JNCC Report No: 536

Reviewing, refining and identifying optimum aggregation methods for undertaking marine biodiversity status assessments

Barnard, S. & Strong, J.

UNIVERSITY OF Hull
Institute of Estuarine and Coastal Studies

December 2014

ISSN 0963 8901

For further information please contact:

Joint Nature Conservation Committee
Monkstone House
City Road
Peterborough PE1 1JY
www.jncc.defra.gov.uk

This report should be cited as:

Barnard, S & Strong, J. 2014. Reviewing, refining and identifying optimum aggregation methods for undertaking marine biodiversity status assessments. JNCC Report **No. 536**. The Institute of Estuarine and Coastal Studies, University of Hull report for JNCC Peterborough.

This project report has been reviewed by a Project Steering Group consisting of experts from the Statutory Nature Conservation Bodies.

Executive Summary

This project identifies, reviews and refines aggregation methods for undertaking marine biodiversity status assessments. For the purposes of this report, aggregation (also referred to as ‘integration’ in some literature) is defined as “any rule or rules which exist to standardise the bringing together of data at different spatial or temporal scales, or across different ecosystem components or aspects of the assessment”.

The appropriate aggregation of information is central to most marine biodiversity status assessment methods. As the choice of aggregation rules can have a significant effect on an assessment outcome, the selection of an appropriate aggregation method represents an important part of developing a robust assessment process. In addition, the potential for harmonisation (where information can be used across multiple obligations to support a coordinated range of separate assessment and reporting cycles) is also an important factor in the selection of an aggregation method.

An initial review of literature covering aggregation methods applied to marine biodiversity assessments, along with a review of key references from within the social sciences, has been undertaken to identify key fundamental methods or learning points. This review supports the identification and description of a series of aggregation methods, and allows an assessment to be made of their data requirements and their relative benefits and limitations. It is noted that the majority of identified references relate to the development of composite indicators (composite indices, CIs), however many of the principles that underpin the development of CIs are applicable across other forms of aggregation. The review of references from the social sciences identifies a series of learning points relating to (amongst other things): transformation of variables; application of normalisation techniques; derivation of weighting values; and output validation.

A range of aggregation methods are described and their associated benefits and limitations are reported. This is done with reference to a quality framework covering four main areas (or performance assessment criteria): the basis for the aggregation; the value of the aggregation; the aggregation process itself; and the subsequent application of methods or data.

The information requirements of different aggregation methods are reviewed and the ability of fundamental aggregation methods to handle each of a range of data types (as described by a broad classification system) is assessed. In addition, the availability of suitable data to support aggregation needs is discussed and a number of potential data gaps identified.

Three practical scenarios are described where use is made of aggregation methods in reporting marine biodiversity assessments against UK or European legislative drivers. These three scenarios included:

(i) the annual assessment and reporting of a single (complex) habitat type (e.g. Annex I reef) at the scale of a single Marine Protected Area and based on the principles of Common Standards Monitoring;
(ii) the assessment of multiple species at the biogeographic (regional sea) scale, reporting every six years (for example under the Marine Strategy Framework Directive or reporting for OSPAR); and
(iii) the assessment of a single, highly mobile species (harbour porpoise *Phocoena phocoena*) at the biogeographic scale (employing the Article 17 aggregation approach used for Habitats Directive reporting).

Assessment of these scenarios suggests that, overall, the existing aggregation structures within the scenarios are fit for purpose and do not provide any significant barriers to the harmonisation of biodiversity monitoring. Although the three scenarios are quite tightly defined, some

'flexibility' in the detail of the approaches used to support aggregation is identified. The consequences of varying the underlying aggregation structures that are adopted are examined, and the consequential need for more structured and detailed process descriptions and aggregation rules for particular scenarios are highlighted. In addition it is suggested that consideration should be given to addressing issues arising from where organisational frameworks give rise to potential spatial discrepancies between monitoring areas and reporting boundaries.

The report concludes with a series of recommendations on future work required to improve the availability of information to support aggregation methods, the development of new aggregation methods, and how the concept of harmonised, aggregated marine biodiversity status assessments might be achieved.

Table of Contents

Optimum aggregation methods for marine biodiversity status assessments

1. Introduction

1.1 Overview

The appropriate aggregation of information is central to most marine biodiversity status assessment methods (e.g. Borja *et al* 2013; Ojaveer & Eero 2011). As the aggregation rules that are used in any given assessment process can have a significant effect on the assessment outcome (e.g. Ojaveer & Eero 2011), the selection of an appropriate aggregation method is an important part of building a robust assessment process. Furthermore, the need to move towards an assessment and reporting cycle where information can be used across multiple obligations means that the concept of harmonisation is also an important factor in the selection of an aggregation method.

To date, marine biodiversity status assessments in the UK have been responsive to particular reporting requirements, resulting in peaks and troughs in resource requirements and hence the need for frequent changes in staff and financial resources. As such, the aspiration is to develop assessment approaches where the results under one marine biodiversity assessment requirement can be easily disaggregated into its component parts so that the information can be reaggregated to meet the needs of a different requirement.

This project aims to review, refine and identify the optimum aggregation methods for undertaking marine biodiversity status assessments. In delivering this aim the project will also identify the resolution of information needed to underpin harmonised assessments of the status of marine biodiversity in the UK.

The project was developed to address a series of objectives and to deliver against a number of associated contributory tasks:

Objective 1: To identify and summarise aggregation methods from environmental assessments and the wider literature that are applicable to marine biodiversity status assessments for all five types of aggregation.

Objective 2: Review, through a critical analysis, the benefits and limitations of the selected aggregation methods identified under Objective 1.

Objective 3: Determine and describe the information requirements of the selected aggregation methods selected under Objective 1, and assess the extent to which these information requirements are met by the existing marine information and evidence in the UK.

- ***Task 3a*** *will determine and describe information requirements of different identified aggregation methods.*
- ***Task 3b*** *will identify commonalties in the information requirements between different aggregation methods.*
- ***Task 3c*** *will assess the extent to which the information requirements of the different aggregation methods are met by the existing marine information and evidence in the UK.*

Objective 4: Develop three conceptual scenarios for undertaking marine biodiversity status assessments, including recommendations on the optimum aggregation methods, the interdependencies between aggregation types and the associated information requirements.

- ***Task 4a*** *will develop and finalise three conceptual scenarios for undertaking marine biodiversity status assessments.*
- ***Task 4b*** *will explore and describe the extent to which the optimum aggregation methods might predetermine or bias the assessment outcome.*

- *Task 4c will explore and describe the interdependencies and/or relationships between the five fundamental assessment types previously identified by JNCC within each conceptual scenario, and between conceptual scenarios.*

Objective 5: Provide detailed recommendations on future work that would be required to improve the availability of information to support aggregation methods, develop new aggregation methods, if necessary, and how the concept of harmonised, aggregated marine biodiversity status assessments could be achieved.

1.2 Terminology

Aggregation is defined here as "any rule or rules which exist to standardise the bringing together of data at different spatial or temporal scales, or across different ecosystem components or aspects of the assessment". Aggregation is referred to in some literature as 'integration'. Whilst the term 'aggregation' is used preferentially here, any use of the term 'integration' is (unless otherwise noted) applied in a synonymous context.

JNCC has identified five different types of aggregation that could be used when undertaking an assessment of the state of biodiversity (see Box 1).

Box 1. JNCC aggregation types

Type 1: Aggregation over different spatial scales or assessment units (i.e. assessment areas, regions, water bodies, Marine Protected Areas). *Combines data taken from different locations, and/or combines data collated at different spatial resolutions.*

Type 2: Aggregation across different temporal scales (e.g. classifying continuous data, comparisons of annual versus continuous data). *Combines data from different temporal scales (e.g. annual data and continuous recording) and may require derivation of mean (annualised) values from long term datasets. Strictly refers to situations were multiple (temporal) data for a given feature at a given location are aggregated to provide a single indicator value (e.g. a representative annual statistic).*

Type 3: Aggregation across biodiversity components (i.e. across species, across functional groups, across habitats, within functional groups, within habitats). *Combines data on (for example) separate species or habitats to derive an overarching indicator. Would include aggregation for Annex I reef where there is a need to combine information on biogenic reef, stony reef and bedrock reef.*

Type 4: Aggregation across indicators/indices within assessment criteria (e.g. Common Standards Monitoring, CSM, for marine features). *Under CSM, for example, multiple attributes under the 'structure & function' parameter need to be combined whilst, for the 'future prospects' parameter, estimated future states of the 'range', 'area' and 'structure & function' parameters need to be combined.*

Type 5: Aggregation across assessment criteria of different biodiversity components to deliver an overall assessment (e.g. for delivering Good Ecological Status for the Water Framework Directive[1], Good Environmental Status for the Marine Strategy Framework Directive[2] or Favourable Conservation Status for the Habitats Directive[3]). *This aggregation type would apply, for example, to the integration of multiple composite indicators (for example, combining an assessment for birds with an assessment for (one or more) highly mobile species).*

[1] The Water Framework Directive (European Council Directive 2000/60/EC)
[2] The Marine Strategy Framework Directive (European Council Directive 2008/56/EC)
[3] The Habitats Directive (European Council Directive 92/43/EEC)

In the context of this report the information available for assessing biological systems is considered to be available, or reported, at three different (hierarchical) levels:

- The lowest level is represented by 'indicators' (I), observation or measurement of parameters (such as species presence/absence, counts, contaminant concentrations, *etc*.) – in this context, indicators may be fully quantitative (e.g. species counts), or semi-quantitative (e.g. classes of counts, from 'Low' to 'High' where values have defined quantitative meanings);
- at the next level, indicators are brought together as 'composite indicators' (CI); CIs may also be termed 'indices'; CIs are likely to be themed (for example an Index of Biotic Integrity);
- finally, at the highest level, are 'assessment indices' (AI) – instances where a set of more than one CI are combined to a single value to support wide scale assessment or reporting requirements.

See Figure 1.1, below, for a generic illustration of these relationships.

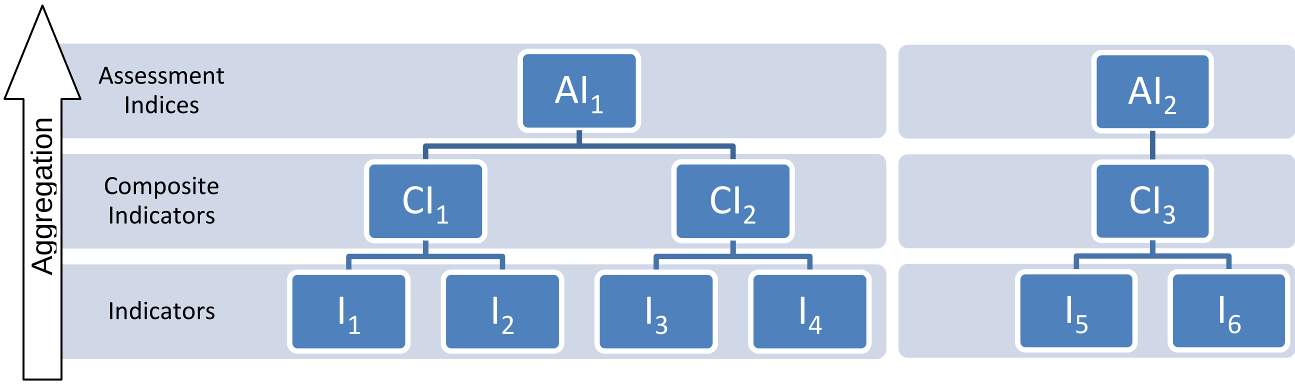


Figure 1.1. Illustration of aggregation relationships between indicators, composite indicators and assessment criteria (see text for details).

Whilst aggregation of information from a lower to a higher level may be either: from indicator to composite indicator (I ⇨ CI); or from composite indicator to assessment index (CI ⇨ AI), there is also the possibility of aggregation direct from indicator to assessment Index (although, in practice, this aggregation is effectively achieved simply by reporting a CI as if it were an AI).

In general terms, CIs are used for management purposes and for reporting at the local (site) scale. In contrast, AIs tend to be used for higher level policy-driven reporting.

It should be noted that the production of CIs is not limited to quantitative data; they may also be produced from qualitative indicator data. Furthermore, it may sometimes be appropriate to consider a hierarchy of CIs (as is the case with the Marine Strategy Framework Directive (MSFD) (2008/56/EC)). Under the MSFD, the reporting of Good Environmental Status (GEnS) will be achieved through the production of an AI that draws together information from each of the 11 descriptors set out in the Directive (biological diversity, non-indigenous species, *etc*.). The Commission Decision of September 2010 on criteria and methodological standards on GES of marine waters (2010/477/EU) describes the criteria and indicators for each MSFD descriptor for which Member States must develop suitable operational indicators and targets in order to assess GEnS. Each descriptor, therefore, is essentially a CI constructed from information on the criteria (e.g. species distribution, population size, *etc*.), which themselves are CIs constructed from the indicators (e.g. distributional range, area covered by the species, *etc*.).

Methods for aggregation are not specific (or restricted) to one or other of the two levels of the process, but may be applied at either level (i.e. they can be used for aggregating across indicators or across composite indicators).

These aggregation levels can be related to the five aggregation types identified by JNCC (Table 1.1). As can be seen, whilst four of the five aggregation types relate to the production of AIs, one (Type 3) relates solely to the production of CIs.

Table 1.1. Aggregation types and levels.

Aggregation type (*sensu* JNCC)	Aggregation level(s) I = Indicator CI = Composite Indicator AI = Assessment Index
Type 1	I ⇨ CI and/or CI ⇨ AI
Type 2	I ⇨ CI and/or CI ⇨ AI
Type 3	I ⇨ CI
Type 4	CI ⇨ AI
Type 5	CI ⇨ AI

The bases for aggregation are examined more closely in subsequent sections of this report including, for example, a consideration of the differing requirements of aggregations over space, over time and over indicators.

2. Understanding Aggregation: needs and application

2.1 Introduction

This section, together with Section 3, addresses Objective 1, identifying and summarising aggregation methods that are applicable to marine biodiversity status assessments.

It is important to recognise that aspects of spatial data aggregation from various information sources are not specific to marine biodiversity assessments but are in fact used extensively in many other disciplines. All data aggregation methods must be thoroughly understood so that input information is compatible for summarisation (incorporating, for example, combined analyses of semi-quantitative data and values collected with differing standardising units), that output statistics are faithful to the input data and that opportunities for spin-off assessments of confidence, fidelity and power are not missed.

In order to capture these various aspects, a three pronged approach has been employed:

(i) The role of, or need for aggregation within the reporting requirements of the main EU legislation pertinent to marine biodiversity assessments is summarised;
(ii) the use of composite indicators in economics and the social sciences[4] is outlined; and
(iii) instances of aggregation of reporting indicators or indices from the peer reviewed literature are summarised.

From these elements a summary overview of the process of aggregation is presented and key methodologies (and supporting studies) identified. References to where aggregation or integration within marine biodiversity assessments has been addressed within the 'grey' literature have also been noted and have been used to inform the study as appropriate.

2.2 Biodiversity assessment and reporting: UK obligations

There is an inherent need for aggregation in support of marine biodiversity assessments in the UK. In part, these needs arise from a range of reporting obligations under various instruments of European Union (EU) legislation, International Conventions, UK legislation, and UK or EU policy or policy instruments. In some cases such obligations require the application of aggregation methods even though they may not be explicitly described in the associated documentation. These reporting obligations include:

- EU legislation:
  - MSFD - Marine Strategy Framework Directive (2008/56/EC)
  - HD - Habitats Directive (92/43/EEC)
  - BD - Birds Directive (2009/147/EC)
  - WFD - Water Framework Directive (2000/60/EC)

- International Conventions:
  - CBD - Convention on Biological Diversity
  - OSPAR - Convention for the protection of the marine environment of the North East Atlantic
  - CMS - Convention on the Conservation of Migratory Species of Wild Animals (Bonn Convention)
  - UNCLOS - United Nations Convention on the Law of the Sea

[4] There is a wealth of information regarding the construction and use of composite indicators available from the fields of economics and social sciences, and aggregation to support high level reporting in these fields appears to be widely used and accepted. It is likely that review of this information will help inform thinking around the topic of aggregation to support marine biodiversity assessments.

- UK Legislation:
  - WCA - Wildlife and Countryside Act (1981)
  - CSA - Conservation of Seals Act (1970)
  - MCAA - Marine and Coastal Access Act (2009)
  - MSA - Marine Scotland Act (2010)

- Policies / Policy Instruments (UK & EU)
  - HLMO - High Level Marine Objectives (2009)
  - Gov HBDS - Government's vision for Healthy and Biologically Diverse Seas (HBDS) (2002)
  - MPS - Marine Policy Statement (2011)
  - EUBS - European Union Biodiversity Strategy (2011)

The need to aggregate data or information can be problematic as the method of aggregation can influence the subsequent assessment outcome. This could potentially lead to a situation whereby, although Member States may assess marine biodiversity against the same Directive, the resultant answers may be different depending on the particular aggregation method used. Therefore, there is a clear need to better understand the nature and application of different aggregation methods and to identify optimum techniques and approaches.

In addition it should be recognised that aggregation to support different obligations can operate at a range of scales, with the consequence that aggregated assessments derived to meet one obligation may not necessarily be able to be directly transposed to inform another obligation. An overview of the underlying obligations that drive the need for the application of aggregation methods, and a summary of the aggregation needs and types associated with each line of reporting (based on information presented in Hinchen 2013), is provided as Table A.1 in Appendix A.

2.3 Composite indicators in economics and the social sciences: their construction, use and development

Economists and social scientists have invested significant effort in developing the use of both composite indicators and (high level) assessment indices. Examples include the index of sustainable economic welfare, the genuine progress indicator, the human development index and the net national product. Indicators with environmental components include the pollution index, the unified global warming index and the index of captured ecosystem value.

Approaches to data aggregation taken from the economic and social sciences have a relevance to marine environmental and biodiversity aggregations for a number of reasons. In particular:

- They make regular use of considered, generic frameworks and construction tools, e.g. partial order analysis, for the structuring of aggregations (a process which is not as well developed in marine environmental assessments);
- the advanced nature of aggregation within this discipline provides examples of best practice that might benefit marine biodiversity assessments;
- the breadth of use within the social sciences highlights novel data integration methods from diverse sources; and
- the quality assessments generated in some social science case studies which assess differing aggregation methods and variable data use within aggregation types, are equally applicable to marine data aggregations.

One additional, and valuable, contribution from the economic and social sciences regarding the use of composite indicators relates to the development of quality assessment frameworks. Assessing quality is a difficult but important ancillary task that supports the use of composite indicators. The use of a quality assessment framework provides an objective basis for assessing the benefits and limitations of different aggregation methods. The development and application of such a framework for use in this project is discussed further under Section 3.

