards as under the WFD, species or habitats failing favourable conservation status under Birds or Habitat Directives, commercial fish stocks failing targets under the Common Fisheries Policy).

Note that rare species cannot easily be monitored, and consequently should be excluded from an OOAO approach.

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.
Sometimes multiple indicators are used to describe state/impact. While not all of those indicators may be equally important, this is done to include indicators that are used as supportive indicators, where P-S-I relations are uncertain. In those cases other aggregation rules than OOAO should be applied.

Borja *et al.* (2009b) 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 logical approach as a precautionary rule, in an ideal world where the status based on each BQE can be measured without error. 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 (Borja and Rodriguez 2010; Caroni *et al.* 2013). The OOAO rule results in very conservative assessments with a full implementation of the precautionary principle (Ojaveer and Eero 2011). In the case of the MSFD, with 11 descriptors and more than 50 indicators, the probability of not achieving good status becomes very high (Borja e*t al.*, 2013; Borja and Rodriguez, 2010; Caroni e*t al.*, 2013; Ojaveer and Eero, 2011) and, probably, unmanageable in practical managerial terms (Borja *et al.* 2013). Alternative methods for integrating multiple BQEs in the WFD are currently being considered (Caroni e*t al.*, 2013).

4.3.2 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 a good opportunity to use expert judgment when combining indicators, in a transparent way.

4.3.3 Averaging approach

The averaging approach is the most commonly used method to combine indicators (Shin *et al.* 2012) and consists of simple combinations of indicators, by using calculation methods like 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 for applying 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, application of the median of the indicator values resulted in very different assessment results compared to assessments based on averages.

The way the indicators are hierarchically arranged influences the assessment results as well, but 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 averages or medians. An adequate basis for assigning weights is not always available. 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.

4.3.4 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 weighted 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 (Birk *et al.* 2012) (see also next approach).
Another example is the method developed by Borja *et al*. (2010; 2011) for a cross-descriptor aggregation, combining the 11 descriptors of MSFD based on the WFD, HELCOM and OSPAR 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.

4.3.5 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 Task Group 6 report (Rice *et al.* 2010) recommends the use of multimetric indices or multivariate techniques for integrating indicators of species composition attributes of D6 such as diversity, distinctness, complementarity/(dis)similarity, 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 into 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 intercorrelated (see e.g. the discussion on the TRIX index in Primpas and Karydis (2011)).

4.3.6 Multidimensional approaches

The Task Group 6 report (Rice *at al.*, 2010) discusses multivariate methods as an alternative for multi-metric methods to combine a number of parameters. Multivariate methods, such as Discriminant Analysis or Factor Analysis combine parameters in a multi-dimensional space. For assessment purposes, such multidimensional spaces need to be classified into groups of GES and non-GES.

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). Distance from target is not easily expressed.

4.3.7 Decision tree

Decision trees provide the opportunity to apply different, specific, rules to combine individual assessments into an overall assessment, and give room for using expert judgment in a transparent way.

Borja *et al.* (2009a) describe a methodology that integrates several biological elements (phytoplankton, benthos, algae, phanerogams, and fishes), together with physico-chemical 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 those individual 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.

4.3.8 High-level aggregation

An example of a high-level aggregation, where assessments for several ecosystem components are merged into a final assessment, is the HELCOM-HOLAS project (HELCOM 2010a). 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 tool on the assessment results of those three groups for the final assessment (Figure 4.5). This approach could be considered a pragmatic compromise, reducing the risks associated with OOAO while still giving an overall assessment.

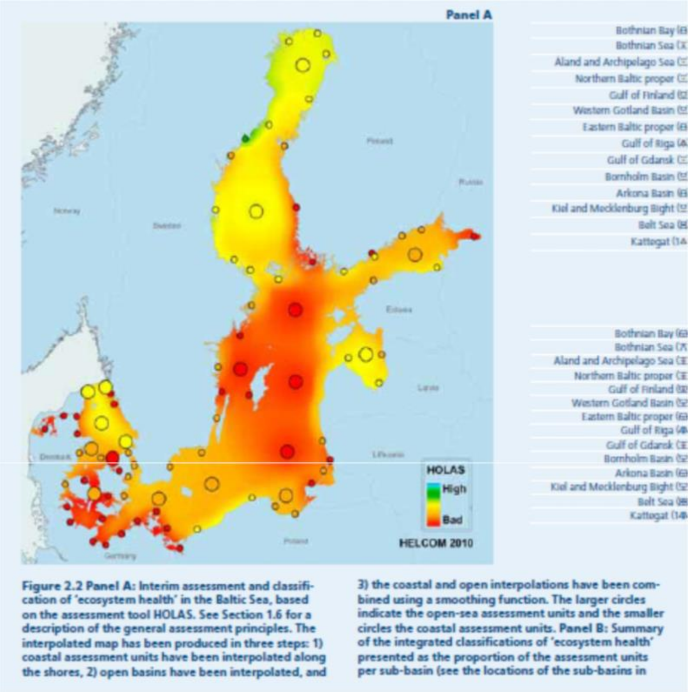


Figure 2.2 Panel A: Interim assessment and classification of 'ecosystem health' in the Baltic Sea, based on the assessment tool HOLAS. See Section 1.6 for a description of the general assessment principles. The interpolated map has been produced in three steps: 1) coastal assessment units have been interpolated along the shores, 2) open basins have been interpolated, and 3) the coastal and open interpolations have been combined using a smoothing function. The larger circles indicate the open-sea assessment units and the smaller circles the coastal assessment units. **Panel B:** Summary of the integrated classifications of 'ecosystem health' presented as the proportion of the assessment units per sub-basin (see the locations of the sub-basins in

Figure 4.5. Aggregation in HOLAS tool (HELCOM, 2010).

Borja et al. 2010 proposed an integrative method where indicators were weighted according to their importance in a certain area. In an application to the southern Bay of Biscay (Borja et al. 2011b), indicators were integrated within each descriptor, and each descriptor was weighted based on their relation to the most important human pressure in the area. The values of all descriptors, multiplied by their respective weighting factors, were combined to obtain a final value between 0 and 1, with 0 being the worst environmental status and 1 the best.

Another method, based more upon human activities and pressures, was developed by Halpern *et al.* (2012), and presents a high-level aggregation, at country level, using internationally available datasets (Ocean Health Index http://www.oceanhealthindex.org/).

Finally, there is a recent high-level aggregation example in Tett *et al.* (2013) for the North Sea. They identified ecosystem properties that can be considered equivalent to good environmental status (MSFD) or good ecological status (WFD). These properties include structure or organization, vigour, resilience and hierarchy. All the information from the different components are combined and synthesized for a holistic approach to assess the ecosystem health. To quantify this, they propose a method to follow the changes in ecosystem condition, taking into account the spatial extent of the system, aggregation across spatial and temporal scale, and following the development of selected variables in a multidimensional state space. This allows for changes in variables within defined boundaries and possibly different configurations (of variable values) representing a healthy ecosystem.

4.4 Criteria when to use specific rules

As shown in the previous section, the criteria to be used in aggregating values and assessing the environmental status are not easily defined.

From the lessons learned above, some guidance when using specific rules can be offered:

- Using OOAO:
    - It can be used when criteria from other EU directives are involved, e.g. contaminants exceeding legal quality standards, species or habitats failing favourable conservation status under Birds Directive or Habitats Directive, commercial fish stocks failing to achieve MSY under Common Fisheries Policy; however, lack of reliable assessment results, for example in the case of rare species under the BHD, or commercial fish stocks under the CFP[1] may limit the applicability of OOAO.
    - It can be used when the precautionary principle is applied (e.g. in the case when little information from only a few indicators is available)
    - It can be used when different pressures are addressed (but in that case other methods can be also used); it should not be used when various indicators are sensitive to the same pressure
    - It should not be used in cases where indicators have a high level of uncertainty. 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.
    - Consider using a less restrictive approach, like a ‘two out, all out’ approach in cases where several methods are combined in one assessment; for example, 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)
- Using the averaging, the scoring or the decision tree approaches:
    - Consider giving different weights for individual indicators/descriptors taking into account the relationship with the pressures within the assessment region/subregion. For example, in an area with high fishing pressure the indicators or descriptors that are most likely to be affected could be given a higher weight.
    - The decision tree approach can be used 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.
- Using multimetric and multivariate methods:
    - A multimetric method can be used when integrating several indicators of

---

[1] ICES has prepared draft advice on the application of Descriptor 3: http://www.ices.dk/sites/pub/Publication%20Reports/Advice/2014/Special%20Requests/EU_Draft_recommendations_for_the_assessment_of_MSFD_Descriptor3.pdf

    species composition or several indicators of eutrophication (e.g. in D1, D5, D6)
  - When using multivariate methods 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
- Using any of the described methods:
  - Using as many ecosystem components/indicators/criteria as reasonable and available will make the analysis more robust
  - Aggregation across state indicators and Descriptors (D1, D3, D4, D6) could be done differently than across pressure indicators or Descriptors (D2, D5, D7, D8, D9, D10, D11), giving higher weight to state-based descriptors.
  - State indicators could be distinguished in indicators for the species level, habitat level and ecosystem level.

4.5 Application of aggregation rules in assessments

In the WFD, the focus lies on a limited number of ecosystem components (the Biological Quality Elements), that are combined in a final assessment of ecological status through the OOAO approach. This can be considered a precautionary approach (Borja *et al.* 2010). In contrast to the WFD, the MSFD can be considered to follow a ‘holistic functional approach’, as it takes into account structure, function and processes of the marine ecosystem. The MSFD uses descriptors that not only relate to biological and physico-chemical indicators but also to pressure indicators (Borja *at al.,* 2010, 2013). The MSFD concentrates on the set of 11 descriptors which together summarize the way in which the whole system (ecosystem components, pressures) functions. The MSFD requires the determination of GES on the basis of the qualitative descriptors in Annex I, but does not specifically require one single GES 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 that are 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 *at al.*, 2013; Caroni *at 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, double counting should be avoided. For example, the assessments under D5 should be indicative of the level of eutrophication, or the assessments under D6 indicative of physical disturbance of the seafloor. The assessments of species or habitats under D1 take account of all impacts to which a species or habitat is subject, so the assessments for D5 or D6 would be an element of the overall assessment, but there may be others as well. For example, an assessment of shelf muds in the Baltic needs to take account of the extent of physical damage and of oxygen depletion from eutrophication, in order to assess whether the total extent of impact is within/outside acceptable levels set for GES. In this way the assessments of pressure-based descriptors can be an essential contribution to the biodiversity assessments and truly avoid double counting.

4.6 Aggregation of indicators and criteria (within a descriptor)

When choosing a level of aggregation and an aggregation method, the objective of the assessment and the level of information needed (as discussed in Chapter 2) have to be taken into consideration. Aggregation should not obscure understanding of the cause-and-effect relation between pressures and environmental state, and should result in assessment results that are informative to management and policy purposes.

The management group report (Cardoso *et al.* 2010) summarizes the methods in the ICES/JRC Task Group reports for a 'within Descriptor' integration, 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 GES will be set at the environmental status of the indicator and/or attribute assessed at the worst state for the area of concern).

Table 4.2 summarizes the approaches suggested in the individual Task Group reports to integrate attributes within each Descriptor; information on methods for integration of indicators can be found in the Task Group reports. In some cases, when proposing aggregation rules, the Task Groups deconstructed the ecosystem into ‘descriptor indicators’ and then recombined them to give a pass/fail for the GES, using in four of the cases the OOAO principle (Table 4.2). 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.

The method proposed by Tett *et al.* (2013), who assess the ecosystem health of the North Sea, using different attributes and components of the ecosystem in a combined approach may give a more accurate representation of the functioning of the ecosystem.

Table 4.2. Summary of Task Group approaches to Integrate Attributes within a Descriptor (Cardoso et al., 2010).

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

A range of methods for aggregation is discussed in §4.3 (and see Appendix B for further details). For some descriptors, aggregation of indicators and criteria may be relatively straightforward. This is probably the case for descriptors that consist of Pressure and Impact indicators and criteria, like D5, D7, D8, D9, D10 and D11 (EC 2011). For those descriptors, that can be linked to specific pressures, aggregation at descriptor level results in integrative assessments that still can be interpreted unequivocally. There are already existing assessment tools that perform such an aggregation, like assessment tools for eutrophication (Bricker *et al.* 2003, OSPAR 2008, HELCOM 2009a, Ferreira *et al.* 2011) and for contaminants (OSPAR 2009, HELCOM 2010b).

For other descriptors, in particular D1 and D4 that combine many different features and characteristics, aggregation of indicators and criteria is more complicated, as is already illustrated by the examples for D1 in §4.2. It requires careful consideration what level and method of aggregation supplies useful information.

If there is a need in those cases to integrate information at the level of a descriptor, there are some examples of possible approaches. One example is the biological valuation approach (Borja *et al.* 2011b) used to assess biodiversity by integrating several components (zooplankton, macroalgae, macroinvertebrates, fish, mammals 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). 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 MSFD requires that communities are in line with the prevailing physiographic, geographic and climatic conditions. Some habitats typically have highly diverse communities, other habitats harbour communities with a low diversity (for example, shallow habitats with a high level of natural disturbance through wave action and tidal currents). Those intrinsic differences caused by natural conditions between habitats can be incorporated in the valuation approach. The advantage of this method is that the current information used to valuate biodiversity can be adapted to the requirements of the MSFD indicators (probably using some consensus workshops to fix the terms of integration). Moreover, this method can avoid duplication of indicators in two descriptors (e.g. D1 and D6), since the metrics used could be different.

4.7 Aggregation across descriptors

The last level of integration relates to the methods that could be used to integrate the results across all Descriptors. Discussion on how to combine or integrate the results of each Descriptor into an overall assessment of GES 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 Cardoso *at al.* (2010) pointed out, cross-descriptor aggregation at the scale of marine (sub)regions runs the risk of blending and obscuring the information that is necessary to follow progress towards GES and to inform decision-makers about the effectiveness of policies and management. It may lead to masking of problems within specific descriptors, or to a high probability of not achieving GES if OOAO is used. The use of the OOAO principle across (or even within) descriptors also limits the information on the distance from target (GES), which may be crucial to assess the effectiveness of policies and the progress towards GES.

Table 4.3. Pros and cons of the decision of aggregating the information across descriptors

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

In the WG GES workshop on 23rd October 2013 the usefulness of aggregation of descriptors to one single value (overall GES assessment based on combination of the 11 descriptors) was discussed. Arguments against aggregating across descriptors raised during the workshop were that it may not be informative and may result in loss of information.

Again, as in all aggregation steps, a decision on aggregation across descriptors should be made taking into account the information needs and the level of aggregation that is adequate to fulfil those needs. There are clearly advantages and disadvantaged in aggregating across descriptors (Table 4.3).

Borja *at al.* (2013) describe 8 options to determine GES in a regional sea context (Table 4.4). The authors detail the concept behind these options, and propose the decision rule that is most adequate for the method to be implemented. In addition, they consider what type and amount of data are required, and 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 status in a robust way. Hence, the options from 1 to 8 proposed in Table 4.4 are sequentially less demanding of new data, and the degree of detailed ecological assessment is also lower.

As such, Option 1, which is most similar to the WFD approach, deconstructs GES into the 11 descriptors and then into the component indicators, assessing each for each area before attempting to produce an overall assessment (Table 4.4). However, having a complete dataset covering all descriptors and indicators for the assessment is difficult, and the use of pressure maps as a proxy of the status and impacts to marine ecosystems could be considered. Option 7, in contrast, only uses published data for the activities, and then infers a relationship between activity, pressures and impacts both on the natural and anthropogenic system. Between these extremes, there are several options to integrate and present information, each with its own requirements, pros and cons.

Table 4.4. Options for determining if an area/marine region is in Good Environmental Status (GES) (modified from Borja et al., 2013). Key: OOAO: 'one out, all out' principle.

Option	Decision rule	Data requirements	Pros	Cons	Examples in place
Either: 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: 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 2010a; Borja *et al.* 2011b; Aubry and Elliott 2006
Or: 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	None
Or: 4.create a synthesis indicator which takes the view that 'GES 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 on the way of integrating 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: 5.have a check-list (ticking boxes) of all the aspects needed	then if an area has e.g. more than 60% of the boxes ticked then it is in GES	An expert judgement approach, based on 'probability of evidence'	It may reflect the state of the science; if done rigorously then it may be the easiest to implement	It may be too subjective (i.e. based on soft intelligence)	Bricker *et al.* 2003; Ferreira *et al.* 2011
Or: 6.have a summary diagram such as a spiders-web diagram showing the 'shape of GES	The shape of the diagram		Easy to understand and show to managers	The decision on when GES is achieved	Halpern *et al.* 2012

Option	Decision rule	Data requirements	Pros	Cons	Examples in place
according to several headline indicators'					
Or: 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: 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

4.7.1 Application of OOAO for aggregation across descriptors

It could be argued that the 11 Descriptors together summarize the way in which the ecosystem functions. As MS have to consider each of the descriptors to determine good environmental status, this could be interpreted as a requirement to achieve GES for each of these descriptors. In that case, applying OOAO is the only aggregation method that can be applied to arrive at an overall assessment of GES.

This assumes that the 11 descriptors, and the indicators associated with this, can be considered a coherent and consistent framework that adequately reflects the environmental status and the MSFD Art. 3(5) definition. In that situation, state descriptors not achieving GES would always be accompanied by pressure descriptors not achieving GES. If this is not the case, for example if a descriptor like D5 or D8 that includes both Pressure and Impact indicators (EC 2011) shows that the level of the pressure is too high to achieve GES, while state descriptors like D1 or D4 do not reflect this, there is clearly an inconsistency in the assessment framework. 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. However, our current state of knowledge on quantitative causal relations between pressures, state and impacts in the marine environment is limited. In addition, nearly all ecosystem components are subject to the cumulative effects of many pressures related to a range of human activities (Knights *et al.* 2013). This means that, for some descriptors at least, there is a large scientific uncertainty associated with the definition and determination of GES. Consequently, developing a consistent assessment framework for all descriptors and indicators is an extremely challenging task.

4.7.2 Alternative approaches for aggregation across descriptors

In the October 2013 WG GES workshop aggregation of descriptors was discussed. Some Member States have suggested that an aggregation across the “biodiversity” descriptors (i.e. D1, D3, D4, D6) while splitting those descriptors in various groups (for example functional or species groups) might be an option. 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).

If integration across descriptors is decided as being necessary, Borja *et al.* (2010) suggest that the 11 descriptors are hierarchical and could have different weighting when assessing the overall GES. Borja *et al.* (2013) state that, for the descriptor Biodiversity to achieve GES, it requires all others to meet GES. Similarly, if one of the pressure-related descriptors fails then by definition the biodiversity descriptor should be adversely affected.

In addition to aggregating indicators and descriptors, there may be a need to integrate and geographically scale-up the assessments for some purposes, like reporting at marine (sub)region or EU level. This requires that assessments by MS are comparable in order to enable integration of the assessments into a region-wide assessment and to avoid cross-border anomalies (Borja *et al.* 2013). Comparable methods and aggregation rules to ensure minimum standards for GES reporting across MS could benefit from common principles (expanded from Claussen *et al.* (2011), as shown in Borja *et al.* (2013)):

I. The integration across levels of different complexity should accommodate different alternatives, i.e., integration below Descriptor level (across indicators within criteria, and criteria within Descriptors, as shown in the previous section) could differ from Descriptor level integration;

II. Integration across state Descriptors (D1, D3, D4, D6) could be done differently than across pressure Descriptors (D2, D5, D7, D8, D9, D10, D11), but avoid double counting of indicators in different descriptors (e.g. phytoplankton under D1 and D5, macroinvertebrates under D1 and D6). However, different aspects may be used, like for macroinvertebrates under D1 (e.g. rarity of species, endangered species, engineer species presence, etc.) and under D6 (e.g. ratio of opportunistic/sensitive, multimetric methods to assess the status, etc.).

III. Consideration of a different contribution of the two types of Descriptors for the overall GES evaluation – giving state Descriptors a higher weight, as receptors of the impacts produced by pressures. The rationale for this, as recognized by Claussen *et al.* (2011), is that "in principle, where GES for state-based Descriptors (D1, 3, 4, 6) are achieved it follows that GES for pressure-based Descriptors should also be met"; this makes the assumption that if the state is satisfactory then the pressures must be having a limited (or mitigated) impact. (see also discussion in §4.7.1).

Notwithstanding, in general and independently 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).

The outlined alternative approach 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 2010a) or radar schemes (Halpern *et al.* 2012) is still helpful and informative for ecosystem management, despite the involved loss of information. Information can be retained when always presenting the aggregated result together with the main underlying data, ideally visualizing the different levels of aggregation, allowing the tracing back of the status at any level and relating the status with the actual pressures that lead to the synthesized value.

An example of a way to present aggregated information is the Ocean Health Index (Halpern *et al.,* 2012), that provides weighted index scores for environmental health, both a global area-weighted average and scores by country (see www.oceanhealthindex.org). The index consists of a radar plot showing the maximum possible score for each goal, and a goal's score and weight. This kind of integration could be adapted for the MSFD, integrating at the level of region or subregion, but also showing the values within each descriptor. Methods are in development in the FP7 EU project DEVOTES: www.devotes-project.eu). This presentation would still allow managers to get information and take actions at different levels: small (or local) scale, (sub)region scale, information for each descriptor, etc., while also providing integrated overviews.

Another example, applied specifically for the MSFD, using all descriptors and most of the indicators, can be consulted in Borja *et al.* (2011b). The authors studied a system in which the main pressure for the whole area is fishing, whilst some other pressures such as waste discharges act at a local level. Although the global environmental status of the area can be considered good after the integration of all indicators and descriptors, two of the descriptors (fishing and food webs) are not in good status (Table 4.5). Interestingly, biodiversity is close to the boundary of the good status, meaning that the system could be unbalanced by fishing, affecting in different degree several biological descriptors. This means that the pressure must be managed to avoid problems in the future, especially because the descriptors already in not good status show a negative trend.

Table 4.5. 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). Key: EQS- Environmental Quality Standards; EQR-Ecological Quality Ratio, both based upon the Water Framework Directive (WFD); NA: not available. Trends: red color, negative; green color, 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	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	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**

4.8 Proposed steps for aggregation

As a possible approach for the aggregation of assessments we propose the following steps (Figure 4.5):

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 many levels of spatial assessment that was described as a nested approach (Step 1 - spatial scales) (see chapter 3.4 for a stepwise definition of assessment areas scales).

Within one descriptor, this could result in a number of assessments for the different indicators, that all use the same scales for assessment areas. This could be the case for descriptors like D5 and D8 (see for example OSPAR and HELCOM assessments for eutrophication and hazardous substances). 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). Rules for this aggregation step are discussed in Chapter 3. These steps are already commonly used procedures, for example in OSPAR and HELCOM assessments for eutrophication and contaminants.

For some of the other descriptors, the spatial scales for indicators may not be the same for all indicators. This could be the case for biodiversity, for example, where depending on the species, habitat or functional group a different spatial scale may be used. In that case, a lower integration level than the descriptor level could be chosen. Integration of different ecosystem components and functional groups in an overall assessment for biodiversity is an issue where methods need further development.

Aggregation up to this level (Step 2) 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).
At the level of marine regions and subregions and at European level, there is a need to present information at a higher level of aggregation, to provide an overview of the current status of the environment and the progress towards GES. The following aggregation steps could provide this higher level of information:

Within a descriptor, the assessment results of all assessment areas within a (sub)region can be presented in a more integrated way (<u>Step 3 - spatial aggregation</u>). This can be done in different ways, e.g. (see chapter 3, analytical report for spatial aggregation rules)

- Use OOAO (if one assessment area fails GES, the whole (sub)region fails)
    - Not useful, as it gives a very conservative result and is not informative
    - In some cases, for example if a pressure is more or less homogeneous across a whole (sub)region (fishing, shipping), it could be useful to apply OOAO
    - If the pressure is highly localized this approach is not adequate, since the whole (sub)region could fail GES for a single location (which, of course will need specific management measures).
- Percentage of surface area achieving GES
    - 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 (sub)region, but the surface area not achieving GES is only 2%, it could be concluded that the (sub)region does not achieve GES in 2% of its area, where management measures are needed.
        - For some descriptors, surface area may be a good measure to express status at the level of a (sub)region: for example, habitats under D1, D5, D6, D8, and D10.
- Other metrics
    - For other descriptors, surface area may not be suitable for all criteria and indicators, but other metrics should be considered, e.g.:
        - D1: numbers of species failing to achieve favourable conservation status
        - D3: number of stocks failing to meet MSY

The end result of <u>Step 3</u> could present the level at which GES is achieved at the scale of a marine (sub)region as a pie chart or something equivalent. In some cases, assessment results may only be represented at the level of regions and subregions. An example could be the percentage of commercial fish stocks achieving MSY. In that case, steps 1 and 2 could be skipped.

The aggregation results of Step 3 could be aggregated across descriptors in a final presentation per (sub)region, using methods like radar plots, or methods similar to the Ocean Health Index (<u>Step 4 - aggregation across descriptors</u>). In this step, weighted approaches as suggested in Chapter 4.6 could be considered.

An important point in all those aggregation steps, is that the aggregation methods should be transparent, and it should always be possible to “disaggregate”, i.e. go back from aggregated levels at larger spatial scales and higher levels of integration. This is necessary to trace down the causes of an assessment result.

