WP 6 Deliverable 6.2

Potential Definition of Good Environmental Status

Deliverable 6.2

Dissemination level

Public

LEAD CONTRACTOR

AZTI

AUTHORS

Angel Borja (AZTI-Tecnalia), Mike Elliott (University of Hull), Jesper H. Andersen (University of Aarhus), Ana C. Cardoso (JRC), Jacob Carstensen (University of Aarhus), João G. Ferreira (IMAR), Anna-Stiina Heiskanen (SYKE), João C. Marques (IMAR), João Neto (IMAR), Heliana Teixeira (JRC), Laura Uusitalo (SYKE), María C. Uyarra (AZTI-Tecnalia), Nikolaos Zampoukas (JRC), Theo Prins (DELTARES), Nomiki Simboura (HCMR), Torsten Berg (MARILIM), Nadia Papadopoulou (HCMR), Johnny Reker (EEA), Iratxe Menchaca (AZTI-Tecnalia)

SUBMISSION DATE

31 | January | 2015

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under the grant agreement n° 308392

www.devotes-project.eu

Citation

Borja A, Elliott M, Andersen JH., Cardoso AC, Carstensen J, Ferreira JG, Heiskanen A-S, Marques JC., Neto JM, Teixeira H, Uusitalo L, Uyarra MC, Zampoukas N, Prins T, Simboura N, Berg T, Papadopoulou N, Reker J, Menchaca I (2015) Report on potential Definition of Good Environmental Status Deliverable 6.2, DEVOTES Project. 62pp.

Abstract

The European Marine Strategy Framework Directive (MSFD) requires EU Member States to achieve Good Environmental Status (GEnS) of their seas by 2020. Also, the environmental status assessment of marine ecosystems is useful for communicating key messages to policymakers or the society, as it reduces the complex information of the multiple ecosystem and biodiversity components, and their important spatial and temporal variability, into manageable units.

The present deliverable addresses the **question of what GEnS entails especially with regard to the level at which targets are set** (descriptors, criteria, indicators), to **scales** for assessments (regional, sub-divisions, site-specific), and to **difficulties in putting into practice the GEnS concept**. In this sense, a **refined and operational definition of GEnS is proposed by DEVOTES**, indicating the data and information needed for all parts of that definition. The options for determining when GEnS has been met, acknowledge the data and information needs for each option, and a combination of existing quantitative targets and expert judgement is recommended.

The **approaches for aggregating and integrating assessments, currently available** for marine status assessment in Europe and other regions of the world, **have been reviewed,** from ecological and management perspectives, highlighting the advantages and shortcomings of the different alternatives. Furthermore, some guidance on the steps towards defining rules for aggregation and integration of information at multiple levels of ecosystem organization and some recommendations on when using specific rules in the assessment have also been provided.

The present Deliverable concluded that the MSFD implementation needs to be less complex in its implementation than shown for other similar directives (e.g. Water Framework Directive), based largely on existing data and centred on the activities of the Regional Seas Conventions. Moreover, any integration principle used should be ecologically-relevant, transparent and well documented, in order to make it comparable across different geographic regions.

This Deliverable is based upon two scientific publications made by the following task teams:

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

- Borja, A., T. Prins, N. Simboura, J. H. Andersen, T. Berg, J. C. Marques, J. M. Neto, N. Papadopoulou, J. Reker, H. Teixeira, L. Uusitalo, 2014. Tales from a thousand and one ways to integrate marine ecosystem components when assessing the environmental status. *Frontiers in Marine Science*, 1. DoI: 10.3389/fmars.2014.00072.

Contents

1. Introduction

Marine waters have traditionally been used by society for different activities (e.g. fishing, aquaculture, shipping, tourism, discharges from agriculture and urban areas). Currently, new activities are being developed or increasing (e.g. renewable energies, extraction of minerals, etc.), and thus competing for space with traditional uses, causing many spatial conflicts and increasing human impacts on marine ecosystems (Ban and Alder, 2008; Halpern et al., 2008).

The different legislative mandates to asses status coming from the Marine Strategy Framework Directive (MSFD) (covering all marine waters up to the limit of the Exclusive Economic Zone (EEZ) and extended continental shelf), Water Framework Directive (WFD) (for transitional and coastal waters up to 1 nm from the continental baseline) (2000/60/EC) and Habitats Directive (92/43/EEC) and other international initiatives have produced numerous methodologies that can be applied to different ecosystem components, such as various taxonomic or functional groups, habitats, traits, physical features, or to the whole ecosystem (Birk et al., 2012; Halpern et al., 2012). In addition, there is a proposed Directive for Maritime Spatial Planning and Integrated Coastal Management (European Commission, 2013) which will integrate management and planning. While the legal framework for 'maritime spatial planning and coastal development' is relatively new (Ehler and Douvere, 2009; European Commission, 2013), most of the legislation to protect, conserve or enhance marine ecosystems is based upon the United Nations Convention on the Law of the Sea (UNCLOS, 1982).

The MSFD aims to achieve or maintain 'good environmental status' (GEnS) in marine waters, by 2020 (European Commission, 2008), whilst the WFD aims to achieve 'good status' for all surface waters by 2015. 'Good status' means both 'good ecological status' (GEcS) and 'good chemical status'. Following the recommendation from Mee et al. (2008), the GEnS and GEcS acronyms are used because the meaning of 'environmental' and 'ecological' is different (see Borja et al. (2010), for differences between both concepts), implying a different emphasis between these two major pieces of legislation.

It has been argued that GEcS, as required by the WFD, focuses more on ecological structure, i.e. at a given time the abundance, presence, cover, etc., of ecological components, referred to as biological quality elements (Heiskanen et al., 2004; Borja et al., 2010), i.e. phytoplankton, macrophytes, macroinvertebrates and fish (the latter not in marine waters). In turn, the concept of environmental status, as defined by the MSFD, takes into account the structure, function and processes of marine ecosystems, bringing together natural physical, chemical, physiographic, geographic and climatic factors, and integrates these conditions with anthropogenic impacts and activities carried out in the area of concern (European Commission, 2008). Hence, it has been argued that, in using wider descriptors which relate to pressures, the MSFD provides a functional approach to measuring ecosystem health, where 'functional' refers to rate (i.e. time-dependent) processes (Borja et al., 2010).

According to the MSFD, the environmental status is defined by 11 descriptors, and forms a proposed set of 29 associated criteria and 56 indicators that include biological, physico-chemical indicators as well as pressure indicators —including hazardous substances, hydrological alterations, litter and noise, an biological disturbance such as introduction of non-indigenous species (Cardoso et al., 2010; European Commission, 2010) (Table 1). Following the implementation schedule of the MSFD, EU Member States (MS) started to assess the environmental status of their marine waters in 2012. This assessment should be carried out in an integrative way, including measurement of many ecosystem components together with physicochemical parameters and elements of pollution (Borja et al., 2009a).

Despite this wealth of methods, determining environmental status and assessing marine ecosystems health in an integrative way is still one of the grand challenges in marine ecosystems ecology research and management (Borja, 2014). In most countries, the precise means of implementing the MSFD are yet unclear. In most cases MS are focussing on individual descriptors and then criteria and indicators within the descriptors, with apparently little or no attention being paid to the means of combining the indicators, criteria and descriptors into a holistic assessment of the environmental status (http://ec.europa.eu/environment/marine/public-consultation/index_en.htm; Palialexis et al., 2014). Although several attempts have been made to assess the environmental status of marine waters in an integrative manner (HELCOM, 2010; Borja et al., 2011b), there are still significant gaps regarding:

(i) the understanding of marine ecosystems and their responses to human activities, including climate change;

(ii) the baseline knowledge required to define GEnS in an adequate and operational way;

(iii) the meaning of GEnS itself;

(iv) the approaches to determine targets or reference conditions for the indicators used in assessing the environmental status; and

(v) the identification, measurement and weighting of the components of the different indicators, synthesizing these aspects into a single value. For example, one indicator of biodiversity is the distribution range of species. However, the questions on how many species should be taken into account, whether all species are equally important, whether they should be considered as groups or as species and on a seasonal or annual basis, are yet to be answered (Borja et al., 2010).

Table 1. Qualitative descriptors, criteria and indicators, selected by the European Commission (2010), and to be used in the assessment of the environmental status of marine waters, in the context of the Marine Strategy Framework Directive.

DESCRIPTOR	CRITERIA	INDICATOR
1: Biological diversity	1.1 Species distribution	1.1.1 Distributional range
		1.1.2 Distributional pattern within the latter
		1.1.3 Area covered by the species (for sessile/benthic species)
	1.2 Population size	1.2.1 Population abundance and/or biomass
	1.3 Population condition	1.3.1 Population demographic characteristics
		1.3.2 Population genetic structure
	1.4 Habitat distribution	1.4.1 Distributional range
		1.4.2 Distributional pattern
	1.5 Habitat extent	1.5.1 Habitat area
		1.5.2 Habitat volume, where relevant
	1.6 Habitat condition	1.6.1 Condition of the typical species and communities
		1.6.2 Relative abundance and/or biomass, as appropriate
		1.6.3 Physical, hydrological and chemical conditions
	1.7 Ecosystem structure	1.7.1 Composition and relative proportions of ecosystem components (habitats, species)
2: Non-indigenous species	2.1 Abundance and state of non-indigenous species, in particular invasive species	2.1.1 Trends in abundance, temporal occurrence and spatial distribution of non-indigenous species
	2.2 Environmental impact of invasive non-indigenous sp.	2.2.1 Ratio between invasive non-indigenous species and native species
		2.2.2 Impacts of non-indigenous invasive species at the level of species, habitats and ecosystem
3: Exploited fish and shellfish	3.1 Level of pressure of the fishing activity	3.1.1 Fishing mortality (F)
		3.1.2 Catch/biomass ratio
	3.2 Reproductive capacity of the stock	3.2.1 Spawning Stock Biomass (SSB)
		3.2.2 Biomass indices
	3.3 Population age and size distribution	3.3.1 Proportion of fish larger than the mean size of first sexual maturation
		3.3.2 Mean maximum length across all species found in research vessel surveys
		3.3.3 95 % percentile of the fish length distribution observed in research vessel surveys
		3.3.4 Size at first sexual maturation
4: Food webs	4.1 Productivity of key species or trophic groups	4.1.1 Performance of key predator species using their production per unit biomass
	4.2 Proportion of selected species at the top of food webs	4.2.1 Large fish (by weight)
	4.3 Abundance/distribution of key trophic groups/species	4.3.1 Abundance trends of functionally important selected groups/species
5: Human-induced eutrophication	5.1. Nutrients levels	5.1.1 Nutrients concentration in the water column
		5.1.2 Nutrient ratios (silica, nitrogen and phosphorus)
	5.2 Direct effects of nutrient enrichment	5.2.1 Chlorophyll concentration in the water column
		5.2.2. Water transparency related to increase in suspended algae
		5.2.3 Abundance of opportunistic macroalgae
		5.2.4 Species shift in floristic composition such as diatom to flagellate ratio, benthic to pelagic shifts, as well as bloom events of nuisance/toxic algal blooms caused by human activities

	5.3 Indirect effects of nutrient enrichment	5.3.1 Abundance of perennial seaweeds and seagrasses impacted by decrease in water transparency
		5.3.2 Dissolved oxygen changes and size of the area concerned
6: Seafloor integrity	6.1 Physical damage, having regard to substrate characteristics	6.1.1 Type, abundance, biomass and areal extent of relevant biogenic substrate
		6.1.2 Extent of the seabed significantly affected by human activities for the different substrate types
	6.2 Condition of benthic community	6.2.1 Presence of particularly sensitive and/or tolerant species
		6.2.2 Multi-metric indices assessing benthic community condition and functionality, such as species diversity and richness, proportion of opportunistic to sensitive species
		6.2.3 Proportion of biomass or number of individuals in the macrobenthos above specified length/size
		6.2.4 Parameters describing the characteristics of the size spectrum of the benthic community
7: Hydrographical conditions	7.1 Spatial characterisation of permanent alterations	7.1.1 Extent of area affected by permanent alterations
	7.2 Impact of permanent hydrographical changes	7.2.1 Spatial extent of habitats affected by the permanent alteration
		7.2.2 Changes in habitats, in particular the functions provided due to altered hydrographical conditions
8: Contaminants	8.1 Concentration of contaminants	8.1.1 Concentration of the contaminants measured in matrices such as biota, sediment and water
	8.2 Effects of contaminants	8.2.1 Levels of pollution effects on the ecosystem components concerned, having regard to the selected biological processes and taxonomic groups where a cause/effect relationship has been established
		8.2.2 Occurrence, origin, extent of significant acute pollution events and their impact on biota physically affected by this pollution
9: Contaminants in fish and seafood	9.1 Levels, number and frequency of contaminants	9.1.1 Actual levels of contaminants that have been detected and number of contaminants which have exceeded maximum regulatory levels
		9.1.2 Frequency of regulatory levels being exceeded
10: Litter	10.1 Characteristics of litter in the marine and coastal environment	10.1.1 Trends in the amount of litter washed ashore and/or deposited on coastlines, including analysis of its composition, spatial distribution and source
		10.1.2 Trends in the amount of litter in the water column and deposited on the seafloor
		10.1.3 Trends in the amount, distribution and composition of micro-particles
	10.2 Impacts of litter on marine life	10.2.1 Trends in the amount and composition of litter ingested by marine animals
11: Energy and noise	11.1 Distribution in time and place of loud, low and mid frequency impulsive sounds	11.1.1 Proportion of days and their distribution within a calendar year over areas of a determined surface, as well as their spatial distribution, in which anthropogenic sound sources exceed levels that are likely to entail significant impact
	11.2 Continuous low frequency sound	11.2.1 Trends in the ambient noise level within the 1/3 octave bands 63 and 125 Hz (centre frequency) measured by observation stations and/or with the use of models

To assist MS in determining GEnS, several principles were highlighted in a Common Understanding document (Claussen et al., 2011). However, these do not provide an operational approach nor tackle the fundamental question of how to integrate all the aspects of assessment within and among marine areas.