Five examples or reviews of approaches to aggregation taken from the economic and social sciences are presented as Appendix B, together with an indication of how methods may be applicable to the biological sciences in general and marine biodiversity assessment in particular. The following selected examples are presented:

- The OECD (Organisation for Economic Co-operation and Development) construction framework for composite indicators (Nardo *et al* 2005);
- the Human Development Index (HDI) (United Nations Development Programme, various years) and the inequality-adjusted HDI (Neumayer 2001);
- a multifunctional land use policy composite indicator (Paracchini *et al* 2011);
- the aggregation of environmental-economic indicators for 12 OECD countries (Bergh & Veen-Groot 1999); and
- the New Zealand eco-efficiency composite indicator (Jollands *et al* 2003).

This restricted assessment highlighted some useful learning points and methods from the economic and social sciences which, although they may not currently be used for marine biodiversity status assessments, are nevertheless applicable. It also provided information on quality dimensions that can be used to underpin the review of the benefits and limitations of selected aggregation methods.

Nardo *et al* (2005) provide a valuable framework for the construction of composite indicators and ways in which the quality of the existing aggregations can be judged. The development of a construction framework provides us with both a useful terminology and structure for aggregation. Whether for marine biodiversity assessments or social science applications, it is apparent that the same methodological components are used in different aggregations. The structure provided by Nardo *et al* (2005) facilitates the deconstruction of existing methods, stripping away much of the apparent complexity and allowing the identification of common aggregation components. Furthermore, this structure supports the design of new aggregations with a consistency of structure, terminology and approach. These issues are equally relevant for both social science applications and marine environmental assessments.

The HDI (United Nations Development Programme, various years) is a basic and widely reported composite indicator. The simple construction of this composite indicator provides transparency and great interpretability. These attributes are of equal validity and importance both in the social science and marine environmental aggregations. Also, the use of rankings, although reducing the information content in the output, greatly increases the confidence in the overall product. This study provides evidence of the interplay between aggregation design, information representation and overall confidence. Assessments of confidence are increasingly important in marine data representation, whether this be habitat maps or biodiversity aggregations. The discussion by Neumayer (2001) regarding the inclusion of environmental indicators into the aggregation highlights important considerations when constructing and modifying composite indicators. These include, for example, understanding and substantiating the mechanism between indicator inclusion and response of the ultimate aggregation, i.e. the critical requirement for all composite indicators to have a proven relationship between indicators and the ultimate aim of the composite indicator. Neumayer also highlights the fact that new indicators cannot be added to a composite indicator unless the dynamics of response are understood.

Paracchini *et al* (2011) and Jollands *et al* (2003) both provide useful case studies on the calculation and application of different weighting schemes, whilst the review by Bergh and Veen-Groot (1999) clearly identifies the composite phases undertaken at most of the vital stages of the aggregation process.

The selection and processing of indicators, and appropriate methods for their normalisation, weighting and aggregation, represent key phases common to the production of all robust composite indicators. This consideration of five examples from economics and the social sciences has highlighted some of the differing methods applied during these phases.

2.4 Examples of aggregation from the wider literature

A series of literature searches were undertaken with the aim of identifying aggregation methods that have been applied to aspects of marine biodiversity assessment. Details of these searches are presented in Appendix C as Table C.1. Although a relatively large number of papers were initially identified, subsequent review identified only a small proportion as being of potential value to the project. Given the relatively poor returns from this formalised search process, the references that had been identified were used as the basis for a recursive ‘paper chase’ search process whereby referenced papers would be followed up. By repeating this recursive process it was hoped that the majority of key, relevant publications would be identified.

Initially, the online literature searches, plus additional *ad hoc* searches (‘paper chase’, *etc*.) yielded 80+ papers. These were briefly reviewed to assess the range of aggregation methods that have been applied, and to provide an indication of the potential benefit in carrying them forward to Phase 2 of this project.

Given the nature and scope of the papers that were identified it was felt likely that most of the relevant material had been collated. However, where further references were identified or became available during subsequent phases of this project, reasonable efforts were made to incorporate them.

2.5 Overview of aggregation described in the literature

The following table (Table 2.1) identifies a number of references that provide material (e.g. some discussion of the process of aggregation) that may be of direct value to this phase of the project. It provides a brief synopsis indicating the value (or otherwise) of each paper and shows, where appropriate, the nature or application of any aggregation methodology that is outlined. References that have that have not been carried forward in their own right but which nevertheless may provide useful insights and might help inform subsequent work are presented in Appendix C as Table C.2.

Table 2.2. Selected references discussing aggregation.

		Aggregation across:				
Reference	**Background to methods, approaches, *etc.***	**1. different spatial scales or assessment units**	**2. different temporal scales**	**3. different biodiversity components**	**4. different indicators/indices**	**5. different assessment criteria of biodiversity components**
ALDEN *et al* (2002)	Verification of benthic index of biotic integrity (B-IBI) developed for Chesapeake Bay. Although some single metrics performed as well as the multi-metric B-IBI, it provided stable 'weight of evidence' instilling confidence in predictions.			X		
ALDER *et al* (2010)	Indices of ecosystem management from across 53 countries are aggregated as an unweighted average. Issues of deriving weights are discussed.				X	
AUBRY & ELLIOTT (2006)	Presents a nine point scale for each component of an Environmental Integrative Indicator (EEI) aggregated weighted arithmetic mean.			X		
BASSET *et al* (2012)	Discusses development of multi-metric index of size spectra sensitivity (ISS) to differentiate between impacted and unimpacted conditions based on benthic macro-invertebrate populations within 12 Mediterranean and Black Sea transitional water bodies (i.e. coastal lagoons). Metrics aggregated as weighted average sensitivity across size classes.			X		
BORJA *et al* (2009a)	Integrative approach uses weighting by area and a rule-based 'decision-tree' to combine spatial data and provide an assessment of the integrative ecological status.	X				X
BORJA *et al* (2013)	Discusses range of options for aggregating across indicators and criteria in assessing GEnS under the MSFD.					X
BROOKS *et al* (2009)	Stream–Wetland–Riparian (SWR) index derived as aggregation of (0-1) scores for separate metrics using averaging. Applied to assessing condition of aquatic ecosystems in small watersheds in US.			X		
CARONI *et al* (2013)	Examined different aggregation methods using simulated data for Swedish surface water bodies (used One Out, All Out (OOAO) and averaging).			X		

Reference	Background to methods, approaches, *etc.*	Aggregation across: 1. different spatial scales or assessment units	2. different temporal scales	3. different biodiversity components	4. different indicators/indices	5. different assessment criteria of biodiversity components
CERTAIN *et al* (2011)	Produces set of scaled indicators (S.*ijktn*). Once calculated, they can be averaged across any of their axes *i*, *j*, *k*, or *t* (indicator *i*, major ecosystem *j*, spatial unit *k*, or date *t*) or any combination of axes. Allows for weighting. Uses three possible models for scaling metrics.	X	X	X		
DALTON *et al* (2012)	Marine Protected Area (MPA) performance in the Caribbean: calculated log-response ratios (lnRR) for each distinct geomorphic zone within an MPA, where: lnRR = ln (no take area/control), and these values were then averaged (no weighting) to compute an overall lnRR for the MPA. This approach was undertaken for three separate ecological measures of MPA conditions – NOT aggregated across these indices. A visual assessment (by box plot) was made of the values for each of the three indices as calculated for all geomorphic zones.	X				
HALPERN *et al* (2012)	Methodology for weighting indices. Overall combination by weighted sum – weightings.				X	X
HARRISON & KELLY (2013)	Used 14 metrics describing four fish community attributes. Each metric scored 1-5 according to degree of deviation from a reference condition, with final index calculated as sum of metric scores.				X	
JOLLANDS *et al* (2004)	Uses Principal Components Analysis (PCA) to aggregate.			X		
LLANSÓ *et al* (2002)	Combined several metrics of benthic community structure and function into a single index - approach based on B-IBI from Chesapeake Bay. Aggregation by averaging the scores of the individual metrics (range: 1–5).			X		
OJAVEER & EERO (2011)	Uses six different indicator aggregation methods, pulling together 142 indicators within both hierarchical and non-hierarchical (flat) structures.			X		
PAUL *et al* (2001)	Selected 3 from 48 metrics (based on attributes of the macrobenthos) for combination into a benthic index of estuarine condition for the Virginian Biogeographic Province (US), using linear discriminant analysis to combine the metrics into an index.			X		

Reference	**Background to methods, approaches, *etc.***	**Aggregation across:** 1. different spatial scales or assessment units	2. different temporal scales	3. different biodiversity components	4. different indicators/indices	5. different assessment criteria of biodiversity components
PRIMPAS *et al* (2010)	Used PCA to remove inter-correlation between variables (e.g. nutrients and phytoplankton biomass or diversity), then used first Principal Component as eutrophication index.			X		?
RODRIGUEZ-RODRIGUEZ & MARTINEZ-VEGA (2012)	Terrestrial Spanish Protected Areas (PA): 43 indicators selected from a long list of 105 that might be meaningful for PA effectiveness. Indicators scaled to 0, 1 or 2 and weighted (1, 1.5 or 2) - 43 indicators integrated into 6 indices.			X		
SIMBOURA *et al* (2005)	Applies One Out, All Out (OOAO) rule to three indices.				X	
SKARPAAS *et al* (2012)	Norwegian Nature Index (NI): discusses flat- and weighted-aggregation and uses Monte Carlo simulation to derive values from distributions with means/quartiles matching supplied data.			X		X
STODDARD *et al* (2008)	Discusses the development of a Multi-Metric Index (MMI) for sites in freshwater aquatic surveys in western US. Metrics identified on a number of factors including metric type (classification), range of metric values, reproducibility (based on signal to noise ratio), adjustment for natural gradients, responsiveness and possible redundancy. All metric values were scaled linearly (0-10) and the final MMI for a site is calculated as the sum of its scored metrics (re-scaled to 0-100).				X	
WEISBERG *et al* (1997)	Benthic Index of Biological Integrity (B-IBI) developed for macrobenthos in Chesapeake Bay. Metric scores were combined into an index by calculating the mean score across all metrics for which thresholds were developed.				X	
ZHOU *et al* (2006)	Compares three aggregation methods (Simple Additive Weighting (SAW), Weighted Product (WP) and World Development Indicator (WDI)) and uses Shannon-Spearman measure to compare them on the basis of information loss. It suggests that WP may be most appropriate.	X				

In terms of literature dealing specifically with marine biodiversity, the distribution of references over the different scales or types of aggregation (*sensu* JNCC) is shown below (Figure 2.1).

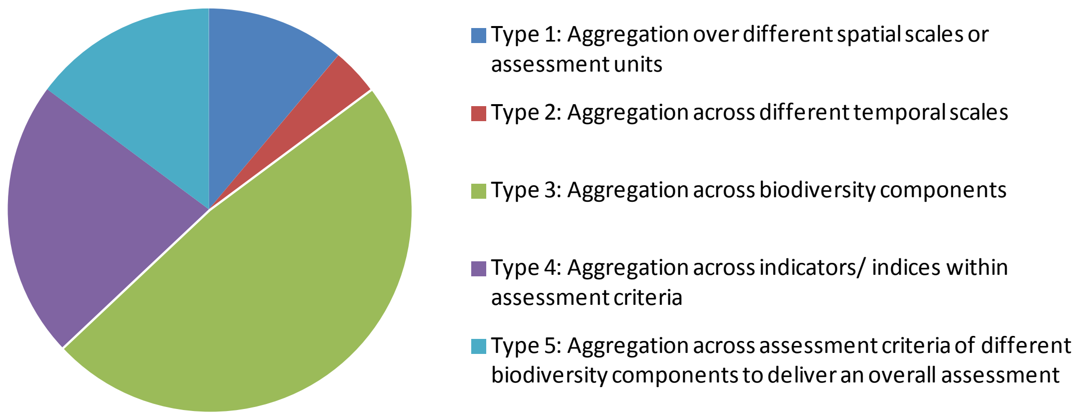


Figure 2.2. Relative frequencies of different aggregation types within identified literature.

From this initial review it is apparent that, with reference to the topic of data aggregation, the majority of published studies relate to the development of Type 3 aggregations, reflecting the production of CIs (composite indicators or indices). Whilst this, in some instances (e.g. the production of various indices of benthic integrity), might be considered a trivial form of aggregation, many of the principles that underpin the development of CIs are applicable across other forms of aggregation and it is likely that some sound approaches can be determined from them.

2.6 The use of composite indicators

The application of aggregation methods to produce CIs potentially has a number of associated pros and cons (see Box 2). More fundamentally, there are opposing ideological views surrounding whether or not indicators should be aggregated (to produce CIs) or left separate and presented in a non-aggregated form. The former view identifies two major benefits associated with the combination of indicators to produce a bottom line: (i) summary statistics are meaningful and can capture reality, and (ii) stressing the bottom line is often extremely useful in garnering media interest and hence the attention of policy makers. The latter view, that of the non-aggregators, identifies a key objection to aggregation, which surrounds what they see as the arbitrary nature of the weighting process by which the variables are combined (OECD & JRC 2008, citing Sharpe 2004).

The use of composite indicators makes it easier to summarise complex, multi-dimensional systems. It represents a means by which information can be quantified so that its significance is more readily apparent, and where information relating to complex phenomena or processes can be simplified, offering invaluable support to decision makers and facilitating communication with the public and the media (e.g. Hammond *et al* 1995). However the potentially arbitrary nature of the weighting process by which individual variables are combined (and the consequent risk that misleading policy messages may be generated if the underlying model is poorly constructed or misinterpreted) is often cited as a principle reason why individual indicators should be kept separate (OECD & JRC 2008, citing Sharpe 2004).

In addition, Hammond *et al* (1995) state that, whilst CIs can improve communication, they can play a useful role only where communication is welcomed and where decision-making is responsive to (for example) new social issues or the effectiveness of current policies. They outline several characteristics of successful indicators: as well as being analytically sound and based upon a fixed methodology, indicators should also be:

- user-driven - aggregated indicators (CIs) must be useful to their intended audience. They must convey information that is meaningful to decision-makers and in a form they and the public find readily understandable. Similarly, they must be crafted to reflect the goals a society seeks to achieve.
- policy-relevant - CIs should be pertinent to policy concerns. For example, for national level indicators, policy-relevant means not just technically relevant, but also easily interpreted in terms of environmental trends or progress toward national policy goals.
- highly-aggregated: CIs may have many components, but the final indices (AIs) should be relatively few in number, otherwise decision-makers and the public will not readily embrace them. The degree of aggregation is inevitably dependent on who is to use them and for what purpose.

The development of a formalised system for aggregating relatively complex and often diverse strands of marine biodiversity data must allow individual data (or indicators) to be combined to provide a single CI, where the CI refers to a measure that is derived from specific individual indicators (e.g. quantitative or a qualitative measures) and which is designed to measure or describe the aggregated performance of a multi-dimensional issue (e.g. Zhou & Ang 2009). The use of CIs should ideally be restricted to the measurement of multi-dimensional concepts which cannot be readily captured by single indicators (OECD & JRC 2008).

Box 2. Pros and Cons of Composite Indicators (after OECD & JRC 2008).

Pros	**Cons**
• Can summarise complex, multi-dimensional realities with a view to supporting decision makers; • are easier to interpret than a battery of many separate indicators; • can assess progress of countries over time; • can reduce the visible size of a set of indicators without dropping the underlying information base; • thus making it possible to include more information within the existing size limit; • place issues of country performance and progress at the centre of the policy arena; • facilitate communication with general public (*i.e.* citizens, media, *etc.*) and promote accountability; • help to construct/underpin narratives for lay and literate audiences; • enable users to compare complex dimensions effectively.	• May send misleading policy messages if poorly constructed or misinterpreted; • may invite simplistic policy conclusions; • may be misused, *e.g.* to support a desired policy, if the construction process is not transparent and/or lacks sound statistical or conceptual principles; • the selection of indicators and weights could be the subject of political dispute; • may disguise serious failings in some dimensions and increase the difficulty of identifying proper remedial action, if the construction process is not transparent; • may lead to inappropriate policies if dimensions of performance that are difficult to measure are ignored.

3. Aggregation methods for marine biodiversity assessments

3.1 Approaches to aggregation – an overview

Prins *et al* (2013) provide a useful working summary of aggregation types used to support biological assessments. Review of the available literature in this study has not identified any additional aggregation methods that were not considered by Prins *et al* (see Table 4.3, Prins *et al* 2013). However, their classification of methods was not felt to be fully exhaustive in the context of the current study which, in part, seeks to examine the relevance and application of particular aggregation methods and to consider the relationship and relative merits of alternative methods. Table 3.1 (below) is based on information presented by Prins *et al* (2013) but the content has been augmented and restructured to assist in developing key themes for consideration in the context of high level (Assessment Index, AI) development.

In particular, a distinction is made in Table 3.1 between approaches that apply qualitative approaches ('conditional rules' - effectively variations on the one out all out theme, see below) and quantitative approaches (e.g. averaging techniques).

Table 3.3. Summary of aggregation approaches (adapted from Prins *et al* 2013).