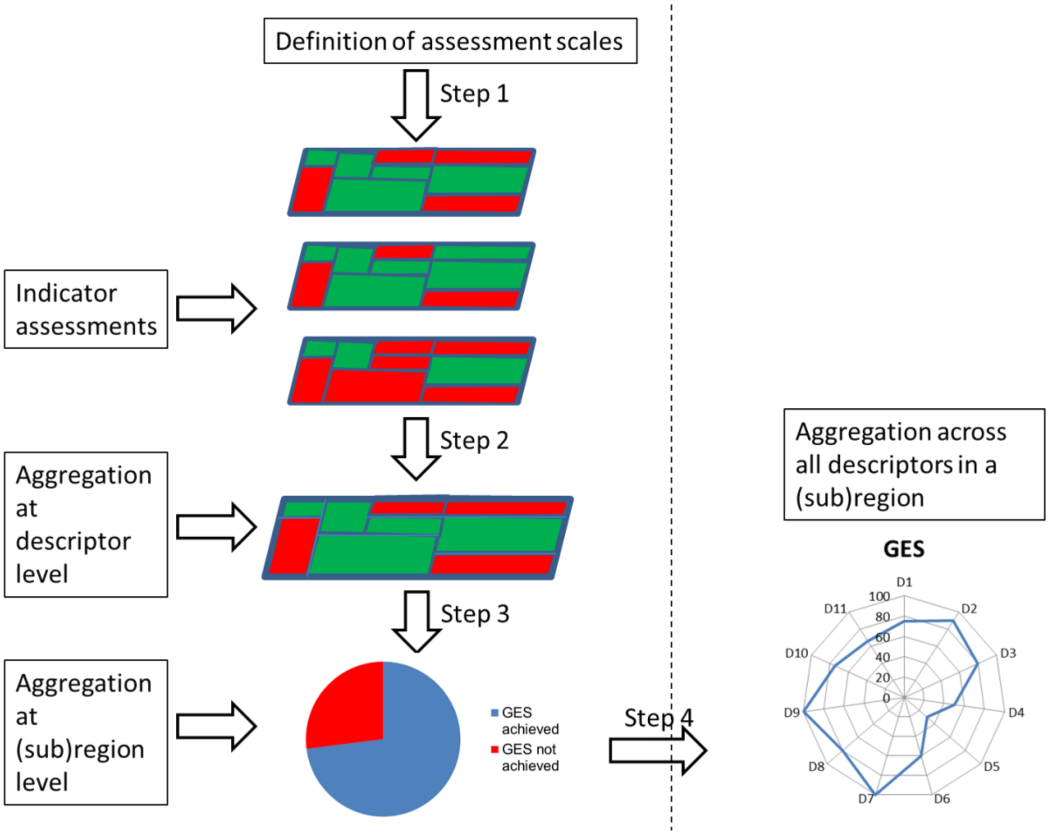


Figure 4.5. Schematic view of steps for aggregation towards an assessment at the level of a marine (sub)region

4.9 Recommendations

- The choice of a method for aggregation depends on the objective of the assessment.
- A decision is needed in what cases aggregation of assessments up to descriptor level can be done.
- A decision should be taken about the need and objective (compliance, monitoring progress, communication, etc.) for aggregation of assessments across descriptors.
- Regional coordination is needed to improve comparability of assessments and enable aggregation of assessments at the scale of a (sub)region.

5 Recommendations for further work

This report has identified generic approaches and criteria to deal with the spatial scales of assessments and the aggregation of assessments. There are still many questions and knowledge gaps that need to be answered before a further specification of approaches for geographic assessment scales and aggregation rules is possible.

Consequently, this report forms a first step towards the development of guidelines that can be applied in practice. The large diversity between member states in both the approach towards assessment scales as well as the use of indicators and aggregation methods stress the need for further work.

We recommend carrying out a number of pilot projects where practical approaches for the definition of assessment scales and for the application of aggregation rules can be further developed and tested. It will be useful to apply these pilot projects in different marine regions and focussed on different types of indicators and descriptors.

Below, a number of specific issues is discussed where further work is needed.

Aggregation of biodiversity related indicators
The indicators under the Biodiversity related descriptors address many different features. For Descriptor 1 this concerns characteristics for species (distribution, population size, population distribution), for habitats (distribution, extent, condition) and for the ecosystem (structure). For Descriptor 4, which is supposed to concern functional aspects of the food web, the indicators address an even wider variety of functional and structural aspects. In addition, the descriptors deal with many different species and functional groups. Although there are examples of indicators to describe biodiversity for specific functional groups (e.g. phytoplankton, benthic fauna), and there have been some attempts to integrate biodiversity elements within one assessment (see §4.5), there is still considerable development needed to solve questions on when and how aggregation is useful.

Time scales
In this report we have not considered the issue of time scales. In the design of monitoring programmes and in the assessment of environmental status, temporal scales are just as important as spatial scales, and choices for both scales can strongly influence the outcomes of the assessment. Time scales and spatial scales of assessments are closely related, and the choice for a specific time scale may have consequences for the spatial scale. In practice, it will be difficult to decide on appropriate spatial scales without considering the temporal scales at the same time. This topic clearly requires more work.

Application of the nested approach of assessment areas
We propose a nested approach for the definition of spatial assessment scales as a way to develop a framework of assessment areas that can be adapted to the specific needs of an indicator or descriptor and the specific characteristics of a regional sea, and can help to develop a coherent approach within a regional sea. This approach has been developed by HELCOM and has been applied in some assessments. However, the nested approach has not been developed yet to suit all indicators and descriptors of the MSFD in the Baltic Sea. The approach has not been developed in the other regional seas, either. The practical implementation of this approach will need further development of criteria and methods, to

promote a common approach that improves coherence across Europe while leaving enough room to take into account regional differences.

Uncertainty in data and assessment results

The applicability of an aggregation method is not only determined by the characteristics of the indicators that are involved. The reliability of the underlying data and methods, and thus the uncertainty in assessment results, should be considered as well. There is a risk of misclassification associated with the uncertainty in assessment results, and the consequences of specific aggregation methods for this risk should be evaluated. Within the scope of this project it has not been possible to deal with this topic, but it is recommended to include this in further work.

Metrics to represent GES at an aggregated level and large spatial scale

For an aggregated representation of GES at a large geographic scale, for example at the level of a subregion, metrics are required that are informative about the achievement of GES. As discussed in §4.7, if OOAO is applied a whole subregion would be flagged as not achieving GES if one of the assessment units within the area would fail to achieve GES. However, other methods are probably more informative. As suggested, the percentage of the total surface area that has achieved GES may be a useful metric as it indicates the extent to which GES is achieved. This approach is probably less useful in those cases where environmental problems can be linked less clearly to specific areas. An example of the latter case could be some of the criteria and indicators under Descriptor 3. If various commercial fish stocks fail to achieve MSY, but these fish stocks have different geographic ranges, the percentage of surface area within a (sub)region achieving GES would not be a suitable metric. As an alternative in this case, we suggest to use the percentage of stocks meeting MSY as a more suitable metric at (sub)region level. Similar questions concern other descriptors, such as D2, D4, D7 and D11. In general, further work is needed to explore whether an approach focusing on surface area or alternatives using other metrics gives the most adequate description of GES at a large spatial scale.

Additionally, various other options exist to combine descriptors and represent GES, as discussed in §4.6. The potential of these methods needs further exploration.

6 References

Andersen, J.H., A. Stock, S. Heinänen, M. Mannerla and M. Vinther (2013). Human uses, pressures and impacts in the eastern North Sea. *Technical Report from DCE – Danish Centre for Environment and Energy,* Aarhus University, DCE – Danish Centre for Environment and Energy, 136 pp.

Aubry, A. and 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.

Birk, S., W. Bonne, A. Borja, S. Brucet, A. Courrat, S. Poikane, A. Solimini, W. van de Bund, N. Zampoukas and 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., V. Valencia, J. Franco, I. Muxika, J. Bald, M.J. Belzunce and 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., J. Bald, J. Franco, J. Larreta, I. Muxika, M. Revilla, J.G. Rodriguez, O. Solaun, A. Uriarte and V. Valencia (2009a). Using multiple ecosystem components, in assessing ecological status in Spanish (Basque Country) Atlantic marine waters. *Marine Pollution Bulletin* 59: 54-64.

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

Borja, A., M. Elliott, J. Carstensen, A.S. Heiskanen and 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. and J.G. Rodriguez (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., E. Barbone, A. Basset, G. Borgersen, M. Brkljacic, M. Elliott, J.M. Garmendia, J.C. Marques, K. Mazik, I. Muxika, J. Magalhaes Neto, K. Norling, J.G. Rodriguez, I. Rosati, B. Rygg, H. Teixeira and 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 and 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. Neto, H. Teixera, L. Uusitalo, M.C. Uyarra and N. Zampoukas (2013). Good Environmental Status of marine waters: What is it and how do we know when we have attained it? *Marine Pollution Bulletin* 76: 16-27.

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

Cardoso, A.C., S. Cochrane, H. Doerner, 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. Tasker and W. van de Bund (2010). Scientific Support to the European Commission on the Marine Strategy Framework Directive – Management Group Report.

EUR – Scientific and Technical Research series, Luxembourg, JRC/ICES, EUR 24336 EN, 57 pp.

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

Clausen, U., G. Casazza, D. Connor, L. De Vrees, J.M. Leppänen, J. Percelay, M. Kapari, O. Mihail, G. Ejdung, J. Rendall and R. Emmerson (2011). Draft Common Understanding of (Inital) Assessment, Determination of Good Environmental Status (GES) & Establishment of Environmental Targets (Articles 8, 9 & 10 MSFD). Version 2 - 23 September 2011, 58 pp.

Claussen, U., G. Casazza, D. Connor, L. de Vrees, J.M. Leppänen, J. Percelay, M. Kapari, O. Mihail, G. Ejdung, J. Rendall and R. Emmerson (2011). Common Understanding of (Initial) Assessment, Determination of Good Environmental Status (GES) and Establishment of Environmental Targets (Art. 8, 9 & 10 640 MSFD). Brussels, WG GES Drafting Group, Version 2, 23 September 2011, 58 pp.

Cochrane, S.K.J., D.W. Connor, P. Nilsson, I. Mitchell, J. Reker, J. Franco, V. Valavanis, S. Moncheva, J. Ekebom, K. Nygaard, R. Serrão Santos, I. Naberhaus, T. Packeise, W. van de Bund and A.C. Cardoso (2010). Marine Strategy Framework Directive – Task Group 1 Report Biological diversity. *EUR – Scientific and Technical Research series,* Luxembourg, JRC/ICES, EUR 24337 EN, 111 pp.

Cumming, G.S., D.H.M. Cumming and C.L. Redman (2006). Scale mismatches in social-ecological systems: causes, consequences, and solutions. *Ecology and Society* 11: 14. [online].

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 and S. Degraer (2007). A concept for biological valuation in the marine environment. *Oceanologia* 49: 99-128.

Druon, J.N., W. Schrimpf, S. Dobricic and A. Stips (2004). Comparative assessment of large-scale marine eutrophication: North Sea area and Adriatic Sea as case studies. *Marine Ecology Progress Series* 272: 1-23.

EC (2005). Overall Approach to the Classification of Ecological Status and Ecological Potential. *WFD CIS Guidance Document,* Luxembourg, 13, 48 pp.

EC (2010). Commission Decision of 1 September 2010 on criteria and methodological standards on good environmental status of marine waters. E. Commission. 2010/477/EU.

EC (2011). Commission Staff Working Paper. Relationship between the initial assessment of marine waters and the criteria for good environmental status. Brussels, 14.10.2011, SEC(2011) 1255 final, 95 pp.

EC (2012a). Guidance for 2012 reporting under the Marine Strategy Framework Directive, using the MSFD database tool. . Brussels, Version 1.0. DG Environment, 164 pp.

EC (2012b). Commission staff working document. European Overview, accompanying the document Report from the Commssion to the European Parliament and the Council on the Implementation of the Water Framework Directive (2000/60/EC) River Basin Management Plans. Brussels, 14.11.2012, SWD(2012) 379 final 257 pp.

EEA (1999). Environmental indicators: Typology and overview. Copenhagen, European Environment Agency, Technical report No 25, 19 pp.

EEA (2014). Marine messages. Our seas, our future - moving towards a new understanding. Copenhagen, European Environment Agency, TH-02-14-122-EN-C, 29 pp.

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 and U. Claussen (2010). Marine Strategy Framework Directive - Task Group 5 Report Eutrophication. *EUR - Scientific and Technical Research series,* Luxembourg, JRC/ICES, EUR 24338 EN, 49 pp.

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 and 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.

Gabrielsen, P. and P. Bosch (2013). Environmental Indicators: Typology and Use in Reporting. Copenhagen, European Enmvironment Agency, EEA internal working paper, 20 pp.

Galgani, F., D. Fleet, J. Van Franeker, S. Katsanevakis, T. Maes, J. Mouat, L. Oosterbaan, I. Poitou, G. Hanke, R. Thompson, E. Amato, A. Birkun and C. Janssen (2010). Marine Strategy Framework Directive - Task Group 10 Report Marine litter. *EUR - Scientific and Technical Research series,* Luxembourg, JRC/ICES, EUR 24340 EN, 48 pp.

Giakoumi, S., S. M., V. Gerovasileiou, T. Mazor, J. Beher, H.P. Possingham, A. Abdulla, M.E. Çinar, P. Dendrinos, A.C. Gucu, A.A. Karamanlidis, P. Rodic, P. Panayotidis, E. Taskin, A. Jaklin, E. Voultsiadou, C. Webster, A. Zenetos and S. Katsanevakis (2013). Ecoregion-based conservation planning in the Mediterranean: dealing with large-scale heterogeneity. *PLoS One* 8(10): doi:10.1371/journal.pone.0076449.

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 and 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 and 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 Environment Proceedings,* Helsinki, No. 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 Environment Proceedings,* Helsinki, No. 116B, 192 pp.

HELCOM (2010a). Ecosystem Health of the Baltic Sea 2003–2007: HELCOM Initial Holistic Assessment. . *Baltic Sea Environment Proceedings,* Helsinki, HELCOM, No. 122., 63 pp.

HELCOM (2010b). Hazardous substances in the Baltic Sea – An integrated thematic assessment of hazardous substances in the Baltic Sea. *Baltic Sea Environment Proceedings,* Helsinki, 120B, pp.

Knights, A.M., R.S. Koss and 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.

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

Korpinen, S., M. Meidinger and M. Laamanen (2013). Cumulative impacts on seabed habitats: An indicator for assessments of good environmental status. *Marine Pollution Bulletin*.

Law, R., G. Hanke, M. Angelidis, J. Batty, A. Bignert, J. Dachs, I. Davies, Y. Denga, A. Duffek, B. Herut, K. Hylland, P. Lepom, P. Leonards, J. Mehtonen, H. Piha, P. Roose, J. Tronczynski, V. Velikova and D. Vethaak (2010). Marine Strategy Framework Directive – Task Group 8 Report Contaminants and pollution effects. *EUR – Scientific and Technical Research series,* Luxembourg, JRC / ICES, EUR 24335 EN, 161 pp.

Maes, J., A. Teller, M. Erhard, C. Liquete, L. Braat, P. Berry, B. Egoh, P. Puydarrieux, C. Fiorina, F. Santos, M.L. Paracchini, H. Keune, H. Wittmer, J. Hauck, I. Fiala, P.H. Verburg, S. Condé, J.P. Schägner, J. San Miguel, C. Estreguil, O. Ostermann, J.I. Barredo, H.M. Pereira, A. Stott, V. Laporte, A. Meiner, B. Olah, E. Royo Gelabert, R. Spyropoulou, J.E. Petersen, C. Maguire, N. Zal, E. Achilleos, A. Rubin, L. Ledoux, C. Brown, C. Raes, S. Jacobs, M. Vandewalle, D. Connor and G. Bidoglio (2013). Mapping and Assessment of

Ecosystems and their Services. An analytical framework for ecosystem assessments under action 5 of the EU biodiversity strategy to 2020. Luxembourg, Publications office of the European Union, pp.

Maxim, L., J.H. Spangenberg and M. O'Connor (2009). An analysis of risks for biodiversity under the DPSIR framework. *Ecological Economics* 69: 12-23.

OECD (1993). OECD core set of indicators for environmental performance reviews: A synthesis report by the group on the state of the environment. Paris, Organisation for Economic Co-operation and Development, OECD Environment Monographs No. 83, , 39 pp.

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

Olenin, S., F. Alemany, A.C. Cardoso, S. Gollasch, P. Goulletquer, M. Lehtiniemi, T. McCollin, D. Minchin, L. Miossec, A. Occhipinti Ambrogi, H. Ojaveer, K. Rose Jensen, M. Stankiewicz, I. Wallentinus and B. Aleksandrov (2010). Marine Strategy Framework Directive – Task Group 2 Report Non-indigenous species. *EUR – Scientific and Technical Research series,* Luxembourg, JRC/ICES, EUR 24342 EN, 44 pp.

OSPAR (2008). Eutrophication Status of the OSPAR Maritime Area. Second OSPAR Integrated Report. London, OSPAR, Publication Number 372/2008, 108 pp.

OSPAR (2009). CEMP assessment report: 2008/2009. Assessment of trends and concentrations of selected hazardous substances in sediments and biota. *Monitoring and Assessment Series,* London, OSPAR, Publication Number: 390/2009, 79 pp.

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 and 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.

Piet, G.J., A.J. Albella, E. Aro, H. Farrugio, J. Lleonart, C. Lordan, B. Mesnil, G. Petrakis, C. Pusch, G. Radu and H.-J. Rätz (2010). Marine Strategy Framework Directive – Task Group 3 Report Commercially exploited fish and shellfish. *EUR – Scientific and Technical Research series,* Luxembourg, JRC/ICES, EUR 24316 EN, 82 pp.

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

Rice, J., C. Arvanitidis, A. Borja, C. Frid , J. Hiddink , J. Krause, P. Lorance, S.Á. Ragnarsson, M. Sköld and B. Trabucco (2010). Marine Strategy Framework Directive – Task Group 6 Report Seafloor integrity. *EUR – Scientific and Technical Research series,* Luxembourg, JRC/ICES, EUR 24334 EN, 73 pp.

Rogers, S., M. Casini, P. Cury, M. Heath, X. Irigoien, H. Kuosa, M. Scheidat, H. Skov, K. Stergiou, V. Trenkel, J. Wikner and O. Yunev (2010). Marine Strategy Framework Directive – Task Group 4 Report Food webs. *EUR – Scientific and Technical Research series,* Luxembourg, JRC/ICES, EUR 24343 EN, 55 pp.

Shin, Y.-J., A. Bundy, L.J. Shannon, J.L. Blanchard, R. Chuenpagdee, M. Coll, B. Knight, C. Lynam, G. Piet and 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.

Solheim, A.L., K. Austnes, P. Kristensen, M. Peterlin, V. Kodeš, R. Collins, Semerádová, S., A. Künitzer, R. Filippi, Prchalová, H., C. Spiteri and T. Prins (2012). Ecological and chemical status and pressures in European waters. Thematic assessment for EEA Water 2012 Report. ETC./ICM Technical Report, vol. 1, p. 146. Prague, European Topic Centre / Inland, Coastal and Marine Waters, ETC./ICM Technical Report, 146 pp.

Spalding, M., H.E. Fox, G.R. Allen, N. Davidson, Z.A. Ferdaña, M. Finlayson, B.S. Halpern, M.A. Jorge, A. Lombana, S.A. Lourie, K.D. Martin, E. McManus, J. Molnar, C.A. Recchia and J. Robertson (2007). Marine ecoregions of the world: A bioregionalization of coastal and shelf areas. *BioScience* 57: 573-583.

Spangenberg, J.H., J. Martinez-Alier, I. Omann, I. Monterroso and R. Binimelis (2009). The DPSIR scheme for analysing biodiversity loss and developing preservation strategies. *Ecological Economics* 69: 9-11.

Swartenbroux, F., B. Albajedo, M. Angelidis, M. Aulne, V. Bartkevics, V. Besada, A. Bignert, A. Bitterhof, A. Hallikainen, R. Hoogenboom, L. Jorhem, M. Jud, R. Law, D. Licht Cederberg, E. McGovern, R. Miniero, R. Schneider, V. Velikova, F. Verstraete, L. Vinas and S. Vlad (2010). Marine Strategy Framework Directive – Task Group 9 Report Contaminants in fish and other seafood. *EUR – Scientific and Technical Research series,* Luxembourg, JRC/ICES, EUR 24339 EN, 36 pp.

Tasker, M.L., M. Amundin, M. André, A. Hawkins, W. Lang, T. Merck, A. Scholik-Schlomer, J. Teilmann, F. Thomsen, S. Werner and M. Zakharia (2010). Marine Strategy Framework Directive - Task Group 11 Underwater noise and other forms of energy. *EUR - Scientific and Technical Research series,* Luxembourg, JRC/ICES, EUR 24341 EN, 55 pp.

Tett, P., R. Gowen, S. Painting, M. Elliott, R. Forster, D. Mills, E. Bresnan, E. Capuzzo,, J.F. T. Fernandes, R. Geider, L. Gilpin, M. Huxham, A. McQuatters-, S.M. Gollop, S. Saux-Picart, T. Platt, M.F. Racault, S. Sathyendranath, J. and M.W. van der Molen, M. (2013). A Framework for understanding marine ecosystem health. *Marine Ecology Progress Series*.

Turner, M.G., W.H. Romme, R.H. Gardner, R.V. O'Neill and T.K. Kratz (1993). A revised concept of landscape equilibrium: Disturbance and stability on scaled landscapes *Landscape Ecology* 8: 213-227.

Turner, M.G., R.H. Gardner and R.V. O'Neill (2001). *Landscape ecology in theory and practice : pattern and process*. New York, Springer-Verlag. pp.

Van Beurden, A.U.C.J. and W.J.A.M. Douven (1999). Aggregation issues of spatial information in environmental research. *International Journal of Geographical Information Science* 13: 513–527.

Vermaat, J.E., F. Eppink, J.C.J.M. van den Bergh, A. Barendregt and J. van Belle (2005). Aggregation and the matching of scales in spatial economics and landscape ecology: empirical evidence and prospects for integration. *Ecological Economics* 52: 229-237.

Zenetos, A., S. Gofas, C. Morri, A. Rosso, D. Violanti, J.E. Garcia Raso, M.E. Cinar, A. Almogi-Labin, A.S. Ates, E. Azzurro, E. Ballesteros, C.N. Bianchi, M. Bilecenoglu, M.C. Gambi, A. Giangrande, C. Gravili, O. Hyams-Kaphzan, P.K. Karachle, S. Katsanevakis, L. Lipej, F. Mastrototaro, F. Mineur, M.A. Pancucci-Papadopoulou, A. Ramos Espla, C. Salas, G. San Martin, A. Sfriso, N. Streftaris and M. Verlaque (2012). Alien species in the Mediterranean Sea by 2012. A contribution to the application of European Union's Marine Strategy Framework Directive (MSFD). Part 2. Introduction trends and pathways. *Mediterranean Marine Science* 13: 328-352.

A *Glossary of geographical terms used in this report*

Geographic term	
(Marine) region	Defined in MSFD Art. 4(1): Baltic Sea, North-east Atlantic Ocean, Mediterranean Sea and Black Sea
(Marine) subregion	Defined in MSFD Art 4(2) for the NE Atlantic Ocean and Mediterranean Sea
(sub)region	Marine region or subregion
subdivision	Spatial delimitation of a part of a member states' marine waters in a smaller spatial unit (MSFD Art. 4(2)
Regional sea	Sea areas falling under the Regional Sea Conventions HELCOM, OSPAR, UNEP/MAP or BSC
Sub-basin	HELCOM's division of the Baltic Sea into smaller spatial units
Region I-V	OSPAR's division of the NE Atlantic into five areas; OSPAR regions do not fully match the marine subregions in the Nort-east Atlantic
Geographical Sub-area	Area defined by FAO for assessment of commercial fish stocks in the Mediterranean Sea and Black Sea
Biogeographical region	The Habitats Directive divides the EU into 9 ecologically coherent "biogeographical" regions. For the marine environment, the following biogeographical regions are relevant: the Atlantic, Boreal, Continental, Macaronesian, Mediterranean and Black Sea region The Baltic Sea is divided in two biogeographical regions (Boreal, Continental)
Biogeographic zone	Used in ICED/JRC Task group 1 report (Cochrane *et al.* 2010) without definition. Spalding *et al.* (2007) distinguish Northern European seas (contains marine subregions: Baltic Sea, Greater North Sea, Celtic Seas), Lusitanian (contains marine subregions Bay of Biscay and Iberian coast, Macaronesia), Mediterranean Sea and Black Sea
Ecoregion	Defined by Spalding et . (2007) as the smallest-scale units in marine ecoregions of the world: *Areas of relatively homogeneous species composition, clearly distinct from adjacent systems. The species composition is likely to be determined by the predominance of a small number of ecosystems and/or a distinct suite of oceanographic or topographic features. The dominant biogeographic forcing agents defining the eco-regions vary from location to location but may include isolation, upwelling, nutrient inputs, freshwater influx, temperature regimes, ice regimes, exposure, sediments, currents, and bathymetric or coastal complexity.* The boundaries of ecoregions do not fully match the boundaries of marine (sub)regions

B Annex - Analytical report

Appendix: Analytical report

Coherent geographic scales and aggregation rules

Theo Prins (Deltares)
Myra van der Meulen (Deltares)
Arjen Boon (Deltares)
Nomiki Simboura (HCMR)
Catherine Tsangaris (HCMR)
Angel Borja (AZTI)
Iratxe Menchaca (AZTI)

1207879-000

Title
Appendix: Analytical report

Client	**Project**	**Reference**	**Pages**
European Commission DG Environment Unit D2	1207879-000	1207879-000-ZKS-0014	59

Keywords
MSFD, monitoring, assessment

Summary

An overview is given of methods for the definition of spatial scales in environmental assessments of the marine environment, and of aggregation methods in assessments. An analysis is presented of the methods applied by member states in the 2012 reporting on the Initial Assessment for the Marine Strategy Framework Directive.