Most of the information collated in the current report has been already published by DEVOTES partners in two scientific papers:

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

- Borja, A., T. Prins, N. Simboura, J. H. Andersen, T. Berg, J. C. Marques, J. M. Neto, N. Papadopoulou, J. Reker, H. Teixeira, L. Uusitalo, 2014. Tales from a thousand and one ways to integrate marine ecosystem components when assessing the environmental status. *Frontiers in Marine Science*, 1. DoI: 10.3389/fmars.2014.00072

2. Objectives

The **main objective** of this DEVOTES Deliverable 6.2 (D6.2) is focus on integration of different biodiversity components for an operational GEnS definition. This document includes the results from Task 6.1.5 (T615).

In addition, the operational objectives of this document are:

- to explain a conceptual definition of GEnS and related problems (e.g. the different steps for selecting criteria/indicators and the proposal of minimum common approaches);
- to show the role of setting of targets to determine GEnS, at indicator level, providing some criteria on setting them;
- to propose an operational definition of GEnS, with the challenges for deriving and using it;
- to present the existing approaches for integrating assessment results, from the indicator level up to the level of GEnS of a marine area, presenting an overview of the different methods currently available to synthesize the ecosystem complexity, by aggregating and integrating information when assessing the status, focusing mostly on the descriptors related to biodiversity, namely D1, D2, D4, D6 (Cardoso et al., 2010; Prins et al., 2013). This overview would assist managers, through the guidelines provided, in taking decisions for a better management of the marine ecosystems.

3. Conceptual definition of Good Environmental Status (GEnS)

The MSFD defines GEnS as ***"the environmental status of marine waters where these provide ecologically diverse and dynamic oceans and seas which are intrinsically clean, healthy and productive, and the use of the marine environment is at a level that is sustainable, thus safeguarding the potential for uses and activities by current and future generations"*** (European Commission, 2008). Hence, this implicitly requires ecosystem services and societal benefits to be delivered (Atkins et al., 2011a, b) even though the descriptors and their proposed criteria (European Commission, 2010) do not mention these aspects.

The concept of environmental status and the normative definitions of GEnS relate to the structure and successful functioning of ecosystems which allow them to maintain their resilience to human-induced environmental change (Elliott et al., 2007). This means that marine species and habitats are protected, human-induced decline of biodiversity is prevented, and diverse biological components are in balance. In addition, hydrographical, physical and chemical properties of the ecosystems, including those which result from human activities, support the ecosystems as described above. Finally, anthropogenic inputs of substances and energy, including noise, into the marine environment do not cause significant adverse effects to marine environment and allow ecosystem services and societal benefits to be maintained and delivered (Atkins et al., 2011a, b). There is, however, little understanding of what to consider as a meaningful quantitative definition of GEnS for a marine area, despite the stipulation in the MSFD that GEnS is an expression of the desired condition of the environment.

Background work to define GEnS at individual descriptor level was undertaken by different Task Groups for 10 out of the 11 descriptors (Table 2): biodiversity (Cochrane et al., 2010); non-indigenous species (Olenin et al., 2010); exploited fish (Piet et al., 2010); food-webs (Rogers et al., 2010); human-induced eutrophication (Ferreira et al., 2010); seafloor integrity (Rice et al., 2010); contaminants (Law et al., 2010; Swartenbroux et al., 2010); litter (Galgani et al., 2010) and noise (Tasker et al., 2010). The remaining descriptor, alterations of hydrographical conditions, was addressed in a more general way. For some descriptors, the Task Group report included a clearer definition of GEnS than for others (Table 2). This lack of clarity means that cross-border harmonisation is needed to ensure that environmental status does not change abruptly at the maritime boundary of two states.

Table 2. Normative definition of Good Environmental Status as suggested by Task Groups for each descriptor and modified from Cardoso et al. (2010). Key: OOAO: 'one out, all out' principle; GEnS: good environmental status; IAS: Invasive alien species; NIS: non-indigenous species; EnQS: Environmental Quality Standards; MSFD: Marine Strategy Framework Directive

DESCRIPTOR	GOOD ENVIRONMENTAL STATUS DEFINITION	AGGREGATION RULES
1: Biological diversity	GEnS will be achieved if there is no further loss of the diversity of genes, species and habitats/communities at ecologically relevant scales and when deteriorated components, where intrinsic environmental conditions allow, are restored to target levels. Target levels are defined as being such that "the quality and occurrence of habitats and the distribution and abundance of species are in line with prevailing physiographic, geographic and climatic conditions". Some deviation from reference conditions, as a result of human use of the marine environment, is acceptable, providing the terms of the Descriptor are still met.	GEnS will be achieved when each of the targets established by the Member States for all the attributes and components of biological diversity have been met (OOAO)
2: Non-indigenous species	IAS cause adverse effects on environmental quality resulting from changes in biological, chemical and physical properties of aquatic ecosystems. These changes include, but are not limited to: elimination or extinction of sensitive and/or rare populations; alteration of native communities; algal blooms; modification of substrate conditions and the shore zones; alteration of oxygen and nutrient content, pH and transparency of water; accumulation of synthetic pollutants, etc. The magnitude of impacts may vary from low to massive and they can be sporadic, short-term or permanent. The degradation gradient in relation to NIS is a function of their relative abundances and distribution ranges, which may vary from low abundances in one locality with no measurable adverse effects up to occurrence in high numbers in many localities, causing massive impact on native communities, habitats and ecosystem functioning.	Methods for aggregating indicators for GEnS assessments need to take into account the known IAS effects in other world regions or in neighbouring areas.
3: Exploited fish and shellfish	Since there is broad scientific evidence that GEnS cannot be achieved for all stocks simultaneously, a realistic threshold for the proportion of stocks with GEnS needs to be established above which the descriptor has achieved GEnS.	GEnS is achieved for a particular stock only if criteria for all attributes are fulfilled (OOAO).
4: Food webs	GEnS should ensure that populations of selected food web components occur at levels that are within acceptable ranges that will secure their long-term viability.	GEnS will therefore be achieved when the indicators describing the various attributes of the descriptor reach the thresholds set for them (OOAO)

5: Human-induced eutrophication	GEnS has been achieved when the biological community remains well-balanced and retains all necessary functions in the absence of undesirable disturbance associated with eutrophication (e.g. excessive algal blooms, low dissolved oxygen, declines in seagrasses, kills of benthic organisms and/or fish) and/or where there are no nutrient-related impacts on sustainable use of ecosystem goods and services.	No specific method is recommended, but those used must be robust, integrated, sufficiently sensitive, comparable, and with recognized scientific merit.
6: Seafloor integrity	The standard for GEnS should reflect the goals for management of the impacts of human activities on the sea floor. It is explicit in the definition of the descriptor that human uses of the ocean, including uses that affect the sea floor, are consistent with the MSFD, as long as those uses are sustainable. Sustainability is achieved when the pressures associated with all those uses cumulatively do not hinder the ecosystem components to retain their natural diversity, productivity and dynamic ecological processes. Perturbations due to use must be small enough that recovery is rapid and secure if a use ceases.	No single algorithm for combining indicator values will be appropriate for evaluating GEnS or providing a meaningful index of GEnS. It may be possible to conduct such analytical syntheses of Indicators for individual attributes on local scales. However, across attributes and on even moderate scales expert assessments rather than algorithmic formulae will be needed for evaluation.
8: Contaminants	Achievement of EnQS. Biological effects should be assessed against threshold levels of response that are indicative of significant harm to the organisms concerned.	Integration is greatly facilitated by coherent and consistent sets of assessment thresholds (EnQS).
9: Contaminants in fish and seafood	GEnS would be achieved if all contaminants are at levels below the levels established for human consumption or showing a downward trend (for the substances for which monitoring is ongoing but for which levels have not yet been set). However, it is generally felt that GEnS for descriptor 9 must be judged in view out the monitoring of descriptor 8, also dealing with contaminants in marine environment	OOAO
10: Litter	Definitions of the acceptable levels of harm and GEnS must consider impacts as assessed by the amount of litter in different compartments of the marine environment (seabed, sea surface, water column, coastline), ecological effects of the litter (e.g. plastics ingested by marine organisms; entanglement rates) and problems associated with degradation of litter (microparticles) as well as social and economic aspects. Tourism is strongly negatively affected by the presence of litter. An overriding objective will be a measurable and significant decrease (e.g. 10%/year for litter on coastlines) in the total amount of litter in the environment by 2020.	No proposal
11: Energy and noise	GEnS occurs when there is no adverse effect of energy inputs on any component of the marine environment. However, such an objective is probably not achievable if, for instance, behavioural disturbance or mortality of plankton (including planktonic larvae) is considered an adverse effect. Such an objective is probably not also measurable for a very large proportion of organisms in the marine environment.	No proposal

The Commission (COM) Decision 2010/477/EU on criteria and methodological standards on good environmental status of marine waters, supported by the documents prepared by the Working Groups, provided a range of criteria and indicators based on which GEnS should be defined. This Decision acknowledges that the application of criteria and identification of indicators should take into account the essential features and characteristics, pressures and impacts identified by MS during the initial assessment of the marine waters (MSFD Article 8, Annex I and III). It gives flexibility for MS to select those criteria and associated indicators that address the most important impacts and threats to a particular marine ecosystem. It also allows for the use of limited criteria/indicators across a wide marine area, leaving the application of additional criteria/indicators to specific subareas. Consequently, it is important to emphasise that the central task for implementing the MSFD is to determine how the many different criteria/indicators should be combined into an integrative assessment framework.

Although the latest available Common Understanding document (Claussen et al., 2011), drafted by a group assigned to develop a common understanding of the main normative concepts of the MSFD, mentions that "*assessments following the first MSFD cycle should move towards full consideration of the relevant criteria and indicators as laid down in COM Decision 2010/477/EU*", it is still unclear if this requires 'full' use of the criteria and indicators as laid down in the COM Decision or only those 'relevant' to each marine region.

This ambiguity is likely to lead to different approaches (and hence possible confusion) within and between MS, and should be accounted for in any proposal for integrating at the level of criteria and indicators. Therefore, although the choice of criteria/indicators and aggregation rules are the responsibility of MS, coherence of frameworks within the different marine regions or sub-regions and across the Community should be ensured. Accordingly, we question whether minimum requirements should be proposed, and if so, at which level should those minima be set: indicators, criteria or descriptors.

4. Setting of targets to quantify GEnS

After agreeing on what to include for definition of GEnS, the next question, if we are to implement it, would be: **on which level do we define it and how do we quantify it**? The MS Initial Assessments (available at http://cdr.eionet.europa.eu/recent_etc?RA_ID=608) made it clear that these were ambiguous steps, since both GEnS and Environmental Targets have been defined and set at different levels across MS, from descriptor level to indicator level (Palialexis et al., 2014). Hence, we question the implications of this for GEnS comparability.

Many environmental initiatives worldwide, and especially European Directives, have the fundamental requirement for an authority to determine whether an area/ecological component/habitat is as expected given the prevailing environmental conditions and, if not and the changes are due to human activities, then which actions are required to remediate any problems. Hence, assessing the current environmental status requires a comparison between an expected/usual/normal state and an impacted one. Thus, it is necessary to determine a reference/baseline/threshold/trigger condition against which the actual or potentially changed situation can be compared (Borja et al., 2012). Therefore by definition, the determination of good status implies a condition which has been or can be compared against that anthropogenically-altered state. For example, the WFD indicates that there are four ways of determining that reference condition (Hering et al., 2010):

(i) to find an area similar to the one under study, but without the pressures (de facto, a control area). However, it is easy to understand that, in an industrial and populated Europe, it is hard to find areas without human pressure;

(ii) to hindcast to a time before pressures exerted an influence;

(iii) to numerically model an unimpacted condition for comparison (some of these possibilities are tested in DEVOTES WP4), and, if none of these are possible

(iv) to use expert judgement, which has been demonstrated in some cases as a useful tool (Teixeira et al., 2010; Borja et al., 2014).

Each of these alternatives poses challenges:

(i) the unavailability or difficulty of finding unimpacted conditions, especially in the highly developed areas or within eco-regions;

(ii) the question of the baseline conditions for hindcasting (and the basis that some unimpacted prior utopia may be unattainable) (Duarte et al., 2009);

(iii) the uncertainty or unavailability of numerical models for unbounded marine areas or moving baselines (caused by climate change), and;

(iv) the perceived reluctance to rely on expert judgement if there is the likelihood of any legal challenge to management mechanisms such as sanctions to industries likely to impact an area (Duarte et al., 2009; Carstensen et al., 2011; Borja et al., 2012).