General approach	Method	Relevant to production of: CIs Composite Indicators	Relevant to production of: AIs Assessment Indicators	Examples:
Qualitative approaches				
Conditional Rule (CR) methods	*One Out, All Out (OOAO)* - all variables have to achieve good status.		✓	• EC 2005 • CARONI *et al* 2013 • BORJA *et al* 2013 • SIMBOURA *et al* 2005
	Hierarchical application of One Out, All Out (OOAO) – 'high-level integration' assessment results for three groups: biological indicators, hazardous substances indicators and supporting indicators, each applying OOAO.		✓	• HELCOM 2010
	Two-out all-out (TOAO) - if two variables do not meet the required standard, good status is not achieved.		✓	• OSPAR 2009 • TUEROS *et al* 2009
	Threshold - a specific proportion of the variables have to achieve good status.		✓	• SIMBOURA *et al* 2012 • PIET *et al* 2010 • BORJA *et al* 2013
	Decision tree – uses specific decision rules to integrate elements into a quality assessment.		✓	• BORJA *et al* 2008 • BORJA *et al* 2009 • OSPAR 2008

General approach	Method	Relevant to production of: CIs Composite Indicators	Relevant to production of: AIs Assessment Indicators	Examples:
Quantitative approaches				
(i) Averaging Approach (AA) methods (NB can be arithmetic or geometric)	*Non-hierarchical, non-weighted averaging* – combination of variables/indicators into a flat structure with no intermediate aggregation. Weightings are equal for all indicators and is a typical approach used when there is not enough information on the influence of individual indicators. This method is the most basic of quantitative aggregations, and is more common for indicator production.	✓	✓	• CARONI *et al* 2013 • ALDEN *et al* 2002 • ALDER *et al* 2010 • BROOKS *et al* 2009 • DALTON *et al* 2012 • HARRISON & KELLY 2013 • LLANSÓ *et al* 2002 • OJAVEER & EERO 2011 • SKARPAAS *et al* 2012 • STODDARD *et al* 2008 • WEISBERG *et al* 1997 • ZHOU *et al* 2006
	Non-hierarchical, weighted averaging - combination of variables/indicators into a flat structure with no intermediate aggregation. Weightings are variable between indicators and can be allocated according to multivariate analysis, expert judgement or based on theoretical assumptions regarding value.	✓	✓	• OJAVEER & EERO 2011 • CARONI *et al* 2013 • BORJA *et al* 2013 • JOLLANDS *et al* 2003 • AUBRY & ELLIOTT 2006 • BASSET *et al* 2012 • CERTAIN *et al* 2011 • HALPERN *et al* 2012 • SKARPAAS *et al* 2012
	Hierarchical, non-weighted averaging – the use of hierarchical approaches to structure indicator inclusion and group is very common. The added structure provides the ability to output intermediate CIs that aid in the interpretation of the overall CI/AI. The nesting of associated indicators into clusters greatly improves the clarity of the aggregation process. Weightings are equal for each indicator and typical of when there is not enough information on the influence of individual indicators.	✓	✓	• OJAVEER & EERO 2011 • BORJA *et al* 2013 • United Nations Development Programme various years • OJAVEER & EERO 2011

General approach	Method	Relevant to production of: CIs Composite Indicators	AIs Assessment Indicators	Examples:
	Hierarchical, weighted averaging - hierarchical layers and clustering of input indicators is used to structure and order the aggregation. Weightings are variable between indicators and can be allocated according to multivariate analysis, expert judgement or based on theoretical assumptions regarding value. Weights can be applied to either individual indicators or to clustered indicators.	✓	✓	• AUBRY & ELLIOT 2006 • RODRIGUEZ-RODRIGUEZ & MARTINEZ-VEGA 2012
Quantitative approaches				
(ii) Non-averaging Approach (NAA) methods	*Multi-metric indices* - often hierarchically-structured and have inputs clustered by metric. Weights can be variable or equal. Calculation is undertaken with complex approaches such as summation, multiplication or bespoke formulae operations.	✓	✓	• VOLLENWEIDER *et al* 1998 • BORJA *et al* 2013 • ROSENBERG *et al* 2004 • PAUL *et al* 2001 [used Linear Discriminant Analysis] • JOLLANDS *et al* 2004 [used Principal Components Analysis] • PRIMPAS *et al* 2010
	Multivariate analyses - use predefined statistical procedures. Commonly applied methods include Factor Analysis, Discriminate analysis and Principal Components Analysis.	✓	✓	• RICE *et al* 2010 • TETT *et al* 2007 • BORJA *et al* 2013 • BRIGGS *et al* 2006

3.2 Summary of aggregation methods

This section briefly summarises the range of aggregation methods that might be applied to marine biodiversity assessments, and outlines the benefits and limitations of each method, building on the initial overview reported in Section 2 and addressing the requirements of Objective 2. Each of the methods is presented and discussed in more detail under Appendix D.

The assessment of each method is reported with reference to a derived quality framework. This framework was based on two prominent quality frameworks from the wider literature (OECD & JRC 2008; Nardo *et al* 2005) and provides a standard framework against which different aggregation methods can be summarised. This derived quality framework is summarised below and is presented in greater detail within Appendix E. The framework considers four principal dimensions:

(i) the basis for the aggregation;
(ii) the value of the aggregation;
(iii) the aggregation process; and
(iv) the subsequent application of methods or data.

3.2.1 The basis for the aggregation

The basis for aggregation considers the principles that have been used within the aggregation method (e.g. the fundamental statistical basis) and includes the **accuracy and reliability of the method or tool**. Where (internationally) agreed methodologies have been employed (rather than novel untested approaches) they are more likely to be considered to be **methodologically sound**. This quality dimension also assesses whether the method has made use of **weighting** or similar methods (available methods for generating appropriate weights will be discussed separately).

3.2.2 The value of the aggregation

The value of the aggregation method considers the underlying **transparency and comprehensibility** to local, national and international stakeholders (including UK and EU Member State Governments). This essentially requires that it should be possible to provide a simple explanation of the underlying method(s) which is accessible to a lay audience. In this context, both the construction process and aggregation output should be **easily interpreted** (including, for example, the **appropriateness** of the methods for capturing the process of interest), and the resulting aggregated indicator should be **accessible and clear**. Note that, to an extent, this facet of method quality may be application dependent. That is, whilst the aggregation process itself may be sound, an absolute assessment of the accessibility of the approach in terms of the transparency and comprehensibility of outputs is likely to change according to the specific application.
The method should also demonstrate **integrity in its outputs**, producing objective outputs through the robust application of formalised routines.

3.2.3 The aggregation process

The aggregation process itself can be assessed in terms of the **robustness and practicality** of the method(s). This includes the **serviceability of the underlying method(s)**, reflecting the suitability of the aggregated indicator computation for the phenomenon of interest; and the **accessibility of the approach** (for example does the process provide outputs that can be usefully deployed for other purposes, and can the methodologies be readily applied by other researches for repeat application).

This quality dimension also considers **data requirements** - a suite of factors that relate to the nature of the underlying data that are used[5]. These include the **quality levels** that are required or are imposed on the data used (e.g. the **accuracy of the data**; the **relevance** and **credibility** of the data sources) as well as the **timeliness and punctuality** of data delivery and of aggregation and reporting.

In addition, the way **uncertainty and confidence** levels are considered, including the **ability to quantify the confidence** associated with the final aggregated outputs, should be noted.

3.2.4 The subsequent application of methods or data

Finally, methods can be assessed on the basis of their repeat application. Consideration should be given to **repeatability** (whether the method can be applied to different assessment requirements) and to the potential for disaggregation of the information to meet other assessment purposes (**coherence**)[6].

3.3 Overview of performance and applicability

The following tables present (i) a summary of the assessment of the identified aggregation methods against the derived quality framework (Table 3.2), (ii) a summary of the advantages and disadvantages of each method (Table 3.3) and (iii) their applicability to each of the five types of aggregation (*sensu* JNCC) (Table 3.4). As noted above, further detail pertinent to these tables is presented under Appendix D and Appendix E.

[5] This particular aspect of the overall quality framework is considered further under Section 4.2.

[6] In the context in which it is used in this study, disaggregation might be seen as a misnomer. Rather than referring to the ease with which a high level aggregated indicator can be disassembled or 'unpicked' to derive building blocks that could then be used for a different aggregation, it relates more to the 'universality' of the components required by a particular methodology. Where a method uses readily available information that is in a format that is common to the needs of a range of other aggregation techniques, then the disaggregation of the method would be high. In effect, methods that show high levels of disaggregation use a common set of contributory building blocks; monitoring programs developed to collect these common building block data will service a higher number of such potential aggregation needs.

Table 3.4. Summary of aggregation method performance according to derived quality framework.

Performance assessment criteria:	**One out, all out (OOAO)**	**Hierarchical application of OOAO**	**Two-out all-out (TOAO)**	**Threshold methods**	**Decision tree approach**	**Non-hierarchical, non-weighted averaging**	**Non-hierarchical, weighted averaging**	**Hierarchical, non-weighted averaging**	**Hierarchical, weighted averaging**	**Multi-metric indices**	**Multivariate analyses**
The basis for the aggregation											
Methodologically sound	2	2	2	2	2	0	1	0	1	2	2
Accurate/reliable	1	1	1	2	2	0	1	1	2	2	2
Weighting applicable	0	0	0	0	1	0	2	0	2	2	1
The value of the aggregation											
Interpretability/transparency & comprehensible[7]	2	2	2	2	2	2	1	1	1	1	1
Accessibility/clarity[7]	2	2	2	2	2	2	1	1	1	1	1
Appropriate methods[7]	1	1	1	1	2	0	1	1	2	2	2
Objective outputs demonstrating integrity	2	2	2	1	1	1	1	1	1	1	2
The aggregation process											
Robustness and practicality	1	1	1	1	1	0	1	0	2	2	1
Accessibility	2	2	2	2	2	1	1	1	1	2	2
Handling of uncertainty and confidence	0	0	0	0	1	1	1	1	1	1	1
Subsequent application of methods or data											
Repeatability	2	1	2	2	1	1	1	1	1	1	1
Disaggregation (coherence) potential	2	2	2	2	2	2	1	2	1	1	2

Key:
0 = poor........... the method performs poorly, or does not satisfy or address the performance assessment criteria, or fails to provide a mechanism for facilitating their consideration
1 = moderate... the method satisfies the performance assessment criteria to a limited extent but there may be outstanding issues regarding their accommodation
2 = good.......... the method satisfies or addresses the performance assessment criteria, or facilitates their accommodation

[7] Assessment against these criteria is generally application-specific - a generalised view of performance against these criteria has been considered here.

Table 3.5. Summary of advantages and disadvantages of different aggregation methods (adapted from Prins *et al* 2013).

Method	Advantages	Disadvantages	Applicable for production of:
Qualitative approaches - Conditional Rule (CR) methods			
One Out, All Out (OOAO) principle - all variables have to achieve good status.	• Most comprehensive approach; • follows the precautionary principle.	• Trends in quality are hard to measure; • does not consider weighting of different indicators and descriptors; • chance of failing to achieve good status is very high; • may include double counting.	Assessment Indicators
Hierarchical application of One Out, All Out (OOAO) – for example 'high-level integration' assessment results for three groups: biological indicators, hazardous substances indicators and supporting indicators, each applying OOAO.	• Reduces the risks associated with OOAO while still giving an overall assessment.	• Identification of appropriate hierarchies may reduce overall transparency of process.	Assessment Indicators
Two-out all-out (TOAO) - if two variables do not meet the required standard, good status is not achieved.	• More robust compared to OOAO approach.	• See above.	Assessment Indicators
Threshold - a specific proportion of the variables have to achieve good status.	• Can focus on key aspects (e.g. biodiversity descriptors).	• Assumes that overall assessment (e.g. GEcS[8]) is well represented by the specific selection of variables that are available or are used.	Assessment Indicators
Decision tree – uses specific decision rules to integrate elements into a quality assessment.	• Possible to combine different types of elements; • allows combination of different approaches (e.g. OOAO & threshold approaches) within the same framework; • flexible approach.	• Only quantitative up to a certain level (needs to incorporate qualitative assessments but can do so in a sensitive manner).	Assessment Indicators

[8] GEcS – Good Ecological Status (under the Water Framework Directive, WFD)

Method	Advantages	Disadvantages	Applicable for production of:
Quantitative approaches - (i) Averaging Approach (AA) methods (NB can be arithmetic or geometric)			
Non-hierarchical, non-weighted averaging – combination of variable (parameter) values (may use mean or median).	• Simple and clear structure; • objective; • doesn't require prior knowledge of variable influence; • can be relative or absolute values.	• Equal weighting often unrealistic; • intermediate indices can't be generated easily.	Composite Indicators Assessment Indicators
Non-hierarchical, weighted averaging - as above, but with different weights assigned to the various variables.	• Simple and clear structure; • weighing represents influence; • can be relative or absolute values.	• Higher data requirements; • problem of setting weights; • intermediate indices can't be generated easily.	Composite Indicators Assessment Indicators
Hierarchical, non-weighted averaging – variables defined at different hierarchical levels.	• Reflects the hierarchy among descriptors and avoids double counting; • different calculation rules can be applied at different levels; • intermediate indices can be generated; • can be relative or absolute values.	• Equal weighting often unrealistic; • problem of agreeing on hierarchy.	Composite Indicators Assessment Indicators
Hierarchical, weighted averaging - variables defined at different hierarchical levels, each with different weights.	• Reflects the hierarchy among descriptors and avoids double counting; • different calculation rules can be applied at different levels; • weighing represents influence; • can be relative or absolute values.	• Higher data requirements; • problem of agreeing on hierarchy; • problem of setting weights.	Composite Indicators Assessment Indicators
Quantitative approaches - (ii) Non-averaging Approach (NAA) methods			
Multi-metric indices - often hierarchically-structured and have inputs clustered by metric. Weights can be variable or equal. Calculation is undertaken with complex approaches such as summation, multiplication or bespoke formulae operations.	• Integrates multiple indicators into one value; • may result in more robust indicators, compared to indicators based on single parameters; • can be relative or absolute values; • can use aggregation functions other than averaging; • variable weightings can be applied.	• Higher data requirements; • autocorrelation between variable; • structure and calculation can be difficult to interpret; • metrics may not be sensitive to the same pressures.	Composite Indicators Assessment Indicators
Multivariate analyses - use predefined statistical procedures. Commonly applied methods include Factor Analysis, Discriminate analysis and Principal Components Analysis.	• No need to set rigid target values, since values are represented within a domain; • highly objective.	• Higher data requirements; • results are hard to communicate to managers; • not accessible to non-statisticians.	Composite Indicators Assessment Indicators

Table 3.6. Application of aggregation methods to different aggregation types.

	Aggregation Methods:										
	Qualitative					Quantitative					
Aggregation type (*sensu* JNCC)	**One Out, All Out (OOAO)**	**Hierarchical application of OOAO**	**Two-out all-out**	**Threshold methods**	**Decision tree approach**	**Non-hierarchical, non-weighted averaging**	**Non-hierarchical, weighted averaging**	**Hierarchical, non-weighted averaging**	**Hierarchical, weighted averaging**	**Multi-metric indices**	**Multivariate analyses**
Type 1: Aggregation over different spatial scales or assessment units	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Type 2: Aggregation across different temporal scales	-	-	-	-	-	✓	✓	✓	✓	✓	✓
Type 3: Aggregation across biodiversity components	-	-	-	-	-	✓	✓	✓	✓	✓	✓
Type 4: Aggregation across indicators/ indices within assessment criteria	✓	✓	✓	✓	✓	?	?	✓	✓	✓	-
Type 5: Aggregation across assessment criteria of different biodiversity components to deliver an overall assessment	✓	✓	✓	✓	✓	?	?	✓	✓	✓	-

As Table 3.2 indicates, aggregation performance (as assessed against the derived quality framework) varies across the different aggregation methods. No single methodology scores consistently well across all performance criteria, with each methodology displaying specific strengths and weaknesses. Of particular note are the relative weaknesses of the qualitative methods such as OOAO and its derivatives (including threshold approaches) in accounting for weightings in the aggregation process or for handling measures of uncertainty or confidence; and the general reduction in transparency and accessibility associated with the more ‘complex’ quantitative methods such as hierarchical weighted averaging and multivariate analyses.

Similarly, the summary of pros and cons associated with each methodology (Table 3.3) fails to identify a clear optimal approach. The obvious advantages of the qualitative methods such as OOAO and its derivatives (hierarchical OOAO, TOAO or threshold methods) lie in their simplicity (and relatively low demands on data) and their subsequent communicability. Within this suite of methods there is a clear hierarchy in terms of the degree to which they align with the precautionary approach. For example, the OOAO method provides the most precautionary aggregated assessment. The TOAO method is slightly more relaxed and less precautionary, whilst threshold methods are (depending on the actual thresholds set) potentially more relaxed still. In terms of the nature of their (aggregated) outputs, the range of qualitative methods considered are only suited to the production of assessment indicators, whilst the more quantitative approaches can be used to derive both assessment indicators and composite indicators (such as indices of benthic integrity).

The clearest differentiation between methods is provided in the summary given in Table 3.4 which indicates that, in general terms, there is a clear presumption against the use of qualitative aggregation methods (OOAO etc.) for Type 2 or Type 3 aggregation applications (aggregation across different temporal scales, or across biodiversity components). In addition, multivariate analyses are not felt to be appropriate for Type 4 or Type 5 aggregation applications (aggregation across indicators or indices within assessment criteria, or across assessment criteria of different biodiversity components to deliver an overall assessment). Notwithstanding the above, there are no clear patterns evident that restrict the choice of method for any given aggregation type to just one or two options.

Given the findings of this initial assessment, it is suggested that the each of the aggregation methods outlined and discussed in this section has a potential role in producing an aggregated output for reporting marine biodiversity assessments, with the final choice of method being dependant on specific circumstances and needs.

4. The information requirements of aggregation methods

4.1 Introduction

This section addresses Objective 3 (and, more specifically, Tasks 3a-3c):

Objective 3: Determine and describe the information requirements of the selected aggregation methods selected under Objective 1, and assess the extent to which these information requirements are met by the existing marine information and evidence in the UK.

- ***Task 3a*** *will determine and describe information requirements of different identified aggregation methods.*
- ***Task 3b*** *will identify commonalties in the information requirements between different aggregation methods.*
- ***Task 3c*** *will assess the extent to which the information requirements of the different aggregation methods are met by the existing marine information and evidence in the UK.*

In so doing, rather than restricting assessment to a subset of methodologies, the information requirements of all of the aggregation methods reported and discussed in the preceding sections has been considered.

Data aggregation methods make use of information in two distinct ways. Firstly, and most obviously, they aggregate data. That is, they process complex data and produce simpler (aggregated) outputs. Secondly they make use of supporting or ancillary information as part of the method or process of aggregation. Different types of aggregation (*sensu* JNCC) may potentially make use of different aggregation methods whilst, in turn, data availability will have an obvious impact on the types of aggregation method that might be considered for operational use (i.e. for management or reporting).

4.1.1 Data types

With regard to marine biodiversity assessments a wide range of different data types may, for management or reporting purposes, need to be aggregated. Figure 4.1 provides a schematic representation of the classification of data types (or categories of data) that might potentially be considered for aggregation. Each aggregation method that has been identified is able to handle one or more of these different data types. In turn, each of these data types may be collated from one or more of a number of possible sources. As well as handling available data, certain methods may be able to cope with gaps in the source information, i.e. missing data.

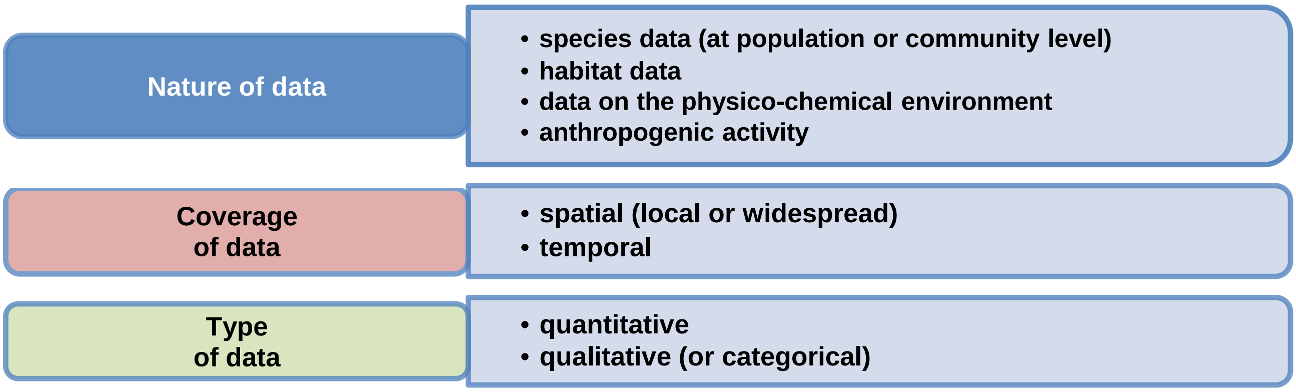


Figure 4.3. Proposed structure for classifying data types.