References
SFRA0019-SCALES

Version	Date	Author	Initials	Review	Initials	Approval	Initials
2	Nov 2013	Theo Prins					
		Myra van der Meulen					
		Arjen Boon					
		Nomiki Simboura					
		Catherine Tsangaris					
		Angel Borja					
		Iratxe Menchaca					
3	July 2014	Theo Prins					

State
final

Contents

1 Introduction

1.1 General background

The 2012 reporting for Marine Strategy Framework Directive (MSFD) Articles 8, 9 and 10 constitutes three important steps in the 1st six-year management cycle of the MSFD. With the reporting on the initial assessment of the marine waters (Art. 8), the determination of Good Environmental Status (GES, Art. 9) and the identification of environmental targets and associated indicators (Art. 10) the Member States (MS) should have identified all relevant issues concerning drivers, pressures, state and impacts in the marine environment.

Article 3(5) of the Marine Strategy Framework Directive (MSFD, 2008/56/EC) requires that good environmental status is determined at the level of the marine region or subregion as referred to in Article 4, on the basis of the qualitative Descriptors in Annex I to the Directive. This means that the MSFD operates at a different geographic scale than existing EU legislation such as the Water Framework Directive (WFD, 2000/60/EC) for coastal and transitional waters, which considers ecological and chemical status at the level of estuarine and coastal water bodies. /It also means that national approaches to determining GES need to ensure that together they articulate GES for a marine region or subregion. The geographical scale to be used for assessments is not well defined in the MSFD. Consequently, in this first cycle of implementation the geographical scales adopted for the assessment of GES may vary considerably between descriptors, and may differ widely among MS.

Assessments of the marine environment need to be carried out for a specific area, which may differ between descriptors or even between criteria and indicators within a descriptor. Therefore, the first question that needs to be addressed is:

- *What is the appropriate spatial scale for the assessment of the marine environment?*

When assessment scales have been defined, the question of scaling up from the individual, specific or sectorial assessments to an assessment for the whole (sub)region needs to be considered:

- *How to scale up from assessment areas to larger geographic scales?*

A third question deals with the aggregation of the various assessments at different levels:

- *How to aggregate indicators within a criterion, or criteria within a descriptor, or all the descriptors to come to a comprehensive and balanced judgement of the status of marine waters through GES?*

In January 2013, the European Commission put out a Service Request, asking for an analysis of national approaches that Member States have taken in their reporting under Articles 8, 9 and 10 of the MSFD, with respect to geographical scaling and aggregation rules, and the development of broad EU guidance for coherent geographic scales in assessment and monitoring of GES and for sets of aggregation rules.

1.2 Objectives of this report

The objective of the Service request is to develop guidance on the application of geographic scales and aggregation rules in the assessment of the marine environment under the MSFD.

The objectives mentioned in the Service Request are to:

- assess the electronic and text reporting undertaken by Member States (MS) under Articles 8, 9 and 10 of the MSFD with the aim to analyse and compare the national approaches taken per descriptor regarding the scales for the assessment of the environmental state of their marine waters, determining GES and setting environmental targets.

- analyse which aggregation rules have been applied, if any, by MS in their reports. Based on the results of these analyses and further comparison with regional approaches and methods applied in research projects, identify issues that require further consultation by MS, Regional Sea Conventions (RSC) and the European Commission.
- develop broad EU guidance for coherent geographic scales in assessment and monitoring of GES and for sets of aggregation rules and organize a debate with MS on this.

This report gives the first results of the project, and presents an analysis of national approaches on geographic scales and aggregation rules in the MSFD reporting, and an analysis of approaches taken in the framework of RSC.
The report addresses three issues that are related to the questions of geographic scaling and aggregation:

a) the definition of the geographic scale for assessments that provide meaningful information for management
b) the geographic scaling up of assessments for an assessment of GES at the scale of a (sub)region
c) the aggregation of individual assessments to an overall assessment of GES

For this analysis, we used information on the implementation of the MSFD by Member States (electronic reporting, parts of the national expert reviews, and a limited number of national reports), EU documents and reports from RSC, peer-reviewed literature, results from research projects and personal communication with the European Commission.

The report is part of the Service Contract SFRA0019 - SCALES under the agreement of the ‘Framework contract for services related to development of methodological standards in relation to good environmental status of the seas under MSFD (ENV.D.2/FRA/2012/0019)’ between the European Commission/DG Environment and Deltares, as lead partner of a consortium with AZTI, HCMR and SYKE.

The report will provide a basis for a consultation paper and a draft guidance document for the European Commission on how to deal with spatial scales and aggregation of the MSFD in a European context.

1.3 Report outline

Chapter 3 provides an overview of approaches for the definition of geographic scales, Chapter 4 discusses the scaling up of assessments, and Chapter 5 gives an overview of aggregation methods. Chapter 6 gives information on the current approaches of MS regarding scales. Chapter 7 discusses the results of the analyses and provides a basis for the consultation paper and the draft guidance.

2 Definition of spatial assessment scales

This chapter deals with the question how to define the spatial scale for assessments that leads to meaningful information to support the management of marine areas.

The general principles and considerations for the application of rules for geographic scaling have been discussed in a number of documents relating to the MSFD implementation and assessments of the marine environment, like the Task group reports drafted in 2010 (Cardoso *et al.* 2010, Cochrane *et al.* 2010, Ferreira *et al.* 2010, Galgani *et al.* 2010, Law *et al.* 2010, Olenin *et al.* 2010, Piet *et al.* 2010, Rice *et al.* 2010, Rogers *et al.* 2010, Swartenbroux *et al.* 2010, Tasker *et al.* 2010) and other documents (OSPAR 2011), and are summarized here. Furthermore, we give an overview of existing approaches to define assessment scales and lessons learned from RSC.

2.1 MSFD requirements

The MSFD requires that good environmental status is determined at the level of the marine region or subregion (Art. 3.5), on the basis of the qualitative descriptors in Annex I of the MSFD. For the Baltic Sea and Black Sea GES will be determined at the level of the marine region. The marine regions North-East Atlantic Ocean and the Mediterranean Sea have each been divided in 4 subregions where GES will be determined:

a) the Baltic Sea
b) the North-east Atlantic Ocean
  - Macaronesia
  - Bay of Biscay and the Iberian coast
  - Celtic Seas
  - Greater North Sea
c) the Mediterranean Sea
  - Western Mediterranean Sea
  - Adriatic Sea
  - Ionian Sea and the Central Mediterranean Sea
  - Aegean-Levantine Sea
d) the Black Sea

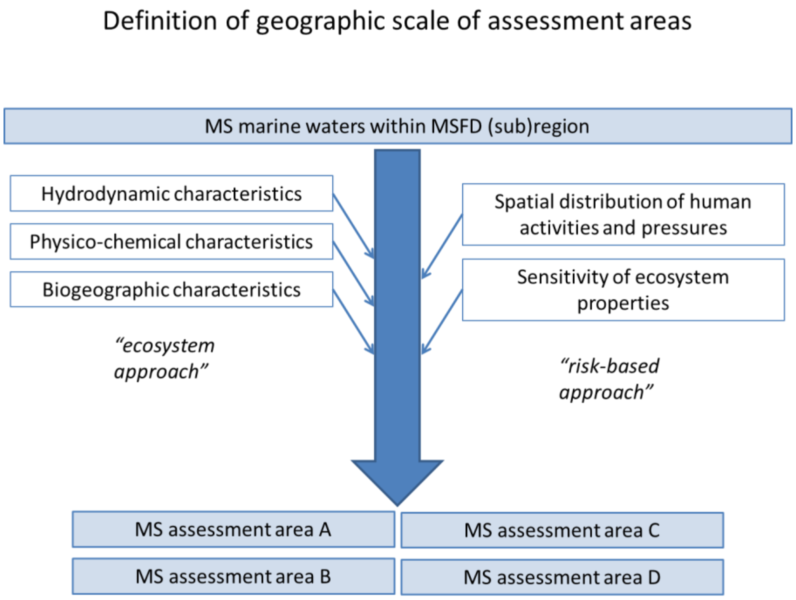


Figure 2.1 Schematic picture of the definition of assessment areas within a (sub)region. (MS: Member State)

2.2 General principles for the definition of assessment areas

The definition of assessment areas needs to address spatial scales at different levels. The highest level is the level of the marine (sub)region. However, in many cases the scale of the regions and subregions is too large for meaningful environmental assessments, as too large assessment areas will mask the more local pressures and their impacts. Further subdivisions may therefore be necessary, depending on the topic. According to Art. 4(2) Member States may, in order to take into account the specificities of a particular area, implement subdivisions within the region and subregions.

At a smaller spatial scale, various spatial units within the larger ecosystem may be distinguished. These units can generally be differentiated on the basis of their physical, chemical and biological characteristics. For environmental assessments, a definition of smaller assessment areas at the level of metrics, indicators, criteria or descriptors may be necessary (Figure 2.1).

In addition to these ecological considerations, there are management-related issues as the areas have to be chosen in such a way that assessments provide the right information to the process of policy development and management of marine areas. At the end of the process, it is crucial that assessment areas are defined that provide a robust and adequate assessment of environmental state, and that enable the evaluation of the effectiveness of management measures.

An example of an approach to deal with this scaling issue is provided by work done by HELCOM in the Baltic Sea. A hierarchical approach was used by HELCOM, where the whole Baltic Sea has been divided in 19 sub-basins on the basis of water exchange characteristics. Within these sub-basins, assessment units are distinguished by a further division in offshore waters and coastal waters that are subdivided in the water bodies defined under the WFD. This approach will be discussed more extensively below.

Ecological considerations

Hydrodynamic characteristics, like transport patterns, freshwater input and mixing/stratification of the water column, are important physical features that define the boundaries of ecologically relevant areas. Biogeographic distribution patterns, related to benthic or pelagic habitats or marine populations are another important feature that needs to be taken into consideration when defining these areas.

Those features are not all equally important for all descriptors, criteria and indicators. For descriptors like D5 (Eutrophication), D8 (Contaminants) and D9 (Contaminants in seafood) with (often) clearly localized sources of pollution (e.g. rivers or other point sources), hydrodynamic characteristics play an important role. For descriptors like D1 (Biodiversity), D3 (Commercial fish and shellfish), D4 (Food webs) and D6 (Seafloor integrity) habitat patterns and biogeographic characteristics are often more important. For ecologically relevant scales of the latter descriptors, the assessment should cover the entire range of the species and/or of discrete populations (e.g. for large/mobile species). For habitats/communities it is most appropriate to assess the status within biogeographic zones, as functionally similar habitats can have wider distributions (Cochrane *et al.* 2010).

Activities may result in different types of pressures, e.g. both localised pressures and pressures operating at a larger spatial scale. For example, pressures and impacts arising from fisheries operate both at the larger scale of stocks of commercial species and at smaller, patchy scales in relation to physical impacts on the marine environment, like in the case of bottom trawling.

Concluding, ecological assessment areas must be defined in a way to adequately reflect both the ecological scales exhibited in each (sub)region and the links to areas which are effective for management measures. Size may vary from small areas of a specific biological feature to large areas relating to highly mobile species, homogenous habitats or large-scale food webs. This means that on the basis of ecological considerations, assessment areas may be different between indicators and descriptors.

Policy and management consideration

Assessment areas must be designed in relation to risks for the marine environment, caused by the main drivers (D) and human activities. The impacts (I) of pressures (P) are generally larger near the source (either land-based or sea-based) and decrease with distance from the source. For land-based sources, this means that there is a gradient of decreasing pressures and impacts from the coast to offshore areas. The density and intensity of human activities is generally higher near the coast as well. Consequently, a finer spatial resolution of assessment areas may be required in coastal areas than in offshore areas where less human activities take place.

Assessments should make it possible to inform managers and policymakers on the environmental impacts of human activities, and link these impacts to pressures and activities. Through this link between pressures, state (S) and impacts, management measures and responses (R) can be identified. Consequently, the spatial scale of assessments must reflect those D-P-S-I-R relationships. Too large areas can mask local pressures and their impacts, and are therefore not suitable for management purposes. On the other hand, too small areas result in a high monitoring burden, and may lead to inadequate assessments as the spatial distribution of ecosystem components is not sufficiently covered, and an evaluation of the wider effects or the cumulative effects of local pressures is not possible.

Consequently, geographic scales need to be chosen to ensure that local impacts remain detectable, in order to inform measures. Smaller assessment areas are better suited for this purpose than larger scales. Scaling up of such areas to larger areas needs to ensure that relevant smaller scale impacts remain captured and are not lost in the overall state.

Risk-based approach

Section 6 of Annex Part A of the EU COM Decision 2010/477/EU, provides that: "*A combined assessment of the scale, distribution and intensity of the pressures and the extent, vulnerability and resilience of the different ecosystem components including where possible their mapping, allows the identification of areas where marine ecosystems have or may have been adversely affected. It is also a useful basis to assess the scale of the actual or potential impacts on marine ecosystems. This approach, which takes into account risk-based considerations, also supports the selection of the most appropriate indicators related to the criteria for assessment of progress towards good environmental status*".

A risk-based approach (Fig. 2.1) helps to prioritize areas and indicators for monitoring and assessment. Assessments of GES should begin with areas of both greatest vulnerability and highest pressures. If the environmental status in these areas is good, then it can be assumed that the status over a larger area is 'good' (Cardoso *et al.* 2010).

2.3 Overview of methods to define assessment areas

2.3.1 Assessment areas based on hydrological, oceanographic and biogeographic criteria.

Description

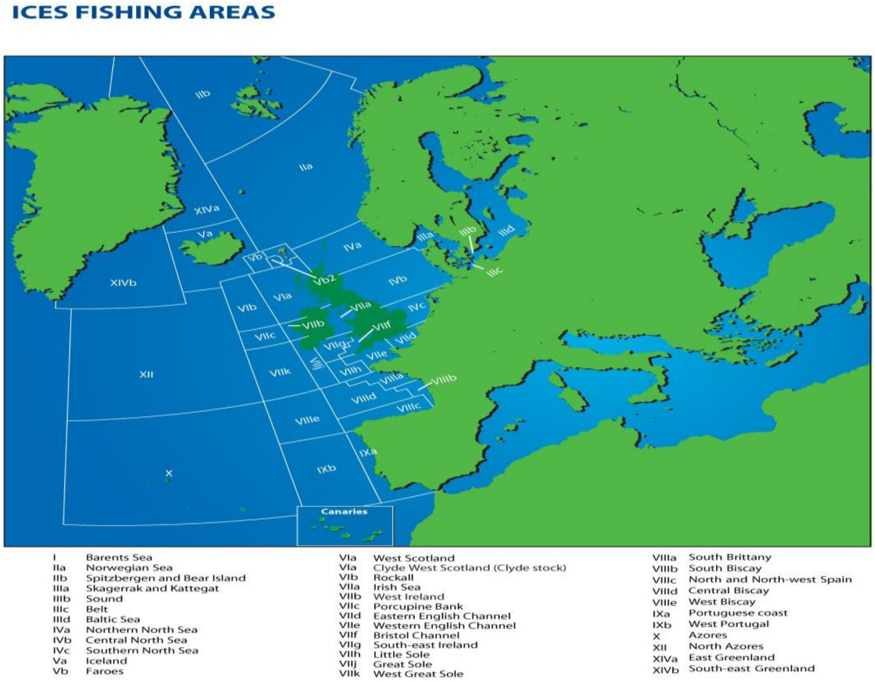
The MSFD indicates that hydrological, oceanographic and biogeographic features should be taken into account in defining the marine regions and marine subregions (Art. 3.2). Ecological assessment areas within those region and subregions can be further specified using hydrological and oceanographic characteristics, in particular sea temperature, salinity, mixing characteristics, frontal systems and turbidity/water clarity (but also depth, currents, wave action and nutrient characteristics where appropriate) to define water masses of similar overall character within each (sub)region. The boundaries between such areas should wherever possible be based on marked changes in these parameters, but where changes are more gradual, more pragmatic factors such as the physiographic shape of the coastline and administrative boundaries may also be used, provided that the set of areas within a (sub)region overall are ecologically-based (Cochrane *et al.* 2010;OSPAR 2012a). These subdivisions (where formally defined), or other informal ecological assessment areas, should allow, where possible, ecosystem-based assessments.

The identification of a set of ecological assessment areas within a (sub)region provides specific geographical areas in which to determine the extent of impacts and whether GES and associated targets have been met. The examples below refer to individual descriptors, however in determining GES for a whole (sub)region it could be argued that ecological assessment areas should differ between descriptors. A nested, hierarchical approach (comparable to the HELCOM approach with spatial scales at different, nested, levels) might be a way to establish the link between local pressures and the status of larger areas

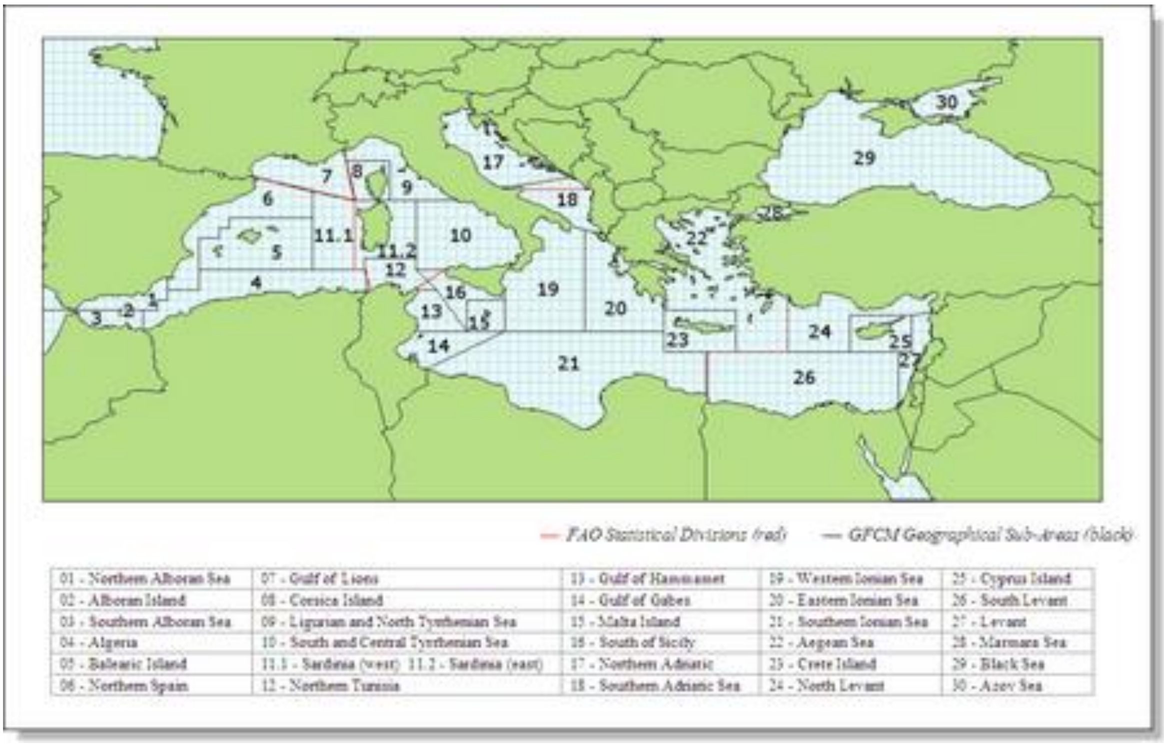
Example, Box 1

Assessments of commercial fish stocks

For the assessment of commercial fish stocks geographical sub-areas have been defined, for example in the Northeast Atlantic, the Mediterranean Sea and Black Sea, based on the spatial distribution of stocks of commercial fish species. These sub-areas may be considered for the determination of GES for descriptor D3.



ICES fishing areas in the NE Atlantic



01 - Northern Alboran Sea	07 - Gulf of Lions	13 - Gulf of Hammamet	19 - Western Ionian Sea	25 - Cyprus Island
02 - Alboran Island	08 - Corsica Island	14 - Gulf of Gabes	20 - Eastern Ionian Sea	26 - South Levant
03 - Southern Alboran Sea	09 - Ligurian and North Tyrrhenian Sea	15 - Malta Island	21 - Southern Ionian Sea	27 - Levant
04 - Algeria	10 - South and Central Tyrrhenian Sea	16 - South of Sicily	22 - Aegean Sea	28 - Marmara Sea
05 - Balearic Island	11.1 - Sardinia (west) 11.2 - Sardinia (east)	17 - Northern Adriatic	23 - Crete Island	29 - Black Sea
06 - Northern Spain	12 - Northern Tunisia	18 - Southern Adriatic Sea	24 - North Levant	30 - Azov Sea

Geographical sub-areas for stock assessments in the Mediterranean Sea and Black Sea (FAO statistics).

Example, Box 2: OSPAR biodiversity assessment in the North Sea

In its Advice manual for Biodiversity (OSPAR 2012a) OSPAR recommends to define assessment scales for habitats that are nested within subregions to:

1) Reflect the changes in ecological character of communities within the same abiotic habitat across a subregion (due to changes in temperature, salinity and other factors across subregions);
2) Better accommodate links to management of human activities and their pressures, which can differ significantly across a subregion;
3) Facilitate aggregation of assessments up to the level of subregions.

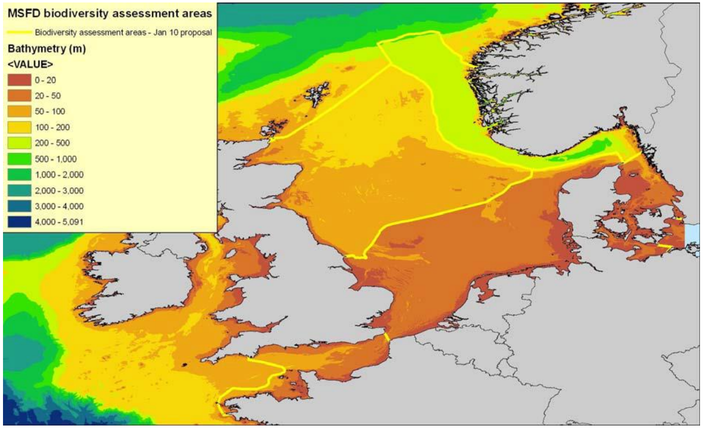
Assessment areas should be defined following the recommendations of the Task Group 1 report (Cochrane *et al.* 2010), by taking into account hydrological and oceanographic conditions that should be reflected in similarities in community compositions of benthic and pelagic habitats.
The boundaries between such areas should be based on marked changes where possible, but more pragmatic factors may be used, as long as areas within a subregion overall are ecologically-based.
The identification of a set of ecological assessment areas within a subregion provides the basis for assessment of the habitats occurring within the area (see Annex 8.6 for a list), as it provides a specific geographical area in which to determine the extent of impacts and whether GES and associated targets have been met. Assessment of ecological status for

WFD (water bodies) and favourable conservation status for Habitats Directive (bioregions of Member States waters) use a defined spatial scale (area) for all assessments. As such areas may span several Member States waters, there is a need to develop practical approaches to undertaking the assessments, as are currently applied for some wide-ranging species (e.g. harbour porpoise in the North Sea), to meet the requirements for an assessment of GES at the scale of a (sub)region.

Applied to the North Sea, this resulted in a provisional division of the North Sea into five areas for biodiversity assessment.

The preferred area in which the assessment is made can differ among descriptors. For the biodiversity descriptors (Descriptors 1, 2, 4 and 6) OSPAR (2012a) notes that the choice of an assessment scale is very important, as different scales may lead to very different outcomes of the status assessment for a particular ecosystem component. The scale used should be meaningful from a biodiversity perspective (i.e. taking into account the scales at which species, populations and communities occur), but also from a management perspective (i.e. a scale at which management measures are effective).

Map showing provisional biodiversity assessment areas of the Greater North Sea (Source:OSPAR 2012a)



2.3.2 Risk-based approach

In the risk-based approach (Cardoso *et al.* 2010) a pragmatic prioritization is made, which enables general statements about environmental status at large scales while keeping monitoring requirements manageable.

This risk-based approach is particularly effective for Descriptors that are spatially patchy and where pressures are applied at specific locations. It is recommended to map the pressures that most likely have the largest impacts, and the vulnerability of various properties of the ecosystem. Cardoso *et al.* (2010) recommend prioritization by prior assessment of:

i. the distribution of the intensity or severity of the pressures across the region at large;
ii. the spatial extent of the pressures relative to the ecosystem properties possibly being impacted;
iii. the sensitivity/vulnerability or resilience of the ecosystem properties to the pressures;
iv. the ability of the ecosystem properties to recover from impacts, and the rate of such recovery;
v. the extent to which ecosystem functions may be altered by the impacts; and
vi. where relevant, the timing and duration of the impact relative to the spatial and temporal extent of particular ecosystem functions (e.g. shelter, feeding, etc.).