It is axiomatic that 'you cannot manage without measurement' and so numerical targets are central in determining whether MS succeed or fail to attain GEnS (Borja et al., 2012). The description of GEnS and the establishment of environmental targets (under Articles 9 and 10 of the MSFD, respectively) need a clear understanding for making this GEnS assessment operational. Claussen et al. (2011) suggest that the MSFD depends on measurable environmental targets and thus is linked to relevant indicators. Furthermore, they expect that the indicators and environmental targets will be related to the 11 Descriptors and thus directly linked to the marine environmental pressures as indicated in the Annex III of the MSFD. Most importantly they suggest that "*… to articulate quantitatively what GES looks like and/or set appropriate environmental targets it will be necessary to define for each of the criteria and, where appropriate, the indicators in COM Decision 2010/477/EU, environmental boundaries or thresholds above or below which GES is considered to have been met. … To that effect, a boundary between success and failure to achieve or maintain GES should be established. Thresholds/levels/limits in this sense represent that boundary between an acceptable and unacceptable status*".

These comments indicate the central problem being addressed here – that in the discussions of the Common Implementation Strategy of the MSFD, GEnS is still contemplated as a combination, as yet undefined, of indicators, targets, thresholds, levels, limits, criteria, and descriptors. Hence, we need to question whether these "*environmental thresholds/levels/limits*" are equivalent to environmental targets and, if so, does this mean we would have GEnS/non-GEnS at indicator level. It is therefore necessary to clarify whether an 'Environmental Target', as described in Art.10 of the MSFD, is equivalent to an 'Environmental threshold/level/limit' proposed at the level of indicators to make them operational by monitoring the changes in attributes (Claussen et al., 2011). Furthermore, if they are, how do these environmental targets at the indicator level relate to a broader GEnS description for any given marine region? Annex IV of the MSFD presents the term Environmental Target as a far more complex and integrative concept (Figure 1) that is unlikely to be set at the indicator level or informed effectively by a single indicator.

Claussen et al. (2011) recognize four types of environmental targets: pressure-, state-, impact-based- and operational targets whose adequacy is largely dependent on the robustness of the evidence available and the nature of any Descriptor. However, their examples indicate that the type of targets differ essentially according to how the target is expressed (as state indicators, identified pressures, or impacts on biodiversity components) rather than in the approaches used to measure achievement of targets (see examples in Claussen et al., 2011).

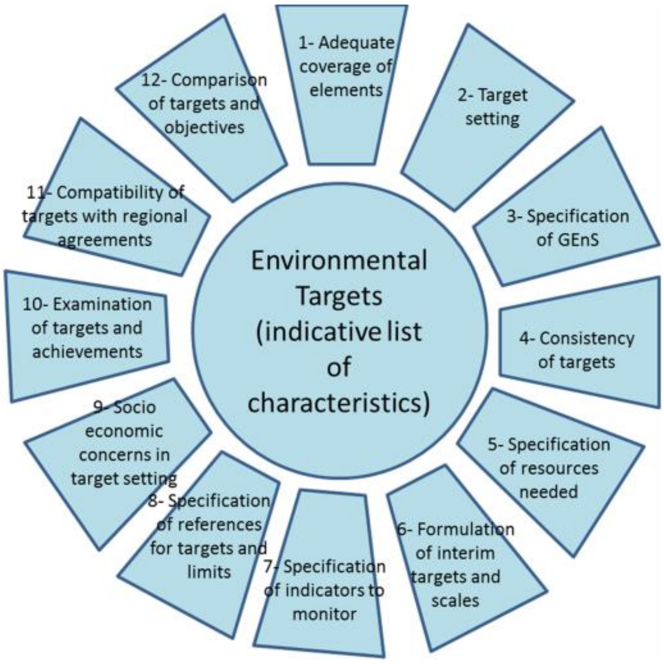

Figure 1. The indicative list of 12 characteristcs of environmetal targets as outlined in Annex IV Marine Strategy Framework Directive (modified from Claussen et al. (2011)).

Although the MSFD does not require defining reference conditions, guidelines from HELCOM and OSPAR propose the use of reference conditions as the starting point for defining targets when historical data or reference sites are available (Jonhson et al., 2013). In addition, the MSFD does not rule out the use of trend-based targets, i.e. to express GEnS as a direction of change. Thus, the MSFD accepts new approaches to define the desirable state of the environment. Some countries (e.g. in Sweden: Jonhson et al., 2013) have decided that, when the biological elements from the WFD used in coastal waters are also applied in the assessments of the marine environment, then the boundary between good and moderate status of the element will represent the boundary between GEnS and sub-GEnS.

However, **some criteria could be used when trying to set reference conditions or environmental targets for the MSFD**:

- Engage stakeholders, scientists and managers, for each relevant indicator or objective for which you desire setting targets or reference conditions, this will allow achieving the goals.
- Ensure that the targets and objectives are consistent with the key commitments established in the MSFD. Targets should be sufficiently clear to answer the question, “Did we achieve our objectives?”, after monitoring.

- Make sure that the objectives and targets are realistic (e.g. if you have determined reference conditions using historical data, are you sure that the target are achievable taking into account shifting baselines?).
- Determine how you will measure progress towards achieving the targets and objectives.
- Know that if an environmental aspect of the MSFD is not significant within your ecoregion, then it does not need an objective and target.

5. Proposal of an operational definition of GEnS

Elliott (2011) suggested that the only main aim in marine management is to protect and maintain the natural ecological functioning and at the same time deliver the benefits for society (based on ecosystem services and societal benefits) – this is *de facto* the Ecosystem Approach. Hence, we suggest that this should be taken as a starting point to define GEnS – in essence that GEnS is achieved if the conservation objectives, the ecosystem services and the societal benefits are delivered sustainably. If this is accepted then the challenge is to make this operational and indicate how it can be linked to monitoring, management and spatial planning.

The essence of the GEnS concept is to determine what type of ecological change is possible as the result of human pressures. If the primary aim is to ensure 'healthy, productive and safe seas' then we could use the *7 indicators for general application* for the diagnosis of ecosystem pathology proposed by Harding (1992; see also Elliott 2011): primary production, nutrients, species diversity and abiotic zones, instability and biotic composition, disease prevalence, size spectrum, bioaccumulation (and effects) of contaminants. Hence, an adverse change in any of these indicates a deviation from GEnS and thus a reduction in health of the system (see Tett et al., 2013) – thus equating GEnS with 'healthy'. This also implies that a healthy system is robust to the effects of stressors (Tett et al., 2013; Mouillot et al., 2013). We acknowledge that these 7 indicators are in someway already covered by criteria in the 11 Descriptors of the MSFD.

If GEnS is regarded as being synonymous with naturally 'productive' then we need to include the unifying currency for determining whether the marine system is delivering what society wants (an anthropocentric view) (Atkins et al., 2011a, 2011b). That is, whether the seas are maintaining the ecosystem services and delivering societal benefits (in a sustainable way). For example, if the marine system delivers the relevant ecosystem services without compromising the natural ecosystem functioning then, is that sufficient to say an area is achieving GEnS? There can be also trade-offs between ecosystem services: for example, a cold-water coral reef can provide high value for the 'regulation and maintenance service' by for example providing the nursery function for some fish species, and at the same time the 'provisioning service', the capture of wild fish of other species is decreasing because of a legal regulation (e.g. establishment of Marine Protected Area, or closure for deep water fisheries) (Foley et al., 2010). In such cases, there is the need to identify beneficiaries, and to evaluate how society, as a whole, would benefit from protecting the reef. If only economic use values were used then this emphasise an anthropocentric definition of GEnS, as above (i.e. 'seas are maintaining the ecosystem services and delivering societal benefits in a sustainable way') but we also have to incorporate the non-use values (Atkins et al., 2011a). In addition, this has to encompass the intrinsic conservation value, i.e. the value of the ecosystem for its own sake (Pascual et al., 2012).

We consider that the definition (or definitions) of GEnS not only has (or have) to summarise the status but that they have to be able to be used in real-time assessments and the operational management of the marine system.

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

The main challenge is to translate this definition into terms suitable to provide an operational tool. Based on the previous section (integration of the assessments), we propose that there are at least 8 options to determine the GEnS in a regional sea context (Table 3). Hence, we detail the concept behind the options, then select the decision rule for the method to be implemented, consider what type and amount of data are required, and then consider the pros and cons of the different options. The options from 1 to 8 are sequentially less demanding of new data, and the degree of detailed ecological assessment. As such, Option 1, which is most similar to the WFD approach, deconstructs GEnS into the 11 descriptors and then into the component indicators, assessing each for each area before attempting to produce an overall assessment (Table 3). 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 (Aubry and Elliott, 2006; Halpern et al., 2008; Korpinen et al., 2012). 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 (see Borja et al., 2011b; Halpern et al., 2012), each with its own requirements, pros and cons (Table 3). Most of the Options have been used in quality assessments of estuaries, lagoons, coastal and marine areas (see Table 3 references).

Determining the quality of any ecological element (individual, community, population or ecosystem), in relation to single or combined pressures and accounting for the inherent variability in the system is expensive and, in the current economic climate, this may be prohibitive (Borja and Elliott, 2013). Therefore, as suggested with Option 7, it is likely to be easier and cheaper to determine the GEnS as the 'absence of pressures' in a region rather than the 'presence of good environment'. The former has been used to good effect to create Quality Status Reports from OSPAR (2010) and HELCOM (2010) and can be obtained from maps and aerial photos, databases of users, automatic measuring such as Vessel Monitoring Systems (VMS) in the open sea, and limited modelling exercises. In contrast, detecting the 'presence of a good environment' requires a large amount of monitoring, i.e. each ecological and physical component needs to be assessed for its 'goodness', in itself a human (and thus subjective) construct (Mee et al., 2008). Despite this, detecting severe impacts is straightforward. However, in most seas surrounding developed countries, these impacts are likely to be on a small scale and/or produced by point-source pressures and thus controlled by licensing. This leads to a problem of whether to judge the quality of a sea area according to the many pressures, which may each only affect a small area and cumulatively affect a small proportion of a sea area but still resulting in a label of 'Poor Environmental Status'. Also this would require agreeing a pressure classification and indices that can show the strength and spatial extent of, and time for exposure to the pressure (e.g. using a pressure classification: low pressure, medium pressure, high pressure). For instance, based on the WFD pressure data reported by MS, it is possible to present the percentage of all classified water bodies affected by certain types of pressures (Solheim et al., 2012). The pressures included in the assessment are point and/or diffuse pollution per marine region, hydro-morphological pressures and/or altered habitats per marine region. However, the intensity and extent of different pressures is not evaluated – only presence or absence. As an example, the Baltic Sea Pressure Index (Korpinen et al., 2012) provides a dimensionless value of the indicator with a gradient from an absence of pressures to cases (areas) where virtually almost all pressures are present. However, the problem of boundary setting still remains including how to determine what is good and what is not good (Mee et al., 2008). One danger of focussing on pressures is that many are point-source in nature and so trying to extrapolate to whole eco-regions and sub-regions, as required by the MSFD, is difficult. Furthermore, diffuse sources may be difficult to identify and tackle, especially if having a trans-boundary effect.

Option 5, suggesting the use of a tick-list approach, takes a quasi-legal approach based on the probability of evidence, some or much of which may be based on informed judgement or based on precedent elsewhere. For example, even if there are insufficient data for all aspects, if the qualitative trends (even using expert judgement) in many of the aspects all point to an area being in good status then by definition the area is in good status (Ferreira et al., 2011). This probability of evidence approach is used in common law (i.e. as the best available evidence) and was the approach taken by Bricker et al. (2003) to rigorously take 'soft intelligence' from many US estuaries relating to eutophication effects and combine this to form 'hard data', hence producing an assessment which had the confidence of managers and stakeholders alike.

Table 3. Options for determining if an area/regional sea is in Good Environmental Status (GEnS) (modified from Borja et al., 2013). Key: OOAO: 'one out, all out' principle.

Option	Decision rule	Data requirements	Pros	Cons	Examples in place
Either: 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 (2010); 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	Feary et al. (2014)
Or: 4.create a synthesis indicator which takes the view that 'GEnS is the ability of an area to support ecosystem services, produce societal benefits and still maintain and protect the conservation features'	Integration of the information from different descriptors and indicators, and evaluation of the overall benefits	Data needed for the indicators included in that synthesis indicator, valuation of the ecosystem services and benefits	Fulfils the main aim of marine management (see text)	Requires a new indicator and an agreement in the way of integrate the information; trade-offs between ecosystem services and their beneficiaries require either economic, ethical or political evaluation and decision, and cannot be based only on ecological knowledge	Borja et al. (2011b)
Or: 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 GEnS	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)

Option	Decision rule	Data requirements	Pros	Cons	Examples in place
Or: 6.have a summary diagram such as a spiders-web diagram showing the 'shape of GEnS according to several headline indicators'	The shape of the diagram		Easy to understand and show to managers	The decision on when GEnS is achieved	Halpern et al. (2012)
Or: 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) See also Solheim et al. (2012) for the analysis of WFD pressures and impacts
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

By definition, because of the way the MSFD has been constructed, GEnS requires a multi-metric approach which can accommodate or encompass the descriptors and criteria/component indicators. As suggested by Option 6, a bottom-up approach whereby each is assessed and combined either numerically or, for example, as an ‘amoeba-type’ representation (Ten Brink et al., 1991), would be of value. Although such a diagram, encompassing the descriptors, criteria and indicators, has not yet been constructed, it is expected that it would indicate some kind of ecosystem ‘deformation’ from a circle or polygon as done in the Ocean Health Index by Halpern et al. (2012). It is especially important to incorporate and quantify the pressure and resilience directly into an assessment, although this is always one of the most difficult aspects to achieve, assuming this direct linkage and accounting for these “un-measurable” aspects.