It is not suggested that this classification framework is fully definitive, but rather that it provides a pragmatic structure to support a brief discussion of the nature of data that may be used in aggregation exercises.

This classification is outlined further below under a series of headings that address:

- The nature of the data; what do the data relate to (e.g. species, habitats, the physic-chemical environment, anthropogenic activities)?
- The coverage or distribution of the data; fundamentally, is it spatial data or temporal data?
- The type of data; is it quantitative, semi-quantitative or qualitative?

The source of the data (i.e. whether it is empirical/observed data or derived from expert judgement) may be considered as representing the fundamental element of this classification. Note that, in practical terms, quantitative data is derived from empirical observation or monitoring, whilst qualitative data is sourced either from an integration of quantitative data or through expert judgement methods. It is possible, although not common, for expert judgement to generate quantitative predictions (for example through the use of methods such as Analytic Hierarchy Process (AHP); Barnard & Boyes 2013).

Spatial data needs to be ‘fit for purpose’ in terms of its scale or extent. For example, an assessment of an Annex I feature within an MPA will need to make use of indicators derived from spatially intensive local data, whilst an assessment of GEnS at the scale of a biogeographic region is likely to make use of indicators from more spatially extensive widespread datasets.

The capability of different aggregation methods to make use of these various data types is discussed under Section 4.2.2. This assessment facilitates an objective consideration of the applicability of each aggregation method to the five identified forms of aggregation. The results from this exercise are outlined and summarised under Section 4.2.4.

4.1.2 Requirements for supporting information

Certain aggregation methods require supporting or ancillary information; this may be in the form of:

- information used as part of the method application (e.g. baseline data, targets or reference values); or
- information that supplements the input data and which ultimately informs the interpretation of the aggregation outputs (e.g. information such as confidence levels).

Such data may represent an intrinsic part of the aggregation process and may be as significant within the overall process as the spatial and temporal components described by the fundamental ‘indicator’ data.

i. Supporting or ancillary information to support or facilitate the aggregation process

Baseline data, targets and reference values are not considered as a fundamental input to the aggregation process but instead are considered as being an intrinsic part of the underlying method. However, it may be necessary to have appropriate sources of such information to permit a methodology to be applied (the application of the One Out, All Out (OOAO) methodology, for example, requires that target values describing thresholds for different quality classes are available, whilst weighted methods require quantitative values for the weightings that are applied).

Baseline data are historical monitoring or survey data representing a known state from which subsequent deviation can be measured; these data are usually taken from the same (or a parallel, representative) system and, in this sense, can be considered to be ‘internal’ to the system being assessed. Baseline data may be required to set a ‘norm’ against which deviation may be assessed, or may be used to set targets, in the form of category thresholds. In this context baseline data can be used to provide a basis for converting quantitative data to qualitative data (e.g. by employing 20 percentile values from baseline data as class boundaries for the categorisation of new data). Targets may be based on systems that are remote from the one under consideration, although in such cases there may be some attempt made to ensure that the targets are relevant. Targets may also be externally imposed, for example through legislation. In terms of their potential roles in aggregation; targets most obviously feed into approaches such as OOAO, where an observed value is categorised (e.g. to a particular quality class) through comparison to a threshold (target) value.

Reference values are data that reflect the condition of one or more parameters in a system, relating to a point in time when the system was believed to be in some form of ‘ideal’ or reference state; they are typically used where a current state or performance needs to be normalised against a de facto standard. They can be based on baseline data from the same system or, equally, on information derived from outside of the system.

Particular quality standards, such as Environmental Quality Standard (EQS) values as defined under the Water Framework Directive, may be established for parameters such as contaminant concentrations in water or sediments. Such EQS values represent a specific type of target that can, for example, be readily used to represent a pass/fail criterion.

Reference lists represent another form of data that is utilised within certain aggregation processes. They may identify, for example, the range of species, habitats, activities or pressures that are considered within an aggregation – as might be the case for a defined Index of Benthic Integrity. Their role in the aggregation process is to determine the *scope* of the aggregation (e.g. which species should be considered) rather than representing the disaggregated (indicator) information upon which the aggregation is based, and as such are not considered further here.

ii. Ancillary information to support interpretation of aggregation outputs

Information on the accuracy or veracity of the original disaggregated (indicator) information represents another form of supporting information. Data that are used as indicators are subject to a number of sources of variability (due both to sampling error and bias, and to natural, stochastic variability). Measures of confidence in the original data (which may frequently be based on pragmatic rules or expert judgement, for example as in the case of pressure sensitivity assessments made under MB0102 or the confidence data associated with the available data on modelled seabed habitat types), are likely to be relatively coarse (e.g. ‘high’, ‘medium’, ‘low’). Such data can nevertheless be used to provide a qualitative measure of the general quality of aggregated outputs.

4.2 Data types and the applicability of different aggregation methods

4.2.1 Data types

In general terms, quantitative data is derived from empirical observation or monitoring, whilst qualitative data is sourced either from an integration of quantitative data or through expert judgement methods. For use in many aggregation methods, the source of the information, whether it is empirical or from expert judgement does not always change its compatibility but does have important implications for the overall assessment of confidence and accuracy.

A brief introduction to the range of data types within the generalised classification is presented below as Table 4.1

The spatial range of data is important and will have a bearing on the overall quality of an aggregated assessment of data for a given area. Data might be local (e.g. from within a defined MPA) or large scale (e.g. from across a wide biogeographic reporting area). In addition, the nature of the spatial data is of importance, for example whether it is at the level of point data or is available only as wider scale area data (e.g. presented as GIS shapefiles that indicate the spatial extent of a particular species as presence/absence).

Whilst temporal data will have an associated spatial component (for example: continuous or regular repeat monitoring at a single point location; repeat monitoring at a wider scale such as repeat observations of seabirds; or remote monitoring of chlorophyll a levels as a surrogate for phytoplankton concentrations) the temporal component of data can be most practically subdivided to temporal frequency (e.g. number of sampling occasions per unit time) and regularity. Sampling may be, for example, annual, seasonal, or part of a more frequent (e.g. daily) time series.

Table 4.7. Data types within the general classification structure.

Data type			Notes
Species data (including both population and community data)	Spatial	Quantitative	This grouping probably represents one of the main data types that is collected to support monitoring and reporting, and which is consequently frequently considered for aggregation. It might include, for example, data derived from subtidal grab sampling programs or intertidal transect surveys. Quantitative data is likely to be derived empirically (i.e. through observation/monitoring) rather than by expert judgement (although it is possible to employ expert judgement techniques to produce quantitative estimates).
		Qualitative	Qualitative spatial data is likely to be derived either indirectly, from the categorisation of empirical (observed) quantitative data, or directly, through expert judgement.
	Temporal	Quantitative	This grouping includes, for example, data derived from larval herring survey work and from the annual Clyde trend monitoring programme, examining levels of chlorobiphenyls (CBs) and polycyclic aromatic hydrocarbons (PAHs).
		Qualitative	Qualitative temporal data is likely to be derived either indirectly, from the categorisation of empirical (observed) quantitative data, or directly, through expert judgement.
Habitat data	Spatial	Quantitative	Most habitat data is in the form of categorical (classified) habitat types, although some quantitative data is available, such as that derived from certain sublittoral surveys.
		Qualitative	This category, which presents data in categorical terms, probably represents the more usual form of spatial habitat data, and includes sources such as the EUSeaMap EUNIS level 3 data.
	Temporal	Quantitative	This would include, for example, the time series outputs from Special Area of Conservation habitat condition monitoring.
		Qualitative	Although in theory this represents a potential category of data it is likely to be a 'null' category, with no practical examples of this particular data type. It is possible, however, that (categorical) expert judgement data on long term trends might be included within this category.
Physico-chemical data	Spatial	Quantitative	This would include, for example, data collected in the sea lough surveys in Northern Ireland.
		Qualitative	As for qualitative spatial habitat data, this category, which presents data in categorical terms, probably represents the more usual form of spatial physico-chemical data.
	Temporal	Quantitative	Data within this category is likely to be derived from longitudinal studies employing automatic data recording equipment at fixed stations.
		Qualitative	Although in theory this represents a potential category of data it is likely to be a 'null' category, with no practical examples of this particular data type. It is possible, however, that (categorical) expert judgement data relating to trends in physic-chemical parameters (e.g. sea surface temperature) might be available in certain studies and such data would be included within this category.
Information on anthropogenic data	Spatial	Quantitative	This grouping would include, for example, VMS data used to indicate fishing activity and GIS (shapefile and vector) data on marine infrastructure (e.g. wind turbines, cables, pipelines).
		Qualitative	This category might include, for example, expert judgement sourced information on activities. One example of this is the FisherMap (and subsequently StakMap) product that was used to support the English Marine Conservation Zone project. These GIS-based products were underpinned by information derived from stakeholder interview. Ultimately, the data presented was qualitative (or, at best, semi-quantitative) in nature.
	Temporal	Quantitative	Whilst this category might include, for example, repeat information on leisure activities in coastal waters, it is unlikely that there would be a need to aggregate such temporal data.
		Qualitative	Although in theory this represents a potential category of data it is likely to be a 'null' category, with no practical examples of this particular data type (although it is possible that (categorical) expert judgement data trends in activity might be included within this category).

4.2.2 Ability to aggregate different data types

By tabulating the applicability of different aggregation methods to different data types it is possible to identify the apparent commonalities (and potential redundancy) between methods (Table 4.2). Note that Table 4.2 does not differentiate between different data types based on their underlying nature (i.e. whether they are species data, habitat data, physic-chemical data or anthropogenic data), but only compares the methods' ability to accommodate data types which differ according to 'coverage' (spatial vs temporal) and the 'type of data' (quantitative vs qualitative).

Table 4.8. Ability of different aggregation methods to accommodate different data types.

		Methods:										
		Qualitative					Quantitative					
Data type		One Out, All-Out (OOAO)	Hierarchical application of OOAO	Two-out all-out	Threshold methods	Decision tree approach	Non-hierarchical, non-weighted averaging	Non-hierarchical, weighted averaging	Hierarchical, non-weighted averaging	Hierarchical, weighted averaging	Multi-metric indices	Multivariate analyses
Spatial	**Quantitative**	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Qualitative	✓	✓	✓	✓	✓	-	-	-	-	-	-
Temporal	**Quantitative**	?	?	?	?	?	✓	✓	✓	✓	✓	✓
	Qualitative	✓	✓	✓	✓	✓	-	-	-	-	-	-

Key:

✓	Method able to accommodate data
?	Method may possibly be applied to data
-	Method probably not suitable

4.2.3 Requirements for supporting information by method

The requirements for each of the aggregation methods for different types of supporting information (both for facilitating the aggregation process and for interpreting or augmenting the outputs) are presented below as Table 4.3. A differentiation is made between where methods require supporting or ancillary information (i.e. obligated or mandatory inclusion) and where the accommodation of such information is optional. In addition, instances where there is no option for including such supporting information are noted.

Table 4.9. Requirements for supporting information by method.

		Methods:										
		Qualitative					**Quantitative**					
		One Out, All Out (OOAO)	**Hierarchical application of OOAO**	**Two-out all-out**	**Threshold methods**	**Decision tree approach**	**Non-hierarchical, non-weighted averaging**	**Non-hierarchical, weighted averaging**	**Hierarchical, non-weighted averaging**	**Hierarchical, weighted averaging**	**Multi-metric indices**	**Multivariate analyses**
Information to facilitate aggregation	**Baseline data**	o	o	o	o	o	o	o	o	o	o	o
	Reference values	o	o	o	o	o	o	o	o	o	o	o
	Targets	M	M	M	M	M	o	o	o	o	o	o
	Reference lists	o	o	o	o	o	/	/	/	/	/	/
	Formal quality standards	o	o	o	o	o	o	o	o	o	o	o
Information to support interpretation	**Data confidence**	/	/	/	/	/	o	o	o	o	o	o
Ability to handle missing data		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Key:

M	mandatory
o	Optional or application specific
/	no option for inclusion
✓	possible

4.2.4 Data types and aggregation methods

The brief assessment presented above as Table 4.2 and Table 4.3 highlights the limitations of certain aggregation methods regarding their ability to cope with different data types.

It is possible to ascribe quantitative scores to qualitative data (e.g. scoring 1, 2 or 3 for qualitative data classified as 'low', 'medium' or 'high') and subsequently apply quantitative methods. Such an approach is not, however, statistically robust. Consequently, the range of quantitative methods that are considered (e.g. the various forms of averaging) cannot be applied to qualitative data in a meaningful manner. For qualitative data therefore, aggregation methods are effectively restricted to OOAO, TOAO, threshold and decision tree approaches. Whilst any of the available aggregation methods could be applied to quantitative data, the applicability of qualitative approaches (OOAO, TOAO, threshold methods or decision tree approaches) to quantitative temporal data is less sound.

In terms of the need for ancillary or supporting information, one of the obvious differences between the approaches relates to the absolute requirement for some form of target value seen across all of the qualitative methods. Clearly, without the option to identify a target value it is not possible to make pass/fail assessments that are central to the application of these methods.

Supporting information relating to the level of confidence in underlying data can only really be handled by the quantitative methods. Whilst this type of information is not mandatory for these methods, where it is important to retain a measure of confidence in final outputs or classifications, it is generally only the quantitative methods (e.g. the averaging and multi-metric approaches) that are able to make formal use of information on uncertainty or confidence.

All methods that have been reviewed are able to handle missing data. In this context the use of the OOAO method is likely to be most problematical (given its extreme precautionary basis); however, the implications of missing data for the resilience of assessments that are based on any of the methods will be specific to the conditions under consideration.

In summary therefore, the choice of method for undertaking a data aggregation exercise relates to a number of factors and there is no simple mapping that relates one type of data to one optimal method. Consideration should be given not just to the type of data that are being aggregated, but also, *inter alia*, to the incidence of missing data, the availability of supporting or ancillary data, and the need to account for variable levels of confidence in the data that are available.

4.3 Data availability

A list of over 80 data sources, covering areas throughout the UK and suitable for use in marine biodiversity aggregations, was compiled. This list (presented in Appendix F as Table F.1) was originally constructed within the DEVOTES[9] (EU) project but was enhanced for use here in order to provide an example of the general availability of different data types required to support different aggregation methods. Whilst it covers the majority of marine biodiversity data sources, as well as relevant activity and physico-chemical monitoring, the list and associated assessments are intended to be indicative and should not be considered as being exhaustive or fully definitive.

All of the data sources represent active monitoring programs or monitoring for projects that have only recently expired. All of the data for the identified information sources are freely available from either online sources or the source organisation.

Each information source was attributed as below:
- Number of UK sea sub-units covered;
- frequency of temporal observations;
- duration of temporal observations;
- broad topic: biological, physico-chemical or activity;
- monitoring topic: population, community, habitat, physico-chemical or activity;
- raw data type: point, track (trawls or camera tows) or continuous (remote methods);
- emphasis of observations: spatial or temporal;
- observed or estimated (expert judgement or semi-quantitative method);
- quantitative or qualitative expression of observed or estimated parameter;
- low or high resolution (spatial)/low or high frequency (temporal);
- small or large extent (spatial)/short or long duration (temporal).

[9] DEVelopment Of innovative Tools for understanding marine biodiversity and assessing good Environmental Status (more information can be found here: http://www.devotes-project.eu/.

Pivot table analysis has been undertaken to summarise the profile of the information sources suitable for aggregation by method. This also highlighted the relative availability of differing information sources required for particular aggregation methods.

As no particular aggregation objective has been stated for this stage, the gap analysis was intentionally broad, and considered:

- Whether there is a shortage of data for a particular topic;
- the distribution of information at different levels of biological organisation;
- the availability of data in particular formats, e.g. maps over point data;
- the incidence of any imbalance between spatial and temporal information sources;
- the proportion of quantitative (empirical) data sources to estimated (expert judgement based) information sources;
- the distribution of information sources over different ranges of extent/frequency;
- the spread of information sources between differing study extents/durations; and
- the availability of ancillary information sources used for thresholds, baselines and targets.

4.4 Profile of information sources suitable for aggregation

Most of the available information sources had both spatial and temporal components. Studies with a consistent temporal component were less evident than those with spatial components. This is probably biased by the inclusion of several habitat map information sources which rarely include a temporal (repeated measures) component. In addition, many of the studies provided a number of separate data components (e.g. both species data and habitat data).

The distribution profile of information sources across the different constituent data types is shown in Table 4.4. The vast majority of information sources contain biological data – this partially reflects the selection of datasets relevant to marine biological assessments. Many of the marine biodiversity information sources also provide some physico-chemical variables, thereby increasing the availability of this data type. By comparison, there were few available sources for activities/pressures relevant to understanding patterns and trends in marine biodiversity. The higher proportion of spatial information sources for habitats was a related to the availability of habitat mapping products.

Table 4.10. Overall profile of information sources available for marine biodiversity aggregations, with examples for each information source type (values within cells are the number of available information sources).

Data type (with number of identified data sources)			Example
Species (population)	Spatial 31	Quantitative 31	*Nephrops* camera and trawl stock assessments
		Qualitative 0	None
	Temporal 26	Quantitative 26	Specific species monitoring programmes – e.g. grey seal monitoring, larval herring surveys
		Qualitative 0	None
Species (community)	Spatial 18	Quantitative 18	Clean Seas Monitoring Programme
		Qualitative 0	None
	Temporal 19	Quantitative 19	Clyde trend monitoring
		Qualitative 0	None
Habitats	Spatial 26	Quantitative 19	Northern Ireland sublittoral survey

		Qualitative 7	EUSeaMap EUNIS level 3
	Temporal 14	Quantitative 13	Special Area of Conservation Monitoring
		Qualitative 1	Rocky reef monitoring
Physico-chemical	Spatial 34	Quantitative 32	Northern Ireland sea lough spatial survey
		Qualitative 2	EUSeaMap physiographic maps
	Temporal 38	Quantitative 38	*In situ* instrumentation e.g. Smartbuoy
		Qualitative 0	None
Anthropogenic activities	Spatial 10	Quantitative 10	Vessel Monitoring System
		Qualitative 0	None
	Temporal 13	Quantitative 13	Disposal site monitoring
		Qualitative 0	None

* NB some of the identified studies were attributed as being having both spatial and temporal elements.

Table 4.5 presents the profile of temporal and spatial data by reported unit type (point data, track data and continuous monitoring data) and highlights the greater availability of continuous data within spatial sources. For temporal information sources: population monitoring (typically marine mammals, fisheries and seabirds) and community monitoring (planktonic and infaunal analysis) were well represented. Temporal monitoring of habitats (typically SAC monitoring) was less available than spatial sources of habitat information.