If the environmental status in these areas is good, then it can be assumed that the status over a larger area is ‛good‛ (Cardoso *et al.* 2010). In contrast, if the environmental status in an area is not ‘good‛, then monitoring and assessments would be conducted stepwise at additional sites along the gradients of pressure or vulnerability. The size of the appropriate steps along the gradient will depend on the nature of the gradient and the way the environmental conditions are degraded. It may vary significantly with different cases (Cardoso *et al.* 2010). A risk-based approach to prioritize areas where human pressures are highest and/or ecosystem components are most vulnerable, is recommended by OSPAR (2012a). This approach may be helpful to prioritize monitoring efforts, but still requires the definition of scales for assessment of GES.

Example, Box 3: Task Group 6 report on seafloor integrity

The Task Group 6 report (Rice *et al.* 2010) proposes a risk based approach for assessing environmental status at scale of the marine region and subregion. The information on how risk is distributed in space provides a basis for **assessing environmental status** from either of two directions. The assessment can start with specific **human activities** of particular concern. Alternatively, assessments can start with specific **attributes** of the sea floor. Assessment of environmental status would start with the highest risk strata, and proceed to progressively low risk strata until areas were found to be in good environmental status. In order to map the spatial distribution of most human activities in the sea, (particularly the ones most likely to cause the largest impacts on the sea floor) and also for assessing key attributes of benthos for GES and vulnerability assessment, it is proposed to construct maps not on a very fine spatial scale, but on the scales characteristic of EUNIS Level 4 classifications of the benthos.

2.4 Analysis of approaches by Regional Sea Conventions

2.4.1 Baltic Sea (HELCOM)

HELCOM developed a division of the Baltic Sea for monitoring and assessment purposes. This division consists of four hierarchical scales (Figure 2.2):

- The whole Baltic Sea
- A division of the Baltic Sea into 19 sub-basins that are divided by sills, and have different physico-chemical (size, volume, depth, salinity) and biological characteristics. Those 19 sub-basins include the Kattegat and the the Sound, which under the MSFD are part of the Greater North Sea.
- A further division of the sub-basins in coastal and offshore areas, including EEZ boundaries between Baltic states
- A further division of the coastal areas into water bodies defined under the WFD.

In recent assessments (HELCOM 2009a; b; 2010b; a) this nested hierarchical approach was used, although there were small differences between assessments.

In the 2009 eutrophication assessment (HELCOM 2009b) used 189 ‘areas’. These areas are a mix of stations, sites and basins, and consist of 17 open water areas and 172 coastal areas. In the most recent eutrophication assessment (Pyhälä *et al.* 2013) the divisions defined by HELCOM were used. The assessment on hazardous substances (HELCOM 2010a) used 144 assessment units, 40 open-sea areas and 104 coastal sites or areas. The thematic assessment on biodiversity (HELCOM 2009a) used 22 national case studies in a test of an indicator-based approach to assess marine biodiversity. In addition, a test assessment for the sub-basin Baltic proper was carried out. It was concluded that biodiversity issues require an assessment at regional scale, for which further development of the methods are necessary

Figure 2.2 Map of the Baltic Sea presenting the HELCOM -division into 17 open sub-basins and 42 coastal areas. EEZs of the countries are shown with a grey dashed line. Note: The final settlement of the border between Great Belt Danish Coastal waters and Kiel Bay German Coastal waters is subject to bilateral consultations between Denmark and Germany. (source: HELCOM 2013)

.

All these assessments were done as case studies, and the spatial scales that were used were determined by the availability of data and the willingness of the Contracting parties in HELCOM to participate in the assessments (J. Andersen, pers. comm.). Hence, there was no strict application of predefined assessment areas.

In HELCOM's view, the various hierarchical sub-division levels can be used depending on the needs. For example, monitoring and assessment of mobile marine mammals such as grey seals may require the whole Baltic Sea scale while assessment of eutrophication indicators may be most relevant at the sub-basin scale in the open sea combined with water body or type level in the coastal zone. HELCOM recommends that the scale to be used should be chosen from the four possible scales (HELCOM 2013).

2.4.2 Mediterranean Sea (Barcelona Convention)

In the process of the application of the Ecosystem Approach (ECaP) adopted by the Barcelona Convention in 2008, the Mediterranean was divided into four geographic areas for the identification of the important ecosystem properties and the assessment of ecological status and pressures. These four areas are (1) Western Mediterranean, (2) Adriatic Sea (3). Ionian Sea and Central Mediterranean, (4) Aegean-Levantine Sea. This operational division was the result of a decision by the Contracting Parties based on biogeographical and oceanographic considerations (UNEP/MAP 2008). The division was used to produce four sub-regional assessments (UNEP/MAP 2010) and the Initial Integrated Assessment of the Mediterranean Sea (UNEP/MAP 2012b) that informs on marine and coastal ecosystem status, pressures and impacts.

The division was used for the assessment of hazardous substances using the MEDPOL monitoring Database (UNEP/MAP 2011).

The ECaP approach sets 11 Ecological Objectives (EOs), corresponding to 21 Operational Objectives (OOs) and 61 Indicators concerning biological diversity, non-indigenous species, commercially exploited fish and shellfish, marine food webs, eutrophication, sea-floor integrity, hydrography, coastal ecosystems and landscapes, pollution, marine litter, and energy including underwater noise. At the first Meeting of Ecosystem Approach Coordination Group in May 2012, providing guidance on the ECaP process UNEP/MAP (2012a) noted that geographic integration will be accomplished at various scales but data compatibility should be considered in order to allow integration at the sub-regional and Mediterranean scale. Scale should be addressed in the discussion of each EO, including the question at which scale indicators can be assessed qualitatively or quantitatively. In principle, scales should be national and when possible regional (Mediterranean) and trans-boundary or subregional. GES should be defined at a higher scale (Mediterranean or marine subregion) than the targets (which will be defined at national or sub-national scale).

2.4.3 Black Sea (Bucharest Convention)

For the implementation of the Black Sea Integrated Monitoring and Assessment Program (BSIMAP) as approved by the BSC in 2002, the Black Sea was divided into seven zones of responsibilities: the territorial waters of the six neighbouring countries and the open sea. The BSIMAP is based on National monitoring programmes financed by the Black Sea countries. The main environmental problems in the region are*:* eutrophication, chemical pollution (including oil), biodiversity decline, habitats destruction and overfishing.

According to the Final ¨Diagnostic Report¨ to guide improvements to the regular reporting process on the state of the Black Sea environment, August 2010, biology (biodiversity and fisheries) is regularly monitored in the Black Sea besides chemistry and hydro-physical variables but these efforts are not well integrated and coordinated at the national and regional scales and on the other hand, spatial and temporal coverage especially of the biological variables is a major gap.

2.4.4 Northeast Atlantic (OSPAR)

OSPAR covers nearly the entire marine region of the Northeast Atlantic, with exception of the waters of the subregion Macaronesia south from 36º N. OSPAR distinguishes five regions (OSPAR regions I to V). The OSPAR regions are to a large extent similar to the European marine subregions within the NE Atlantic, but it should be noted that there are differences between MSFD and OSPAR in the boundaries between the areas and in the outer boundaries.

In the latest Quality Status Report (OSPAR 2010) the results of environmental assessments are presented for a number of themes. Well-developed approaches for assessments have been developed for the whole OSPAR area for the topics eutrophication, hazardous substances and radioactive substances.

For eutrophication assessments, the ‘Common Procedure’ is applied that integrates ten indicators for nutrient enrichment, and direct and indirect effects. The application is area-specific following, as a first step, a screening procedure that was completed in 2000. This procedure identified obvious non-problem areas with regard to eutrophication. Features that were taken into account in defining assessment areas were hydrodynamic characteristics and the proximity to nutrient sources. The second step, the Comprehensive Procedure, consisted of assessment and classification of the (potential) problem areas (OSPAR 2003). A second assessment was carried out in 2008 (OSPAR 2008). In the assessments carried out by the OSPAR states, different geographical scales for identifying individual assessment areas were used, ranging from small individual fjords to large coastal strips. A total of 204 assessment areas (Greater North Sea: 93; Celtic Seas: 84; Bay of Biscay and Iberian Coast: 27) were used in the 2008 assessment. The size of the assessment areas increased from inshore waters (estuaries, bights, fjords) to offshore. Parameters used to define assessment areas were hydrographical and physico-chemical characteristics like salinity gradient, depth, mixing characteristics (such as fronts, stratification), transboundary fluxes, upwelling, sedimentation, residence time/retention time, mean water temperature (water temperature range), turbidity (expressed in terms of suspended matter), mean substrate composition (in terms of sediment types) and typology of offshore waters.

The Coordinated Environmental Monitoring Programme (CEMP) provides a common framework for the collection of marine monitoring data by OSPAR countries. Status and trends in pollution are assessed for a number of substances, by monitoring concentrations in water, sediments and biota (OSPAR 2009c). CEMP monitoring is mainly focussed on coastal areas, because these are close to discharge and emission sources. Increasing attention is being paid to monitoring in offshore areas, in relation to activities like oil and gas production and shipping. The assessments are based on a large number of (predominantly coastal) monitoring stations. The results were aggregated for each of the 5 OSPAR regions by grouping stations into coastal stations (<12 nm), likely to be more affected by land-based inputs of contaminants, and offshore stations (Figure 2.3). Further divisions of the coastal stations were made where appropriate (Table 2.1; OSPAR 2009c).

The following map shows divisions of the OSPAR area used for contaminant data assessment (Task Group 8 Report, Law *et al.* 2010).

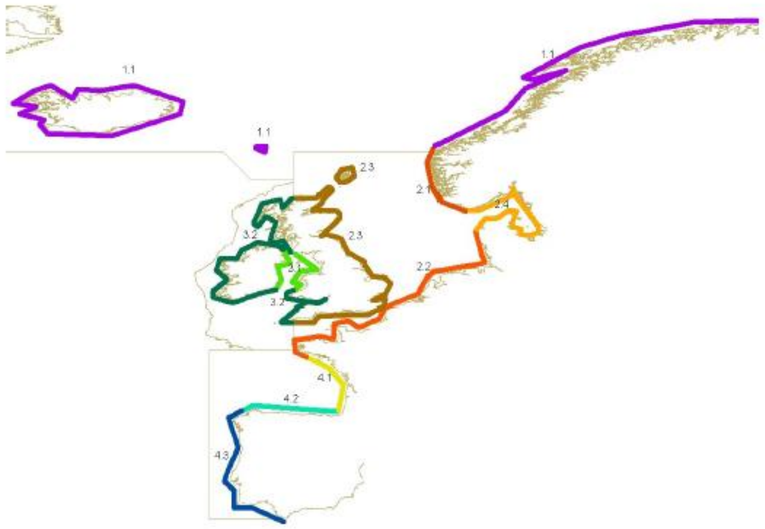
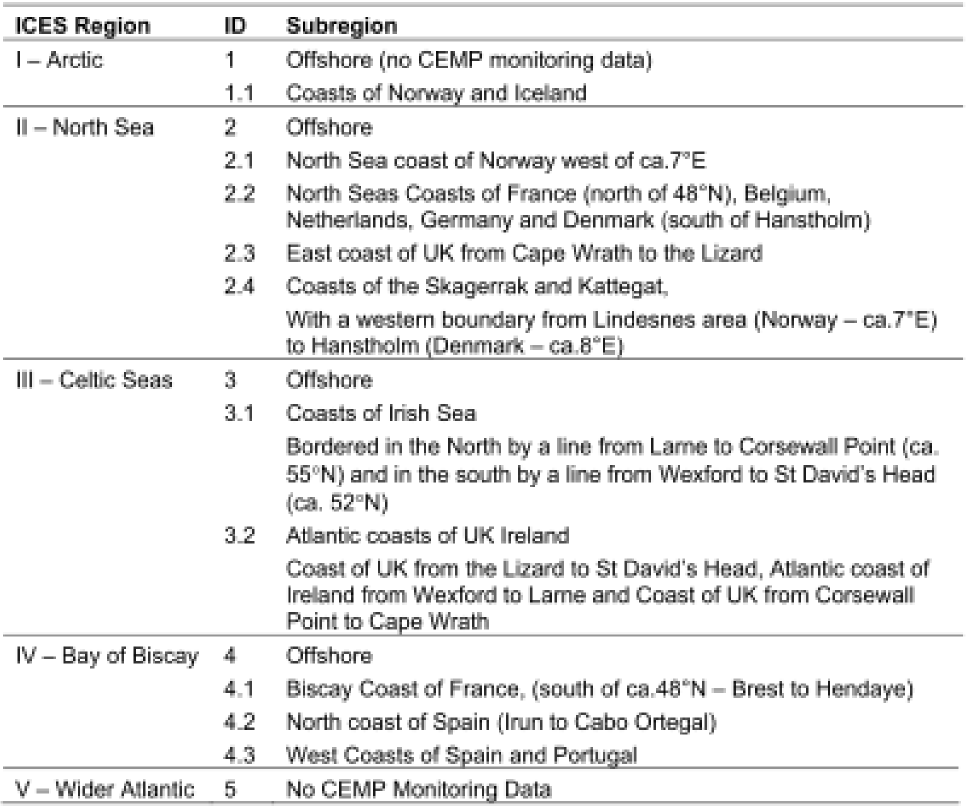


Figure 2. Subdivisions of the OSPAR area used in contaminant data assessment for the QSR 2010, showing offshore areas, and coastal waters defined by the 12 nautical mile limit, in OSPAR Regions I-IV. The colours are of no significance, but are included to improve clarity of the map.

Figure 2.3 Divisions of the OSPAR area used for contaminant data assessment of the QSR 2010.

Table 2.1 OSPAR divison of coastal (<12 nm) and offshore (≥12 nm) waters for assessment of hazardous substances (OSPAR 2009c).



ICES Region	**ID**	**Subregion**
I – Arctic	1	Offshore (no CEMP monitoring data)
	1.1	Coasts of Norway and Iceland
II – North Sea	2	Offshore
	2.1	North Sea coast of Norway west of ca.7°E
	2.2	North Seas Coasts of France (north of 48°N), Belgium, Netherlands, Germany and Denmark (south of Hanstholm)
	2.3	East coast of UK from Cape Wrath to the Lizard
	2.4	Coasts of the Skagerrak and Kattegat,
		With a western boundary from Lindesnes area (Norway – ca.7°E) to Hanstholm (Denmark – ca.8°E)
III – Celtic Seas	3	Offshore
	3.1	Coasts of Irish Sea
		Bordered in the North by a line from Larne to Corsewall Point (ca. 55°N) and in the south by a line from Wexford to St David's Head (ca. 52°N)
	3.2	Atlantic coasts of UK Ireland
		Coast of UK from the Lizard to St David's Head, Atlantic coast of Ireland from Wexford to Larne and Coast of UK from Corsewall Point to Cape Wrath
IV – Bay of Biscay	4	Offshore
	4.1	Biscay Coast of France, (south of ca.48°N – Brest to Hendaye)
	4.2	North coast of Spain (Irun to Cabo Ortegal)
	4.3	West Coasts of Spain and Portugal
V – Wider Atlantic	5	No CEMP Monitoring Data

For the assessment of the concentrations of radioactive substances the OSPAR area was divided into 15 monitoring areas, taking into account hydrodynamic transport patterns, the location of sources and the location of potential impact areas (OSPAR 2009b).

2.4.4.1 *Ecological Quality Objectives*

Ecological Quality Objectives (EcoQOs) are the OSPAR instruments to apply the ecosystem approach in the North-East Atlantic and manage human activities that may have negative effects on the marine environment (OSPAR 2007). A first set of EcoQOs were developed in collaboration with ICES for OSPAR Contracting Parties in the North Sea as part of a pilot project in 2002. An evaluation of the North Sea pilot was conducted in 2006 (OSPAR 2006) and evaluations of the EcoQO system were conducted in 2008 and 2009 as a contribution to the Quality Status Report in 2010. In these documents several suggestions for alterations and refinement of the existing EcoQOs have been made and new EcoQOs have been proposed which are currently under development (i.e. seabird populations, threatened and/or declining habitats, marine beach litter).

For the North Sea, EcoQO's were developed in collaboration with ICES. The EcoQO's set objectives for specific indicators, and are used to indicate the status of specific components of the ecosystem (OSPAR 2010). The EcoQO system is based on data from monitoring. These data are usually provided per area. Several strategies for the application of geo-spatial scales are used. As yet, there is no standardized approach in the application of the EcoQOs with respect to the use of spatial scales. In Annex II, EcoQO's are discussed in more detail in relation to the assessment scales used for determining the ecological status of the OSPAR region.

2.5 Synthesis

The analysis of the approaches by the RSCs can be summarized as follows:

- All RSCs have defined geographic assessment scales by taking into account hydrodynamical and biogeographical characteristics, as well as administrative borders
- HELCOM has followed a nested, hierarchical approach that allows assessments at different spatial scales; this gives a common approach to spatial scales, while at the same time providing the opportunity to choose the most relevant geographic scale, depending on needs
- RSCs have used a risk-based approach, with a higher density of monitoring stations and a smaller spatial scale of assessment areas in the coastal zone

3 Scaling up from assessment areas to larger spatial scales

This chapter deals with the question how to scale up assessment areas to larger spatial scales, for example within a marine (sub)region. The Task group reports (Cardoso *et al.* 2010, Cochrane *et al.* 2010, Ferreira *et al.* 2010, Galgani *et al.* 2010, Law *et al.* 2010, Olenin *et al.* 2010, Piet *et al.* 2010, Rice *et al.* 2010, Rogers *et al.* 2010, Swartenbroux *et al.* 2010, Tasker *et al.* 2010) and other documents (OSPAR 2011) are summarized here. These documents provide a basis for the evaluation of different approaches which are illustrated using examples from research projects and literature.

3.1 MSFD requirements

As mentioned in s*ection 5 of Part A of EU COM Decision 2010/477/EU, "[w]hen the assessment needs to start at a relatively small spatial scale to be ecologically meaningful (for instance because pressures are localised), it could be necessary to scale up assessments at broader scales, such as at the levels of subdivisions, sub-regions and regions".* As discussed above, there are ecological, technical and management-related reasons to define assessment areas at relatively small spatial scales. To establish the status of an assessment areas, spatial aggregation may be necessary to combine results from various monitoring sites. Then to express GES at the (sub)region scale, status assessments per area could be scaled up, where possible. Several methods are available, such as:

- Aggregation, where assessment areas are merged and assessments are summed up to give an assessment for the larger spatial scale (see Figure 3.1). Careful consideration to the method for aggregation is necessary, as it should not result in a biased assessment (under- or overestimating GES) nor in an averaging out of all problems, leading to a lack of meaningful information for management purposes
- Grouping, where assessment areas are clustered within a larger scale; this could also simply be a presentation, for example a map showing the assessment results (GES or not) for each of the various assessment areas within a (sub)region, or a figure showing the proportion of the area where GES is achieved.

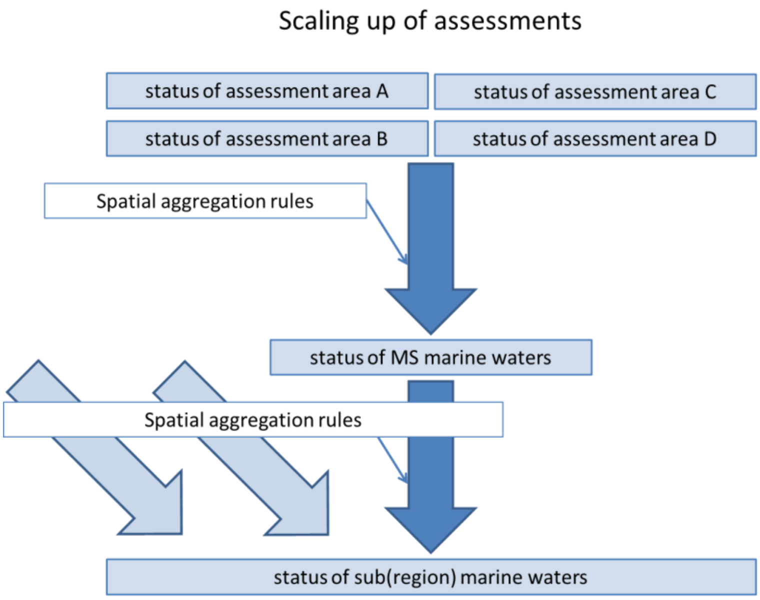


Figure 3.1 Schematic picture of a possible approach for scaling up of assessments for a number of assessment areas

In addition to the question of how to scale up the MSFD assessments comes the question on how to incorporate evaluations under other EU legislation, like the Water Framework Directive or the Bird and Habitat Directives. These Directives operate at different scales and the assessments under these Directives only apply to certain areas (for example, only coastal waters under the WFD), and additionally the Directives cover only some elements of GES. A way must be found to ensure that the MSFD assessments complement the other assessments to ensure an efficient assessment of all Directives.

3.2 Scaling up from assessment areas to larger spatial scales

Various methods have been used in environmental assessments to make the step from assessment results in subunits at a small spatial scale (specific sites or subareas) to an overall assessment of the status for a larger area. These methods can be separated in two different approaches that determine whether ‘good” status has been achieved:

1) All assessments at the small spatial scale have to meet the quality standards according to the one-out all-out (OOAO) approach.
2) The assessments at the small spatial scale are combined applying a specific rule. This can be a simple averaging method or a weighted procedure. In this case it is possible that while in some subunits good status is not achieved, the overall assessment for the larger area shows good status.

Below, examples of some of these approaches are given.

3.2.1 One-out all-out (OOAO)

In the OOAO approach, the status of the spatial subunit with the lowest classification determines the overall status.

3.2.2 Averaging

An overall assessment for a larger area can be constructed by averaging of the underlying assessments for subareas. Averaging can be done by simply calculating the arithmetic means, but it can also be done by assigning weights to areas or metrics.

Example, Box 4: Traffic light system

In the OSPAR assessment of CEMP monitoring data on contaminant concentrations in fish, shellfish and sediment for the Quality Status Report 2010 (OSPAR 2010), assessments for specific areas were expressed in a traffic light system by comparing values for individual contaminants, in individual matrices (e.g. sediment, water, biota) and at individual stations to assessment criteria (OSPAR 2009c).
In a next step, geographical sub-areas were defined for which integrated assessments should be made. These sub-areas consisted of a large offshore areas and a number of near shore waters. For each contaminant within a sub-area, data across stations were combined by calculation of the proportion of blue, green and red station assessments (see figure below). The final step was to combine data across sub-areas within Regions to obtain Region-scale assessments, by averaging the sub-area assessments.
The final presentation used in the draft QSR document allows comparisons to be made of environmental quality for each contaminant within a OSPAR Region, and also individual contaminants across OSPAR Regions (see figure below).

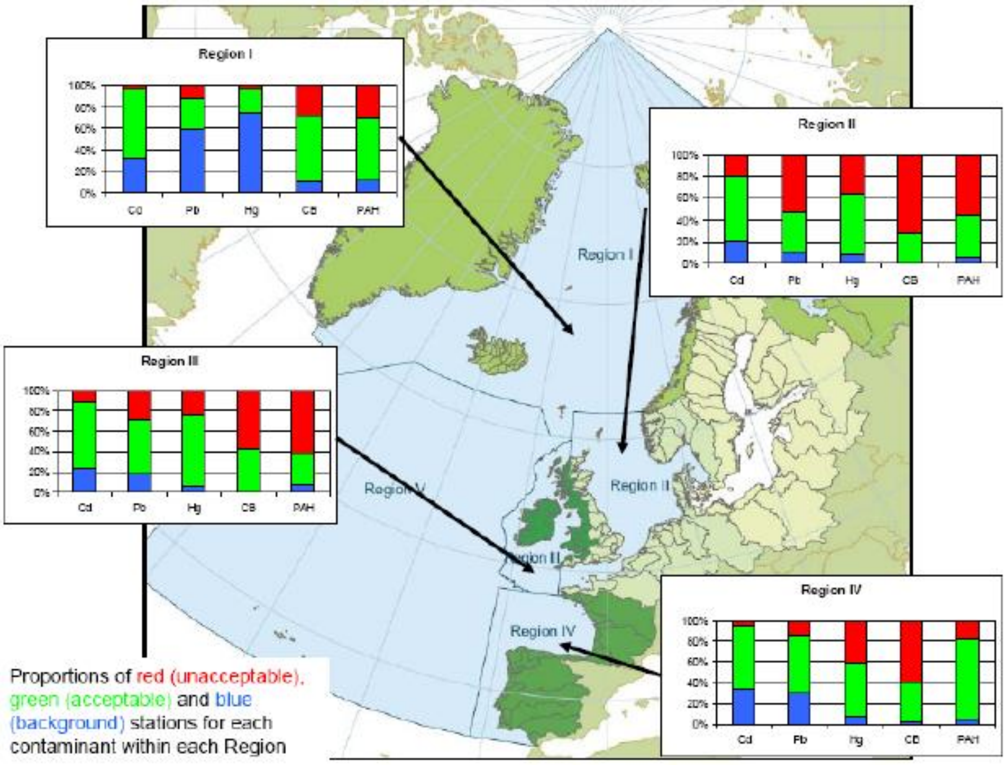


Figure 4. Integrated data presentation used in the draft OSPAR QSR 2010

Integrated data presentation across OSPAR Regions (*Law* et al. *2010*).