The application of an operational definition of GEnS, such as that proposed before, implies many challenges. Firstly, the current economic difficulties may preclude an appropriate amount of monitoring to be carried out if any determination of GEnS requires assessments for all descriptors and for all biodiversity or ecosystem components and indicators of those descriptors in all areas. Hence, the MSFD requires flexibility in selecting the most relevant criteria and indicators. Although this somehow foreseen in the conditions for application of the criteria for good environmental status (European Commission, 2010: Annex - Part A, paragraph 4), such a stepwise approach requires further guidance to ensure a harmonized implementation and a balanced, fair and meaningful final GEnS assessment across marine regions. De Jonge et al. (2006) consider the problems of marine monitoring against changing political imperatives, where the successive revision of monitoring programmes usually equates to reductions, and Borja and Elliott (2013) consider the problem of a lack of funding for monitoring and thus the potential threat of uninformed marine decision-making. For example, the Danish national monitoring programme was initiated in 1989 to follow responses in the aquatic environment to nutrient reduction plans adopted in 1987 (Carstensen et al., 2006). This monitoring programme was ambitious from the outset but revisions have gradually reduced monitoring efforts, and for the first WFD planning period (2009-2015) and with the start of the MSFD monitoring (from 2014), the sampling effort is at its lowest for more than two decades. Thus, we are concerned that the information base on which decision-making should be made, potentially with large economic consequences, is gradually eroding due to reduced budgets for marine monitoring.

Secondly, it is questioned whether there should be a continuing emphasis on the physical and chemical approach against a prevailing desire (and philosophy) to monitor ecological health. If there are accepted links between deterioration in physical and chemical parameters and the reduction in ecological health then the former parameters can be used as a proxy for the latter and may be easier to monitor. However, if such links are not reliable then the more complex and expensive ecological monitoring will be required. This also relates to the difficulty of determining and implementing what is meant by pass and fail in an operational context, especially against a background of high variability in all aspects of the biological, physical and chemical system in the marine environment. The high variability may require ecologists to acknowledge that some of the components cannot be used operationally in monitoring for good environmental status, for example, trying to determine if a low and highly variable abundance of a marine mammal conforms to an accepted level. Hence, as suggested above, it may be easier to determine 'an absence of pressures' rather than a 'presence of good physical, chemical and biological features' and rely on adequate, remotely collected data and information.

In addition, further research is required to determine the relative merits of measuring each ecological component, which are surrogates or proxies for others, in which case is it better and more cost-effective to measure an absence of pressures rather than a presence of good ecology, and which components are most suitable and cost-effective to assess which pressure effects.

Thirdly, there are spatial considerations and anomalies including the problems created by the overlap in space of the WFD, MSFD, Habitats Directive and the proposed Maritime Spatial Planning and Integrated Coastal Management Directive, as well as the Common Fisheries Policy (to be published under Deliverable D.2.2 Report on the key barriers of achieving GES of DEVOTES project). This also needs consideration of which piece of legislation takes precedence in an area and so whether the definitions of 'good status', GEnS, GEcS or FCS (Favourable Conservation Status, under the Habitats Directive), are in agreement. It has been suggested that the WFD takes precedence in the near-shore area and the Habitats Directive in designated areas of conservation. In sum, as far as coastal waters are concerned, the MSFD applies only to the extent that activities are not covered by the WFD or other Community legislation (Hildering et al., 2009). However, there are likely to be spatial anomalies in the proposed system whereby marine areas may fail under GEnS and yet still be given the status as Special Protection Areas (SPA) and Special Areas for Conservation (SAC), under the Habitats and Wild Birds Directives. It is true that the opposite example can be also occur: e.g. an important bird area, located in the migration route of certain species, and which happens to be degraded – i.e. under GEnS, and therefore would be in need of recovery. That area would still be a SPA, in terms of its importance for bird species, but it just happens to be in a degraded condition. Its SPA status of protection just reinforces the need to prioritize such an area in terms of management and recovey.

6. Integration of GEnS assessement[1]

After setting reference values for the indicators, and determining the status of individual indicators and descriptors, **a new problem emerges when attempting to integrate all this information into a unique status assessment**. Whereas the WFD uses the so-called rule 'one out, all out' (OOAO), which means that the element with the worst status determines the global status and hence is regarded as a precautionary level, **no specific rule has yet been proposed for the MSFD**. However, when proposing aggregation rules, the Task Groups for the implementation of the MSFD deconstructed the ecosystem into 'descriptor indicators' and then recombined them to give a pass/fail for the GEnS, using in four of the cases the OOAO principle (Table 2). 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 (Borja et al., 2010). It implies that recombining a set of structural attributes gives an accurate representation of the ecosystem functioning.

Our detailed analysis of the work carried out by the descriptor-specific working groups suggests there has been a lot of initial activity, but little attention was paid to how the data/information will eventually be used. Despite their importance, combination rules for the MSFD descriptors and indicators were excluded from the remit of the Task Groups; in our view, an omission which needs to be addressed. For example, if the OOAO principle was to be applied to the 11 descriptors and 56 indicators, the probability to fail one or more indicators, even through analytical error, would be very high at any studied location, as demonstrated for the WFD with fewer components, the 5 biological quality elements (Borja and Rodríguez, 2010). This is especially problematical when different elements address the same pressure (Caroni et al., 2013) which, if both are combined, leads to 'double-counting' and thus an over-emphasis on the resulting status. Although Annex 1 of the MSFD describes the GEnS individually for each of the 11 descriptors, this does not imply the ability to have GEnS at the level of all the descriptors, nor does it mean that each descriptor should necessarily be graded individually in a binary way (i.e. good or not good environmental status). For example, the HELCOM (2010) tool for assessment of 'ecosystem health' (in practice equivalent to GEnS) groups indicators into 3 categories: biology, hazardous substances, and supporting indicators; applying then the OOAO rule at the 3 category level. Hence, reducing the MSFD descriptors to 3 groups, and then using the OOAO approach for those groups, may be a pragmatic compromise which minimises the probability of failure, due to analytical errors or statistical chance, while still giving one overall assessment of GEnS. This is analogous to the Environmental Integrative Indicator (EII) approach of Aubry and Elliott (2006) in which 27 indicators were grouped and weighted into 3 EII.

[1] This section has been done in collaboration between the research teams of DEVOTES and that of the Framework contract No ENV.D2/FRA/2012/0019 (Coherent geographic scales and aggregation rules in assessment and monitoring of Good Environmental Status—analysis and conceptual phase), of the European Directorate General of Environment (European Commission). The opinions expressed in this document are the sole responsibility of the authors and do not represent the official position of the European Commission.

The integration issue is further complicated by the fact that some of the descriptors function as pressures for other descriptors; e.g. alien species can be a threat to biodiversity and food web functioning; alterations in the hydrological regime can be a threat to seafloor integrity. Thus, we suggest that the 11 descriptors are hierarchical and do not have an equal weighting when assessing the overall GEnS (Borja et al., 2010). We further suggest that we take the philosophical view that for the descriptor Biodiversity to be fulfilled requires all others to be met and similarly if one of the stressor or pressure-related descriptors (e.g. energy including noise) fails, then by definition the biodiversity will be adversely affected.

In addition to this problem of aggregating indicators and descriptors, whilst the WFD centres on quality assessments within a small area, albeit extrapolated to a water body (Hering et al., 2010), the MSFD requires MS to integrate and geographically scale-up the assessments, at the level of an eco-region (Borja et al., 2010). This means that the GEnS assessments of the MS need to be comparable in order to enable integrating their assessments into an ecoregion-wide assessment and to avoid cross-border anomalies. This requires that comparable methods and aggregation rules are needed to ensure minimum standards for GEnS reporting across MS and as such we advocate a set of common principles (expanded from Claussen et al., 2011):

- The integration across levels of different complexity should accommodate different alternatives, i.e., integration below Descriptor level (across biological component within indicators, across indicators within criteria, and criteria within Descriptors) could certainly differ from Descriptor level integration (see Table 3, for different options at this level);
- Integrate across state Descriptors (D1, D3, D4, D6) differently than across pressure Descriptors (D2, D5, D7, D8, D9, D10, D11) (see Table 1, for identification);
- Consider a different contribution of the two types of main Descriptors for the overall GEnS 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 GEnS for state-based Descriptors (D1, 3, 4, 6) are achieved it follows that GEnS 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.

Independently of which aggregation proposal(s) is adopted and at which level, the precautionary principle should always prevail in the absence of more robust knowledge.

6.1. Combination methods

Combining the different values from indicators into an assessment of GEnS can be done in various ways. Indicators of the same descriptor can be combined as well as criteria. These can be combined within descriptors and across descriptors to eventually result in an assessment for a specific geographic area. Combination either involves numerical calculations or logical operations. If similar or comparable elements are combined (for example, different biodiversity state indicators targeting species diversity), the term “aggregation” is used. This indicates the applicability of numerical operations but is not limited to this kind of combination. If different elements are combined, we use the term “integration” to indicate that logical operations typically are applied (for example, combining the results from different descriptors).

Based on a literature review, a number of different approaches have been identified for combining a number of variables (which could be metrics, indicators, or criteria) into an overall assessment. Some of them have been used within the WFD, others within the RSCs and some others in the MSFD. An overview of the methods is given in Table 4.

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

6.1.1. One-out, all-out (OOAO)

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

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

Borja et al. (2009a) discussed the challenge of assessing ecological integrity in marine waters, and suggest that simple approaches, such as the 'OOAO' principle of the WFD, may be a useful starting point, but eventually should be avoided. The ecological integrity of an aquatic system should be evaluated using all information available, including as many biological ecosystem elements as is reasonable, and using an ecosystem-based assessment approach. The OOAO rule can be considered a rigorous approach to the precautionary rule, in an ideal world where the status based on each BQE can be measured without error. It results in very conservative assessments (Ojaveer and Eero, 2011). In practice, the inevitable uncertainty associated with monitoring and assessment for each metric and BQE leads to problems of probable underestimation of the true overall status.

The OOAO principle has therefore been criticized as it increases the probability of committing a false positive error, leading to an erroneous downgrading of the status of a water body as it has been observed especially within the WFD (Borja and Rodríguez, 2010; Ojaveer and Eero, 2011; Borja et al., 2013; Caroni et al., 2013). In the case of the MSFD, with such large number of descriptors, criteria and indicators, the probability of not achieving good status becomes very high and, probably, unmanageable in practical terms (Borja et al., 2013). Alternative methods for integrating multiple BQEs in the WFD are currently being considered (Caroni et al., 2013).

6.1.2. Averaging approach

The averaging approach is the most commonly used method to aggregate indicators (Shin et al., 2012) and consists of simple calculations, using methods such as arithmetic average, hierarchical average, weighted average, median, sum, product or combinations of those rules, to come up with an overall assessment value.

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

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

Table 4. Approaches for combining different metrics, indicators or criteria to assess the status, including the advantages and disadvantages of each approach, as considered by the authors. Key: GEnS: Good environmental status.