Table 4.11. Profile of temporal and spatial information sources by reported unit type.

Distribution	Point	Track	Continuous
Temporal	74%	25%	1%
Spatial	63%	23%	15%

The vast majority of information sources used in the analysis were observed data. This is likely a reflection of the selection process used to find studies for this analysis. There are many information products generated from estimates/expert judgement such as sensitivity matrices, conservation status reports and distribution maps. These information sources are mandatory for some aggregation processes and are used in variable weighting and the calculation of thresholds or baselines. However, many of the ‘off the shelf’ expert judgement information sources are fixed over time and therefore have reduced worth for inclusion as a source data input into an aggregation process. This however does not exclude the value of repeated assessments, for a particular aggregation objective, using expert judgement. This process is obviously a bespoke process and therefore not readily available as an immediate information source.

Most of the available temporal and spatial information sources were observed data (Figure 4.2). Furthermore, this information was typically expressed in a quantitative format. The reported larger proportion of spatial data sources using categorical expression is likely to be associated with the habitat mapping products contained in this category.

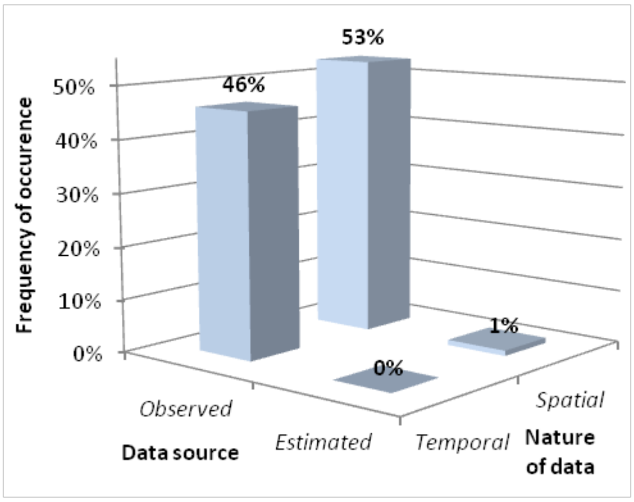
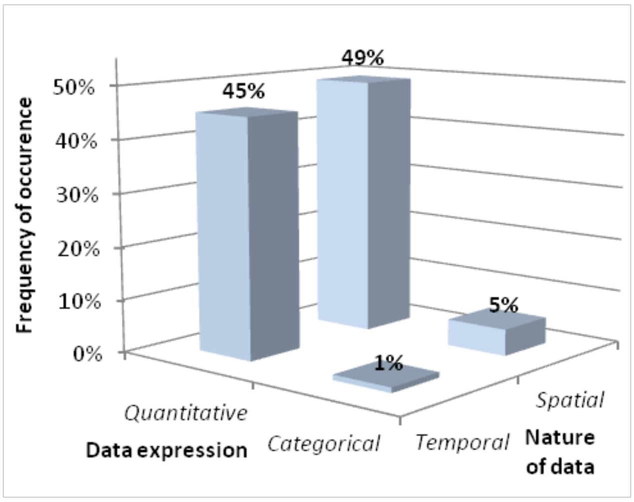
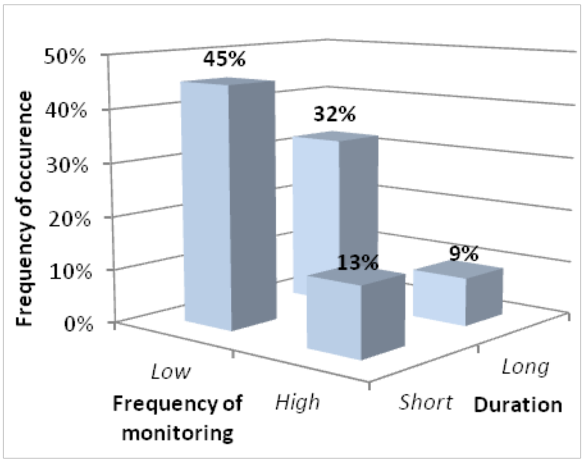
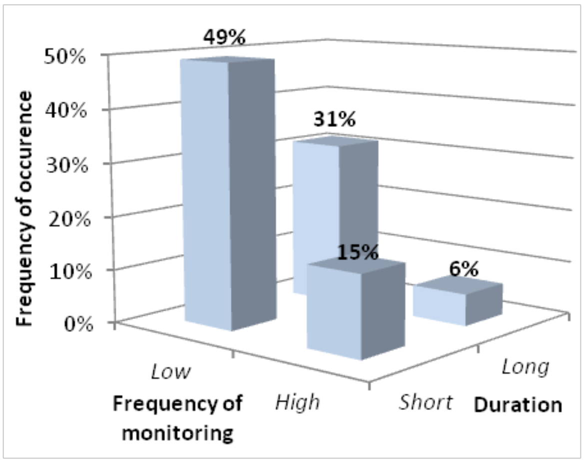
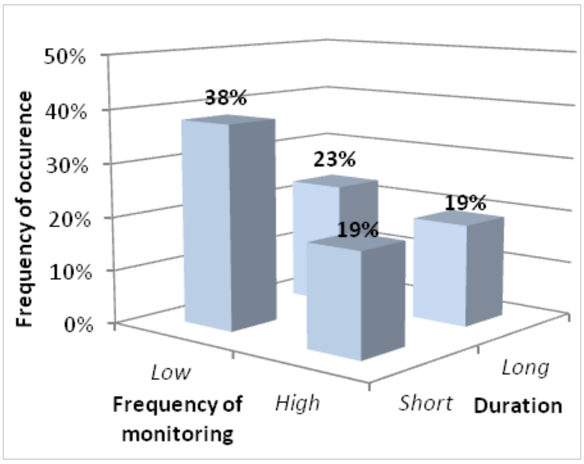
<table>
<tr><td>
</td><td>
</td></tr>
<tr><td>Data source</td><td>Data expression</td></tr>
</table>

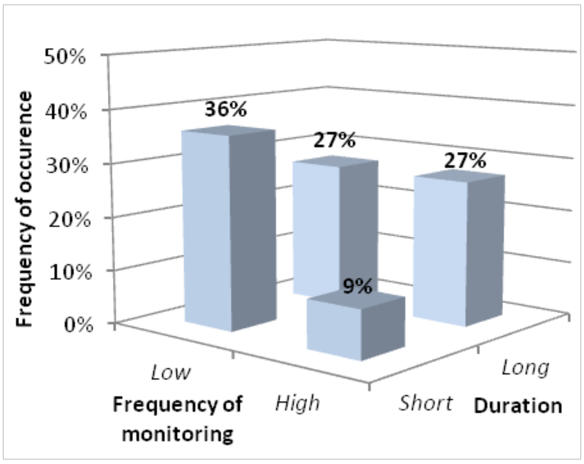
Figure 4.4. Profile of temporal and spatial information sources by data source and by expression.

Most of the temporal data sources were typically low frequency (annual or greater) and of a short duration (less than 10 years in length) (Figure 4.3). Only 32% were high frequency data sources – these studies typically used *in-situ* instrumentation. Less than 10% of the information sources were both high frequency and long duration. Examining just the biologically-focused temporal studies revealed a similar pattern (again, see Figure 4.3). The physico-chemical and activity focused monitoring was better spread between the four combinations of duration and frequency although as seen above, the total number of sources available is significantly less when compared with biological monitoring (Figure 4.3).

<table>
<tr><td>
</td><td>
</td></tr>
<tr><td>Distribution of all temporal information</td><td>Distribution of biologically-focussed temporal information</td></tr>
</table>



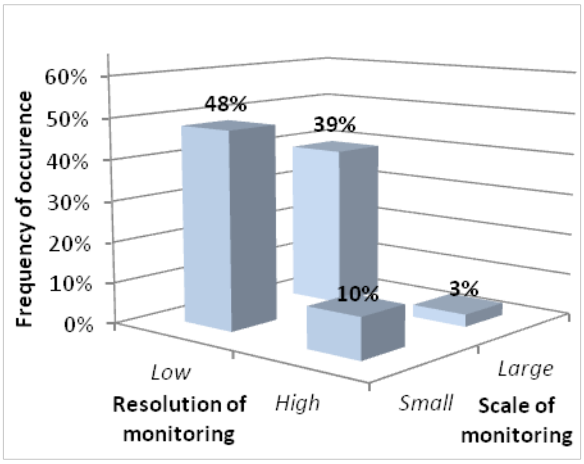
Distribution of physico-chemical temporal information



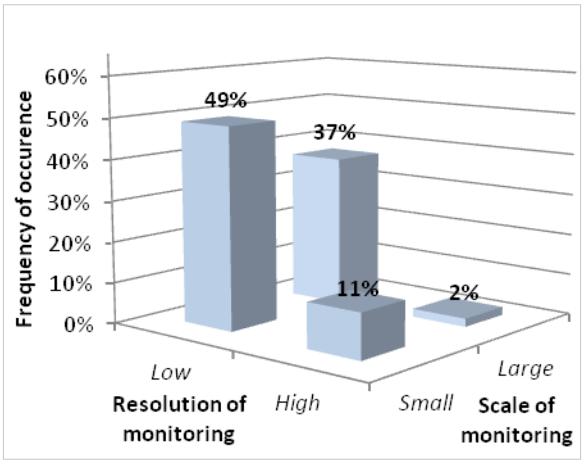
Distribution of activity-pressure temporal information

Figure 4.5. Relative frequency of occurrence of temporal data sources as defined by frequency of monitoring and monitoring duration.

The majority of the spatial information sources contained relatively low resolution data (Figure 4.4). High resolution data were typical of habitat mapping projects using marine acoustics. The spatial information sources were roughly split between relatively small and large extents. Based on the likely cost of generating high resolution large extend data sets, it is not surprising that this category is poorly represented. The sources that contribute to this category include the collective catalogue of acoustic habitat maps (patchy in distribution) and some forms of marine mammal monitoring. The biological sources also reflected the same distribution of study types which is a reflection of this category being the most abundant component of the spatial studies (Figure 4.4). The majority of the spatial physico-chemical information sources were low resolution and small extent studies (Figure 4.4). This is probably due to the use of ship-based point sampling for many studies. The inclusion of VMS monitoring within the activity/pressures studies provides an example of a high resolution/large extent activity data source (Figure 4.4).



Distribution of all spatial information



Distribution of biologically-focussed spatial information

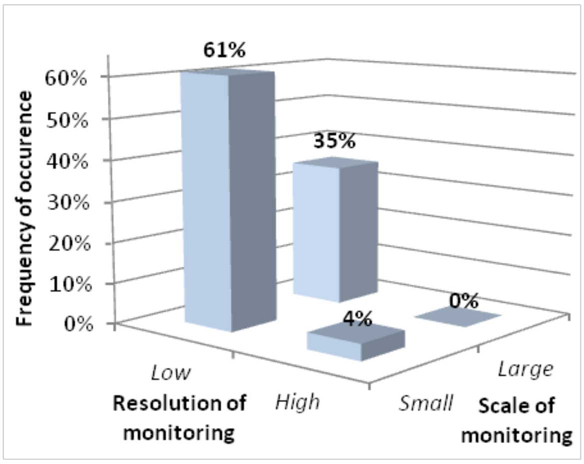
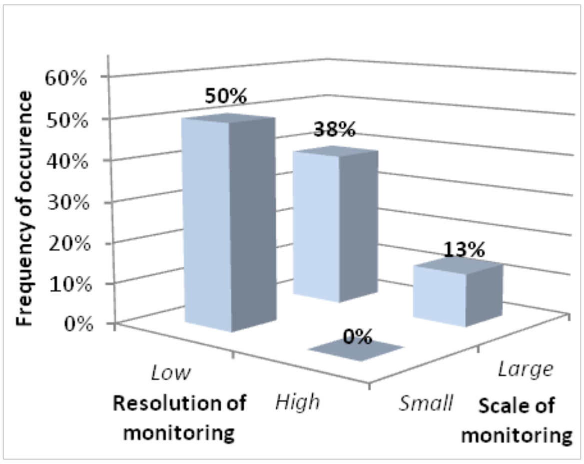
	
Distribution of physico-chemical spatial information	***Distribution of activity-pressure spatial information***

Figure 4.6 Relative frequency of occurrence of spatial data sources as defined by extent (scale) and resolution of monitoring.

4.5 Ancillary information

The information sources above provide the input data for the iterative calculation of the differing aggregation methods. Some aggregations, especially the non-quantitative forms, require additional information (e.g. for standardisation/normalisation). This is often in the form of baselines, thresholds and target or reference values. Establishing these values is critical for the sensitivity, response and repeatability of the aggregation. Often these values will remain set over several iterations of the aggregation. A number of the information sources that are used to produce indicator values may provide the basis for deriving these values but information sources that provide exact values for practical use within aggregations may not exist.

Baselines can be established by using any one of four general methods (e.g. Hill *et al* 2012):
- Observation of pristine conditions;
- historical observations;
- modelled predictions; and
- best professional judgement (expert judgement).

In the absence of pristine or historical data, baselines and target or reference values (that may be required by some aggregation processes to normalise observed data) can be predicted through the use of modelling techniques. Modelling can simulate both the spatial extent of the feature of interest as well as quantitative aspects of the population. Modelling outputs containing multiple components, such as for a biological community associated with a habitat, are harder to achieve and the overall confidence may be reduced. To achieve robust model predictions, a large data set that is spatially and temporally intensive is required. Such data sets do not exist for many eco-regions (Borja *et al* 2012). The use of models without suitable, high quality input information can lead to unrealistic and incorrect predictions, and superficially convincing outputs that confer more confidence about the baseline setting procedure than actually exists. Furthermore, the inherent complexity of some models and the use of required assumptions can also reduce the transparency, objectivity and comprehensibility of the resulting baseline.

Based on these issues and inadequacies, Borja *et al* (2012) considered ‘best professional judgement’ as the best approach for establishing baselines and target or reference values. This view is echoed by Hill *et al* (2012) who identified the use of expert opinion in combination with whatever suitable data are available as the recommended method for setting reference conditions for the biodiversity criteria of habitat and benthic condition. This is clearly a more subjective and application-specific approach that leads to less ‘off the shelf’ information products that could be used as ancillary inputs for the aggregation methods identified. Based on the overwhelming importance of thresholds and baselines within aggregations such as one-out-all-out, the confident and objective establishment of these values is extremely difficult. The absence of information sources dedicated to providing transferable thresholds for aggregation purposes represents a significant data/methodological gap for most aggregation methods.

4.6 Conclusions

Based on the profile of the information sources and the availability of ancillary inputs, the following gaps have been identified:

- **Standardised methods and information sources for setting thresholds, baselines and targets** (which are a mandatory requirement in many aggregations such as decision trees and OAOO). This may be available immediately but specific cases will require short to medium term studies to develop the required methodologies and data.

- **The majority of the available information sources represent observed data expressed in a quantitative manner**. Some aggregation methods will use categorical or ordinal classes. Some work is likely to be required for the standardised and transparent conversion of quantitative values to classes and ranks (setting thresholds and targets).

- **Paucity of readily available activity/pressure information expressed in a suitable and meaningful manner for use as a surrogate within marine biodiversity aggregations**. It was apparent from the profile that information for activities/pressures relevant to biodiversity modification was less available than biological data. Furthermore, ‘off the shelf’ information sources linking activity surrogates with known conditions of state for multiple spatial scales, habitats and overlapping pressures are also less evident than biological datasets. Numerous matrices exist that link pressures with habitat-specific impacts. However, validated and quantitative methods will be required if activities are to be used as surrogates within aggregation. This work may be currently ongoing (e.g. within MSFD working groups or MarLIN[10]) but otherwise could be made available within the short term (<5 years).

- **Absence of clearly defined methods for the calculation of weightings for aggregated inputs**. Weightings are mandatory in many aggregation methods and overwhelmingly influential for the overall aggregation output. Weightings will be specific for particular aggregation objectives. However, an objective and standardised framework and calculation process is probably required to underpin these processes – guidance on possible approaches is available (Barnard & Boyes 2013). Such guidance could be made available in the short term if discrete work packages were to be taken forward.

- **Shortage of continuous datasets that would allow the footprint size of an activity to be calculated or the change of a species or habitat range to be monitored**. Much of the information sources reviewed are point data; this does not provide the required information on the change in the distribution of a species or habitat for high level marine biodiversity aggregations. Greater efforts must be made to examine the potential of point data sets to be interpolated into suitable surfaces. This could be made available in the short term if discrete work packages were to be taken forward.

- **Reduced availability of high frequency, long duration temporal studies and high resolution, large extent spatial studies**. These temporal information sources are typically the

[10] Marine Life Information Network (http://www.marlin.ac.uk/)

most important for normalising dataset for long-term background change and detection of trends within marine biodiversity. Equally, high resolution and large extent spatial data sets are expensive to collect but provide the best background for nesting and then aggregating site-specific studies. Both the temporal and spatial information sources could be made available in the long term if significant work packages were to be taken forward, e.g. a national seabed mapping program similar to INFOMAR[11] in the Republic of Ireland.

5. Development and assessment of conceptual scenarios

5.1 Introduction

This section of the report details work undertaken to address Objective 4, specifically developing and finalising three conceptual scenarios for undertaking marine biodiversity assessments (task 4a), exploring and describing the extent to which optimum aggregation methods might predetermine or bias the assessment outcome (task 4b), and exploring and describing the inter-dependencies/inter-relationships between the five aggregation types (*sensu* JNCC) both within and between the conceptual scenarios (task 4c).

The three conceptual scenarios considered comprise marine biodiversity status assessments for:

(i) a single benthic habitat feature at a local (MPA) scale, reporting annually under a Common Standards Monitoring (CSM) framework;
(ii) multiple species at the regional (UK regional sea) scale, reporting every six years under MSFD; and
(iii) a single, highly mobile species at the European scale, reporting every six years under the Habitats Directive (HD).

By considering these scenarios it is not intended to redesign established processes or to solve specific issues relating to their application in support of ongoing or anticipated reporting requirements. The intention is rather to examine the application and implications of different aggregation methods within current reporting needs, and to consider the benefits and problems associated with alternative aggregation (or process) options.

Consideration of each scenario is followed by a brief summary of conclusions and learning points.

5.2 Scenario approaches

This section addresses Tasks 4a and 4b. It specifically considers three scenarios that each, potentially, encompass a range of different aggregation types. Accordingly, a quick initial assessment was undertaken of how the five types of aggregation identified by JNCC contribute to each of the three scenarios (summarised as Table 5.1).

Type 1 aggregation (aggregation over different spatial scales or assessment units) will contribute to all three conceptual scenarios[12].