3.2.3 Spatial scale rating

This is an aggregation method where the overall assessment result for an area is calculated by weighting the assessment results of sub-areas in proportion to the surface area of those sub-areas. The assessment value of each sub-area is multiplied by the weight of this area, and the sum of all products (assessment values times weight) are divided by the sum of all weights to arrive at a weighted assessment values for the entire area. This method has been applied for integrative assessments of coastal waters along the Basque coast in northern Spain (Borja *et al.* 2009a, Borja *et al.* 2011b).

Another example of surface or coastline length spatial rating is given by Orfanidis *et al.* 2003. Spatial and temporal changes of benthic macrophytic communities are identified by seasonal sampling along transects or at sampling sites. The ecological status class of each sampling unit is expressed as a number (Ecological Evaluation Index EEI). The overall assessment of an area is calculated by averaging the assessments of the different sites, weighted for their length or surface area (see Figure 1a in Orfanidis *et al.* (2003)).

3.2.4 Minimum proportion achieving target

In this approach a certain percentage of stations should meet an environmental target or threshold. This rule can be applied to stations within an assessment area, but also to assessment units nested within a larger area.

The "percentage of stations "passing thresholds" option was followed in the Greek Initial Assessment report for the environmental targets of D6. The report suggests that if 90% of the monitored stations pass the threshold (Simboura *et al.* 2012) for indicators 6.2.1 and 6.2.2, the assessment area arrives at GES considering these indicators.

3.3 Analysis of approaches by Regional Sea Conventions

3.3.1 Baltic Sea (HELCOM)

In the initial holistic assessment of the ecosystem health of the Baltic Sea (HELCOM 2010b) two different methods were used to integrate the assessments on eutrophication, hazardous substances and biodiversity. In the first approach the assessments on eutrophication, hazardous substances and biodiversity were simply merged to provide a combined overview of the results of each assessment. This resulted in an assessment for approximately 80 sites in the Baltic Sea, Sound and Kattegat (see paragraph 5.3.6). The second approach used an integrative indicator-based assessment tool (HOLAS; (“tool for the Holistic Assessment of ecosystem health”) based on the same indicators as in the three underlying assessment methods. This approach resulted in an interpolated map of ‘ecosystem health’ in the Baltic Sea, linking the various geographic divisions and assessment units.

3.3.2 Mediterranean Sea (Barcelona Convention)

The question how to scale up the analyses for an effective ecosystem approach was raised in the Initial Integrated Assessment report of the Mediterranean Sea (UNEP/MAP 2012b). It was suggested that the three Clusters of the Correspondence Groups addressing 1) Pollution and litter related EOs; 2) Biodiversity and Fisheries related EOs and 3) Integrated Coastal Zone Management and Hydrological Conditions related EOs, should consider thematic integration when targets are being defined, with eventually integration across all EOs. However there was no advice on rules to be applied.

3.3.3 Black Sea (Bucharest Convention)

For the Black Sea, no information was found on methodologies for the upscaling of assessments.

3.3.4 Northeast Atlantic (OSPAR)

In the eutrophication assessment (OSPAR 2008) results are presented for each individual assessment area. No attempt was made to scale up the assessments to a higher geographic level, like for example the OSPAR regions. A qualitative summary for each OSPAR region is presented in the Quality Status Report (OSPAR 2010), which was based on expert judgment.

For the assessment of hazardous substances, a method was developed to scale up assessments from monitoring stations to sub-areas to overall assessments for an OSPAR region (see Example box 5).

3.4 Synthesis

HELCOM and OSPAR present the results of assessments at the scale of the Baltic Sea and the NE Atlantic, respectively (see e.g. HELCOM 2009b; 2010a; 2010b OSPAR, 2010). Examples are maps for the eutrophication assessment of OSPAR or the interpolated maps of HELCOM.

However, in most cases the presentation consists only of a merging of the assessments for sub-areas. The OSPAR approach for the assessment of hazardous substances is the only example where rules for scaling up of the assessments have been applied.

4 Aggregation

This chapter discusses the different methods that can be applied to aggregate criteria and indicators within and among descriptors to eventually come to an assessment of GES for a geographic area. The general principles for aggregation are discussed and current approaches are illustrated using examples from literature.

4.1 MSFD requirements

Article 3(4) of 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*". To assess whether or not GES has been achieved, aggregation within Descriptors is required to move from the evaluation at the level of indicators to an assessment of status within a Descriptor. In addition, Cardoso *et al.* (2010) mention a third level of integration, the assessment of GES across all Descriptors.

4.2 General principles for aggregation

The WFD focuses on a limited number of ecosystem components (the Biological Quality Elements or BQEs), that are combined through a one-out all-out (OOAO) approach which means that the status of the worst element determines the final status of the overall approach. This can be considered a precautionary approach (Borja *et al.* 2010). In contrast to the WFD, the MSFD can be considered to follow a 'holistic functional approach', as it takes into account structure, function and processes of the marine ecosystem, and also uses wider descriptors which not only relate to biological and physic-chemical indicators but also to pressure indicators (Borja *et al.* 2010; Borja *et al.* 2013). The MSFD concentrates on the set of 11 descriptors which together summarize the way in which the whole system functions. The MSFD requires the determination of GES on the basis of the qualitative descriptors in Annex I, but does not specifically require one single GES assessment, in contrast to the WFD.

There are many methodological challenges and uncertainties involved in establishing a holistic ecosystem assessment, if 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 the indicator aggregation rules (Ojaveer and Eero 2011; Borja *et al.* 2013; Caroni *et al.* 2013). Different methodologies can be applied for aggregating indicators, which vary, amongst others, in the way the outliers influence the aggregate value.

There are four levels of aggregation or integration required to move from evaluation of the individual metrics or indicators identified by the Task Groups to an assessment of GES (Cardoso *et al.* 2010);

- Aggregation of metrics/indices within indicators
- Aggregation of indicators within the criteria of a Descriptor (for complex Descriptors)
- Status across all the criteria of a Descriptor
- Status across all Descriptors

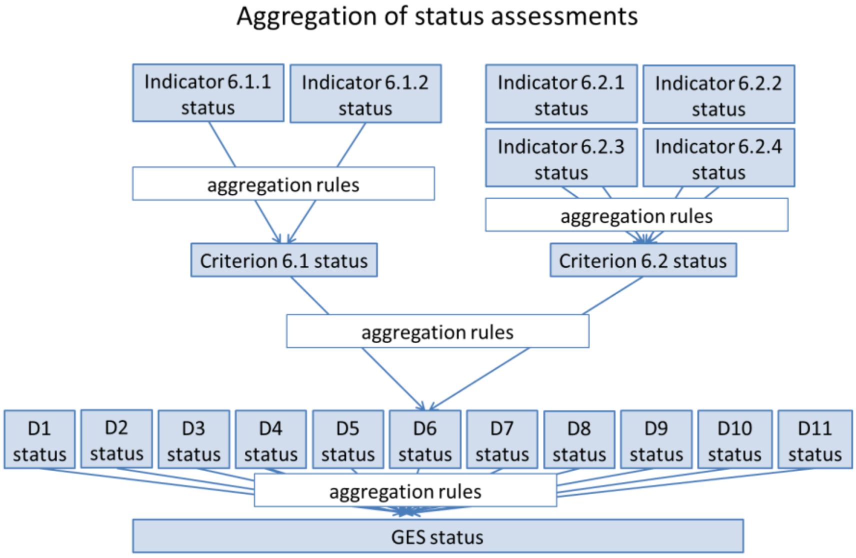


Figure 4.1 Schematic picture of a possible approach for aggregation of indicators, criteria and descriptors.

As one moves up the scale from metric/indicator level to overall GES, the diversity of features that have to be integrated increases rapidly (Figure 4.1). This poses several challenges arising from the diversity of metrics, scales, performance features (sensitivity, specificity, etc.) and inherent nature (state indicators, pressure indicators, response indicators) of the measures that must be integrated, that are discussed in the next paragraphs.

4.2.1 Within Descriptor integration

This integration method relates to the methods that might be required within a Descriptor to take account of multiple criteria and indicators, and where not all indicators and/or attributes reach their desired levels or targets. The management group report (Cardoso, 2010) summarizes the methods in the Task Group reports for a within Descriptor integration, 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 GES will be set at the environmental status of the indicator and/or attribute assessed at the worst state for the area of concern.

Table 4.1 summarizes the approaches (based on the individual Task Group reports) to integrate attributes; information on methods for integration of indicators can be found in the Task Group reports.

Table 4.1 Summary of Task Group approaches to Integrate Attributes within a Descriptor (Cardoso et al. 2010).

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

4.2.2 Cross-Descriptor integration

The last level of integration relates to the methods that could be used to integrate the results across all Descriptors. Discussion of how to combine or integrate the results of each Descriptor into an overall judgment of GES 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 scale of marine (sub)regions.

Borja *et al.* 2013) argue that aggregation rules for the integration across indicators and criteria does not necessarily have to follow the same approach as cross-descriptor integration. Also, different approaches could be considered for state descriptors (D1, D3, D4, D6) and pressure descriptors (D2, D5, D7, D8, D9, D10, D11).

As Cardoso *et al.* (2010) pointed out, cross-descriptor aggregation at the scale of (sub)regions runs the risk of blending and obscuring the information that is necessary to follow progress towards GES and to inform decision-makers about the effectiveness of policies and management. It may lead to masking of problems within specific descriptors, or to a high probability of not achieving GES if OOAO is used.

The next paragraph will show examples and discuss two groups of approaches, one where assessments are done based on a “worst case” approach and one where integrative assessments are made.

4.3 Overview of current approaches

Based on a literature review, we identified a number of different approaches for aggregation rules that combine a number of variables (which could be metrics, indicators, or criteria) in an overall assessment. An overview of the methods is given in the table below. The following paragraphs describe the methods in more detail.

General approach	Details of method	Advantages	Disadvantages	References
One-out all-out (OOAO) principle	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 May include double-counting	EC 2005 Caroni *et al.* 2013 Ojaveer and Eero 2011 Borja *et al.* 2013
	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	OSPAR 2009c Tueros *et al.* 2009
Conditional rules	A specific proportion of the variables have to achieve good status	Focuses on the key aspects (i.e. biodiversity descriptors)	Assumes that GES is well represented by a selection of variables	Simboura *et al.* 2012 Piet *et al.* 2010 Borja *et al.* 2013
Averaging approach	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	Ojaveer and Eero 2011 Caroni *et al.* 2013
	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	Ojaveer and Eero 2011 Caroni *et al.* 2013 Borja *et al.* 2013
	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	Ojaveer and Eero 2011 Borja *et al.* 2013
Scoring or rating	Sum of weighted scores	Different weights can be assigned to the various	Problem of agreeing on weights	Borja *et al.* 2004a Borja *et al.* 2010,

General approach	Details of method	Advantages	Disadvantages	References
		elements	Metrics may not be sensitive to the same pressures	Borja *et al.* 2011b Cyprus IA report
Multimetric approaches	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	Vollenweider *et al.* 1998 Borja *et al.* 2013
Multi-dimensional approaches	Multivariate analyses	No need to set rigid target values, since values are represented within a domain	Results are hard to communicate to managers	Rice *et al.* 2010 Tett *et al.* 2007 Borja *et al.* 2013
Decision tree	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	Borja *et al*.2008 Borja *et al*. 2009 OSPAR 2008
High-level integration	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	HELCOM 2010a

4.3.1 One-out-all-out

The one-out all-out approach is used within the WFD to integrate within and across Biological Quality Elements (BQEs) (EC 2005). This approach follows the general concept that a particular status assigned to a water body depends on the quality element with the lowest status, and consequently, the OOAO approach results in a “worst case” assessment (Figure 4.2).

Several examples of application of the OOAO principle and associated issues are given in Annex III.

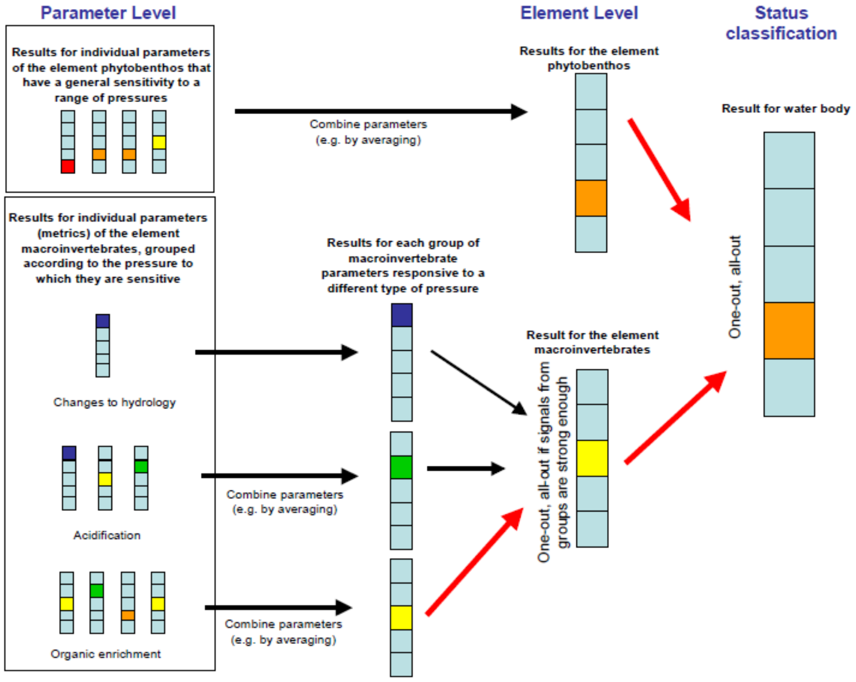


Figure 4.2 Scheme showing the general concept of OOAO principle from WFD classification guidance (EC 2005).

As an alternative to the one-out all-out approach, a “two-out all-out” approach is sometimes used. Status will be considered “not good” when two underlying variables do not meet the quality standard for good status.

An example is the OSPAR assessment for contaminants (OSPAR 2009c; OSPAR 2010), where for some groups of substances (e.g. PAHs) an assessment was based on underlying assessments of single assessments. A “one out all out” approach was considered too sensitive to uncertainties in either the data or the assessment criteria. A “two out all out” approach was found to be more robust.

Another example of "two-out-all-out" approach is presented in Tueros *et al.* (2009) for the assessment of chemical status for the WFD; integrating water, sediments and biomonitor matrices: if more than one variable does not meet the quality standards, good chemical status is not achieved (looking also at the 3 integrated matrices).

4.3.2 Conditional rules

An example of this approach is the application of a biotic index to describe Seafloor integrity in Greek waters, Simboura *et al.* (2012) combined indicators 6.2.1 (presence of sensitive and/or tolerant species), 6.2.2 (indices assessing species diversity (H) and richness (S), and the proportion of opportunistic to sensitive species (Bentix). The rule applied required that at least two indicators should pass the threshold in order to achieve GES for D6.

The Task Group report on commercially exploited fish and shellfish (reference!) presents another example of the conditional rule in which a stock can only achieve GES if all three criteria for the attributes are fulfilled. Criteria used are:

- Exploited sustainably consistent with high long-term yield
- Full reproductive capacity

- Healthy age and size distribution

However, when aggregating across stocks only the sustainable exploitation criterion and full reproductive capacity criterion need to be fulfilled by all stocks (i.e. $F<F_{MSY}$ and $SSB>SSB_{pa}$ for 100% of the stocks).

4.3.3 Averaging approach

The averaging approach is the most common (Shin *et al.* 2012) and consists of simple combinations of indicators, using calculation methods like arithmetic average, hierarchical average, weighted average, median, sum, product or combinations of those rules, to come up with an overall assessment value.
Several examples of applications are given below.

Average/median rules

Individual metrics can be combined into an overall assessment value by averaging. A way to do this is by converting the status class for each into a numeric value (for example, when using WFD status classes high = 1, good = 2, moderate = 3, poor = 4, bad = 5). The arithmetic average can then be calculated and rounded to the nearest class. Similarly, instead of averaging the median value can be calculated.
When metrics are defined at various hierarchical levels, the average or median value of the transformed indicator values can be calculated at each level of aggregation (Ojaveer and Eero 2011).
Caroni *et al.* (2013) performed data analysis with a lake monitoring dataset, and reported that the average rule is recommended when combined BQEs are sensitive to a single pressure (e.g. eutrophication and acidification). Then the best approach is to average the metrics responding to the same pressure within each BQE (group them by pressure, then by BQE).

Ojaveer and Eero (2011) show that a simple average or median of all indicators is not necessarily the best solution in every circumstance, considering that different indicators meet various screening criteria differently. Individual indicators could be weighted differently in the averaging procedure. However, this requires an adequate basis for assigning weights to indicators which is not always available.
In Caroni *et al.* (2013) the weighted average rule was used only for the subset of data having four BQEs (17 lakes). The BQE fish was given a lower weight, while the remaining BQEs had equal weights. The fish BQE was down-weighted because it appeared to be the most stringent among all BQEs, as it classified the highest number of lakes in moderate or worse status.

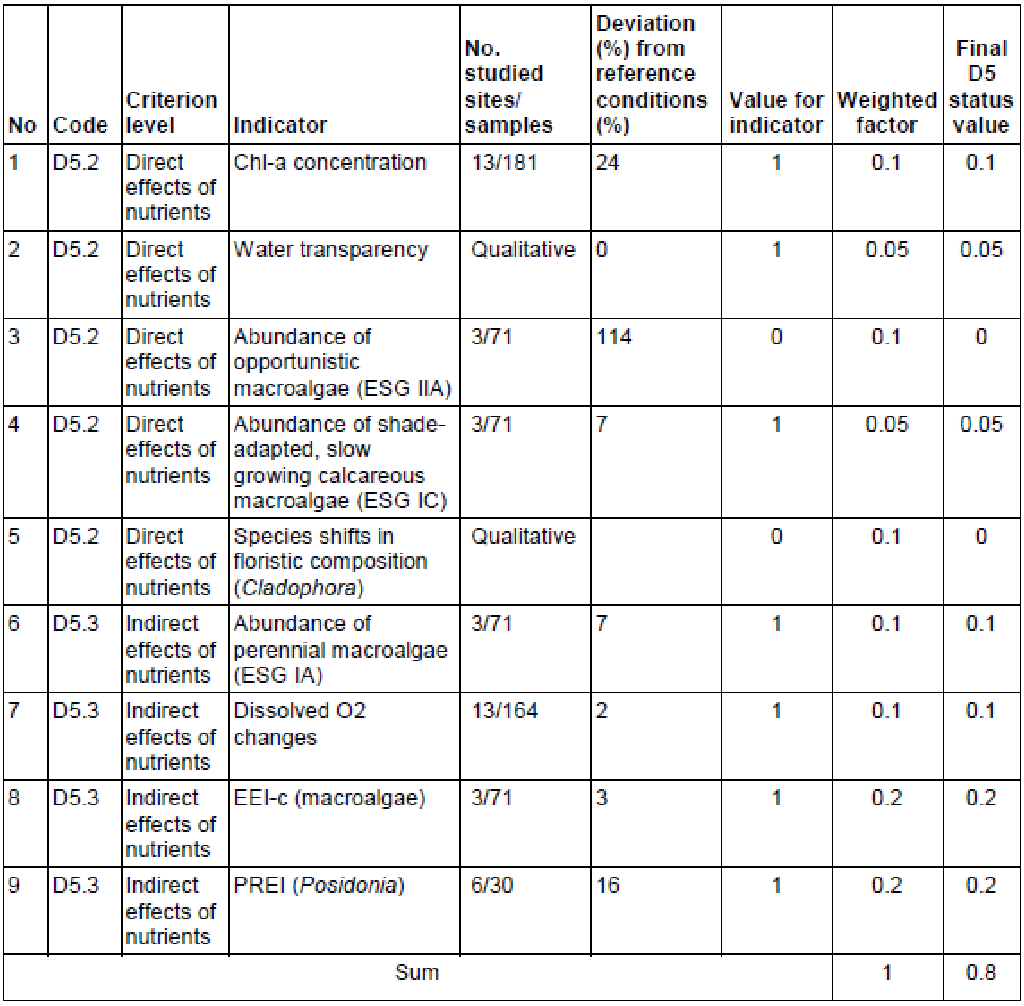
4.3.4 Scoring or rating

In this method 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. (2004b) to calculate an integrative index of quality (IIQ). Elements that were combined consisted of parameters describing water quality, sediment quality and biomonitors.
For a cross-descriptor aggregation, Borja et al. (2010) developed an integrated approach combining the 11 descriptors of MSFD based on the WFD, HELCOM and OSPAR experience. 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 weight assigned to each descriptor, an overall environmental status value was derived.

A similar approach was implemented for an integrative MSFD assessment in parts of the Bay of Biscay along the Basque coast (Borja *et al.* 2011b).

Cyprus used a similar method for aggregating indicators within descriptors in its Initial Assessment report. This aggregation method was followed for descriptors D1 (macroalgae, angiosperms, benthic macroinvertebrates, fishes), for D3 (criteria 3.1, 3.2, 3.4 and for D5 (criteria 5.2, 5.3).
After calculating indicator values based on a spatial comparison with reference conditions or expert judgment, a status value and a weighting factor was assigned to each descriptor. The sum of the weighted status values of each indicator was used to calculate a final assessment value for a Descriptor (see Table 4.2 for the application for Descriptor 5).

Table 4.2 Initial assessment of Cyprus marine waters using eutrophication descriptor No 5 (D5). Note: a threshold value of 0.75 is used to determine whether GES is achieved or not.



No	Code	Criterion level	Indicator	No. studied sites/ samples	Deviation (%) from reference conditions (%)	Value for indicator	Weighted factor	Final D5 status value
1	D5.2	Direct effects of nutrients	Chl-a concentration	13/181	24	1	0.1	0.1
2	D5.2	Direct effects of nutrients	Water transparency	Qualitative	0	1	0.05	0.05
3	D5.2	Direct effects of nutrients	Abundance of opportunistic macroalgae (ESG IIA)	3/71	114	0	0.1	0
4	D5.2	Direct effects of nutrients	Abundance of shade-adapted, slow growing calcareous macroalgae (ESG IC)	3/71	7	1	0.05	0.05
5	D5.2	Direct effects of nutrients	Species shifts in floristic composition (*Cladophora*)	Qualitative		0	0.1	0
6	D5.3	Indirect effects of nutrients	Abundance of perennial macroalgae (ESG IA)	3/71	7	1	0.1	0.1
7	D5.3	Indirect effects of nutrients	Dissolved O2 changes	13/164	2	1	0.1	0.1
8	D5.3	Indirect effects of nutrients	EEI-c (macroalgae)	3/71	3	1	0.2	0.2
9	D5.3	Indirect effects of nutrients	PREI (*Posidonia*)	6/30	16	1	0.2	0.2
Sum							1	0.8

4.3.5 Multimetric indices to combine indicators

The Task Group 6 report (Rice *et al.* 2010) recommends the use of multimetric indices or multivariate techniques for integrating indicators of species composition attribute of D6 such as diversity, distinctness, complementarity/(dis)similarity, species-area relationships.

The TRIX index (Vollenweider *et al.* 1998) is an example of a multi-parameter Eutrophication index combining water column measures (i.e. chlorophyll-a), dissolved oxygen and nutrients. The TRIX index is calculated by the formula:

*TRIX= [log10(CHLa * D%O * N * P) + 1.5]/1.2*

where: Chl a= Chlorophyll a (µg L-1), D%O= Oxygen as an absolute deviation (%) from saturation, N= Dissolved inorganic nitrogen N-NO3+NO2 (µg-at L-1), P= Total phosphorus P-PO4 (µg-at L-1).

In TRIX the various indicators are integrated, and cannot be judged on their own. Scaling of a multi-metric index may be an issue. In the case of TRIX, the combination of CHLa and nutrient levels may complicate the performance of the indicator, as in some cases nutrient and phytoplankton parameters can be negatively correlated (Primpas and Karydis 2011).

Another multimetric index combining chlorophyll and nutrient concentrations was developed by Primpas *et al.* 2010).

Another example of a multimetric index is the Benthic Quality index BQI (Rosenberg *et al.* 2004). This index combines several parameters describing characteristics of the benthic fauna (tolerance value $ES50_{0.05}$ of each species, the mean relative abundance A of each species, and the mean number of species S).

$$\mathrm{BQI} = \left(\sum_{i=1}^{n} \left(\frac{A_i}{\mathrm{tot}A} \times \mathrm{ES50}_{0.05i} \right) \right) \times {}^{10}\log(S+1)$$

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).

4.3.6 Multidimensional approaches

The Task Group 6 report (Rice *et al.* 2010) discusses multivariate methods as an alternative for multi-metric methods to combine a number of parameters. Multivariate methods, such as the Discriminant Analysis or Factor Analysis combine parameters in a multi-dimensional space. For assessment purposes, areas need to be classified into groups of GES and non-GES.

Another example of this approach is a methodology proposed by Tett *et al.* (2007). Various variables are combined in a multi-dimensional presentation and indicate an “undesirable disturbance” when they move outside an area that is considered to be representative for type-specific conditions (Figure 4.3).