General approach	Details of method	Advantages	Disadvantages
One-out all-out (OOAO) principle CIS (2003), Borja et al. (2009a), Borja and Rodríguez (2010), Ojaveer and Eero (2011), Caroni et al. (2013)	All variables have to achieve good status	Most comprehensive approach. Follows the precautionary principle	Trends in quality are hard to measure. Does not consider weighting of different indicators and descriptors. Chance of failing to achieve good status very high.
	As a variation, Tueros et al. (2009) proposed the Two-out all-out: if two variables do not meet the required standard, good status is not achieved	More robust compared to OOAO approach	See above
Averaging approach Ojaveer and Eero (2011), Shin et al. (2012)	<u>Non-weighted</u>: 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
	<u>Weighted</u>: 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
	<u>Hierarchical</u>: With variables defined at different hierarchical levels	Reflects the hierarchy among descriptors and avoids double counting Different calculation rules can be applied at different levels	Problem of agreeing on hierarchy
Conditional rules Tueros et al. (2009), Simboura et al. (2012), Breen et al. (2012)	A specific proportion of the variables have to achieve good status	Focuses on the key aspects (i.e. biodiversity descriptors)	Assumes that GEnS is well represented by a selection of variables
Scoring or rating Borja et al. (2004, 2010, 2011b), Birk et al. (2012)	Sum of weighted scores	Different weights can be assigned to the various elements	Problem of agreeing on weights. Metrics may not be sensitive to the same pressures

General approach	Details of method	Advantages	Disadvantages
Multimetric approaches Rice et al. (2010), Borja et al. (2011a), Birk et al. (2012)	Multi-metric indices	Integrates multiple indicators into one value. May result in more robust indicators, compared to indicators based on single parameters	Correlations between parameters can be an issue. Results are hard to communicate to managers. Metrics may not be sensitive to the same pressures
Multi-dimensional approaches Shin et al. (2012)	Multivariate analyses	No need to set rigid target values, since values are represented within a domain	Results are hard to communicate to managers
Decision tree Borja et al. (2004, 2009b, 2013)	Integrating elements into a quality assessment using specific decision rules	Possible to combine different types of elements, flexible approach	Only quantitative up to a certain level
Probabilistic Barton et al. (2008, 2012), Lehikoinen et al. (2013, 2014)	Bayesian statistics	Produces a probability estimate of how likely the area is in GEnS; managers can decide the acceptable undertainty	Difficult to calculate
High-level integration HELCOM (2010), Borja et al. (2010, 2011b), Halpern et al. (2012), Tett et al. (2013)	Assessment results for three groups: biological indicators, hazardous substances indicators and supporting indicators, each applying OOAO	Reduces the risks associated with OOAO while still giving an overall assessment	Technical details

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

6.1.3. Conditional rules

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

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

6.1.4. Scoring or rating

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

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

6.1.5. Multimetric indices to combine indicators

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

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

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

6.1.6. Multidimensional approaches

Multivariate methods, such as Discriminant Analysis or Factor Analysis combine parameters in a multi-dimensional space. For assessment purposes, areas need to be classified into groups of GEnS and non-GEnS. Multivariate methods have the advantage of being more robust and less sensitive to correlation between indicators. However, interpretation is less intuitive than other methods, as information on individual indicators in each ecosystem is lost (Shin et al., 2012) and links to management options are less obvious.

6.1.7. Decision tree

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

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

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

6.1.8. Probabilistic approach

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

This approach can be combined to one or several of the above-mentioned approaches: for example, conditional rules can be set in addition to the probabilistic integration rule to include expert judgement; and the principles outlined in the decision tree approach can be applied as well.

Barton et al. (2012) demonstrate how to use the probabilistic approach in the DPSIR framework in the case of eutrophication management. There are several other examples in the recent literature about how to evaluate various management measures under uncertainty to optimise one target, such as eutrophication (Barton et al., 2008; Lehikoinen et al., 2014) and oil spill severity (Lehikoinen et al., 2013). This approach could be expanded to include several descriptors or indicators.

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

6.1.9. High-level integration

An example of a high-level integration, where assessments for several ecosystem components are merged into a final assessment, is the HELCOM-HOLAS project (HELCOM, 2010). The report presents an indicator-based assessment tool termed HOLAS ('Holistic Assessment of Ecosystem Health Status'). The indicators used in the thematic assessments for eutrophication (HEAT), hazardous substances (CHASE) and biodiversity (BEAT) were integrated into a Holistic Assessment of 'ecosystem health'. The HOLAS tool presented assessment results for three groups: biological indicators, hazardous substances indicators and supporting indicators, and then applied the OOAO principle on the assessment results of those three groups for the final assessment (Figure 2).

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

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

(a)

(b)

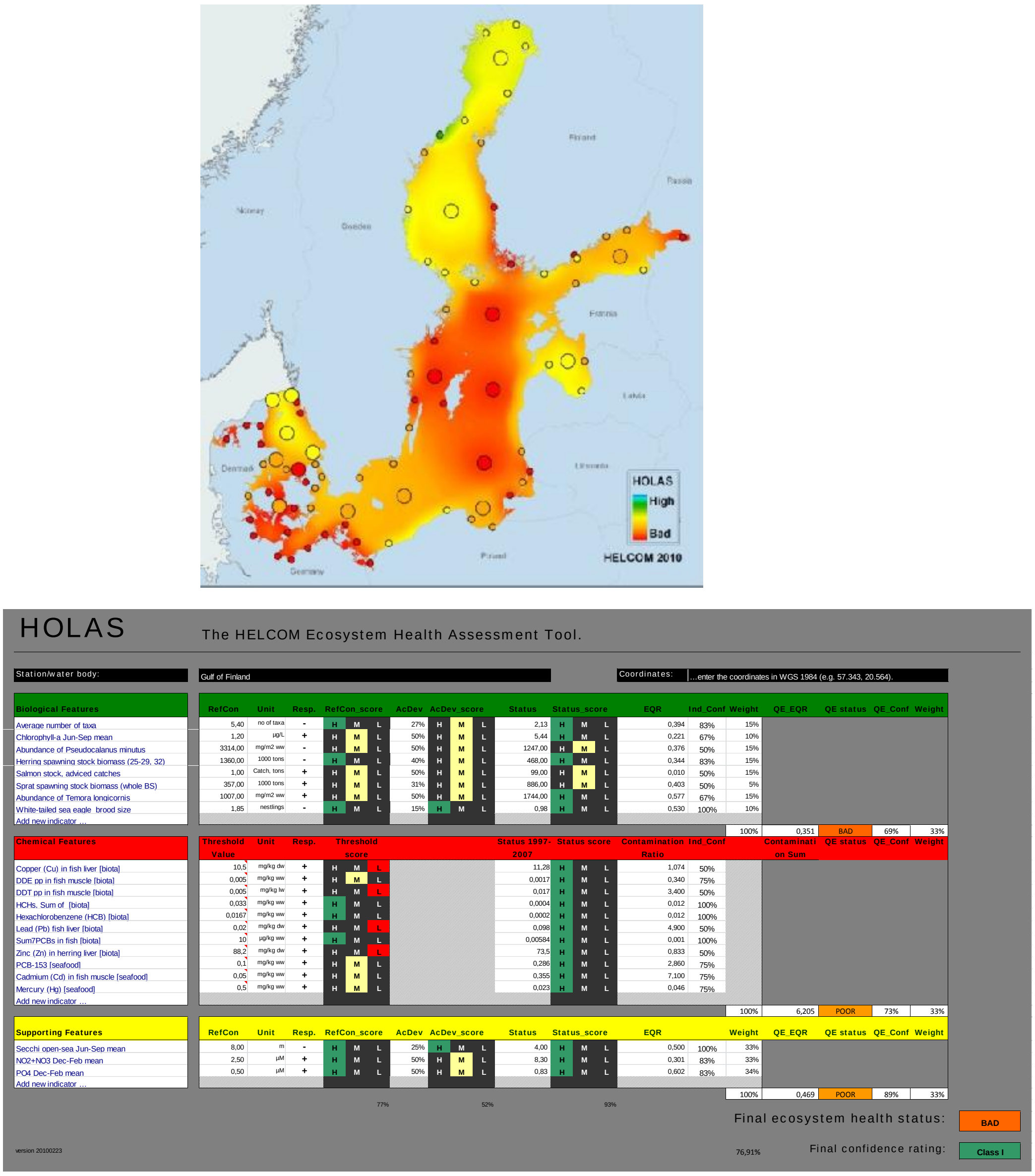

Figure 2. (a) Example of an integrated assessment of ecosystem health in the Baltic Sea 2003-2007 based on the HOLAS tool. (b) Screenshot to illustrate how the HOLAS classification tool for the Gulf of Finland works. See HELCOM (2010) for details. Courtesy by Helsinki Commission.

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

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

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

6.2. Considerations and recommendations when using specific rules

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

(i) OOAO is appropriate when:

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

(ii) OOAO cannot be used:

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

(iii) A ‘two out, all out’ approach can be considered in cases where several methods are combined in one assessment; e.g., when several matrices are used in pollutants to give a broader view of the status (e.g.

pollutants in water for an instant picture, pollutants in sediments or biota for a time-integrated result, Tueros et al. (2009)).

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

(v) Scoring or decision tree approaches are appropriate when:

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

(vi) Probabilistic approach:

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

(vii) Multimetric and multivariate methods are appropriate when:

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

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

6.3. Application of combination rules in assessments

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

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

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

There are at least four levels of combination required to move from evaluation of the individual metrics or indicators identified by the Task Groups to an assessment of GEnS (Cardoso et al., 2010). As an example, using D6 (Seafloor integrity), Figure 3 shows: (i) aggregation of metrics/indices within indicators (see names of indicators in Table 1); (ii) aggregation of indicators within the criteria of a descriptor (for complex descriptors), e.g. criteria 6.1 (physical damage) and 6.2 (condition of benthic community); (iii) status across all the criteria of a descriptor; and (iv) integration of status across all descriptors.

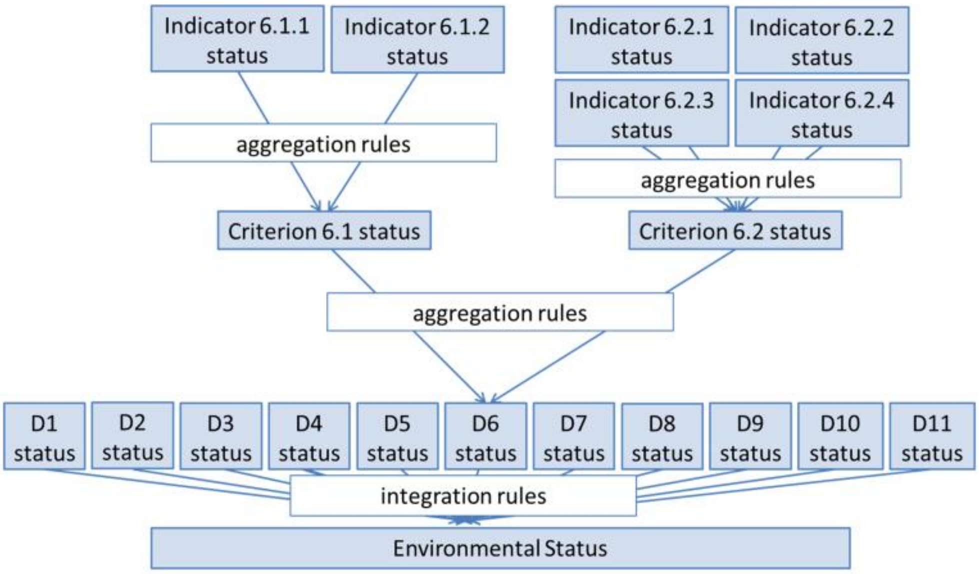

Figure 3. Diagram of a possible approach for aggregation of indicators and criteria and integration of descriptors (D), using D6 as an example. For indicators and criteria description, see Table 1.

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

6.3.1. Aggregation of indicators and criteria (combination within a descriptor)

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

Table 5 summarizes the approaches to aggregate attributes within each descriptor. In some cases the MSFD Task Groups propose deconstructing the ecosystem into 'descriptor indicators' and then recombining them again to give a pass/fail for the GEnS, using (in four cases) the OOAO principle (Table 3). Borja et al. (2013) emphasize that such a 'deconstructive structural approach' makes large assumptions about the functioning of the system and does not consider the weighting of the different indicators and descriptors. It implies that recombining a set of structural attributes gives an accurate representation of the ecosystem functioning.

Table 5. Summary of Task Group approaches to aggregate attributes within a Descriptor (Cardoso et al., 2010).

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

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

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

D6), since the metrics used could be different. This information can be converted into environmental status values, as shown in Borja et al. (2011b).

6.3.2. Integration of descriptors (combination across descriptors)

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

Borja et al. (2013) describe at least 8 options to determine GEnS in a regional sea context (see Table 3). These authors detail the concept behind these options, and propose the decision rule more adequate for the assessment method to be used, depending on the circumstances i.e. data availability, lack of monitoring, etc. In addition, these authors consider what type and amount of data are required, and then discuss the pros and cons of the different options. The implementation of a complex directive, such as the MSFD, requires a high amount of data to assess the environmental status in a robust way. Hence, the options from 1 to 8 proposed in Table 3 are sequentially less demanding of new data, and the degree of detailed environmental assessment is also decreasing.

As such, Option 1, which is most similar to the WFD approach, deconstructs GEnS into the 11 descriptors and then into the component indicators, assessing each of them for each area before attempting to produce an overall assessment (see Table 3). However, having a complete dataset covering all descriptors and indicators for the assessment is difficult, if not impossible to achieve in practical terms. The use of pressure maps as an estimator of the environmental status and possible impacts to marine ecosystems could be considered instead (see Table 3). This would, however, build on the substantial assumption that the level of pressure is adequately representing the current state on all different levels of ecosystem components. Option 7, in contrast, only uses published data for the activities, and then infers a static relationship between activity, pressures, state changes and impacts both on the natural and the human system. Here, the number of underlying assumptions is even larger than using pressure maps, since the method relies on predefined and static DPSIR relations. Between these extremes, there are several intermediate options to integrate and present information, each with its own requirements, pros and cons (see Table 3).

a. One-out, all-out (OOAO)

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

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

b. Alternative approaches

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

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

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

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

The combination across levels of different complexity should accommodate different alternatives, i.e., aggregation below descriptor level (across indicators within criteria, and criteria within descriptors, as shown in the previous section) and can certainly differ from descriptor level integration; Integration across state descriptors (D1, D3, D4, D6) should be done differently than across pressure descriptors (D2, D5, D7, D8, D9, D10, D11), but avoiding double counting of indicators in different descriptors (e.g. phytoplankton under D1 and D5, macroinvertebrates under D1 and D6).