Whilst Type 2 aggregation (aggregation across different temporal scales) may apply to Scenarios B & C (which may potentially have time series of data being available for each six year reporting period), it is not likely that this would apply to Scenario A. Strictly, it is perhaps better to view the need to derive single estimates from temporal replicates of data as being a data processing issue rather than aggregation *per se* (i.e. how should time series of data be converted to metrics that

---

[11] Integrated Mapping for the Sustainable Development of Ireland's Marine Resource (http://infomarupdates.blogspot.co.uk/2013/06/infomar-2013-celtic-voyager-mapping.html)

[12] NB Aggregation over different spatial scales or assessment units would occur even within a single MPA where the site is subdivided into discrete management units.

can, subsequently, be spatially aggregated)[13]. Irrespective, the question of how the processing of temporal and spatial data is handled (in terms of the phasing of temporal versus spatial aggregations) remains important. Alternative scenarios, such as those where both continuous and discrete data are available (e.g. phytoplankton monitoring) may be considered to have more of a temporal aggregation component.

Type 3 aggregation (aggregation across biodiversity components) is likely to apply to Scenarios A & B (e.g. combining different specific habitats that together constitute biogenic reef, or combining different bird species or groups), Scenario C is a single species assessment and it is unlikely that this aggregation type would be involved in the assessment process.

Type 4 aggregation (aggregation across indicators/indices within assessment criteria) will apply to Scenarios A & C. Scenario A incorporates CSM and will therefore look to aggregate multiple attributes under the 'structure & function' parameter. In addition, the 'future prospects' parameter will require the aggregation of estimated future states of the 'range', 'area' and 'structure & function' parameters. For the purposes of this report, consideration of Scenario B is restricted to the production of a population abundance indicator; Type 4 aggregation would apply where such indicators are subsequently aggregated across biological dimensions (for example, species associated with particular habitat types, such as intertidal estuarial areas, sea cliffs, *etc*., or with certain feeding traits).

Finally, Type 5 aggregation (aggregation across assessment criteria of different biodiversity components to deliver an overall assessment) does not strictly apply to any of the three scenarios when considered in isolation. This aggregation type would apply to the integration of more than one separate scenarios (for example, combining an assessment for birds with an assessment for (one or more) highly mobile species. Type 5 aggregations are essentially the top level assessments (for example under MSFD or WFD) – as such, the lower level aggregations discussed here are likely to feed into Type 5 aggregations but are not strictly, in their own right, represented within the Scenarios discussed.

Together with Table 5.1, the earlier assessment of the applicability of aggregation methods across the different aggregation types (Table 3.4) can be used to infer options for aggregation under each of the scenarios that are considered.

[13] Where temporal data are derived from a time series that includes more than one reporting period, it is anticipated that aggregation (processing) should be restricted to those data that relate to the reporting period of interest (effectively excluding data that lie outside of the reporting period timeframe).

Table 5.12 Aggregation types within defined reporting scenarios.

Aggregation type (*sensu* JNCC)	Scenario		
	A – single habitat assessment (e.g. Annex I reefs)	B – multi-species assessment (e.g. seabird abundance)	C – highly mobile species assessment (e.g. harbour porpoise population)
Type 1: Aggregation over different spatial scales or assessment units	✓	✓	✓
Type 2: Aggregation across different temporal scales	-	✓	✓
Type 3: Aggregation across biodiversity components	✓	✓	-
Type 4: Aggregation across indicators/ indices within assessment criteria	✓	-[†]	✓
Type 5: Aggregation across assessment criteria of different biodiversity components to deliver an overall assessment	*	*	*

* Type 5 aggregations are not strictly applicable to the individual scenarios as presented as this aggregation relates more to high level (e.g. overall MSFD-type assessment); see text for details
[†] as discussed in the text, it is acknowledged that Type 4 aggregation may be applied to support certain instances of the reporting of seabird abundance but the focus within this report is on an example of single species reporting

5.3 Scenario A: single benthic habitat assessment

5.3.1 Scenario description

This section considers the annual assessment and reporting of a single (complex) habitat type at the scale of a single MPA, based on the principles of Common Standards Monitoring (CSM). Specifically, it considers the reporting of Annex I reef habitat - rocky marine habitat or biological concretion that rises from the seabed. Reefs are very variable in form and in the communities that they support, although two main subtypes (rocky reefs and biogenic reefs) are recognised (Jackson & McLeod 2000, 2002)[14].

Under CSM reporting the overall condition of a feature such as an Annex I reef is based on the condition of constituent sub-features, for example *Modiolus modiolus* biogenic reef and bedrock reef (JNCC 2004a, 2004b). In turn, these sub-features are assessed on the basis of defined attributes (derived from a focussed and prioritised list of all attributes for the feature that most efficiently define its expected condition at a site) and associated attribute targets (e.g. Table 1 in JNCC 2004b). The generic attributes that should be used to define the condition of littoral rock and inshore sublittoral rock features (including, for example, the Habitats Directive Annex I habitat type H1170 Reefs) are listed within JNCC 2004b as:

- **Mandatory attributes:**
  - Extent;
  - biotope composition of the littoral rock and inshore sublittoral rock;
  - distribution of biotopes: spatial arrangement of biotopes at specified locations;

[14] JACKSON, D.L. & McLEOD, C.R. (Editors), 2000, 2002 Handbook on the UK status of EC Habitats Directive interest features: provisional data on the UK distribution and extent of Annex I habitats and the UK distribution and population size of Annex II species, Revised 2002, JNCC Report 312, 180 pp.

- **Optional site-specific attributes[15]:**
  - Extent of sub-feature or representative/notable biotopes;
  - presence of representative/notable biotopes;
  - species composition of representative or notable biotopes; and
  - presence and/or abundance of specified species.

The suggested decision processes to assess sub-feature condition (JNCC 2004a) requires the condition of attributes to be defined (‘Favourable condition’ or ‘Unfavourable condition’) and these then to be aggregated to determine the condition of the sub-feature based on a default-based approach (i.e. an OOAO aggregation). The sub-feature condition assessment is then refined by a second decision process to determine the status of the sub-feature (‘Partially destroyed/destroyed’, ‘Unfavourable declining’, ‘Unfavourable maintained or recovering’, ‘Favourable recovered’ or ‘Favourable maintained’) (JNCC 2004a).

It is obvious, however, that reporting the condition status of Annex I reef potentially requires the consideration and aggregation of assessment of a number of separate (composite) habitat types. Two distinct paradigms are presented that both address the aggregation of independent sub-feature/attribute assessments, an aggregation that, in turn, underpins the overall assessment of condition status for certain habitat features:

- Option A.1 (a schematic overview of which is presented as Figure 5.1) is based on JNCC guidance (JNCC 2004a). It represents a process whereby condition assessments are undertaken independently for each of the sub-features that contribute to the main feature of interest, with an ultimate (spatial) aggregation of condition status assessments across sub-features to derive an overall feature condition assessment.
- Option A.2 presents an alternative approach that supports the same overall assessment, but which aggregates condition assessments for each attribute across the range of sub-features that have been identified, producing condition assessments for each attribute at the level of the overall feature (Figure 5.2). Subsequent overall feature condition assessment is made based on these aggregated attribute condition assessments.

i. Option A.1: detail

Under Option A.1, the attribute condition for each sub-feature is defined with reference to stated targets and is classified as ‘Favourable’ or ‘Unfavourable’. This is a largely qualitative (or at best semi-quantitative) process which is based to a great extent on expert judgement.

Aggregation of the condition assessments across the attributes for each sub-feature is currently undertaken on the basis of OOAO, producing a sub-feature condition assessment (‘Favourable’ or ‘Unfavourable’).

Each derived sub-feature condition assessment is then refined by a second decision process (described within JNCC 2004a), again making extensive use of expert judgement, to determine conservation status (‘Destroyed/partially destroyed’, ‘Unfavourable declining’, ‘Unfavourable maintained or recovering’, ‘Favourable recovered’ or ‘Favourable maintained’).

Individual sub-feature conservation status assessments are then aggregated to derive an overall assessment of feature condition (i.e. a spatial aggregation). There is no formal guidance on this stage of the process, but it is anticipated that some element of weighting would be involved (correcting, on an area-based basis, for the relative contributions made by different sub-features, and taking into the account the intrinsic importance or value of each sub-feature within the MPA site).

---

[15] Used only where they reflect the conservation interest of the individual site.

ii. Option A.2: detail

As for Option A.1, the first stage of the process as presented under Option A.2 is to classify the attribute condition for each sub-feature with reference to stated targets; classification is ‘Favourable’ or ‘Unfavourable’. This is a largely qualitative (or at best semi-quantitative) process which is based to a great extent on expert judgement.

Aggregation of the condition assessments across the sub-features for each attribute (i.e. the spatial aggregation stage) is then undertaken (for example on the basis of OOAO or a similar approach), producing an attribute condition assessment (‘Favourable’ or ‘Unfavourable’). Although there is no formal guidance to base this stage of the process on, it is anticipated that some element of weighting would be involved. This would correct, on an area-based basis, for the relative contributions made by different sub-features, and would take into the account the intrinsic importance or value of each sub-feature within the MPA site. In this context, although it may embody OOAO, this stage of the process would be likely to involve expert judgement and a more detailed decision tree approach.

Each derived attribute condition assessment is then refined by a second decision process (as described within JNCC 2004a), again making extensive use of expert judgement, to determine a conservation status assessment (‘Destroyed/partially destroyed’, ‘Unfavourable declining’, ‘Unfavourable maintained or recovering’, ‘Favourable recovered’ or ‘Favourable maintained’) for each attribute under consideration.

The conservation status assessments for the attributes are then aggregated to derive an overall assessment of feature condition. With each attribute being classified to any of five different categories it is suggested that this stage could be best based on the application of a decision tree approach.

Sub-feature	Attributes		Attribute condition		Sub-feature condition		Sub-feature status		Feature condition
X (e.g. rocky reef)	$A_{1.X}$ (e.g. extent)	→	$C_{A1.X}$	→	C_X	→	S_X	→	C_F (feature condition)
	$A_{2.X}$ (e.g. biotope composition)	→	$C_{A2.X}$						
	$A_{3.X}$ (e.g. distribution of biotopes)	→	$C_{A3.X}$						
Y (e.g. biogenic reef)	$A_{1.Y}$ (e.g. extent)	→	$C_{A1.Y}$	→	C_Y	→	S_Y		
	$A_{2.Y}$ (e.g. biotope composition)	→	$C_{A2.Y}$						
	$A_{3.Y}$ (e.g. distribution of biotopes)	→	$C_{A3.Y}$						

Key (see text for more detail):

→	Expert judgement assessment with respect to defined targets; attribute condition defined as: *'Favourable'* or *'Unfavourable'*
→	Aggregation using OOAO (although can be over-ridden or modified with expert judgement weighted for area/importance/value of separate contributions – see main text); sub-feature condition defined as: *'Favourable'* or *'Unfavourable'*
→	Expert judgement assessment based on defined process presented as flow chart in JNCC 2004a; sub-feature 'reporting status' defined as: *'Destroyed/partially destroyed', 'Unfavourable declining', 'Unfavourable maintained or recovering', 'Favourable recovered'* or *'Favourable maintained'*
→	Spatial aggregation; no clear process guidance available although likely to involve some form of weighting; final overall assessment of feature condition defined as: *'Destroyed/partially destroyed', 'Unfavourable declining', 'Unfavourable maintained or recovering', 'Favourable recovered'* or *'Favourable maintained'*

Figure 5.7 Schematic representation of Annex I reef feature assessment process under CMS (Option A.1) based on composite habitat features ('X' and 'Y').

Attributes	Attribute condition at the level of:		Status	Feature condition
	Sub-feature	Feature		
A_1 (e.g. extent)	→ $C_{A1.X}$ → $C_{A1.Y}$	→ $C_{A1.F}$	→ $S_{A1.F}$	
A_2 (e.g. biotope composition)	→ $C_{A2.X}$ → $C_{A2.Y}$	→ $C_{A2.F}$	→ $S_{A2.F}$	→ C_F (feature condition)
A_3 (e.g. distribution of biotopes)	→ $C_{A3.X}$ → $C_{A3.Y}$	→ $C_{A3.F}$	→ $S_{A3.F}$	

Key (see text for more detail):

→	Expert judgement assessment with respect to defined targets; attribute condition defined as: *'Favourable'* or *'Unfavourable'*
→	Spatial aggregation using OOAO or decision tree approach (to assist with inclusion of expert judgement and to incorporate weighting for area/importance/value of different sub-features – see main text); sub-feature condition defined as: *'Favourable'* or *'Unfavourable'*
→	Expert judgement assessment based on defined process presented as flow chart in JNCC 2004a and incorporating expert judgement; sub-feature 'reporting status' defined as: *'Destroyed/partially destroyed', 'Unfavourable declining', 'Unfavourable maintained or recovering', 'Favourable recovered'* or *'Favourable maintained'*
→	No clear process guidance available but potentially achieved through application of decision tree approach; final overall assessment of feature condition defined as: *'Destroyed/partially destroyed', 'Unfavourable declining', 'Unfavourable maintained or recovering', 'Favourable recovered'* or *'Favourable maintained'*

Figure 5.8 Schematic representation of alternative Annex I reef feature assessment process under CMS (Option A.2) based on composite habitat features ('X' and 'Y').

iii. Influencing outcomes: predetermination and potential bias

As noted earlier, the intention here is to examine the application and implications of different aggregation methods within current reporting needs, and to consider the benefits and problems associated with alternative aggregation (or process) options.

In aggregating data there is, as well as the aim of simplifying the available information and producing one or more aggregated metrics that are relatively easily presented and interpreted, a parallel aim is to maintain the maximal information within the aggregated outputs. When alternative choices are available regarding the order in which data are aggregated there should be a presumption in maintaining as much information as far through the course of the process as is possible. This would suggest that, for the initial aggregation, it is more appropriate to consider aggregating across the least variable data.

In the example outlined here, two distinct options have been identified for developing a process to accommodate spatial aggregation across (local scale) features that constitute a habitat feature at the scale of, for example, an MPA. Selection between these options would involve a judgement being made regarding the relative variability of attribute condition at the level of sub features. Where there is more variability in sub-feature attribute condition apparent across sub-features than across attributes, initial aggregation across attributes (i.e. Option A.1) should be considered. Conversely, where it is felt that there is more variability in sub-feature attribute condition apparent across attributes than across sub-features, initial aggregation across sub-features (i.e. Option A.2) should be considered.

For both alternative options however, the principle choices of underlying methods remain the same.

The second stage of both processes requires the use of a qualitative approach such as OOAO to aggregate condition assessments of attribute/sub-feature combinations (whether this is across attributes as in Option A.1, or a spatial aggregation across sub-features as in Option A.2). In both instances, individual elements take one of two possible states, or conditions: ‘Favourable’ or ‘Unfavourable’.

Similarly the final stage of both processes calls for the aggregation of status assessments (either a spatial aggregation across sub-features as in Option A.1, or across attributes as in Option A.2). In both instances, individual elements take one of five possible states: Destroyed/partially destroyed’, ‘Unfavourable declining’, ‘Unfavourable maintained or recovering’, ‘Favourable recovered’ or ‘Favourable maintained’.

The implications of different aggregation approaches for both of these stages are considered briefly below.

Firstly, the issue of aggregating across the condition assessments of attribute/sub-feature combinations (where each combination is assessed as either ‘Favourable’ or ‘Unfavourable’) is considered. This takes an instance where three elements are being aggregated (i.e. either aggregating across three attributes, or across three sub-features). Under these conditions a range of possible outcomes can be envisaged, as shown in Table 5.2. For the purpose of this exercise, two alternative aggregation methods, OOAO and TOAO, are being considered as examples of how the final aggregation might be realised.

Table 5.13 Possible outcomes for assigning each of three elements to one of two categories, together with aggregated assessments under alternative regimes.

Potential combinations of assessments:			Aggregated assessments derived through the application of different aggregation methods	
Element #1	**Element #1**	**Element #1**	**OOAO**	**TOAO**
Fav	Fav	Fav	Fav	Fav
Fav	Fav	Unfav	Unfav	Fav
Fav	Unfav	Fav	Unfav	Fav
Fav	Unfav	Unfav	Unfav	Unfav
Unfav	Fav	Fav	Unfav	Fav
Unfav	Fav	Unfav	Unfav	Unfav
Unfav	Unfav	Fav	Unfav	Unfav
Unfav	Unfav	Unfav	Unfav	Unfav

NB: categorisations in the above table are given as: Favourable (Fav); or Unfavourable (Unfav).

From inspection of Table 5.2, the use of OOAO in this example is clearly precautionary in nature, with just one of the eight possible combinations being promoted as 'Favourable'. A less restrictive rule, such as TOAO, provides a more relaxed assessment, giving a 'Favourable' output 50% of the time. This example assumes that all three attributes, or all three sub-features, are of equal importance. In practice there is likely to be more a role to be played by expert judgement – either weighting for the different areas and relative importance of different sub-features, or for the perceived importance of the different attributes. In this context it is likely that a decision tree approach would be optimal at this stage in the process, but the strong (and perhaps over-precautionary) nature of OOAO should be noted.

Given three elements (which might be attributes or sub-features), each assigned to one of five categorised states, and assuming that all parameters have equal importance, there are 35 potential combinations of categories that need to be considered in terms of an overall (cross-parameter) assessment. The overall assessment outputs for each of these possible combinations are shown in Table 5.3 for an OOAO approach, together with a threshold style approach that seeks to report the mid-way or 50% value (equivalent in this instance to a TOAO approach). It is assumed, for the purpose of this exercise, that the five potential status categories are hierarchical, ordered from 'Destroyed/partially destroyed' at the lowest level, up to 'Favourable maintained' at the highest level.

Table 5.14 Comparison of different methods for deriving overall feature condition assessment through categorisations of three contributary elements.

Number of elements categorised as:					Overall assessment following different aggregation methods:	
D/PD	**UD**	**UMR**	**FR**	**FM**	**OOAO**	**50% threshold**
0	0	0	0	3	FM	FM
0	0	0	1	2	FR	FM
0	0	1	0	2	UMR	FM
0	1	0	0	2	UD	FM
1	0	0	0	2	D/PD	FM
0	0	0	2	1	FR	FR
0	0	2	0	1	UMR	UMR
0	2	0	0	1	UD	UD
2	0	0	0	1	D/PD	D/PD
0	0	1	1	1	UMR	FR
0	1	0	1	1	UD	FR
1	0	0	1	1	D/PD	FR
0	1	1	0	1	UD	UMR
1	0	1	0	1	D/PD	UMR
1	1	0	0	1	D/PD	UD
0	0	0	3	0	FR	FR
0	0	1	2	0	UMR	FR
0	1	0	2	0	UD	FR
1	0	0	2	0	D/PD	FR
0	0	2	1	0	UMR	UMR
0	2	0	1	0	UD	UD
2	0	0	1	0	D/PD	D/PD
0	1	1	1	0	UD	UMR
1	0	1	1	0	D/PD	UMR
1	1	0	1	0	D/PD	UD
0	0	3	0	0	UM	UMR
0	1	2	0	0	UD	UMR
1	0	2	0	0	D/PD	UMR
0	2	1	0	0	UD	UD
2	0	1	0	0	D/PD	D/PD
1	1	1	0	0	D/PD	UD
0	3	0	0	0	UD	UD
1	2	0	0	0	D/PD	UD
2	1	0	0	0	D/PD	D/PD
3	0	0	0	0	D/PD	D/PD

NB: categorisations in the above table are given as: Destroyed/partially destroyed (D/PD), Unfavourable declining (UD); Unfavourable maintained or recovering (UMR), Favourable maintained (FM); or Favourable recovered (FR).