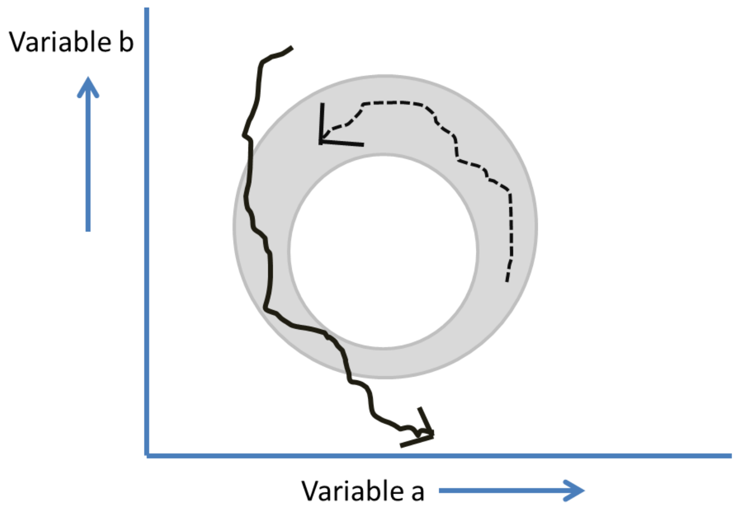


Figure 4.3 Generalized diagram showing a state-space defined by two variables; a 'normal' or 'reference' domain is shown by the grey region; evolution of variable values over time representative of an undisturbed situation are represented by the broken line; a disturbance is a movement outside this region represented by the full line. Redrawn from Tett et al. (2007).

4.3.7 Decision tree

Borja *et al.* (2009a) describe a methodology that integrates several biological elements (phytoplankton, benthos, algae, phanerogams, and fishes), together with physico-chemical elements (including pollutants) into a quality assessment. For each station, decision trees were used to integrate (i) water, sediment and biomonitor chemical data to achieve an integrated physico-chemical assessment and (ii) multiple biological ecosystem elements into an integrated biological assessment. This decision tree was presented by Borja *et al.* (2004a) for the implementation of the WFD, but methodologies have been updated in Borja *et al.* (2009a). The proposed methodologies can accommodate both WFD and the MSFD (see Figure 2 in Borja *et al.* (2009a)) and Figure 4.4 for the basic design of a decision tree.

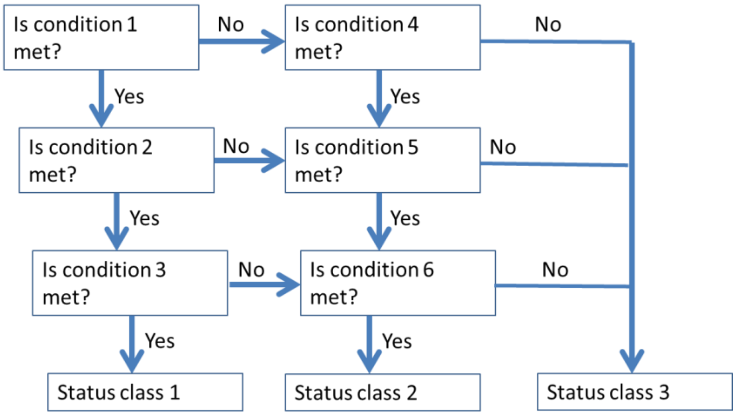


Figure 4.4 Example of decision tree to integrate various conditions in a final assessment result

Borja *et al.* (2008) also present a decision tree for sediment contaminants and biomonitors for the assessment of integrative chemical status (Figure 4.5). This decision tree contained the principle of two-out-all-out rule explained above.

The OSPAR Eutrophication assessment (OSPAR 2008) also applies a decision tree for the final assessment. Eutrophication effects are determined for three categories (degree of nutrient enrichment, direct effects and indirect effects). Within those categories, various indicators are used and scoring applies the OOAO rule. For the second step, a decision tree is used.

4.3.8 High-level aggregation

An example of a high-level aggregation, where assessments for several ecosystem components are merged into a final assessment, is the HELCOM-HOLAS project (HELCOM 2010b). 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 tool on the assessment results of those three groups for the final assessment (Figure 4.5).

This approach could be considered a pragmatic compromise, reducing the risks associated with OOAO while still giving an overall assessment.

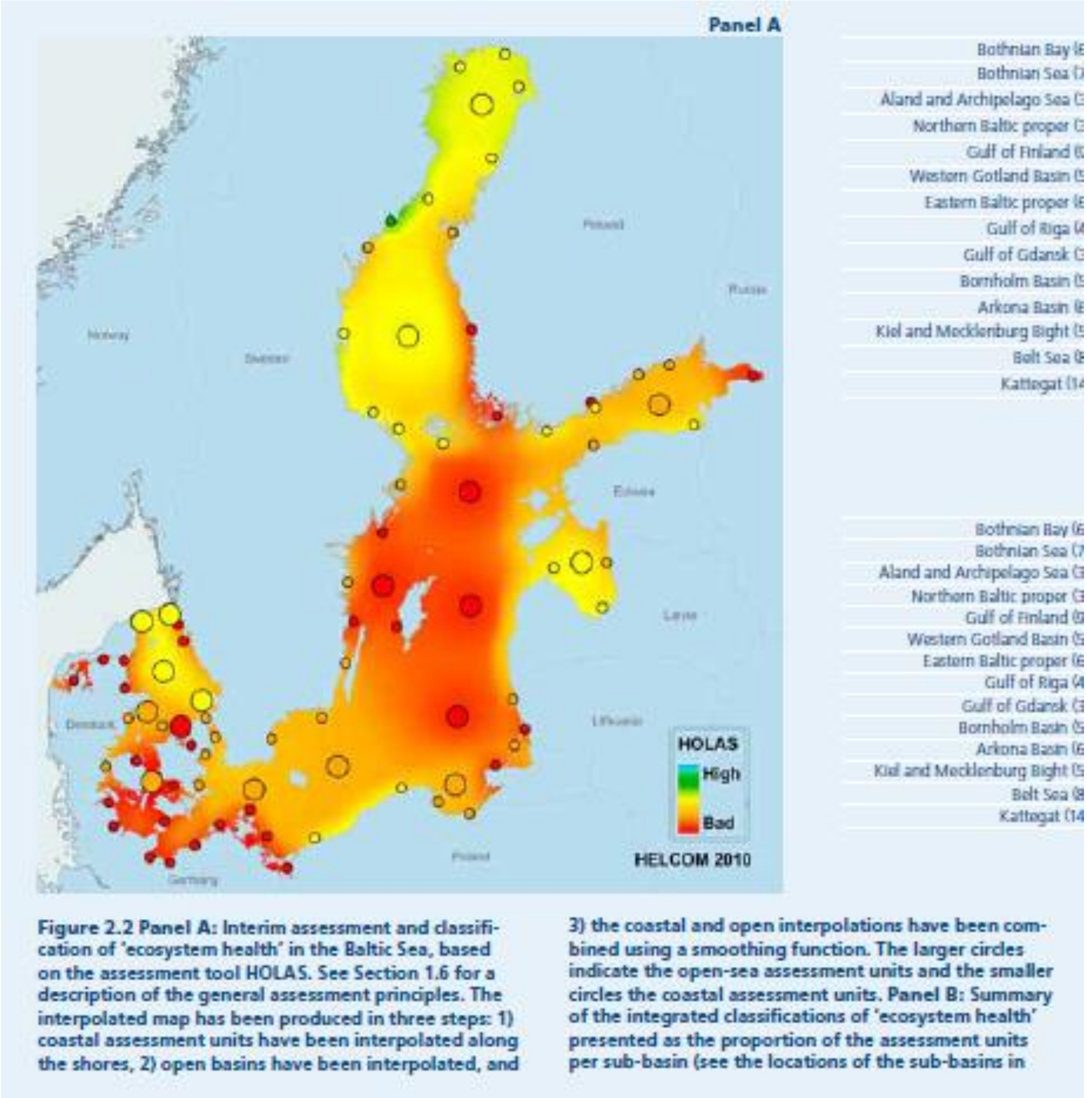


Figure 2.2 Panel A: Interim assessment and classification of 'ecosystem health' in the Baltic Sea, based on the assessment tool HOLAS. See Section 1.6 for a description of the general assessment principles. The interpolated map has been produced in three steps: 1) coastal assessment units have been interpolated along the shores, 2) open basins have been interpolated, and 3) the coastal and open interpolations have been combined using a smoothing function. The larger circles indicate the open-sea assessment units and the smaller circles the coastal assessment units. Panel B: Summary of the integrated classifications of 'ecosystem health' presented as the proportion of the assessment units per sub-basin (see the locations of the sub-basins in

Figure 4.5 Aggregation in HOLAS tool (HELCOM, 2010a).

The "Biopollution level" (BPL) index for the estimation of the magnitude of bio-invasion impacts can be considered as a within Descriptor (2) aggregation holistic tool, as well. The BPL index takes into account the abundance and distribution range of NIS in relation to native biota in the invaded area and aggregates data on the magnitude of the impacts these species have on: native communities, habitats and ecosystem functioning (free access to BPL assessment system is provided at: www.corpi.ku.lt/~biopollution). BPL aggregates the results of the assessment into five categories: “No bioinvasion impact”, “Weak”, “Moderate”, “Strong” and “Massive" (Olenin *et al.* 2010).

4.4 Synthesis

Borja *et al.* (2009b) discussed the challenge of assessing ecological integrity in marine waters, and suggest that simple approaches such as the ‘OOAO’ principle of the WFD, which determines the final status of a water body on the basis of the worst rated element, 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 logical approach as a precautionary rule, in an ideal world where the status based on each BQE can be measured without error. 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 class. The OOAO principle has therefore been criticized as it increases the probability of committing a false positive error, leading to an erroneous downgrading of a waterbody (Borja and Rodriguez 2010; Caroni *et al.* 2013). The OOAO rule results in very conservative assessments with full implementation of the precautionary principle (Ojaveer and Eero, 2011). In the case of the MSFD, with eleven descriptors and more than 50 indicators, the probability of not achieving good status becomes very high (Borja *et al.* 2013).

When the OOAO principle is not applied, but calculation rules are used to combine parameters, the choice of appropriate aggregation rules is very important. Ojaveer and Eero (2011) reported that an important aspect in reference-based assessment is the selection of an indicator aggregation formula. Their analysis showed that the assessment results can be highly sensitive to aggregation rules.

A prerequisite for the combination of various parameters is that they are sensitive to the same pressure (Caroni *et al.* 2013). In that case, different methods can be used to combine parameters (e.g. medians, averages, etc.). Caroni *et al.* (2013) recommend an OOAO approach when aggregation 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 (EC 2005) also advises to use one-out all-out when combining parameters/indicators that are sensitive to different pressures.

Ojaveer and Eero (2011) showed that in cases where a larger number of indicators is available, the choice for applying 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, application of the median of the indicator values resulted in very different assessment results compared to assessments based on averages.

The way the indicators are hierarchically arranged influences the assessment results as well, but 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 averages or medians. An adequate basis for assigning weights is not always available, however (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.

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 intercorrelated (see e.g. the discussion on the TRIX index in Primpas and Karydis (2011).

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

Through the use of the OOAO approach in the WFD it has been recognized that the OOAO rule results in a conservative approach, following the precautionary principle, and with a high probability of a type 1 error, in particular when a large number of variables is involved (Borja *et al.* 2013; Borja and Rodriguez 2010; Caroni *et al.* 2013; Ojaveer and Eero 2011). Alternative methods for integrating multiple BQEs in the WFD are currently being considered (Caroni *et al.* 2013).

5 Analysis of approaches by Member States in the Initial Assessments

In this chapter, an overview is given of the approaches that MS have taken to deal with spatial scales and aggregation in the 2012 reporting for the MSFD. The information comes from the electronic reporting by MS (4GEO.xml files on cdr.eionet.europa.eu), a GIS analysis by the European Topic Centre for Inland, Coastal and Marine Waters, and relevant parts of the national expert reviews.

Good environmental status has to be determined at the level of the marine region or subregion (Art 3.5), but the assessment of environmental status can be done at other spatial scales. In the Article 8 reporting by MS a wide variety of geographical scales has been used. Eight MS have used a geographic scale that is similar to their marine waters within a subregion, i.e. have applied only one assessment area. Fifteen MS have subdivided their marine waters in more than one assessment area. (Table 5.1).

Table 5.1 Overview of the number of assessment areas used by Member states (source: cdr.eionet.europa.eu)

Member state		Number of assessment areas per subregion	Remarks
BE	Belgium	1	
BG	Bulgaria	>1	
CY	Cyprus	1	
DE	Germany	1	
DK	Denmark	1	
EE	Estonia	1	
EL	Greece	>1	EL has >1 assessment area in Aegean-Levantine Sea
ES	Spain	>1	ES uses subdivisions within the subregions, and mentions the use of various scales for analysis, depending on the feature
FI	Finland	>1	
FR	France	1	
HR	Croatia	>1	
IE	Ireland	1	
IT	Italy	>1	
LT	Lithuania	>1	
LV	Latvia	>1	
MT	Malta	>1	
NL	Netherlands	1	
PL	Poland	>1	
PT	Portugal	>1	
RO	Romania	>1	
SE	Sweden	>1	
SI	Slovenia	>1	
UK	United Kingdom	>1	UK mentions informal use of 8 biogeographically defined assessment areas

Table 5.2 Overview of the number and range in surface areas of assessment areas for each subregion (source: electronic reporting by MS).In some cases, information on surface areas was not available yet.

Marine region	Subregion	Member state	Number of assessment units	Surface area of assessment units (min-max) (in 10^2 km^2)
Baltic Sea		DE	1	155
		DK	1	288
		EE	1	365
		FI	7	25-301
		LT	2	1-64
		LV	10	2-290
		PL	8	
		SE	48	<1-872
Black Sea		BG	12	
		RO	3	5-279
Mediterranean Sea	Adriatic Sea	EL	1	23
		HR	6	
		IT	48	15-610
		SI	4	<1-849
	Aegean-Levantine Sea	CY	1	1308
		EL	5	341-1737
	Ionian Sea and Central Mediterranean Sea	EL	1	1821
		IT	48	<1-2195
		MT	34	
	Western Mediterranean Sea	ES	21	259-2313
		FR	1	1108
		IT	72	13-5912
		UK	1	
Northeast Atlantic	Bay of Biscay and Iberian Coast	ES	17	149-3045
		FR	1	1881
		PT	2	
	Celtic Sea	FR	1	284
		IE	1	4888
		UK	5	307-3237
	Greater North Sea	BE	1	35
		DE	1	409
		DK	1	765
		FR	1	441
		NL	1	589
		SE	16	2-143
		UK	3	219-1811
	Macaronesia	ES	6	4851
		PT	2	

Within (sub)regions differences appear in the application of geographic areas by Member states. Large differences in the approach to spatial scales are apparent. While eight MS use one assessment area, twelve MS have used more than one assessment area, and the surface area of the assessment areas ranges from <100 km^2 to >100000 km^2 (Table 5.2). There is no relation between the surface area of the marine waters of a MS, and the number of assessment areas that were reported (Figure 5.1)

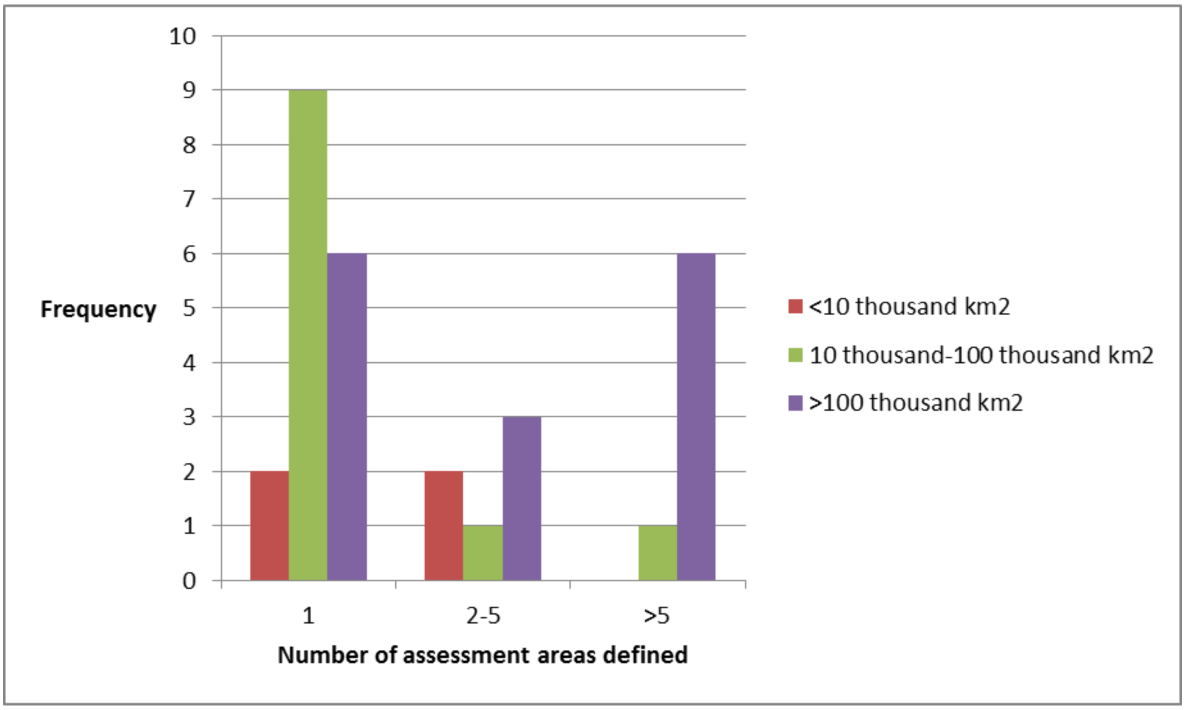


Figure 51 Frequency of number of assessment areas defined by MS. A distinction is made between MS with marine waters with a surface area of <10^4 km^2, 10^4-10^5 km^2 and >10^5 km^2.

Additional information per Member State:

BE, Belgium has defined one assessment area. GES and environmental targets are defined for the entire Belgian part of the North Sea, with the exception of Descriptor 3 which should be implemented at the scale of the (sub)region according to Belgium. No further distinction in assessment areas is made.

BG, Bulgaria has defined the following formal assessment area for the pelagic zone: coastal (0-30 m below sea level), shelf (30-200 m BSL) and open sea (>200 m BSL), based on satellite-derived chlorophyll-a concentrations and review of the existing data/literature, showing correspondence with the previous zonation based on anthropogenic pressure, sea currents, productivity and bathymetry.

The benthic zone is divided into 10 assessment areas based on the substrate characteristics and associated communities. There are five benthic assessment areas identified within the coastal zone, three within the continental shelf, and two in the open sea (outer shelf area and abyssal area).

CY, Cyprus has defined one assessment area.

DE, Germany has defined one assessment area within the Baltic Sea and one area within the Northeast Atlantic. The initial assessment, characteristics of GES and targets and associated indicators have been developed for each marine (sub)region. For a number of descriptors, specific details are provided with regard to the Wadden Sea. In those cases where the Wadden Sea is specifically mentioned, Germany refers to the Trilateral Monitoring and Assessment Programme (TMAP) carried out jointly by the Netherlands, Denmark and Germany for the Wadden Sea. According to the Wadden Sea Secretariat, the standards defined in the TMAP combine the requirements of the EU Water Framework Directive (WFD), Habitats and Birds Directives. The TMAP Common Package parameters, used to make the 2009 Quality Status Report, were revised in 2007, before the adoption of the MSFD. However, the Secretariat notes that the requirements of the MSFD “will have an influence on parameters which are predominantly foreseen to be integrated and adjusted to the TMAP programme”. In those cases where the TMAP is used as reference values/standards (e.g. D8), the standards reflect the EU (e.g. EQS) and OSPAR (e.g. EAC) standards.

DK, Denmark has defined one assessment area within the Baltic Sea and one area within the Northeast Atlantic.

EE, Estonia has not defined specific assessment areas for the purpose of the MSFD reporting. It mentions in its report the typology has been defined on the basis of the Water Framework Directive but no similar typology has been developed for the MSFD.

EL, Greece has distinguished one assessment area in the subregion Adriatic Sea, and one in the Ionian Sea and Central Mediterranean Sea. In the subregion Aegean-Levantine Sea five assessment areas are distinguished. The GES definitions and targets are defined for the whole of the Greek marine waters together, with no distinction of specific assessment areas. For biological features, the areas assessed are the Aegean Sea, and the Adriatic and Ionian Seas together. For some pressures (e.g. physical loss and damage, underwater noise), the assessment is not made in detail for each area but rather in general for the whole of the marine waters (with from time to time specific examples for one or the other area). In terms of aggregation rules, Greece has made a number of aggregated judgements in relation to GES (e.g. for D5, D8) but it has not clearly defined aggregation rules.

ES, Spain distinguished two formal subdivisions within the Mediterranean Sea. In the Northeast Atlantic, two formal subdivisions are distinguished for the subregion Bay of Biscay and Iberian coast, and one for the subregion Macaronesia. In addition to those subdivisions, Spain mentions the use of distinct spatial scales for the analysis of descriptors, indicators, pressures and impacts. For the definition of those assessment areas the most relevant ecological scale was chosen, in most cases within the limits of the formal subdivisions. For some aspects, like ICES rectangles or migratory species, scales beyond the boundaries of the subdivision may be used. In some cases spatial scales were used taking into account the availability of data.

FI, Finland has used eight areas for reporting; those areas match with the Finnish waters within the HELCOM division of the Baltic Sea into sub-basins. Finland has prepared a separate report for the Sea of Åland. In addition, Finland has used the following assessment areas: Bay of Bothnia, Quark, Bothnian Sea, Northern Baltic Proper, Archipelago Sea, Gulf of Finland, Baltic Sea. These assessment areas are used in the initial assessment (but not systematically for all pressures) but they are not used for the definition of GES and the setting of targets, which are done for all areas altogether.

FR, France reports that at this stage, assessment areas are used at the scale of the subregion. No specific (smaller) assessment areas have been defined. Data on more limited areas or assessments of more limited areas will be used for evaluation at the scale of the marine subregion. Aggregation rules at the level of the descriptor will be specified, if necessary, following complementary studies aimed at updating the definition of GES by 2018.

IE, Ireland has used one assessment area for its marine waters, and has not distinguished formal subdivisions or smaller assessment areas.

HR, Croatia has reported six assessment areas, three in coastal waters and three in open waters.

IT, Italy reported a large number of assessment areas in all its marine subregions. However, Italy included a reference to the various reporting sheets as part of the coding of the assessment units. This means that in a number of cases different assessment units are reported (related to different reporting sheets), which in reality refer to the same geographic area. For example, assessment area “Northern Ionian Sea” was reported twice, once as assessment area IT-IMS-8A03-0002 and once as assessment area IT-IMS-8A04-0002, in both cases referring to an assessment area with exactly the same surface area. The total number of different assessment areas used by Italy is therefore lower than the number reported in Table 5.2.

LT, Lithuania defined four subdivisions by reusing WFD coastal water bodies and including the territorial waters and the EEZ. The assessment areas show some overlap. It uses these assessment areas not in a systematic way for its GES definitions, its initial assessment or its environmental targets, but the use depends on the descriptor.

LV, Latvia distinguished several assessment areas, consisting of a number of small coastal waters and overlapping larger areas. Latvia has defined a number of assessment areas, which differ depending on the topic assessed but which can be as many as seven. Latvia also explains that within HELCOM a distinction is already made in the assessments between the open waters in the Baltic Sea and the Gulf of Riga, because of the differences in physical and biological characteristics.

MT, Malta reported eleven assessment areas for physical and biological features, and eleven assessment areas for pressures.

NL, The Netherlands have defined one assessment area for their marine waters. At this stage, no specific assessment area has been defined. There is no indication on aggregation scales.

PL, Poland reported 8 assessment areas that are similar to the subdivisions defined by HELCOM

PT, Portugal has defined four subdivisions.
Within the subregion 'Bay of Biscay and Iberian coast', two subdivisions are distinguished. One is the continental subdivision, which includes all marine waters within the EEZ bordering the mainland. The second is the 'extended continental shelf', which includes the continental platform beyond 200 nautical miles. The Process of Extension of the Continental Shelf is currently ongoing within the framework of the United Nations.
Within the subregion Macaronesia two subdivisions are defined, the Acores subdivision which includes all marine waters next to the Archipelago of Acores, and the Madeira subdivision which includes all marine waters next to the Archipelago of Madeira. For the latter two subdivisions, the extended continental platform is not included.
For the continental division, Portugal has used various assessment areas depending on the descriptor, based on the geographical boundaries and the specific characteristics of the descriptor. For the extended shelf, Portugal chose five areas corresponding to the OSPAR MPAs.

RO, Romania has identified 3 assessment areas, two small areas covering transitional and coastal waters and a large area stretching from 1 nautical mile offshore to the 50m isobaths. It uses these areas for the initial assessment and in certain cases for the definition of GES and for the environmental targets as well. Romania's marine waters extend from the 1 nm line to the outer limit of the EEZ. Coastal waters and transitional waters were delineated according to the WFD.

SE, Sweden Sweden has defined a number of assessment areas in its legislation. For the Baltic Sea, there are nine assessment areas (Sea of Akona and S Øresund, Bornholm Sea and Hanöbukten, E Gotland Sea, W Gotland Sea, N Gotland Sea, Sea of Åland, the Southern part of the Gulf of Bothnia, N Kvarken, Gulf of Bothnia). For the Greater North Sea, three assessment areas are defined, Skagerrak, Kattegat and Øresund. In the electronic reporting, Sweden reported a large number of assessment areas to enable reporting at different geographical levels (e.g. coastal water types, offshore waters, marine basins, administrative areas). For certain descriptors, Sweden has defined different indicators or thresholds for specific assessment areas.