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

c. Visualizing and communicating the status

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

As an example, the Ocean Health Index (Halpern et al., 2012) provides weighted index scores for environmental health, both a global area-weighted average and scores by country (Figure 4).

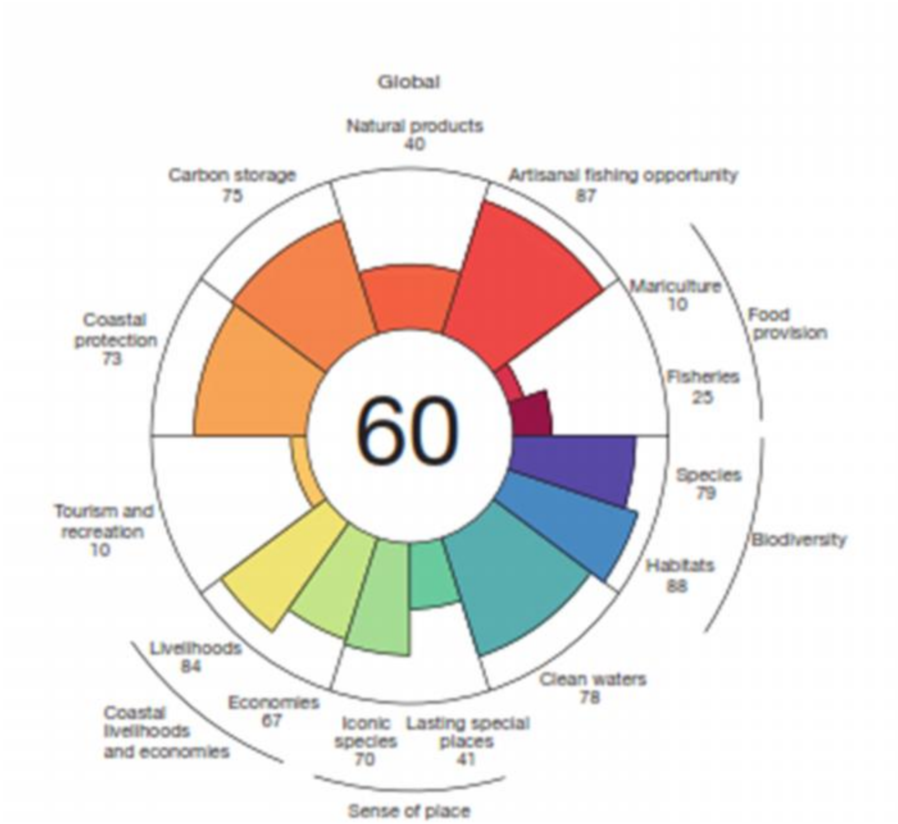

Figure 4. Ocean Health Index scores (inside circle) and individual goal scores (coloured petals) for global area-weighted average of all studied countries (modified from Halpern et al., 2012).

The outer ring of the radar scheme is the maximum possible score for each goal, and a goal's score and weight (relative contribution) are represented by the petal's length and width, respectively. This way of visualizing the integration could be adapted for the MSFD, integrating at the level of region or subregion, but also showing the values within each descriptor. This would still allow managers to extract relevant information and take actions at different levels: small (or local) scale, large (regional) scale, integrative (whole ecosystem status), or for each descriptor.

Another example, applied specifically for the MSFD, using all descriptors and most of the indicators, can be consulted in Borja et al. (2011b). These authors studied a system in which the main driver for the whole area is fishing, whilst at local level some pressures such as waste discharges are important. Although the overall environmental status of the area was considered good, after the integration of all indicators and descriptors, two of the descriptors (fishing and food webs) were not in good status (Table 6). Interestingly, biodiversity was close to the boundary to good status (Table 6), suggesting that the system could be unbalanced by fishing, but affecting various biological descriptors to different degrees. This means that the pressure must be managed to avoid problems in the future, especially because the descriptors already in less than good status showed a negative trend (Table 6).

Hence, from the examples above and the given reasoning, both main choices are still useful: either integrate or not integrate information across descriptors. Irrespectively of which combination proposal(s) is adopted and at which level, the precautionary principle should always be followed in absence of more robust knowledge (Borja et al., 2013). As a summary, the pros and cons of each decision are shown in Table 7.

Table 6. Example of an assessment of the environmental status, within the Marine Strategy Framework Directive, in the Basque Country offshore waters (Bay of Biscay) (modified from Borja et al., 2011b). EQS: Environmental Quality Standards; EQR: Ecological Quality Ratio, both based upon the Water Framework Directive (WFD); NA: not available. Trends: red colour, negative; green colour, positive (in both cases can be increasing/decreasing, depending on the indicator).

Qualitative Descriptors	Explanation of the indicators used	Reference conditions/EQS	Recent trend	Reliability (%)	Weight (%)	EQR	Final Environmental Status	Final Confidence ratio
1.- Biological diversity	integrated biological value		NA	69	15	0.51	0.08	10.35
2.- Non-indigenous species	ratio non-indigenous sp.	OSPAR	▲	80	10	0.98	0.10	8
3.- Exploited fish and shellfish			▼	100	15	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**

Table 7. Pros and cons of the decision of integrating the information across descriptors

Procedure	Pros	Cons
No integration	• Direct detection of problems (management needs) for each descriptor • Useful for local managers (close to specific or local pressures) • Reduces multiple accounting • 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 Member States and regions • Does not fulfil the main aim of marine management in an integrative way
Integration (all descriptors or a subset)	• Progress towards GEnS relevant at regional scale (comparable across regional seas and countries) • Environmental status defined in an integrative way, as health of the ecosystem (full 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 GEnS • Can mask problems from specific descriptors/pressures • May include multiple accounting • May be too subjective, as it typically involves expert judgment

6.4. Proposed steps for combination

As a possible approach for the combination of assessments we propose the following steps (Figure 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 the levels of spatial assessment that was described as a nested approach (Step 1 - spatial scales).

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

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

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

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

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

- Generally, use of OOAO (if one assessment area fails GEnS, the whole subregion fails) is not useful, as it gives a very conservative result and is not informative. Also, if the pressure is highly localized this approach is not adequate, since the whole subregion could fail GEnS due to a single location (which, of course, will need specific management measures).

- In some cases, for example if a pressure is more or less homogeneous across a whole subregion (fishing, shipping), it could be useful to apply OOAO

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

- Other metrics

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

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

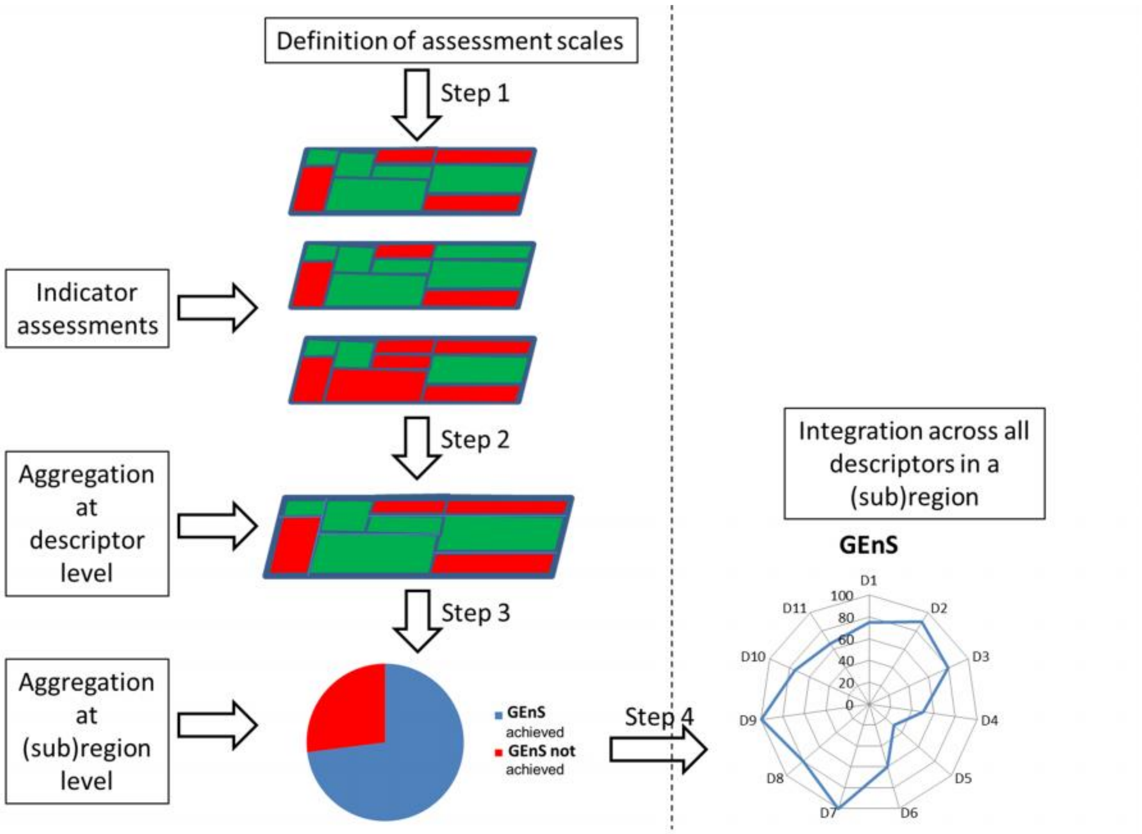

Figure 5. Schematic view of steps for combination towards an assessment at subregional level. GEnS: Good Environmental Status.

7. Conclusion

As discussed in Deliverable 6.2: (i) marine management governance initiatives have to deal with the complexity of the marine and adjoining systems and thus the initiatives become complex in themselves, as with the WFD (Hering et al., 2010); (ii) there is a plethora of such mechanisms such as the Regional Seas Conventions (such as OSPAR, HELCOM), which have been assessing and advising on marine areas since the 1970s, and (iii) it is recognised that there is limited funding for large-scale data gathering exercises. Hence, we advocate that the means of implementing the MSFD have to be simpler than the WFD, to be based on the work of the Regional Seas Conventions and as much as possible be based on existing datasets.

Above **an operational definition of GEnS is proposed** and, based on the collective experience of the authors, this can be accepted even after being refined as necessary. In order to be accepted there is the need to show that such a definition encompasses all descriptors and their criteria, and that we are aware of the data and information required (and either existing or which can be obtained) to ensure the definition is met (see Figure 6). We have made a preliminary attempt at this in Table 8. This shows that whereas we have some data covering the regional seas, we have many data based on pressures within small areas and so any large scale assessment will have to be derived by combining such data. Hence, it illustrates the fundamental challenge of arriving at a regional quality status either by having a broad approach and omitting or down-weighting point-source problems or summing the point-source problems (which may cover only a very small area) to indicate the quality status of the whole area. Furthermore, in keeping with the overall direction of the MSFD, and indeed the recently proposed Maritime Spatial Planning and Integrated Coastal Management Directive (European Commission, 2013), we emphasise in the definition and Table 8 the need for the marine environment to not only protect and enhance the nature conservation features but also to deliver ecosystem services and societal benefits.

While it is important to define what we mean by GEnS, Table 8 still emphasises the challenge of having sufficient methods and measurements to indicate when it has been achieved. While for certain elements we have a plethora of such agreed targets and limits, for example the levels of contaminants that should not be exceeded in water and seafood, for many others such limits are not yet defined. For example, while indicators have been proposed for many ecosystem services and societal benefits (e.g. see the review by Liquete et al., 2013), these are to indicate the direction of trajectories (i.e. is a feature getting better or worse) rather than a level against which successful management is judged.