The data presented in Table 5.3 are further summarised as Table 5.4.

Table 5.15 Summary comparison of outputs from different methods for deriving overall feature condition assessment from categorisations of three contributary elements.

Overall (cross-parameter) condition status assessment	Distribution of range of possible assessments derived through the application of different aggregation methods:	
	OOAO	**50% threshold***
D/PD	15	5
UD	10	8
UMR	6	9
FR	3	8
FM	1	5

NB: categorisations in the above table are given as: Destroyed/partially destroyed (D/PD), Unfavourable declining (UD); Unfavourable maintained or recovering (UMR), Favourable maintained (FM); or Favourable recovered (FR).

Once again, the strong precautionary nature of OOAO can be seen in the results presented in Table 5.4. In the example used, the use of the alternative (50%) threshold method provides a more even distribution of occurrences of different outcomes for the overall condition status assessment. In this context, whilst this latter method (which is less precautionary) might be considered to perform better (in the sense that it a much better distribution of frequency of outcomes), it may not be provide the necessary discrimination required by conservation assessments.

5.4 Scenario B: multi-species assessment

5.4.1 Scenario description

This section considers the assessment of multiple species at the biogeographic (regional sea) scale, reporting every six years (for example under MSFD or OSPAR)[16]. As an example it takes the assessment of trends in the relative abundance of non-breeding and breeding seabirds.

Two fundamental approaches to deriving a spatially aggregated multi-species assessment are considered (Figure 5.3). These represent the two options for combining multi-species assessments across distinct reporting areas (i.e. aggregating across areas first, then across species; or across species and then across areas):

- Option B.1 - for each species, aggregate across subdivisions (smaller areas that are monitored or assessed independently and which, together, represent the full range of the reporting area) to derive a series of species specific aggregations at a wider (biogeographic) spatial scale, with a subsequent aggregation across species to derive an overall seabirds assessment; or
- Option B.2 - for each spatial subdivision, aggregate across species to derive a series of subdivision values, then aggregate across subdivisions to derive an overall assessment at the biogeographic spatial scale.

The current suggested approach for spatial assessments and aggregations for marine birds (as described in the MSFD Indicator Technical Specification 25v.5; OSPAR 2014), follows the first of these two options. For any given year's assessment, this approach calls for a range of information:

[16] Variation in the spatial definition of reporting areas between different reporting frameworks (e.g. OSPAR and MSFD) will mean that, inevitably, the smaller 'subdivisions' that represent the basic monitoring areas will not always coincide (spatially) with reporting areas. This potential mis-match is discussed under Section 5.6.

$A_{i.j}$ = annual abundance estimate of species *i* in subdivision *j* (actually composed of the sum of constituent colony counts from within subdivision *j*). Values for $A_{i.j}$ are (preferentially) based on direct observation. Where data are not available for specific colonies of species *i*, trend models are used to estimate surrogate values.

$B_{i.j}$ = baseline population value for species *i* in subdivision *j* (derived in one of three ways: an historic value taken from when anthropogenic impact was minimal but climatic conditions reflected those that currently prevail; mean of a defined time series; or current values [but subsequently amended as the mean of a new, extending time series]. In practice, the value $B_{i.j}$ is likely to be the sum of constituent colony baselines values from within subdivision *j*).

= $A_{i.j}$ expressed as a percentage of $B_{i.j}$
= $A_{i.j}$ x 100 / $B_{i.j}$

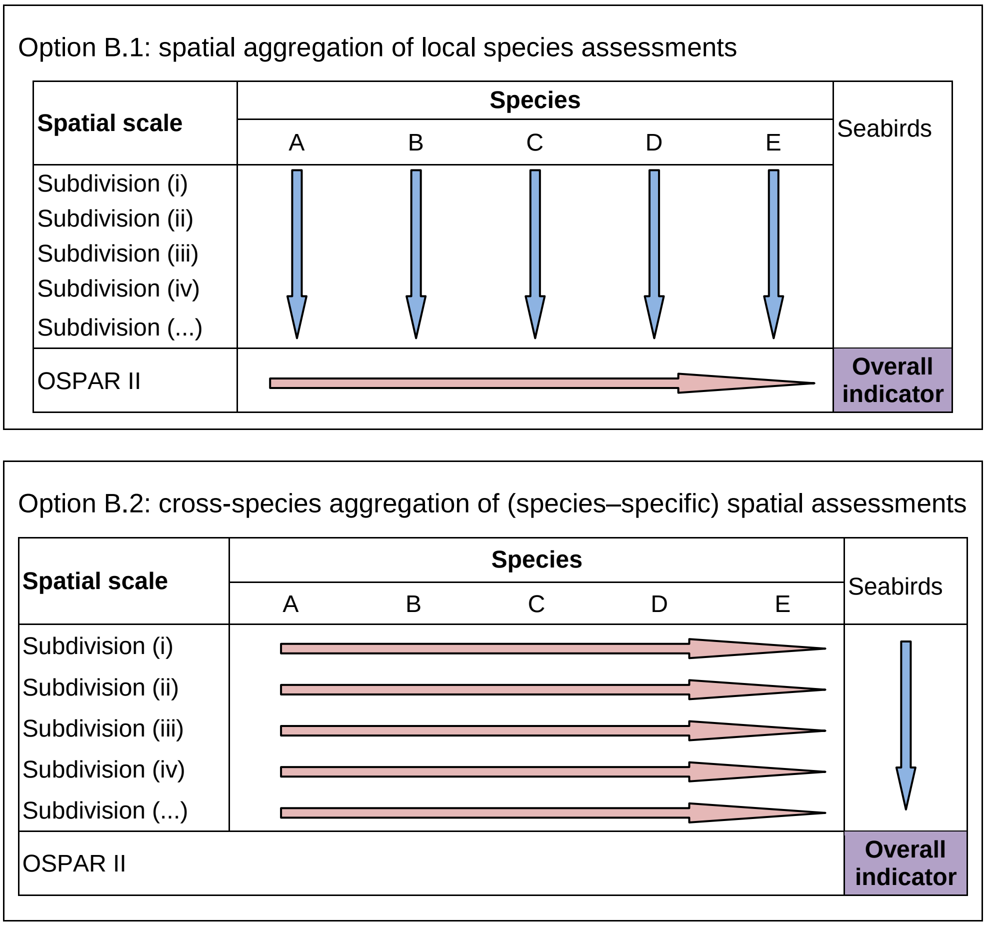


Figure 5.9 Schematic outlining derivation of a spatially aggregated multi-species assessment.

Either approach to multispecies assessment will encompass two of the five aggregation types:

- Type 1, aggregation over different spatial scales or assessment units (aggregating up from local data to the regional scale); and
- Type 3, aggregation across biodiversity components (aggregating over species or species groups).

It may potentially also involve a Type 2 (temporal) aggregation, although this is not considered here. In practical terms, although (annual) time series data may be available, these are not typically used in population abundance assessments (JNCC, pers. comm.). Time series data are used for trend assessment but only the most recent season's data are used for abundance estimates. For breeding success however, the number of years in which target breeding success has been achieved is used as an indicator (reflecting the variability of breeding success due to both anthropogenic and natural factors). This effectively involves a Type 2 aggregation, with data being compared a across a six year period. The actual metric used is simply the number of years where breeding success has met or exceeded the target. Arguably this is not an aggregation *per se*, but rather just a summary statistic. Further discussion on the application of temporal application is presented under Scenario C.

For an aggregated multi-species assessment at the scale of (for example) OSPAR regional subdivisions, the following approach is described in the MSFD Indicator Technical Specification (OSPAR 2014):

- Discrete values for $I_{i.j}$ (the indicator metric for species *i* in subdivision *j*) are calculated for each species within a subdivision. Each value of $I_{i.j}$ is then tested against a target (80% for species that lay their eggs singly, or 70% for species that lay clutches comprising two or more eggs).
- Values for $I_{i.j}$ are also tested against an upper target of 130%; instances where $I_{i.j}$ >130% are considered to meet a predefined target (the Ecological Quality Objective, or EcoQO, target) but are flagged as representing potentially disruptive increases (in species *i*) that may adversely impact on other species.
- The proportion of species exceeding their lower target - i.e. the percentage of values of $I_{i.j}$ within subdivision *j* that exceed the lower target ($I_{j.\mathrm{PASS}}$ %) - is calculated and compared to a GEnS target of 75%.

  $I_{j.\mathrm{PASS}}\ \% = \#(I_{i.j.\mathrm{PASS}}) \times 100 / (\#(I_{i.j.\mathrm{PASS}}) + (\#(I_{i.j.\mathrm{FAIL}}),$

  where:

  $\#(I_{i.j.\mathrm{PASS}})$ = number of values of $I_{i.j}$ exceeding their associated target, and
  $\#(I_{i.j.\mathrm{FAIL}})$ = number of values of $I_{i.j}$ failing their associated target.

- If the calculated value for $I_{\mathrm{PASS}.j}$ meets or exceeds the 75% target, subdivision *j* is held to have achieved GEnS.

For an aggregated multi-species assessment at the wider scale of OSPAR II region (regional sea) the following approach is described in the MSFD Indicator Technical Specification:

- Values for $I_{i.j}$ are calculated for each species within a subdivision.
- Values for $I_{i.j}$ are then aggregated for each species (*i*) across all subdivisions (1-*j*) using $B_{i.j}$ as a weighting (reflecting the relative importance of subdivision *j* to species *i*), producing weighted mean values, $\bar{I}_i$, for each species.
- Each value for $\bar{I}_i$, is assessed against its target (80% for species that lay their eggs singly, or 70% for species that lay clutches comprising two or more eggs) to derive values for the number of species passing, $\#(\bar{I}_{i.\mathrm{PASS}})$, and failing, $\#(\bar{I}_{i.\mathrm{FAIL}})$.
- Values for $\bar{I}_i$ are also tested against an upper target of 130%; instances where $\bar{I}_i$ >130% are considered to meet the EcoQO target but are flagged as they represent potentially disruptive increases (in species *i*) that may adversely impact on other species.
- The final assessment statistic is a calculated as the overall percentage pass rate:

  $\bar{I}_{\mathrm{PASS}}\% = \#(\bar{I}_{i.\mathrm{PASS}}) \times 100 / (\#(\bar{I}_{i.\mathrm{PASS}}) + \#(\bar{I}_{i.\mathrm{FAIL}})),$

where $\bar{I}_{\mathrm{PASS}}$% meets or exceeds the 75% target, the OSPAR region is held to have achieved GEnS.

Figure 5.4 shows the process as a simplified flow chart for the aggregation process(es) underpinning a conceptual multi-species assessment, such as those undertaken for seabirds.

As described above, the process uses data that describe both the current annual abundance and baseline population values for each of species 1-*i* in each of the subdivisions, 1-*j*, of the biogeographic reporting region (e.g. OSPAR II region). The four potential aggregating operation points shown in Figure 5.4 are considered further below.

1 : involves normalising the data, by comparing current data with baseline data to produce relative abundance estimates, effectively providing a standardised measure of population change (trend) for each species in each subdivision. The method currently applied is a simple expression of abundance as a % of the baseline. It is assumed that, in situations where there are repeat observations of abundance within a six year reporting period, multiple values (for a given species in a given subdivision) are averaged.

The only alternatives to this step would be to use non-normalised data – but this would allow population variations in what might be sub-optimal areas of the region (i.e. those with a low baseline population value) to have an inappropriate influence on the final aggregated indicator.

2 : this step is a non-hierarchical weighted averaging process. It involves the aggregation of individual relative species abundance estimates for each subdivision, weighted by their corresponding baseline values, to derive relative species abundance estimates at the OSPAR II regional scale. The baseline values are used a measure of the importance of each of the subdivisions for each species.

As discussed above, the relative abundance estimates need to be weighted to reflect the importance of different subdivisions (or different colonies within a subdivision) – for example by using a weighting index based on a long-term time-series[17] for subdivision population sizes. The only alternative weighting that could be used would be an area-based value, but this would assume that all the species are potentially evenly distributed. The contagious (clumped) distribution of birds indicates that spatial factors such as habitat variability are likely to play a role in determining the distribution of bird populations and that weighting based on subdivision area alone would be inappropriate.

Alternatives to this approach might include the use of modelled weighting factors based, for example, on habitat type and potential for disturbance. Whilst the use of baseline population values is problematic (in terms of how the baseline is derived) they nevertheless provide the most pragmatic estimate of the relative worth of different subdivisions and, consequently, it is suggested that this approach is the most appropriate.

[17] The use of a long time series of data reduces the risk of downgrading the apparent importance of a subdivision that has displayed a relatively rapid change in population size.

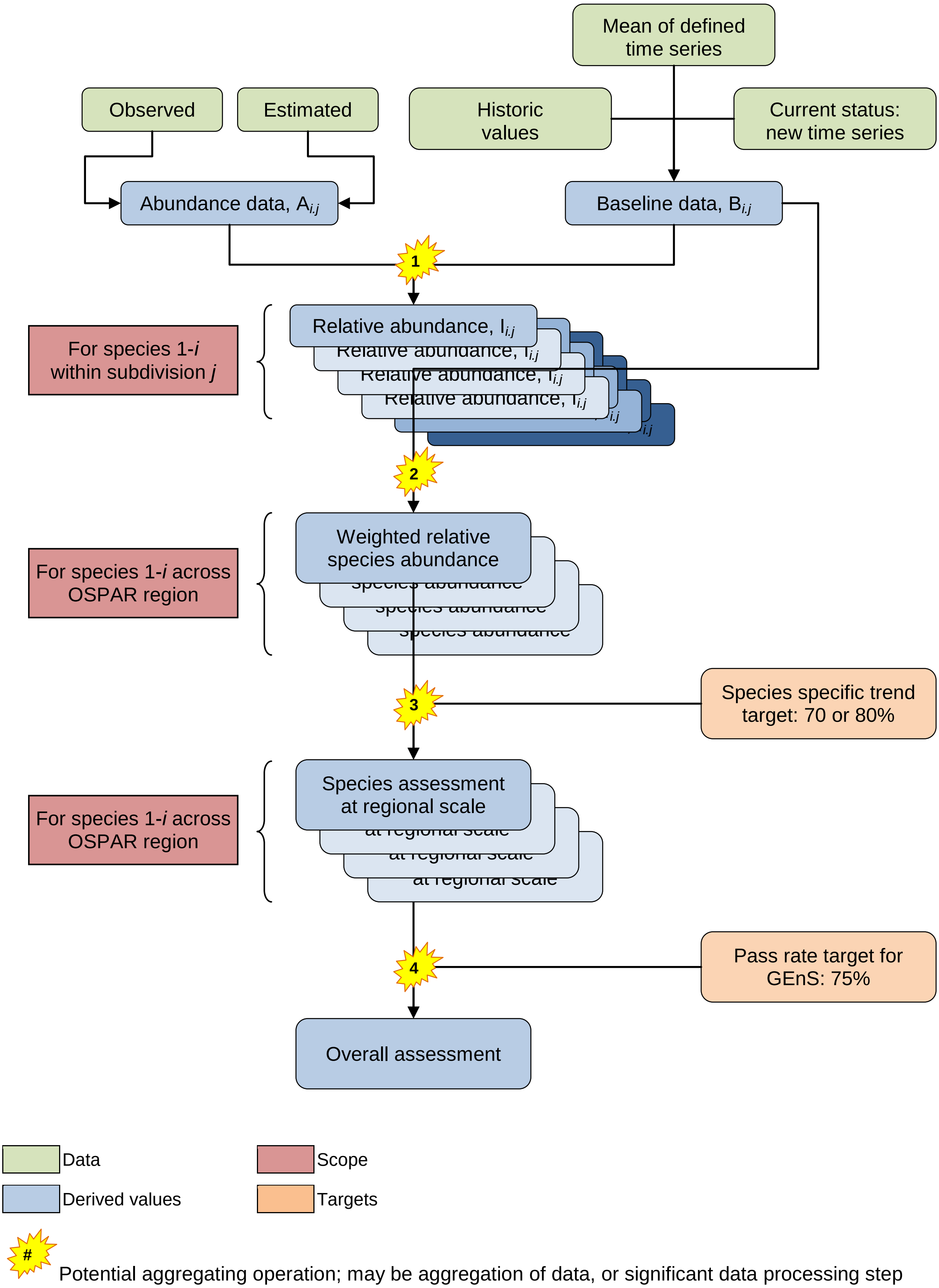


Figure 5.10 Simplified flowchart for multi-species aggregation (see text for details).

3 : involves the comparison of species-specific population (trend) estimates at the OSPAR regional scale with identified thresholds to determine whether the population of each individual species is at an acceptable level (pass), whether it is too low (fail) or whether it is sufficiently high to warrant flagging (caution). It is at this point that that the overall assessment moves from being quantitative to being categorical.

The alternative is to maintain a quantitative approach, for example by expressing the performance of each species (i.e. its weighted abundance performance relative to its corresponding baseline) as a percentage of its associated target.

4 : the final stage involves an assessment of the frequency of target failure. In essence the method currently applied is a qualitative threshold method, allied to the OOAO approach. Where 75% or more of the aggregated values (in this instance individual species as assessed across the entire biogeographic reporting region under consideration) meet (or exceed) their targets then Good Environmental Status (GEnS) is assumed.

The alternative options for this final stage of aggregation might include both quantitative and semi-quantitative approaches. At its simplest it would be possible to calculate some form of mean performance statistic (e.g. taking a non-hierarchical, non-weighted average). Such an approach would, however, lose any information on the relative performance of different species and would potentially mask even catastrophic declines in a small number of species.

5.4.2 Influencing outcomes: predetermination and potential bias

This scenario employs targets at a number of stages through the aggregation and assessment process.

For example, a 75% target is used in the (spatial) assessment of the number of species passing their ‘trend targets’. Is 75% the right target to use (or does it predetermine the outcome of the assessment)? Answering such a question (how many stocks can fail before overall good status is compromised) would require some data modelling (see Section 6.2.3) although some initial considerations are outlined below.

Within the initial assessment of compliance with trend targets, a weighting should be applied to account for a species’ value to, or importance within, each subdivision within the overall biogeographic region. This reflects the concept that a lower abundance (or poor ‘performance’) of a given species within an area that is sub-optimal for that species might be considered as making an equivalent ‘contribution’ to overall status under (for example) MSFD as a higher abundance (or better ‘performance’) of a species within an area that is optimal for the species.