SI, Slovenia's initial assessment, characteristics of GES and associated targets and indicators have been developed for the Slovenian marine waters as a whole. In the electronic reporting Slovenia reported 4 assessment areas.

UK, The **United Kingdom** reported that the boundary between the Greater North Sea and the Celtic Seas subregions has been established on the basis of, oceanographic and biogeographic features. The UK informally distinguishes eight biogeographically defined assessment areas within the NE Atlantic. These biogeographical assessment areas are compatible with the boundaries of the marine subregions defined in UK marine waters and are based on earlier studies, using physical and biological features such as tidal fronts and seabed flora and fauna.
The UK characteristics of GES and associated targets and indicators have been developed for the UK marine waters as a whole. Where there are significant biogeographical differences between the Greater North Sea and the Celtic Seas subregions these have been taken into account. The Initial Assessment makes reference to the status of UK waters at the scale of the subregions and/or the informal assessment areas.

5.1.1 Application of geographic scales for the descriptors.

The application of assessment areas is not necessarily the same for all GES descriptors (Table 5.3). The member states that applied more than one assessment area differentiated the use of assessment areas in some. This was mostly done for Descriptors D1, D5, D6 and D8. Several MS indicate that (informally) various assessment scales are used depending on the Descriptor or the information available

Table 5.3 Overview of the number of assessment areas per descriptor, for those MS that applied more than one assessment area. Empty cells: no specific information available (source: Atkins MSFD database 17 June 2013)

(Sub)region	MS	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11
Baltic Sea	**FI**											
	LT	1	2			2	1					
	LV	7	1	2	1	7	6			1		
	SE	38	1	1	29	38	37	1	38	10	1	1
Black Sea	**RO**	2	1	1		3			1			
Adriatic Sea	**EL**			1		1		1	1	1	1	
	IT											
	SI	2	1	2	3	2	1	1	2	1	1	1
Aegean-Levantine Sea	**EL**			2		1		1	1	1	1	
Ionian Sea and Central Mediterranean Sea	**EL**			1		1		1	1	1	1	
	IT											
Western Mediterranean Sea	**ES**	4	2	2	2	11	2	3	4	2	2	2
	IT											
Bay of Biscay and Iberian Coast	**ES**	3	2	2	2	13	2	4	4	2	2	2
Celtic Sea	**UK**											
Greater North Sea	**SE**	11	1	1	1	11	11	1	3	4	1	1
	UK											
Macaronesia	**ES**	1	1	1	1	1	1	1	1	1	1	1

5.1.2 Comparison with RSC approaches

The Regional Sea Conventions have all developed approaches to spatial assessment scales, although there are significant differences in the level of development between RSCs (see paragraph 2.4). Table 5.4 summarizes the RSC approaches and the approaches that MS within the regional seas have taken in the implementation of the MSFD.

This comparison shows that in many cases already existing approaches of RSCs have not been used. However, at the level of specific assessments (for example, in the case of Descriptors 3, 5, 8) some MS have used assessment methods developed by RSCs, which implies that in those cases also similar spatial scales were used.

Table 5.4. Approaches to spatial scaling by Regional Sea Conventions and by MS in the implementation of the MSFD.

Regional sea	**RSC**	**MSFD implementation**
Baltic Sea	HELCOM developed spatial scales at different hierarchical levels, that are nested within each other	Some MS have reported assessment areas linked to the HELCOM system
Black Sea	Territorial waters and open sea are distinguished	MS use different assessment scales, including a distinction between territorial waters and open sea
Mediterranean Sea	UNEP/MAP has defined sub-basins (similar to subregions; assessment areas at a smaller scale have not been defined yet	MS use different assessment scales within the subregions
NE Atlantic	OSPAR uses different assessment scales on a case-by-case basis.	MS use different assessment scales within the subregions

6 Discussion & conclusions

This report presents an analysis of existing approaches in environmental status assessment of European seas, with respect to the geographic scales of assessments, the scaling up of assessments to larger areas, and the aggregation from indicator level up to overall assessments of environmental status. The analysis presented in this report is based on information from the implementation of the MSFD by the MS, and on information from peer-reviewed literature, research projects and work by the Regional Sea Conventions. A large part of this information relates to environmental assessments carried out in the framework of European directives (in particular the WFD) and assessments by the RSCs.

The analysis of the MS approaches in the implementation of the MSFD is mainly based on the information available through the electronic reporting by MS. Due to time constraints, the paper reporting could not be taken into consideration.

6.1 Spatial scales

The results of our analysis show that there seems to be a common understanding of the general principles for the definition of assessment areas, which were discussed in paragraph 3.2. The scale of assessment areas should be in line with ecological characteristics, should ensure that the assessments are informative for management, and should preferably be based on a risk-based approach that helps to prioritize areas where pressures and impacts are likely to be important. These are basic starting points for the definition of scales that can be found in numerous documents, such as the Task Group reports and the Common Understanding document (Claussen *et al.* 2011).

The RSCs have already developed approaches to define assessment areas for specific purposes (e.g. some biodiversity aspects, fisheries, eutrophication, contaminants). There are differences between regional seas in the level of development. In the Baltic Sea, the most elaborate system with a nested design of assessment areas at different hierarchical level has been developed by HELCOM.

In the analysis of MS approaches in the implementation of the MSFD, large differences in the approach to spatial scales are apparent. Eight MS use one assessment area for the assessment of the environmental state of their marine waters. Eleven MS have used more than one assessment area, for three MS information is not available yet. The scales of the assessment areas range from <100 km^2 to >100,000 km^2. There seems to be no relation between environmental conditions and the scale of the assessment areas, as there are large differences between MS within a (sub)region. In the Baltic Sea, Sweden seems to have taken an approach that markedly differs from the approaches of other MS within this region. In the Mediterranean Sea the approach of Italy is also very different from the approach of Greece, France or Spain. A point of attention is the fact that some MS have made a distinction between formal subdivisions that were defined and informal assessment areas that are used for specific analyses. In the information from the electronic reporting by the MS the distinction between those categories is not always clear, and this may partly explain the large differences in assessment scales between MS.

Nevertheless, it is clear that there are large differences between MS. While some MS have used WFD coastal water bodies and marine waters further offshore as assessment areas, other MS have defined other, and sometimes larger, assessment scales. With a few exceptions, MS do not explicitly mention the use of biogeographical or hydrodynamical criteria or a risk-based analysis when defining assessment areas. The information from the electronic reporting indicates that some MS use more assessment areas for some Descriptors. This is in particular the case for D1, D5 and D6. Several MS indicate that (informally) various assessment scales are used depending on the Descriptor or the information available. This includes the use of spatial scales based on existing approaches for other assessments, such as other European Directives (WFD, Bird and Habitat Directive) or RSC assessments. Decisions on spatial scales seem to have been made on a case-by-case basis.

The information used in our analysis, does not indicate that MS have attempted to scale up from smaller assessments areas to an assessment for their entire marine waters within a subregion (where applicable), nor that a scaling up to a subregion assessment of GES has been done. An analysis by Borja *et al.* (2013) shows that there are many data based on pressures within small areas while only few data cover entire marine (sub)regions, which means that scaling up to derive large scale assessment will depend on combining data with different levels of detail.

6.2 Aggregation methods

In our analysis, we give an overview of existing approaches for the aggregation of assessments. Most of the methodologies can, in principle, be applied at various hierarchical levels, from the level of metrics/indicators up to the level of overall GES. So far, it seems that MS have mainly used existing assessment methods, and development of aggregation methods seems to have been limited.

No specific rules have been proposed for the MSFD, in contrast to the WFD where the OOAO rule is used. The OOAO rule has obvious disadvantages (see paragraph 5.4 and Annex III, for the example of the Basque coast). Various alternative methods exist. Borja *et al.* (2013), expanding on the Common understanding document (Claussen *et al.* 2011) proposes some common principles for the development of aggregation rules:

- Integration across levels of different complexity should accommodate different alternatives (i.e. integration of criteria or indicators could differ from integration of descriptors)
- Integration of state descriptors could differ from integration of pressure descriptors
- Weighting of descriptors could differ (e.g. different weights for state and pressure descriptors)

The management group report (Cardoso *et al.* 2010) points out that what is needed for combining the information available on the diverse attributes of e.g. seafloor integrity is not some fully specified and well-structured analytical method for assessing GES, but a fully specified and well-structured process for conducting assessments of GES. The key design features of reliable, consistent assessments include:

a) Specified objectives and scope of individual assessments;
b) An effective relationship between science and policy;
c) Modalities for stakeholder participation;
d) Nomination and selection of experts;
e) Data and information: sourcing, quality assurance and the availability and accessibility of underlying data and information;
f) Treatment of lack of consensus among experts;
g) Treatment of uncertainty;
h) Peer review;
i) Effective communication;
j) Capacity building and networking;
k) Post-assessment evaluation.

As stated by Cardoso *et al.* (2010), designing a sound assessment process, incorporating those design features in the process and products produced, will provide the only realistic avenue for having regular evaluations of GES at (sub)region scales. The periodic assessments would not have a single specified set of steps that would be the required approach. Rather the process could adapt practice from assessment to assessment with regard to indicators selected, weightings and benchmarks applied, and approaches to integrating local scale evaluations into conclusions at (sub)region scale based on the developing experience and knowledge.

6.3 Next steps

This analytical report provides a basis for the further steps in the project: a consultation on specific issues necessary to improve coherence of future assessments, and the development of a draft guidance on coherent scales and aggregation rules. As part of the process towards developing a draft guidance, a workshop of WG GES was held in October 2013. The guidance document can build on work that has already been done in the preparation for the MSFD implementation, and experiences within RSC's (for example, the OSPAR advise documents (OSPAR 2012f; a; e; d; c; b) and results from recent research projects. The discussions at the WG GES workshop prioritized issues that need guidance.

7 References

Aubry, A. and 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.

Borja, A., F. Franco, V. Valencia, J. Bald, I. Muxika, M.J. Belzunce and O. Solaun (2004a). Implementation of the European water framework directive from the Basque country (northern Spain): a methodological approach. *Marine Pollution Bulletin* 48: 209-218.

Borja, A., V. Valencia, J. Franco, I. Muxika, J. Bald, M.J. Belzunce and O. Solaun (2004b). 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 and 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., J. Bald, J. Franco, J. Larreta, I. Muxika, M. Revilla, J.G. Rodriguez, O. Solaun, A. Uriarte and V. Valencia (2009a). Using multiple ecosystem components, in assessing ecological status in Spanish (Basque Country) Atlantic marine waters. *Marine Pollution Bulletin* 59: 54-64.

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

Borja, A., M. Elliott, J. Carstensen, A.S. Heiskanen and 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. and J.G. Rodriguez (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., E. Barbone, A. Basset, G. Borgersen, M. Brkljacic, M. Elliott, J.M. Garmendia, J.C. Marques, K. Mazik, I. Muxika, J. Magalhaes Neto, K. Norling, J.G. Rodriguez, I. Rosati, B. Rygg, H. Teixeira and 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 and 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. Neto, H. Teixera, L. Uusitalo, M.C. Uyarra and N. Zampoukas (2013). Good Environmental Status of marine waters: What is it and how do we know when we have attained it? *Marine Pollution Bulletin* 76: 16-27.

Cardoso, A.C., S. Cochrane, H. Doerner, 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. Tasker and W. van de Bund (2010). Scientific Support to the European Commission on the Marine Strategy Framework Directive – Management Group Report. *EUR – Scientific and Technical Research series,* Luxembourg, JRC/ICES, EUR 24336 EN, 57 pp.

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

Claussen, U., G. Casazza, D. Connor, L. de Vrees, J.M. Leppänen, J. Percelay, M. Kapari, O. Mihail, G. Ejdung, J. Rendall and R. Emmerson (2011). Common Understanding of (Initial) Assessment, Determination of Good Environmental Status (GES) and Establishment of Environmental Targets (Art. 8, 9 & 10 640 MSFD). Brussels, WG GES Drafting Group, Version 2, 23 September 2011, 58 pp.

Cochrane, S.K.J., D.W. Connor, P. Nilsson, I. Mitchell, J. Reker, J. Franco, V. Valavanis, S. Moncheva, J. Ekebom, K. Nygaard, R. Serrão Santos, I. Naberhaus, T. Packeise, W. van de Bund and A.C. Cardoso (2010). Marine Strategy Framework Directive – Task Group 1 Report Biological diversity. *EUR – Scientific and Technical Research series,* Luxembourg, JRC/ICES, EUR 24337 EN, 111 pp.

EC (2005). Overall Approach to the Classification of Ecological Status and Ecological Potential. *WFD CIS Guidance Document,* Luxembourg, 13, 48 pp.

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 and U. Claussen (2010). Marine Strategy Framework Directive - Task Group 5 Report Eutrophication. *EUR - Scientific and Technical Research series,* Luxembourg, JRC/ICES, EUR 24338 EN, 49 pp.

Galgani, F., D. Fleet, J. Van Franeker, S. Katsanevakis, T. Maes, J. Mouat, L. Oosterbaan, I. Poitou, G. Hanke, R. Thompson, E. Amato, A. Birkun and C. Janssen (2010). Marine Strategy Framework Directive - Task Group 10 Report Marine litter. *EUR - Scientific and Technical Research series,* Luxembourg, JRC/ICES, EUR 24340 EN, 48 pp.

HELCOM (2009a). Biodiversity in the Baltic Sea – An integrated thematic assessment on biodiversity and nature conservation in the Baltic Sea. *Baltic Sea Environment Proceedings,* Helsinki, No. 116B, 192 pp.

HELCOM (2009b). Eutrophication in the Baltic Sea – An integrated thematic assessment of the effects of nutrient enrichment and eutrophication in the Baltic Sea region. *Baltic Sea Environment Proceedings,* Helsinki, No. 115B., 152 pp.

HELCOM (2010a). Hazardous substances in the Baltic Sea – An integrated thematic assessment of hazardous substances in the Baltic Sea. *Baltic Sea Environment Proceedings,* Helsinki, 120B, pp.

HELCOM (2010b). Ecosystem Health of the Baltic Sea 2003–2007: HELCOM Initial Holistic Assessment. . *Baltic Sea Environment Proceedings,* Helsinki, HELCOM, No. 122., 63 pp.

HELCOM (2013). Draft HELCOM Monitoring and assessment strategy. HELCOM, HELCOM MONAS 18/2013, Document 6/6, pp.

Law, R., G. Hanke, M. Angelidis, J. Batty, A. Bignert, J. Dachs, I. Davies, Y. Denga, A. Duffek, B. Herut, K. Hylland, P. Lepom, P. Leonards, J. Mehtonen, H. Piha, P. Roose, J. Tronczynski, V. Velikova and D. Vethaak (2010). Marine Strategy Framework Directive – Task Group 8 Report Contaminants and pollution effects. *EUR – Scientific and Technical Research series,* Luxembourg, JRC / ICES, EUR 24335 EN, 161 pp.

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

Olenin, S., F. Alemany, A.C. Cardoso, S. Gollasch, P. Goulletquer, M. Lehtiniemi, T. McCollin, D. Minchin, L. Miossec, A. Occhipinti Ambrogi, H. Ojaveer, K. Rose Jensen, M. Stankiewicz, I. Wallentinus and B. Aleksandrov (2010). Marine Strategy Framework Directive – Task Group 2 Report Non-indigenous species. *EUR – Scientific and Technical Research series,* Luxembourg, JRC/ICES, EUR 24342 EN, 44 pp.

Orfanidis, S., P. Panayotidis and N. Stamatis (2003). An insight to the ecological evaluation index (EEI). *Ecological Indicators* 3: 27-33.

OSPAR (2003). OSPAR integrated report 2003 on the eutrophication status of the OSPAR maritime area based upon the first application of the Comprehensive Procedure. 59 pp.

OSPAR (2005a). North Sea Pilot Project on Ecological Quality Objectives. Background Document on the Ecological Quality Objective on imposex in dog whelks *Nucella lapillus*. London, OSPAR, Publication Number: 2005/247, 11 pp.

OSPAR (2005b). North Sea Pilot Project on Ecological Quality Objectives. Background Document on the Ecological Quality Objective for Spawning Stock Biomass of Commercial Species in the North Sea. London, OSPAR, Publication Number: 2005/242, 17 pp.

OSPAR (2005c). North Sea Pilot Project on Ecological Quality Objectives. Background Document on the Ecological Quality Objective on Oiled Guillemots. London, OSPAR, Publication Number: 2005/252, 33 pp.

OSPAR (2005d). North Sea Pilot Project on Ecological Quality Objectives. Background Document on the Ecological Quality Objective for Seal Population Trends in the North Sea. London, OSPAR, Publication Number: 2005/245, 13 pp.

OSPAR (2006). Report on North Sea Pilot Project on Ecological Quality Objectives. London, OSPAR, Publication Number: 2006/239, 126 pp.

OSPAR (2007). EcoQO Handbook. Handbook for the application of Ecological Quality Objectives in the North Sea. London, OSPAR, Publication Number: 2007/307, 39 pp.

OSPAR (2008). Eutrophication Status of the OSPAR Maritime Area. Second OSPAR Integrated Report. London, OSPAR, Publication Number 372/2008, 108 pp.

OSPAR (2009a). Evaluation of the OSPAR system of Ecological Quality Objectives for the North Sea. London, OSPAR, Publication Number: 406/2009, 102 pp.

OSPAR (2009b). Towards the Radioactive Substances Strategy objective. Third periodic evaluation. London, OSPAR, Publication number 455/2009, 207 pp.

OSPAR (2009c). CEMP assessment report: 2008/2009. Assessment of trends and concentrations of selected hazardous substances in sediments and biota. *Monitoring and Assessment Series,* London, OSPAR, Publication Number: 390/2009, 79 pp.

OSPAR (2009d). Report of the Utrecht Workshop - Regional assessment. Publication Number: 468/2009, 103 pp.

OSPAR (2010). Quality Status Report 2010. . London, OSPAR Commission, 176 pp.

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

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

OSPAR (2012b). MSFD Advice Manual and Background document on Good environmental status - Descriptor 5: Eutrophication. London, Publication Number: 582/2012, 26 pp.

OSPAR (2012c). MSFD Advice Manual and Background document on Good environmental status - Descriptor 11: Underwater noise. London, OSPAR, Publication Number: 586/2012, 27 pp.

OSPAR (2012d). MSFD Advice document on Good Environmental Status: Descriptor 10: Marine Litter. OSPAR, Publication Number: 585/2012, 6 pp.

OSPAR (2012e). MSFD Advice document on Good Environmental Status: Descriptor 7: Hydrographical conditions. London, OSPAR, Publication Number: 583/2012, 9 pp.

OSPAR (2012f). MSFD Advice Document on Good Environmental Status - Descriptor 8: Contaminants. London, OSPAR, Publication Number: 584/2012, 21 pp.

Piet, G.J., A.J. Albella, E. Aro, H. Farrugio, J. Lleonart, C. Lordan, B. Mesnil, G. Petrakis, C. Pusch, G. Radu and H.-J. Rätz (2010). Marine Strategy Framework Directive – Task Group 3 Report Commercially exploited fish and shellfish. *EUR – Scientific and Technical Research series,* Luxembourg, JRC/ICES, EUR 24316 EN, 82 pp.

Primpas, I., G. Tsirtsis, M. Karydis and G.D. Kokkoris (2010). Principal component analysis: Development of a multivariate index for assessing eutrophication according to the European water framework directive. *Ecological Indicators* 10: 178-183.

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

Pyhälä , M., V. Fleming-Lehtinen, E. Lysiak-Pastuszak, M. Carstens, J.-M. Leppänen, C. Murray and J. Andersen (2013). Eutrophication status of the Baltic Sea 2007-2011. A concise thematic assessment. Helsinki, HELCOM, HELCOM Thematic assessment, 25 pp.

Rice, J., C. Arvanitidis, A. Borja, C. Frid , J. Hiddink , J. Krause, P. Lorance, S.Á. Ragnarsson, M. Sköld and B. Trabucco (2010). Marine Strategy Framework Directive – Task Group 6

Report Seafloor integrity. *EUR – Scientific and Technical Research series,* Luxembourg, JRC/ICES, EUR 24334 EN, 73 pp.

Rogers, S., M. Casini, P. Cury, M. Heath, X. Irigoien, H. Kuosa, M. Scheidat, H. Skov, K. Stergiou, V. Trenkel, J. Wikner and O. Yunev (2010). Marine Strategy Framework Directive – Task Group 4 Report Food webs. *EUR – Scientific and Technical Research series,* Luxembourg, JRC/ICES, EUR 24343 EN, 55 pp.

Rosenberg, R., M. Blomqvist, H. Nilsson and A. Dimming (2004). Marine quality assessment by use of benthic species-abundance distributions: a proposed new protocol within the European Union Water Framework Directive. *Marine Pollution Bulletin* 49: 728-739.

Shin, Y.-J., A. Bundy, L.J. Shannon, J.L. Blanchard, R. Chuenpagdee, M. Coll, B. Knight, C. Lynam, G. Piet and 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., P. Panayotidis and E. Papathanassiou (2005). A synthesis of the biological quality elements for the implementation of the European Water Framework Directive in the Mediterranean ecoregion: The case of Saronikos Gulf. *Ecological Indicators* 5: 253-266.

Simboura, N., A. Zenetos, M.A. Pancucci-Papadopoulou, S. Reizopoulou and 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.

Swartenbroux, F., B. Albajedo, M. Angelidis, M. Aulne, V. Bartkevics, V. Besada, A. Bignert, A. Bitterhof, A. Hallikainen, R. Hoogenboom, L. Jorhem, M. Jud, R. Law, D. Licht Cederberg, E. McGovern, R. Miniero, R. Schneider, V. Velikova, F. Verstraete, L. Vinas and S. Vlad (2010). Marine Strategy Framework Directive – Task Group 9 Report Contaminants in fish and other seafood. *EUR – Scientific and Technical Research series,* Luxembourg, JRC/ICES, EUR 24339 EN, 36 pp.

Tasker, M.L., M. Amundin, M. André, A. Hawkins, W. Lang, T. Merck, A. Scholik-Schlomer, J. Teilmann, F. Thomsen, S. Werner and M. Zakharia (2010). Marine Strategy Framework Directive - Task Group 11 Underwater noise and other forms of energy. *EUR - Scientific and Technical Research series,* Luxembourg, JRC/ICES, EUR 24341 EN, 55 pp.

Tett, P., R. Gowen, D. Mills, T. Fernandes, L. Gilpin, M. Huxham, K. Kennington, P. Read, M. Service, M. Wilkinson and S. Malcolm (2007). Defining and detecting undesirable disturbance in the context of marine eutrophication. *Marine Pollution Bulletin* 55: 282-297.

Tueros, I., A. Borja, J. Larreta, J.G. Rodriguez and V.M. Valencia, E. (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.

UNEP/MAP (2008). Report of the Second Meeting of Government-designated Experts on the Application of the Ecosystem Approach. Athens, Greece, United Nations Environment Programme. Mediterranean Action Plan, pp.

UNEP/MAP (2010). Subregional Assessment Reports. Athens, Greece, United Nations Environment Programme. Mediterranean Action Plan, pp.

UNEP/MAP (2011). Hazardous substances in the Mediterranean: a spatial and temporal assessment. Athens, Greece, United Nations Environment Programme, Mediterranean Action Plan, pp.

UNEP/MAP (2012a). Note on the Ecosystems approach process in 2012-2013. Athens, Greece, United Nations Environment Programme, Mediterranean Action Plan, pp.

UNEP/MAP (2012b). Initial integrated assessment of the Mediterranean Sea: Fulfilling step 3 of the ecosystem approach process. Athens, Greece, United Nations Environment Programme, Mediterranean Action Plan, pp.

Vollenweider, R.A., F. Giovanardi, Montanari and A. Rinaldi (1998). Characterization of the trophic conditions of marine coastal waters with special reference to the NW Adriatic Sea: Proposal for a trophic scale, turbidity and generalized water quality Index. *Environmetrics* 9: 329-357.