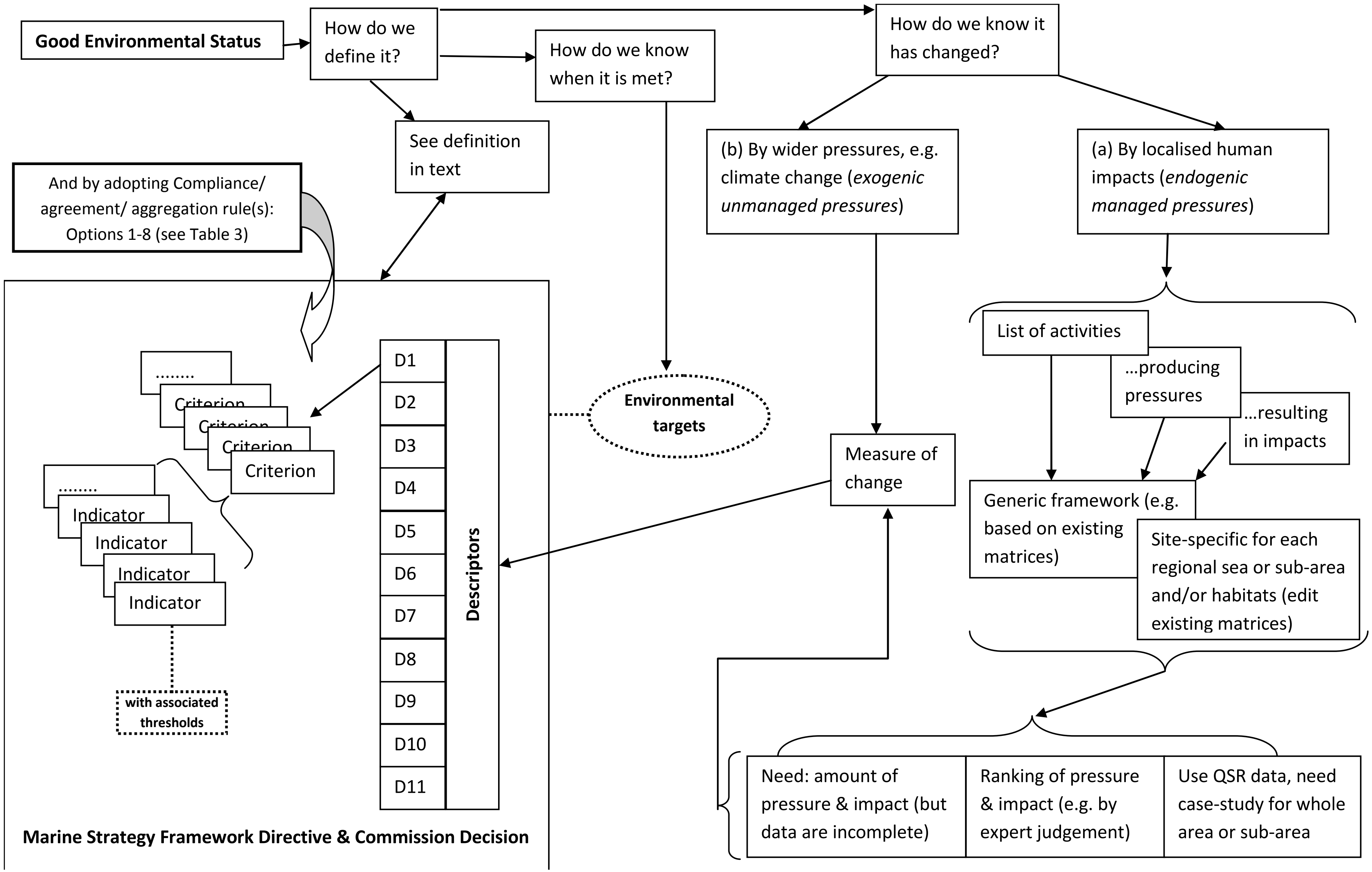

Figure 6. Making operational the definition of Good Environmental Status. QSR: Quality Standard Reports

Table 8. The operational basis of the proposed Good Environmental Status (GEnS) definition. Key: RSC: Regional Sea Conventions; VMS: Vessel Monitoring System; EUNIS: European Nature Information System; WFD: Water Framework Directive; UWWTD: Urban and Waste Water Treatment Directive; QSR: quality status reports from RSC; UKNEA: UK National Ecosystem Assessment; TEEB: The Economics of Ecosystems and Biodiversity; CBD: CBD: Convention on Biological Diversity.

'GEnS is achieved when	Related to which Descriptor?	How is this determined?	What data/information are available?	Which targets or limits to be used?
.....physicochemical (including contaminants, litter and noise)	5, 6. 8, 9, 10, 11	Rapid assessment of pressures using GIS and presented via interactive .pdf documents; traditional sampling, buoys (different sensors, including acoustic)	Use of RSC databases to indicate the quality status of the area (using QSR), summed point-source inputs of contaminants; use of VMS for open sea fishing pressure; RSC eutrophication maps	Targets established in ad-hoc Directives (e.g. UWWTD, WFD, Environmental Quality Standards Directive, etc.) or RSC
.....and hydrographical conditions are maintained at a level	1, 6, 7	Remote measurements and aerial/satellite sensing; habitat surveys; traditional sampling	Seabed maps of selected areas, modelling of current patterns, satellite data for surface conditions (waves, currents, temperature, etc).	RSC, expert judgment
.....**where the structuring components of the ecosystem are present and functioning, **.....	1,3, 4, 6	Habitat maps, habitat suitability modelling, genomics, traditional sampling, abundance estimates of key species disparate survey data combined to give larger assessments; assessments and indices of biological functioning (productivity, competition, bioengineers, etc)	Broad biotope data (EUNIS), regional characterisation from summed local surveys; mammal and bird recording systems; fish stock assessments; specific surveys of ecosystem functioning supported by literature from elsewhere	Habitats and Birds Directives targets; Common Fisheries Policy Targets; expert judgement
.....enabling the system to be resistant (ability to withstand stress) and resilient (ability to recover after a stressor) to harmful effects of human pressures/activities/impacts, ...	1, 3, 4, 5, 6, 8	Multimetric indices, functional indices, size-spectra analyses; evidence of areas or components which have recovered after stressors have been removed	National and RSC databases Evidence from case-studies and Environmental Impact Assessments and extrapolated to wider areas; Alien and invasive species databases	Adapted targets from other directives (e.g. WFD) or RSC; expert judgement

'GEnS is achieved when	Related to which Descriptor?	How is this determined?	What data/information are available?	Which targets or limits to be used?
.....where they maintain and provide the ecosystem services	1, 3, 4, 5, 6, 7,	An analysis of ecosystem services across sea areas, Contingency valuation, biological valuation and other economic valuation techniques	A modelling and linked GIS analysis of habitats and ecosystems to ecosystem services will be needed; data and information from QSR	None available; some indicators of trends (e.g. UKNEA, TEEB, CBD reports) but not of targets (see Liquete *et al.,* 2013)
.....that deliver societal benefits	3, 4, 5, 9,	Economic valuation techniques	Fisheries statistics, monitoring of seafood quality, Databases of human uses; fisheries (VMS data), oil & gas, aggregate returns, etc;	Legal limits for contaminants in seafood, fish stocks under safe limits, seabed extraction within permits, etc.
.....*in a sustainable way (i.e. that pressures associated with uses cumulatively do not hinder the ecosystem components) in order to retain their natural diversity, productivity and dynamic ecological processes,* ***.....***	1, 2, 3, 4, 5, 6	Productivity values and separation of natural from anthropogenic production; alien and introduced species are minimised	National and RSC databases; use of data from small areas extrapolated to larger areas; Alien and Invasive species databases.	Legal limits for contaminants in seafood, fish stocks under safe limits; expert judgement
..... *and where recovery is rapid and sustained if a use ceases'.*	1, 2, 3, 4, 5, 6, 8, 9, 10, 11	Traditional sampling, trend analysis, Evidence of areas or components which have recovered after stressors have been removed	Long-term monitoring series; Alien and Invasive species databases.	Trends showing a tendency towards the previous state (before pressure) (i.e. hindcasting)

Consequently, there are a number of research and management needs. As shown here, there is a continuing and pressing need for scientists and policy makers to clarify the terminology across the different policy drivers, e.g. GEnS in MSFD, Favourable Conservation Status in the Habitats Directive, Good Ecological Status in the WFD, Ecological Quality Objectives of OSPAR, or approaches within the Convention on Biological Diversity. HELCOM (and OSPAR) has discussed this in detail (see HELCOM (2013) TARGREV report).

The relationship between the implementation of the MSFD and integrated maritime spatial planning is as yet poorly defined especially as much marine management is sector-based. Hence there is the need to integrate the monitoring and assessment across the policy drivers/descriptors, etc. It is necessary to assess the costs and benefits of co-location of marine activities and their effect on attaining GEnS and on the maintenance of ecosystem services and the delivery of societal benefits (Christie et al., 2014). Thus, further research is required to produce a better understanding and more comprehensive datasets but also to concentrate on processes and cause and effect and to improve the science of monitoring especially to ask appropriate questions, determine key processes and give an interaction of components and processes.

Whichever way is used to define GEnS it will require an assessment of change against an expected standard or reference condition. While this has not yet been determined, once GEnS is determined then such a reference-deviation will be required as a management mechanism. However, given the difficulties of gathering data and the cost of producing data-rich means for defining reference conditions (i.e. comparisons with natural areas, hindcasting and predictive models), in the first instance reliance on expert judgement will be needed (Borja et al., 2012). An expert system that can capture expert judgement in a robust and defensible manner, as demonstrated across different continents (Teixeira et al., 2010), may be required.

As shown here and elsewhere, there is a plethora of indicators and targets and it is likely to be exceedingly difficult to reconcile the use of all of these, especially across the 11 descriptors—and it may well be unnecessary. They respond differently both in time and space, and might be counter-acting; although they may indicate poor quality in a small area, they will be absorbed across larger spatial scales and thus have little influence in an overall good quality eco-region. Consequently, we take the view that a combination of quantitative indicator targets and, where these are lacking, expert judgement are needed to integrate the natural and social science requirements for the sound implementation of the MSFD.

In relation to the **integrated assessment**, this should be ecologically-relevant, transparent and documented and should not only present a classification result (primary assessment) but also address uncertainties and assess confidence of the classification result (as a secondary assessment). When carrying out an assessment at a specific scale, the decisions made in regard to integration principles/rules should be available as a sort of third assessment or backlog. Moreover, assessments should be planned around the question(s) to be addressed and the tool(s) to be used. Monitoring should subsequently be designed to meet the requirements of the planned assessments.

This DEVOTES Deliverable provides information on combining methods to integrate ecosystem components to assess status and guidelines for scientists and managers on the steps to be followed, when deciding on assessment scales and combination approaches. Integration of taxonomic, functional and key or keystone biodiversity components into an overall biodiversity assessment able to link to GEnS and to ecosystem service provision and the sustainable management of detrimental human activities is the next challenge.

8. References

Atkins JP, Gregory AJ, Burdon D, Elliott M (2011a) Managing the marine environment: is the DPSIR framework holistic enough?. Systems Research and Behavioural Science 28: 497–508.

Atkins JP, Burdon D, Elliott M, Gregory AJ (2011b) Management of the Marine Environment: Integrating Ecosystem Services and Societal Benefits with the DPSIR Framework in a Systems Approach. Marine Pollution Bulletin 62(2): 215-226.

Aubry A, Elliott M (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(1-4): 175-185.

Ban N, Alder J (2008) How wild is the ocean? Assessing the intensity of anthropogenic marine activities in British Columbia, Canada. Aquatic Conservation-Marine and Freshwater Ecosystems 18: 55-85.

Barton DN, Saloranta T, Moe SJ, Eggestad HO, Kuikka S (2008) Bayesian belief networks as a meta-modelling tool in integrated river basin management - Pros and cons in evaluating nutrient abatement decisions under uncertainty in a Norwegian river basin. Ecological Economics 66: 91-104.

Barton DN, Kuikka S, Varis O, Uusitalo L, Henriksen HJ, Borsuk M, de la Hera A, Farmani R, Johnson S, Linnell JD (2012) Bayesian networks in environmental and resource management. Integrated environmental assessment and management 8: 418-429.

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

Borja A, Valencia V, Franco J, Muxika I, Bald J, Belzunce MJ, Solaun O (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, Ranasinghe A, Weisberg SB (2009a) Assessing ecological integrity in marine waters, using multiple indices and ecosystem components: Challenges for the future. Marine Pollution Bulletin 59: 1-4.

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

Borja A, Rodríguez JG (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, Elliott M, Carstensen J, Heiskanen A-S, van de Bund W (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, Barbone E, Basset A, Borgersen G, Brkljacic M, Elliott M, Garmendia JM, Marques JC, Mazik K, Muxika I, Neto JM, Norling K, Rodríguez JG, Rosati I, Rygg B, Teixeira H, Trayanova A (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, Galparsoro I, Irigoien X, Iriondo A, Menchaca I, Muxika I, Pascual M, Quincoces I, Revilla M, Germán J Rodríguez, Santurtún M, Solaun O, Uriarte A, Valencia V, Zorita I (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, Dauer DM, Grémare A (2012) The importance of setting targets and reference conditions in assessing marine ecosystem quality. Ecological Indicators 12: 1-7.

Borja A, Elliott M (2013) Marine monitoring during an economic crisis: the cure is worse than the disease. Marine Pollution Bulletin 68: 1-3.

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

Borja, Á, Marín S, Núñez R, Muxika I (2014). Is there a significant relationship between the benthic status of an area, determined by two broadly-used indices, and best professional judgment? Ecological Indicators 45: 308-312.

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

Bricker SB, Ferreira JG, Simas T (2003) An Integrated Methodology for Assessment of Estuarine Trophic Status. Ecological Modelling 169(1): 39-60.

Cardoso AC, Cochrane S, Doerner H, Ferreira JG, Galgani F, Hagebro C, Hanke G, Hoepffner N, Keizer PD, Law R, Olenin S, Piet GJ, Rice J, Rogers SI, Swartenbroux F, Tasker ML, van de Bund W (2010) Scientific support to the European Commission on the Marine Strategy Framework Directive. Management Group Report. EUR 24336 EN – Joint Research Centre, Luxembourg: Office for Official Publications of the European Communities: 57 pp.

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

Carstensen J, Conley DJ, Andersen JH, Ærtebjerg G (2006) Coastal eutrophication and trend reversal: A Danish case study. Limnology & Oceanology 51: 398-408.