8 Annexes

8.1 ANNEX I

Geographic scaling

Descriptor	Source	Summary of advice on geographic scaling
D1 Biological diversity	TG 1 report	A suitable set of ecological assessment areas should be defined, which can adequately reflect both the ecological scales exhibited by the biodiversity components in each region/subregion and link to areas which are effective for management measures. The outcomes of a status assessment are highly dependent on the geographical scale at which they are undertaken. The assessment scale can be set ecologically or by policy. For ecologically relevant scales, ideally the assessment should cover the entire range of the species or be related to discrete populations (e.g. for large/mobile species). For habitats/communities it is most appropriate to assess within biogeographic zones, as functionally similar habitats can have global distributions In practice policies are often applied at specific geographic scales relating to the scope of the policy or national jurisdictions and thus can lead to different classifications of status for the same species/habitat. In order to facilitate monitoring and management, and to reflect biogeographic and genetic variation, the assessment scale should reflect the variation in biological diversity that operates at a range of spatial scales related to distinct populations or subspecies and, for communities, biogeographic regions.
D2 Non-indigenous species	TG 2 report	The assessment of IAS impacts generally should begin at the local scale, such as “hot-spots” and “stepping stone areas” for alien species introductions or in areas of special interest. Depending on the taxonomic/functional group an IAS belongs to, the assessment can involve areas from confined benthic habitats to the entire water column. Local scale assessments can be further integrated into the next spatial level evaluations at a sub-regional (e.g. Gulf of Finland in the Baltic or Adriatic Sea in the Mediterranean) or a regional sea level. NIS will have impact on the environment at very different spatial scales. Spatial extent and rate of spread will depend on biological traits of NIS and environmental conditions: e.g., species with planktonic phases will have a greater dispersal potential. Determining the size of an assessment area will vary

		according to whether it is a single species under consideration or whether a general study of a region is to be performed. Within defined localities the impact of a species can be more easily assessed, while at larger scale (e.g. sub-regional) the effect will depend on the number of localities impacted.
D3 Commercially exploited fish and shellfish	TG 3 report	For this descriptor the relevance of spatial scale is only apparent for assessed species in the selection of appropriate stocks and for the non-assessed species by the choice of the most appropriate survey for each (sub-)region. For a particular region only those stocks that mostly occur in that region will be selected. The temporal scale is determined by the fact that usually both the analytical assessments as well as the surveys are conducted on an annual basis.
D4 Food webs	TG 4 report	At small spatial scales, such as parts of a MSFD Sub-Region, immigration and emigration by advection and migrations become important components of change. For large, long-lived taxa, spatial scales which integrate over migration ranges may be appropriate, but these scales may span fundamentally different habitats and communities for lower trophic levels, for example plankton or benthos, to the point that a synthesis at this scale becomes questionable. Ultimately, it seems likely that the appropriate spatial scale at which to assess food webs will be set by the purpose for which the assessment is required rather than any ecological considerations. Other practical considerations, such as the availability and spatial extent of monitoring data for key taxa, are also likely to influence the scale at which assessments are made.
D5 Eutrophication	TG 5 report	Breakdown into subunits is expected. These smallest divisions should be defined according to oceanographic characteristics aiming for spatially homogeneous areas. The risk of eutrophication is linked to the capacity of the marine environment to confine growing algae in the well-lighted surface layer. The geographical extent of potentially eutrophic waters along European coasts may vary widely, depending on: (i) the extent of shallow areas, i.e. with depth (ii) the extent of stratified river plumes. The risk increases with increasing water residence time; (iii) extended water residence times in enclosed seas leading to blooms triggered to a large degree by internal and external nutrient pools; (iv) upwelling phenomena leading to autochthonous nutrient supply and high nutrient concentrations from deepwater nutrient pools, which can be of natural or human origin. As the “ecological status” has to be monitored on the whole shelf, huge areas where a MSFD eutrophication

		assessment must clearly delineate the areas potentially subject to detrimental effects. Furthermore, Good Environmental Status (GES) has to be set for these areas based on eutrophication parameters that will be part of the monitoring programmes. Such areal delineation should be based on oceanographic characteristics, such as the Physically Sensitive Area (PSA), the EU TRISK indices developed by the JRC, and the subdivision used by HELCOM and OSPAR.
D6 Sea-floor integrity		Scale for assessing environmental status of the sea floor is particularly challenging for three reasons. First, the wide range of human activities causing pressures that may degrade the status of the sea floor operate at different but always patchy spatial scales. For all pressures resulting from land based activities, there are two intrinsic gradients of their potential pressures. There is an inherent initial gradient from coastal areas to offshore regions. The patchiness of the human activities causing the pressures also means that the scales of initial impacts of those activities are usually also local. A third consideration is that there are many differences between coastal and deeper-water benthic communities.The methodology for assessing environmental status at regional and subregional scales takes a risk-based approach, considering the threats posed by the human activities occurring in the region. It is considered feasible to map the spatial distribution of most human activities in the sea, particularly the ones most likely to cause the largest impacts on the sea floor. Such maps may not be possible on very fine spatial scales, but are likely possible on the scales characteristic of EUNIS Level 4 (or finer, for some sediment types) classifications of the benthos. It is also feasible to tabulate the major pressures caused by various human activities and the vulnerability of various properties of the sea floor to the various pressures. Such cross-tabulations have been developed already for many activities, pressures, and ecosystem features, in fact.

Aggregation rules

Descriptor	Source	Summary of advice on aggregation rules
	Management group report	Within Descriptor integration relates to the methods that might be required within a Descriptor to take account of multiple indicators, and a situation where not all indicators and/or attributes reach their desired levels or targets. For each Descriptor the task groups have outlined in their reports the best approach to be taken. Two approaches are recommended: (i) integrative assessments combining indicators

		and/or attributes appropriate to local conditions and; (ii) assessment by worst case. In this context ‚worst case‘ does not mean the full area of concern is assumed to be at the status of the worst part of the area. Rather, it means that the evaluation of GES will be set at the environmental status of the indicator and/or attribute assessed at the poorest state for the area of concern.
D1 Biological diversity	TG 1 report	Because the different elements of biological diversity may not respond to pressures in a similar manner, or at similar rates, the results of assessments for individual biodiversity components cannot be integrated into a single assessment for Descriptor 1.
D2 Non-indigenous species	TG 2 report	Efforts should be made to record all NIS known in the assessment area; however attention should be paid primarily to assessments of IAS impacts. Methods for aggregating indicators for GES assessments need to take into account the known IAS effects in other world regions or in neighbouring areas. One of the approaches may be estimation of the magnitude of bioinvasion impacts or “Biopollution level” (BPL) index which takes into account the abundance and distribution range of NIS in relation to native biota in the invaded area and aggregates data on the magnitude of the impacts these species have on: native communities, habitats and ecosystem.
D3 Commercially exploited fish and shellfish	TG 3 report	Based on the most robust methodology (comparison of indicators to reference levels and based on stock assessments) but which cover only a limited proportion of the stocks: A stock can only achieve GES if all three criteria for the attributes are fulfilled. Because $SSB>SSB_{MSY}$ cannot be achieved for all stocks simultaneously, the other two criteria should apply to a specific proportion of the stock. Based on the less robust methodology (indicator trends based on surveys and catch statistics) but which covers a much larger proportion of the stocks: A stock can only achieve GES if all three criteria for the attributes are fulfilled. All three criteria should apply to a specific proportion of the stock.
D4 Food webs	TG 4 report	Further work needs to be undertaken to agree how a number of assessments can be combined to achieve an overall assessment of GES for the descriptor. Several methods have been proposed to combine assessments, ranging from those which requires all assessments to be acceptable before agreeing a final status assessment („one out all out‟), to those which provide weightings to give priority to some ecosystem components or attributes over others. Each individual assessment will also be subject to uncertainty in determining the metric and the reference point value. The „fuzzy set‟ approach has been suggested by Silvert (1997;

		2000) as a way of including uncertainty when combining a range of specific ecological assessments. The method relies on scoring assessments based on a combination of their achievement of assessment criteria and certainty of knowledge. However, there is currently no agreed method for aggregating the assessments of Food Web status across attributes and within Regional Seas.
D5 Eutrophication	TG 5 report	The question of aggregation was discussed at two levels: (i) the integration of different indicators into attributes for the descriptor; and (ii) A range of tools was reviewed. No specific method (i.e. tool) is recommended to be used for GES, but those used must be robust, integrated, sufficiently sensitive, comparable, and with recognized scientific merit. Contrary to the WFD, which defines a “one out all out” approach in order to classify a water body, in the MSFD, GES may be envisaged as an integration (e.g. sum, weighted average, or other approaches) of all/most criteria.
D6 Seafloor integrity	TG 6 report	However across attributes and on even moderate scales expert assessments rather than algorithmic formulae will be needed for evaluation of GES of Seafloor Integrity.

8.2 ANNEX II: EcoQO's in OSPAR

Widespread populations and widespread pressures

Widespread populations, such as harbour seals (EcoQO 2.1A), should provide information on wide-spread pressures. Harbour seal

populations are monitored within 15 sub-units. Nine parts of the North Sea are distinguished for the EcoQO on Grey seal pup. This scaling of assessment areas is based on the distribution of the population (OSPAR 2005d). This EcoQO is quite generic, and data are reported in different sub-units. This can be explained by the fact that these animals are not limited by national boundaries and that their location differs per species per country. Even though some countries have multiple populations of harbour seals, these populations are reported on separately. There does not seem to be an integration step within the EcoQO that tries to give a comprehensive overview of marine mammals per (sub-)region. For the EcoQO on oiled guillemots, sub-regions should be chosen for each country to sample the entire coastline appropriately, taking local conditions into account, including the amount of input of oil. The selection of sub-regions will vary per country since local conditions will vary. The OSPAR background document on this EcoQO (OSPAR 2005c) proposes 15 sub-regions within the OSPAR region, not necessarily based on regional or national boundaries, but combining the different needs for data collection and

analysis and building on historical subdivisions. The EcoQO for plastic particles in stomachs of seabirds is also assessed in 15 sub-regions within the OSPAR region (OSPAR 2007).

Localized pressures with wide-spread populations

Fishing is measured through wide-spread populations of fish and local populations of fish. For the Commercial fish stock EcoQO, ICES uses sub-areas and divisions of sub-areas which form the basis for catch-statistics and population monitoring, leading to some 35 areas in total. All areas have set reference points for population size, depending on the fish species (Backgrounddocument EcoQO Commercial fish stocks, OSPAR 2005b).

Localized pressures with localized populations

Localized pressures can also be measured through local populations, such as with the EcoQO for TBT. This substance is monitored using imposex in gastropods. Advice on monitoring for this EcoQO was provided by ICES and is to focus on areas in which high TBT concentrations would be most likely, such as harbours (OSPAR 2005a). The localized pressure of eutrophication is measured in area-specific indicator species, which combined make up the EcoQO of an entire area (OSPAR 2009a). Areas are scored according to the OOAO principle: if one or more of its respective assessment parameters shows an increasing trend, elevated level, shift or change, the whole category is scored as increase. An area is classified as a problem area if it shows an increase in one or more of the categories. An area can be classified as a non-problem area if there are no increases in either of the categories or if the degree of nutrient enrichment does not pose a threat to the area itself, but may contribute to eutrophication problems in other areas.

OSPAR pilot approach

OSPAR piloted a new approach for the assessment of the status of ecosystems (OSPAR 2009d). Overlap in space and time of pressures and species/habitats was mapped, and the degree of impact was estimated. One of the lessons drawn from this exercise was, that future assessments should have a finer resolution, regarding both geographical scale and the level of aggregation of the ecosystem components. It was concluded that there is a trade-off between simple, aggregated 'policy' statements and scientific credibility. Assessments at a very fine scale (for example individual species and habitat types) may be scientifically more desirable but are resource intensive. It was also recognized that such a level of detail would subsequently require aggregation of the results to make broader judgements about GES, and such aggregations can bring their own difficulties (OSPAR 2009d).

8.3 ANNEX III: Examples

Examples of application of the One-Out All-Out approach

1. Caroni et al.(2013) performing data analysis with a lake monitoring dataset provided by SLU (Swedish University of Agricultural Sciences), Sweden, consisting of up to four BQEs: phytoplankton, macroinvertebrates, macrophytes and fish, summarize that in cases where BQEs include metrics sensitive to different pressures (multiple pressure BQEs), or are complementary and when the level of uncertainty in the metrics used in the assessment was not so high, an OOAO approach is recommended both within and between BQE.

2. Ojaveer and Eero 2011) analysing indicators related to biodiversity, eutrophication and hazardous substances from the Baltic Sea showed in their study, that application of the widely used ‘‘one out – all out’’ principle (similar to fuzzy AND rule) could easily result in a fully negative overall evaluation for all objectives. The assessment based on this methodology is certainly very conservative from the management perspective and probably ensures a full implementation of precautionary principles. However, a drawback of this approach is that a few strongly negative indicator values could shadow the potentially generally positive state of a given ecological objective. This would make any progress towards improving the environmental status invisible, as long as at least one indicator is showing poor performance.

3. The HELCOM Eutrophication Assessment Tool (**HEAT**) (HELCOM 2009b) is an example of application of the OOAO rule for aggregating the different elements assessment results (Figure 0.2). HEAT is based on existing indicators, which for this purpose have been grouped as follows: (1) physical-chemical features (PC), (2) phytoplankton (PP), (3) submerged aquatic vegetation (SAV), and (4) benthic invertebrate communities (BIC). Groups 1and 2 (PC and PP) are considered ‘primary signals’of eutrophication, while groups 3 and 4 (SAV and BIC) are considered ‘secondary signals’. Within the four mentioned groups, HEAT allows **weighting** between indicators. Hence, indicators thought to be very good can be given a higher weight than an indicator with a low quality and vice versa. This assessment represents a progression from a single-indicator based assessment of eutrophication status toward an integrated indicator-based assessment of eutrophication status. It uses the same indicators as the single-indicator approach, but applies a HELCOM Eutrophication Assessment Tool (HEAT) for an overall assessment and classification of the eutrophication status. HEAT is a multimetric indicator based tool and makes use of synoptic information in regard to reference conditions, acceptable deviation from reference conditions, and actual environmental status. HEAT also makes use of the ‘One out – All out principle’ *sensu* the Water Framework Directive, which means that the overall classification of an assessed area is based on the most sensitive quality element. In addition, HEAT produces a provisional ‘accuracy assessment’ of the final classification results in order to assess the reliability of the final classification.

4. In the application of the WFD in Greece and specifically in the Saronikos gulf or Athens gulf area impacted by treated sewage outfall the classification of the global ecological status was determined by the element being at the worst class, which is benthic macroinvertebrates or zoobenthos following the principle of one-out-all-out (Simboura *et al.* 2005Simboura *et al.* 2005) as illustrated in the following figure (Figure 0.1).

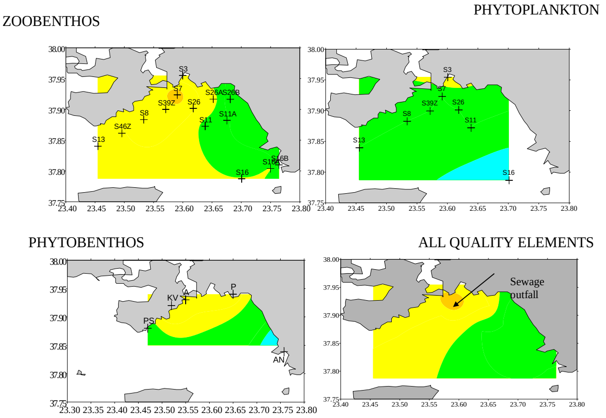


Figure 0.1 Application of the OOAO principle in Saronikos gulf (Greece, eastern Mediterranean).

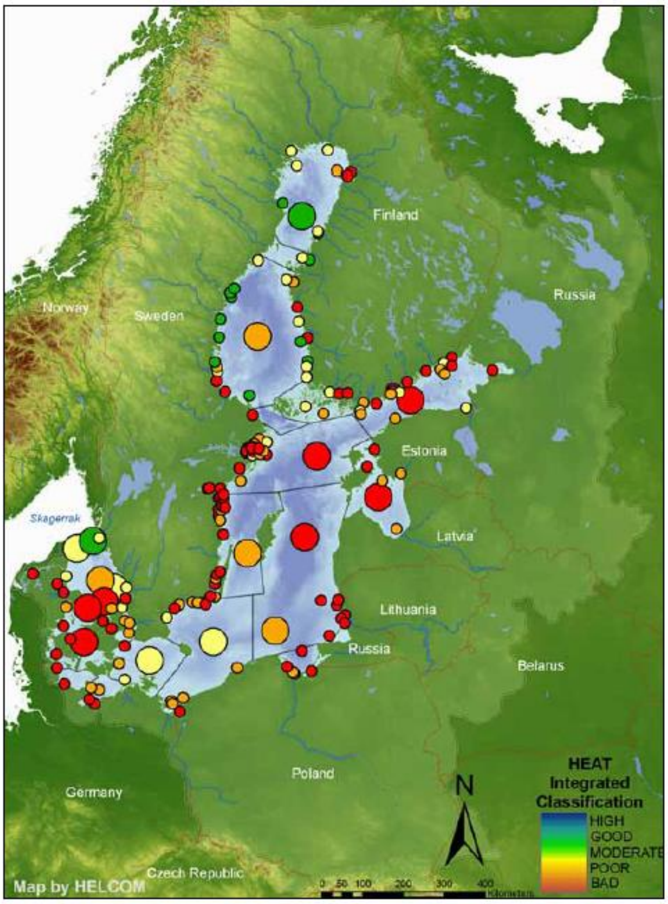


Figure 5.1 Integrated classification of eutrophication status based on 189 areas. Green = good status, yellow = moderate status, orange = poor status, and red = bad status. Good status is equivalent to 'areas not affected by eutrophication', while moderate, poor and bad are equivalent to 'areas affected by eutrophication'. Large circles represent open basins, while small circles represent coastal areas or stations. HEAT = HELCOM Eutrophication Assessment Tool.

Figure 0.2 Integrated classification of eutrophication status based on the HEAT HELCOM eutrophication tool of the Baltic Sea.

5. Another example of using the OOAO rule in the integration of different elements results is the HELCOM hazardous substances assessment tool **CHASE** tool presented in HELCOM (2009). CHASE (Figure 0.3) has been used to integrate the status of contamination by individual chemicals and biological effects at specific sites or areas into a single status value termed the "contamination ratio-CR". The data in the integrated assessment were primarily from biota and only secondarily from sediment or water. The CHASE tool gives each element a status (bad, poor, moderate, good or high) and the final status is defined as the lowest status of the four elements. Thus, the final classification is based on the "one out, all out principle". Moreover, the approach adopted gives **equal weight** to all the elements, i.e., the objectives of BSAP. All common groups of hazardous substances PCBs, dioxins, heavy metals, organometals, alkylphenols, phthalates, brominated substances, polycyclic aromatic hydrocarbons (PAHs), DDTs and chlorinated pesticides as well as the radionuclide cesium-137—were found among the substances with the highest CRs.

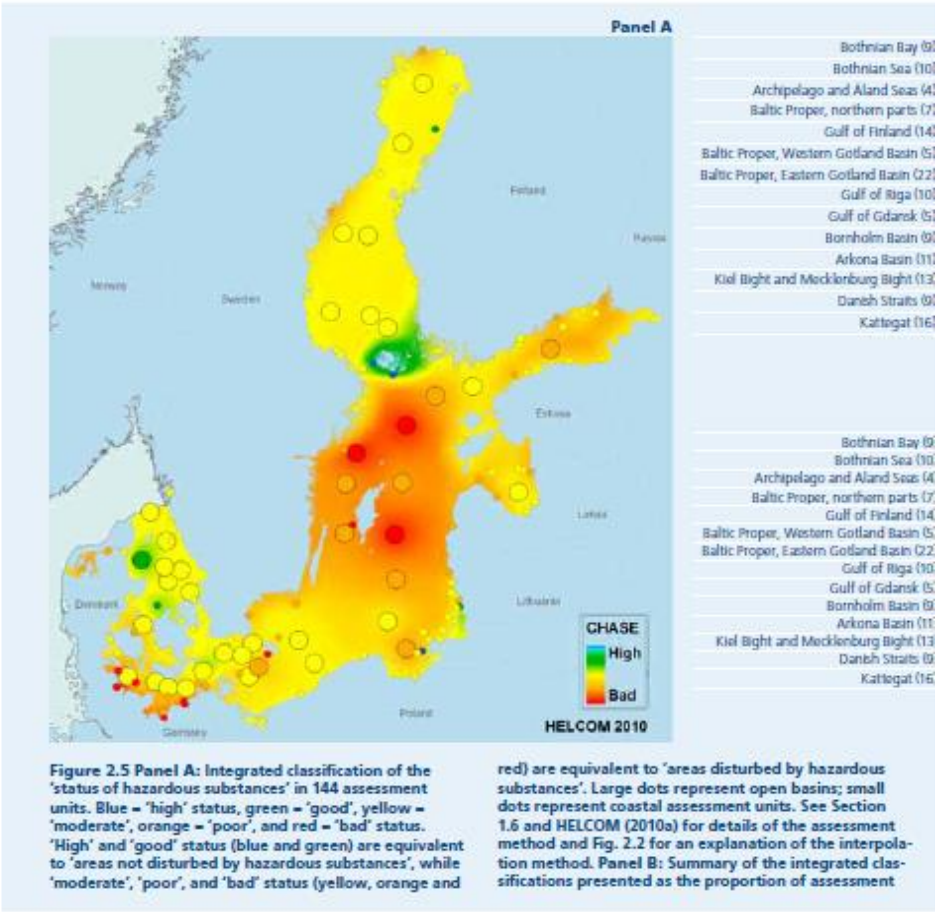


Figure 2.5 Panel A: Integrated classification of the 'status of hazardous substances' in 144 assessment units. Blue = 'high' status, green = 'good', yellow = 'moderate', orange = 'poor', and red = 'bad' status. 'High' and 'good' status (blue and green) are equivalent to 'areas not disturbed by hazardous substances', while 'moderate', 'poor', and 'bad' status (yellow, orange and red) are equivalent to 'areas disturbed by hazardous substances'. Large dots represent open basins; small dots represent coastal assessment units. See Section 1.6 and HELCOM (2010a) for details of the assessment method and Fig. 2.2 for an explanation of the interpolation method. Panel B: Summary of the integrated classifications presented as the proportion of assessment

Figure 0.3 Integrated classification of the hazardous substances based on the CHASE HELCOM Tool for the Baltic Sea.

Task Group 3 Report on commercially exploited fish and shellfish report that a stock can only achieve GES if all three criteria for the attributes are fulfilled. This is **the horizontal rule** applying for criteria of the attributes of D3 for a specific proportion of the stocks **within stocks** and in a way is identical with the OOAO rule because if one criterion is not fulfilling GES, then the rest of the criteria and the whole attribute would fail GES (all out).

6. Finally, an example for the Basque coast (northern Spain) is presented in Figure 0.4. It is based on the WFD, but it is illustrative of the problems when using the OOAO principle: no trend is shown in the quality with this approach. However, when using the decision tree of Borja et al. (2004, 2009a), an improving trend is shown. This improvement is due to the measures undertaken in the area to revert the situation at the beginning of the series. In this and other monitoring networks changes exist (i.e. different number of stations studied; different elements monitored; etc.), making comparisons difficult. However, the methods to be used in the assessment should be able to catch that evolution of the system after taking measures (or increasing pressures), and the OOAO seems to not able doing that. This is because the probability of having an element in less than good status increases with the number of elements/descriptors included. In this way, the MSFD, with 11 descriptors/54 indicators, risks to be always below good status, if the OOAO is applied.

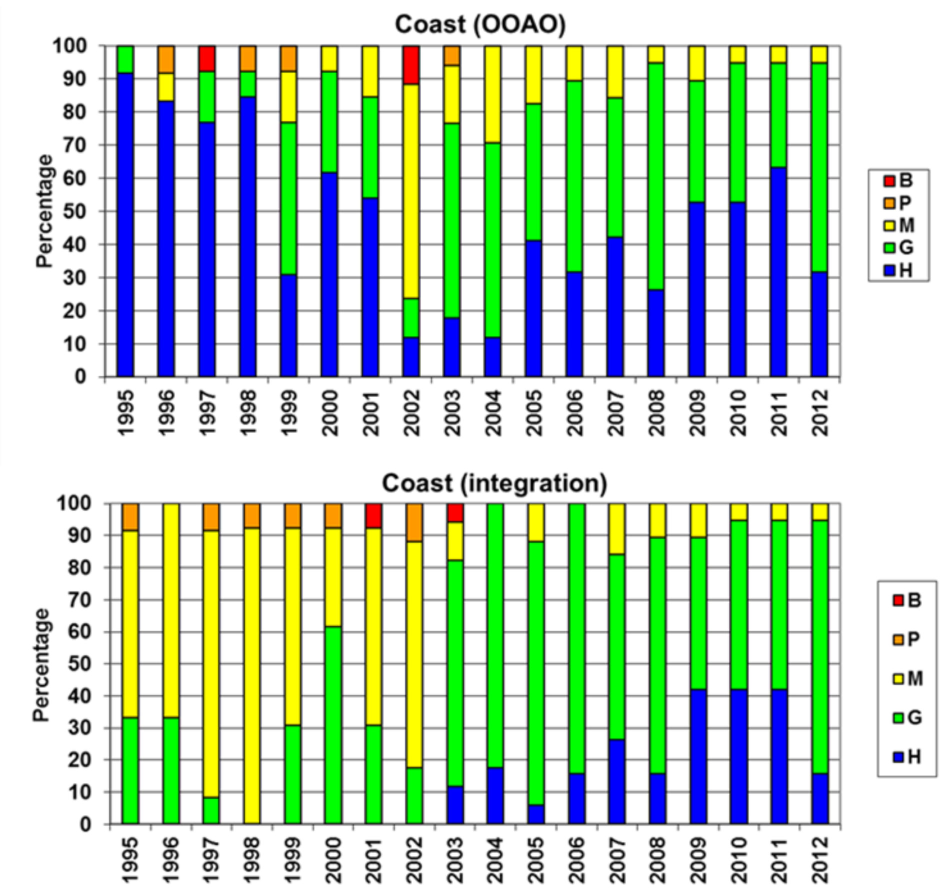


Figure 0.4 Comparison between the results obtained using the 'One out, all out' (OOAO) principle and an Integrated classification (Borja et al., 2004, 2009a) for the same dataset on the Basque coast (northern Spain), within the Water Framework Directive. H: high; G: good; M: moderate; P: poor; B: bad.