Carstensen J, Sánchez-Camacho M, Duarte CM, Krause-Jensen D, Marbà N (2011) Connecting the dots: Responses of coastal ecosystems to changing nutrient concentrations. Environmental Science and Technology 45: 9122-9132.

Christie N, Smyth K, Barnes R, Elliott M, 2014. Co-location of activities and designations: a means of solving or creating problems in marine spatial planning? Marine Policy 43: 254-261.

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

Claussen U, Connor D, de Vrees L, Leppänen J, Percelay J, Kapari M, Mihail O, Ejdung G, Rendell J (2011) Common Understanding of (Initial) Assessment, Determination of Good Environmental Status (GES) and Establishment of Environmental Targets (Art. 8, 9 & 10 MSFD). WG GES EU MSFD (https://circabc.europa.eu/sd/d/ce7e2776-6ac6-4a41-846f-a04832c32da7/05_Info_Common_understanding_final.pdf)

Cochrane SKJ, Connor DW, Nilsson P, Mitchell I, Reker J, Franco J, Valavanis V, Moncheva S, Ekebom J, Nygaard K, Serrao Santos R, Naberhaus I, Packeiser T, van de Bund W, Cardoso AC (2010) Marine Strategy Framework Directive – Task Group 1 Report Biological Diversity. EUR 24337 EN – Joint Research Centre, Luxembourg: Office for Official Publications of the European Communities: 110 pp.

Crain CM, Kroeker K, Halpern BS (2008) Interactive and cumulative effects of multiple human stressors in marine systems. Ecology Letters 11: 1304-1315.

De Jonge VN, Elliott M, Brauer VS (2006) Marine monitoring: its shortcomings and mismatch with the EU Water Framework Directive's objectives. Marine Pollution Bulletin 53: 5-19.

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

Duarte CM, Conley DJ, Carstensen J, Sánchez-Camacho M (2009) Return to Neverland: Shifting baselines affect restoration targets. Estuaries and Coasts 32: 29-36.

Ehler C, Douvere F (2009) Marine Spatial Planning: a step-by-step approach toward ecosystem-based management. Intergovernmental Oceanographic Commission and Man and the Biosphere Programme. IOC Manual and Guides No. 53, ICAM Dossier No. 6. Paris: UNESCO: 99 p.

Elliott M, Burdon D, Hemingway KL, Apitz S E (2007) Estuarine, coastal and marine ecosystem restoration: Confusing management and science - A revision of concepts. Estuarine, Coastal and Shelf Science 74: 349-366.

Elliott M (2011) Marine science and management means tackling exogenic unmanaged pressures and endogenic managed pressures – a numbered guide. Marine Pollution Bulletin 62: 651-655.

European Commission (2000) Directive 2000/60/EC of the European Parliament and of the Council of 23 Octobre 2000 establishing a framework for community action in the field of water policy. Official Journal of the European Union L327: 1-72.

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

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

European Commission (2013) Proposal for a Directive of the European Parliament and of the Council establishing a framework for maritime spatial planning and integrated coastal management 34 pp.

Feary DA, Fowler AM, Ward TJ (2014) Developing a rapid method for undertaking the World Ocean Assessment in data-poor regions – A case study using the South China Sea Large Marine Ecosystem. Ocean & Coastal Management 95: 129-137.

Ferreira JG, Andersen JH, Borja A, Bricker SB, Camp J, Cardoso da Silva M, Garcés E, Heiskanen AS, Humborg C, Ignatiades L, Lancelot C, Menesguen A, Tett P, Hoepffner N, Claussen U (2010) Marine Strategy Framework Directive – Task Group 5 Report Eutrophication. EUR 24338 EN – Joint Research Centre, Luxembourg: Office for Official Publications of the European Communities: 49 pp.

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

Foley NS, van Rensburg TM, Armstrong CW (2010) The Ecological and Economic Value of Deep Water Corals. Ocean and Coastal Management 53: 313-326.

Galgani F, Fleet D, van Franeker J, Katsanevakis S, Maes T, Mouat J, Oosterbaan L, Poitou I, Hanke G, Thompson R, Amato E, Birkun A, Janssen C (2010). Marine Strategy Framework Directive – Task Group 10 Report Marine litter. EUR 24340 EN – Joint Research Centre, Luxembourg: Office for Official Publications of the European Communities: 48 pp.

Halpern BS, Walbridge S, Selkoe KA, Kappel CV, Micheli F, D'Agrosa C, Bruno JF, Casey KS, Ebert C, Fox HE, Fujita R, Heinemann D, Lenihan HS, Madin EMP, Perry MT, Selig ER, Spalding M, Steneck R, Watson R (2008) A Global Map of Human Impact on Marine Ecosystems. Science 319: 948-952.

Halpern BS, Longo C, Hardy D, McLeod KL, Samhouri JF, Katona SK, Kleisner K, Lester SE, O'Leary J, Ranelletti M, Rosenberg AA, Scarborough C, Selig ER, Best BD, Brumbaugh DR, Chapin FS, Crowder LB, Daly KL, Doney SC, Elfes C, Fogarty MJ, Gaines SD, Jacobsen KI, Karrer LB, Leslie HM, Neeley E, Pauly D, Polasky S, Ris B, St Martin K, Stone GS, Sumaila UR, Zeller D (2012) An index to assess the health and benefits of the global ocean. Nature 488: 615-620.

Harding LE (1992) Measures of marine environmental quality. Marine Pollution Bulletin 25: 23–27.

Heiskanen A-S, van de Bund WJ, Cardoso AC, Nõges P (2004) Towards good ecological status of surface waters in Europe – Interpretation and harmonisation of the concept. Water Science and Technology 49 (7): 169-177

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

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

HELCOM (2010) Ecosystem Health of the Baltic Sea 2003–2007: HELCOM Initial Holistic Assessment. Baltic Sea Environment Proceedings No. 122: 63 pp. (www.helcom.fi)

HELCOM (2013) Approaches and methods for eutrophication target setting in the Baltic Sea region. Baltic Sea Environment Proceedings No. 133. (www.helcom.fi)

Hering D, Borja A, Carstensen J, Carvalho L, Elliott M, Feld CK, Heiskanen A-S, Johnson RK, Moe J, Pont D, Solheim AL, van de Bund WJ (2010) The European Water Framework Directive at the age of 10: A critical review of the achievements with recommendations for the future. Science of The Total Environment 408: 4007-4019.

Hildering A, Keessen AM, van Rijswick FMW (2009) Tackling pollution of the Mediterranean Sea from land-based sources by an integrated ecosystem approach and the use of the combined international and European legal regimes. Utrecht Law Review 5(1): 80-100.

Johnson RK, Lindegarth M, Carstensen J (2013) Establishing reference conditions and setting class boundaries. WATERS Report no. 2013:2. Deliverable 2.1-1, 66 pp. Havsmiljöinstitutet, Sweden.

Knights A M, Koss RS, Robinson LA (2013) Identifying common pressure pathways from a complex network of human activities to support ecosystem-based management. Ecological Applications 23: 755-765.

Knights AM, Culhane F, Hussain SS, Papadopoulou KN, Piet GJ, Raakær J, Rogers SI, Robinson LA (2014) A step-wise process of decision-making under uncertainty when implementing environmental policy. Environmental Science & Policy 39 56-64.

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

Law R, Hanke G, Angelidis M, Batty J, Bignert A, Dachs J, Davies I, Denga Y, Duffek A, Herut B, Hylland K, Lepom P, Leonards P, Mehtonen J, Piha H, Roose P, Tronczynski J, Velikova V, Vethaak D (2010) Marine Strategy Framework Directive – Task Group 8 Report Contaminants and pollution effects. EUR 24335 EN – Joint Research Centre, Luxembourg: Office for Official Publications of the European Communities: 161 pp.

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

Lehikoinen A, Helle I, Klemola E, Mäntyniemi S, Kuikka S, Pitkänen H (2014) Evaluating the impact of nutrient abatement measures on the ecological status of coastal waters: a Bayesian network for decision analysis. International Journal of Multicriteria Decision Making 4(2): 114-134.

Liquete C, Piroddi C, Drakou EG, Gurney L, Katsanevakis S, Charef A, Egoh B (2013) Current Status and Future Prospects for the Assessment of Marine and Coastal Ecosystem Services: A Systematic Review. PLoS ONE 8: e67737.

Mee LD, Jefferson RL, Laffoley DdA, Elliott M (2008) How good is good? Human values and Europe's proposed Marine Strategy Directive. Marine Pollution Bulletin 56: 187-204.

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

Mouillot D, Bellwood DR, Baraloto C, Chave J, Galzin R, Harmelin-Vivien M, Kulbicki M, Lavergne S, Lavorel S, Mouquet N, Paine CET, Renaud J, Thuiller W (2013) Rare Species Support Vulnerable Functions in High-Diversity Ecosystems. PLoS Biol, 11: e1001569.

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

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

Olenin S, Alemany F, Cardoso AC, Gollasch S, Goulletquer P, Lehtiniemi M, McCollin T, Minchin D, Miossec L, Occhipinti-Ambrogi A, Ojaveer H, Rose Jensen K, Stankiewicz M, Wallentinus I, Aleksandrov B (2010) Marine Strategy Framework Directive – Task Group 2 Report Non-indigenous species. EUR 24342 EN – Joint Research Centre, Luxembourg: Office for Official Publications of the European Communities: 44 pp.

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

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

Palialexis A, Tornero V, Barbone E, Gonzalez D, Hanke G, Cardoso AC, Hoepffner N, Katsanevakis S, Somma F, Zampoukas N (2014) In-depth assessment of the EU Member States' submissions for the Marine Framework Strategy Framework Directive under articles 8, 9 and 10 JRC Scientific and Technical Reports. Publications Office of the European Union.

Pascual M, Borja A, Eede SV, Deneudt K, Vincx M, Galparsoro I, Legorburu I (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.

Pascual M, Borja A, Franco J, Burdon D, Atkins JP, Elliott M (2012) What are the costs and benefits of biodiversity recovery in a highly polluted estuary? Water Research 46: 205-217.

Piet GJ, Albella AJ, Aro E, Farrugio H, Lleonart J, Lordan C, Mesnil B, Petrakis G, Pusch C, Radu G, Ratz HJ (2010) Marine Strategy Framework Directive – Task Group 3 Report Commercially exploited fish and shellfish. EUR 24316 EN – Joint Research Centre, Luxembourg: Office for Official Publications of the European Communities: 82 pp.

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

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

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

Rogers S, Casini M, Cury P, Heath M, Irigoien X, Kuosa H, Scheidat M, Skov H, Stergiou KI, Trenkel VM, Wikner J, Yunev, O (2010) Marine Strategy Framework Directive – Task Group 4 Report Food Webs. EUR 24343 EN – Joint Research Centre, Luxembourg: Office for Official Publications of the European Communities: 55 pp.

Shin Y-J, Bundy A, Shannon LJ, Blanchard JL, Chuenpagdee R, Coll M, Knight B, Lynam C, Piet G, Richardson AJ, (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, Zenetos A, Pancucci-Papadopoulou MA, Reizopoulou S, Streftaris N (2012) Indicators for the Sea-floor Integrity of the Hellenic Seas under the European Marine Strategy Framework Directive: establishing the thresholds and standards for Good Environmental Status. Mediterranean Marine Science 13: 140-152.

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

Swartenbroux F, Albajedo B, Angelidis M, Aulne M, Bartkevics V, Besada V, Bignert A, Bitterhof A, Hallikainen A, Hoogenboom R, Jorhem L, Jud M, Law R, Licht Cederberg D, McGovern E, Miniero R, Schneider R, Velikova V, Verstraete F, Vinas L, Vlad S (2010) Marine Strategy Framework Directive – Task Group 9 Report Contaminants in fish and other seafood. EUR 24339 EN – Joint Research Centre, Luxembourg: Office for Official Publications of the European Communities: 36 pp.

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

Tasker ML, Amundin M, Andre M, Hawkins A, Lang W, Merck T, Scholik-Schlomer A, Teilmann J, Thomsen F, Werner S, Zakharia M (2010) Marine Strategy Framework Directive – Task Group 11 Report, Underwater noise and other forms of energy. EUR 24341 EN – Joint Research Centre, Luxembourg: Office for Official Publications of the European Communities: 55 pp.

Teixeira H, Borja A, Weisberg SB, Ananda Ranasinghe J, Cadien DB, Dauer DM, Dauvin J-C, Degraer S, Diaz RJ, Grémare A, Karakassis I, Llansó RJ, Lovell LL, Marques JC, Montagne DE, Occhipinti-Ambrogi A, Rosenberg R, Sardá R, Schaffner LC, Velarde RG (2010) Assessing coastal benthic macrofauna community condition using best professional judgement - Developing consensus across North America and Europe. Marine Pollution Bulletin 60: 589-600.

Ten Brink BJE, Hosper SH, Colijn F (1991) A quantitative method for description and assessment of ecosystems: the AMOEBA-approach. Marine Pollution Bulletin 23: 265-270.

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

Tueros I, Borja A, Larreta J, Rodríguez JG, Valencia V, Millán 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.

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

UNCLOS (1982) United Nations Convention on the Law of the Sea, signed at Montego Bay, Jamaica, on 10 December 1982, 202 pp. www.un.org/Depts/los/index.htm.