Another source of potential influence would arise from the use of insufficient data, producing estimates that are subject to unacceptably large errors. In essence, how extensive should the indicator data be (i.e. how many colonies should be included) to ensure that the overall assessment of a given species is valid? The central problem is to identify the number of colonies that should be selected (at random) from within a biogeographic region in order to derive an overall population estimate for a species that deviates from the underlying ‘true’ value by an acceptable amount. This form of assessment is analogous to the power calculations that would be performed in support of classical experimental design. It is likely that such calculations would require information on typical colony sizes (e.g. mean size and variability), and the numbers and distribution of colonies within the reporting region. Despite the fact that the ‘power’ of an assessment might be identified, it must be recognised that ‘available resource’ is a key variable, and one that will inevitably dictate the maximal extent of monitoring that is possible across any given region. As a consequence, the credibility of outputs may ultimately be resource constrained.

As presented, the methods employed in this scenario do not introduce bias to the assessment outcome. The use of the threshold (qualitative) method, although allied to the OOAO approach, results in a less precautionary and more pragmatic assessment and provides a sound approach to the final stage of assessing whether GEnS been attained. The potentially large number of bird species involved reduces the tendency for a small number of ‘failures’ (which may be due to normal random population fluctuations) to trigger a failure to attain GEnS (see, for example, Figure 5.5).

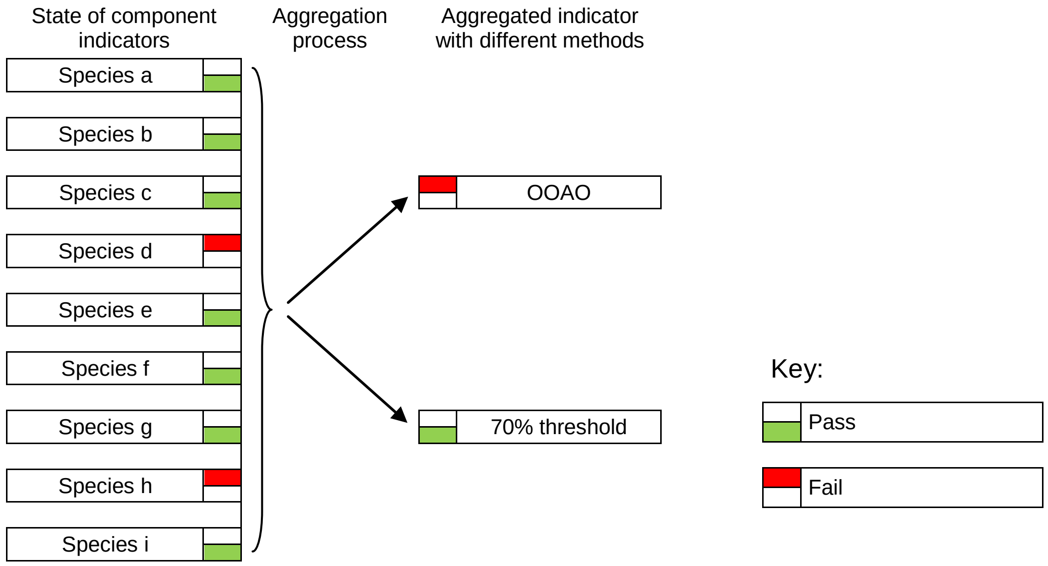


Figure 5.11 Comparision of OOAO and threshold method as applied to final, cross-species aggregation under Scenario B.

The generation of baseline population values is central to this entire scenario and their estimation is likely to have a substantial influence on the outcomes of the (aggregation) assessment.

The scenario is able to accommodate missing data, effectively by leaving out colonies or subdivisions where data are not available in any given reporting period. Whilst this would weaken the overall assessment, missing data would not prevent an assessment being made. In practice, some attempt is made to use trend models to predict values for missing data, but species are removed from the overall assessment if the trend models are felt not to be adequately robust.

To help present the maximum information from aggregated assessment consideration should be given to representing the outputs not just as a single GEnS pass/fail statistic but as a composite indicating GEnS attainment and the individual contributions made by each species. This could be done, for example, by presenting the data as a fan chart (Figure 5.6) as described by Halpern *et al* (2012). In this figure, each species is represented by a separate segment of the fan chart and the segment radius represents percentage attainment of the relevant population target (shown as a red dashed line). The overall (aggregated) assessment is shown as the central figure.

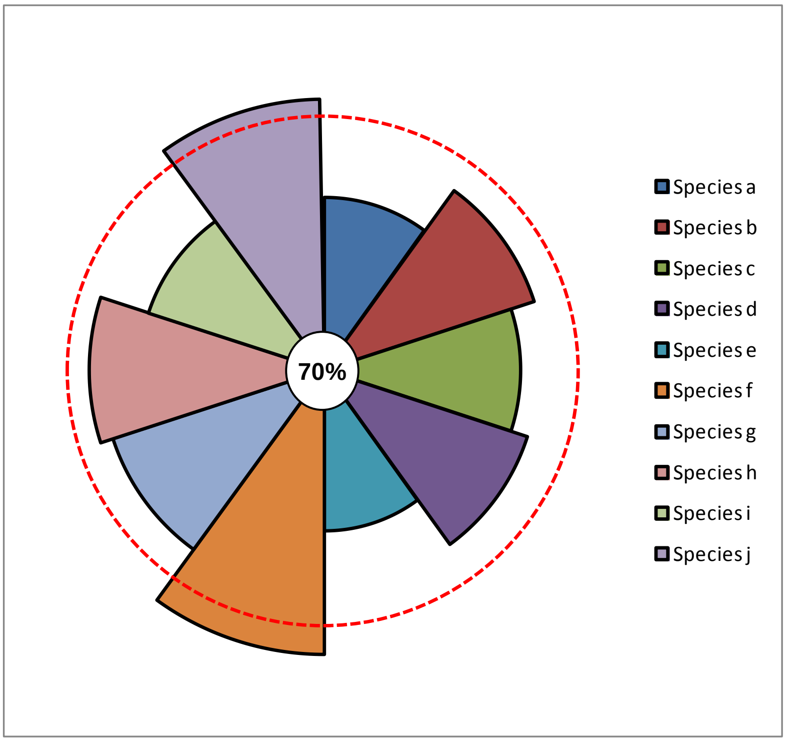


Figure 5.12 Example fan chart (see text for more details).

5.5 Scenario C: single, highly mobile species assessment

5.5.1 Scenario description

This section describes the existing aggregation approach for the assessment of conservation status (also referred to as Habitats Directive (HD) Article 17 reporting) for the harbour porpoise (*Phocoena phocoena*). In doing so the scenario considers the assessment of the harbour porpoise at the Member State (MS) scale and the biogeographical scale.

Any assessment of conservation status, under Article 17 of the HD, is carried out for each biogeographical and marine regions present in a MS on a six yearly cycle. There are five biogeographical marine regions including, the Atlantic – North east, Atlantic – Macaronesia, Baltic, Black Sea and Mediterranean (Figure 5.7). Where a MS is entirely within one biogeographical region, such as the UK, only one report is required for each habitat type and species present. If a MS is in two or more regions a report is required for each region. For example, *Bombina variegata* (yellow-bellied toad) in Germany occurs in the Alpine, Atlantic and Continental terrestrial biogeographical regions and therefore Germany will report separately for all three regions.

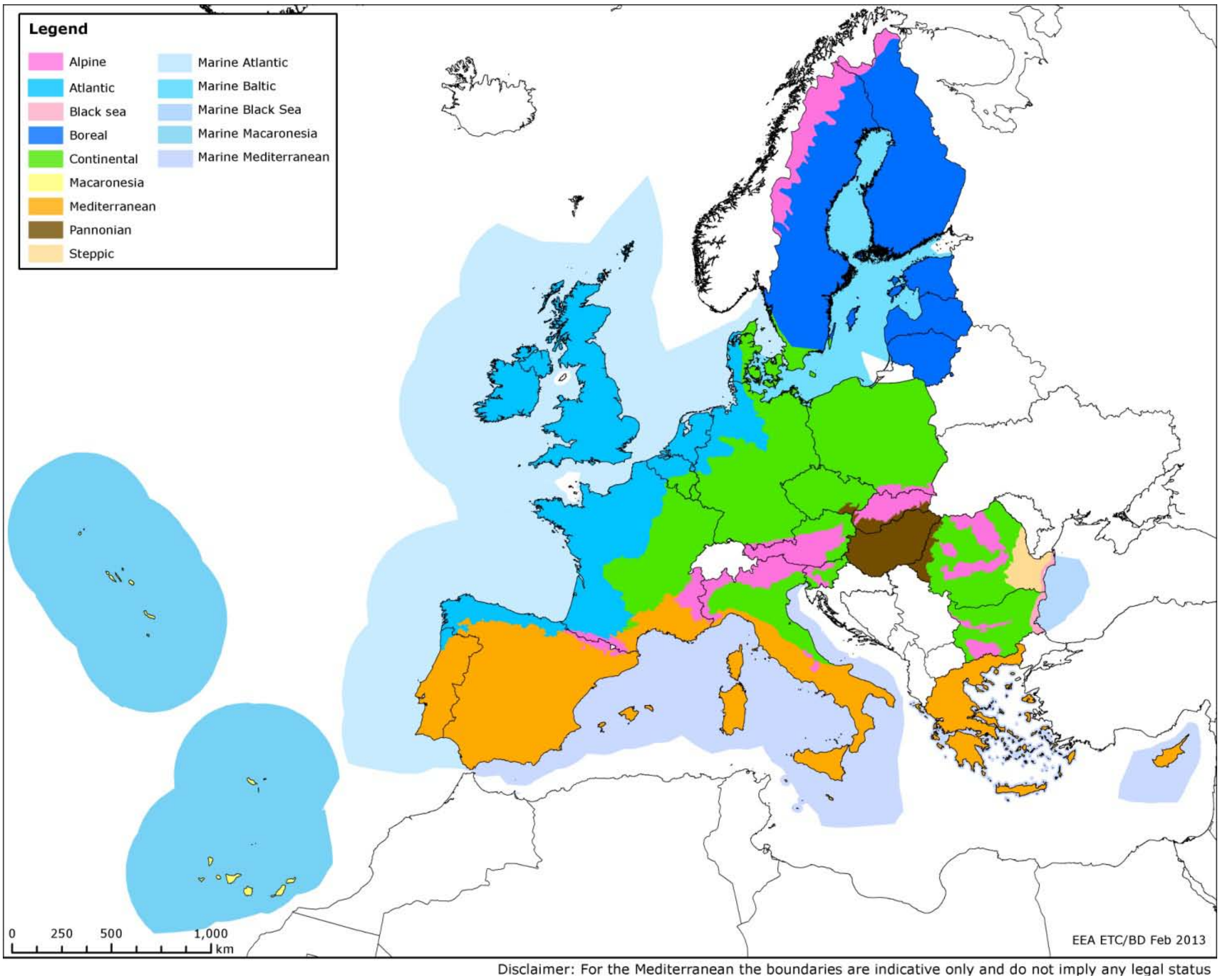


Figure 5.13 The biogeographical and marine regions for reporting under Article 17 of the Habitats Directive (reproduced from ETC/BD 2014).

As a consequence, there are two distinct and separate phases of aggregation in the assessment of conservation status. First, aggregation within MS, to provide species and habitat assessments for each portion of a biogeograhical region present in their territory (European Topic Centre on Biological Diversity, ETC/BD 2011); and second, by the European Environment Agency (EEA), aggregation across MS to provide species and habitats assessments at the scale of biogeographical regions (based on contributory MS assessments) (ETC/BD 2014). This scenario considers both aggregation phases, but temporal aggregation is explored in detail in the MS aggregation phase and spatial aggregation is explored in detail in the EEA aggregation phase.

5.5.2 Member State aggregation phase

The existing Article 17 aggregation for the harbour porpoise combines four separate parameters: range, population, habitat suitability, and future prospects (Figure 5.8). Data for the first two of these parameters (referred to in this report as parameter indicators) are quantitative, and are derived from monitoring programs and/or expert judgement. The indicators for each parameter are used to derive performance classifications (in the case of the species and population indicators these classifications are based on the production of intermediary performance indicators). The performance classifications for each of the four parameters are then used to derive an overall assessment of conservation status.

Whilst the supporting data for the habitat suitability and future prospects parameters may be quantitative, in practice these parameters are amalgamations of data with information on pressures and threats, and combining quantitative empirical data with expert judgement. Consequently supporting indicators for these two parameters should be considered to be semi-quantitative. For all four parameters, information regarding short- and long-term trends (derived, for example, through the review of published studies, assessment of available data, and expert judgement) are also considered.

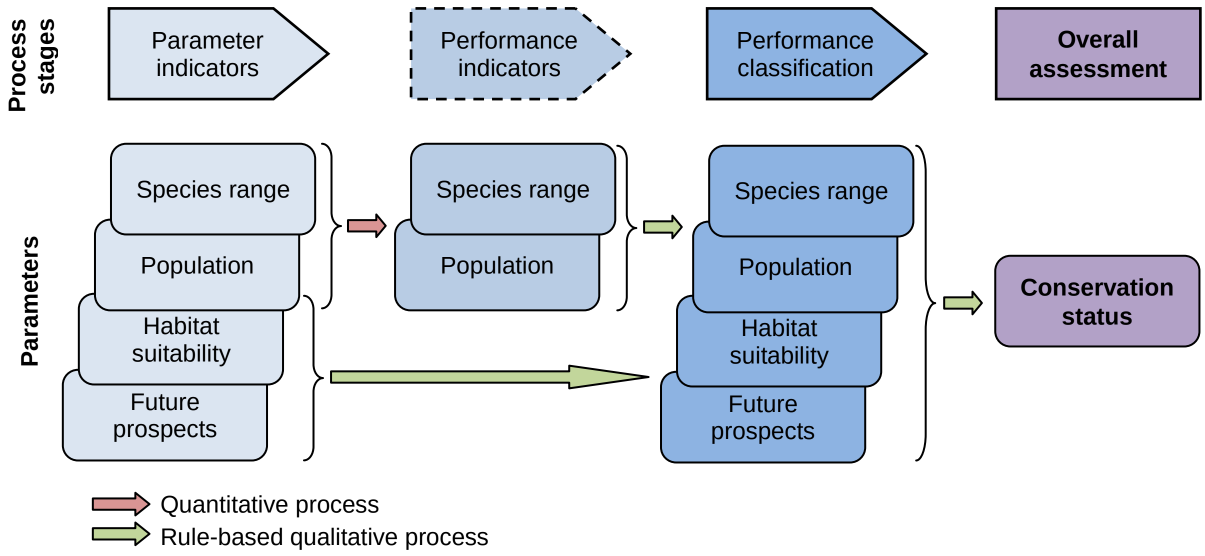


Figure 5.14 Schematic representation of assessment process for Article 17 (species) reporting under Habitats Directive (see text for details).

For the first two (quantitative) parameters there is an associated reference value that equates to the threshold at and above which the condition assessment would be 'favourable'. This value is termed the Favourable Reference Value (FRV)[18]. FRV estimates are quantitative values derived from survey data, review and expert judgement. The second pair of parameters are qualitative (derived from data review and expert judgement); both are categorised using a rule-based approach to provide associated performance classifications. All four performance classifications are then combined to derive an overall assessment (again, following a rule-based aggregation defined in the EC Reporting Guidelines, as outlined in Appendix G of this report) (Type 4 aggregation).

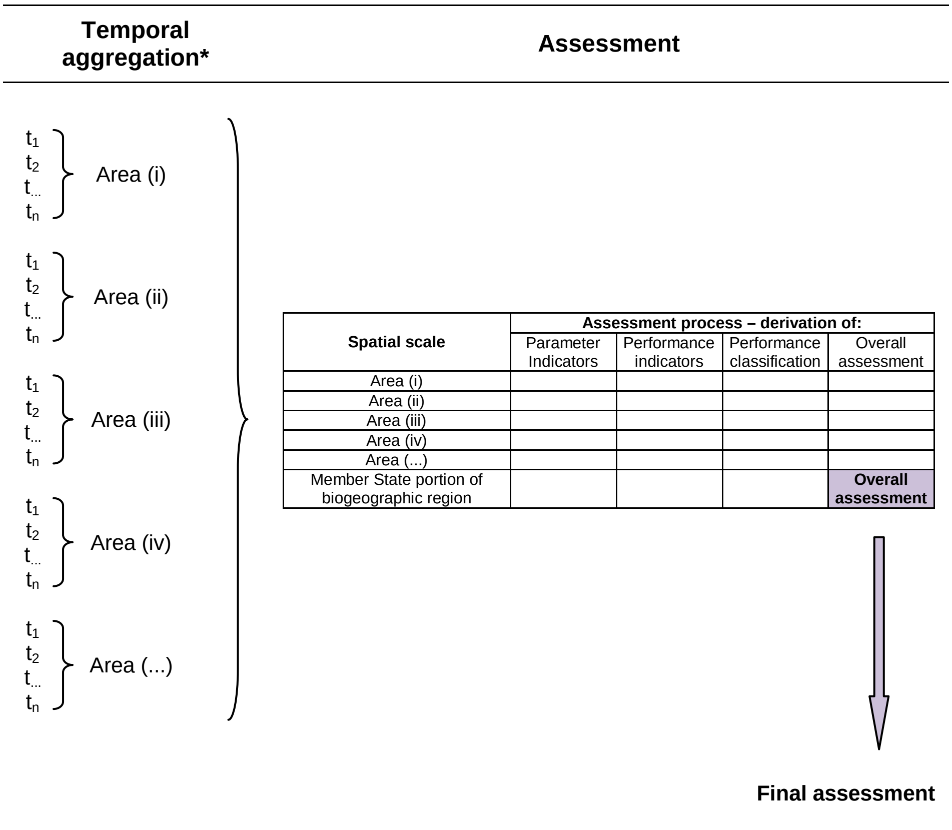
The classification of each parameter represents a form of aggregation, as several distinct pieces of information are drawn together to provide a single assessment. For example, in the case of the Range parameter; the value of the range area relative to the FRR is used together with an estimate of the short-term trend for (changes in) range, to derive a single range parameter categorisation. Although for the first three parameters this is a transparent process, aggregation for the fourth parameter ('future prospects'; assessed almost entirely on expert judgement and interpretation) is less easy to define. However, overall, given the low volume of information types being aggregated to define each parameter the current rule-based approach is likely to be optimal.

i. Temporal aggregation

Where a temporal component to the data available for any given area is available, then the inclusion of temporal aggregation will need to be considered.

[18] The acronym FRR may be used in the case of favourable reference range, and FRP in the case of favourable reference population.

There are two clear options for the inclusion of temporal aggregation within the overall assessment process: either at the beginning, undertaking the temporal aggregation of available data at the smallest spatial (local) and temporal scales followed by an assessment of the overall condition status (see Figure 5.9); or at the end, undertaking an independent assessment of overall condition status for each separate temporal component, followed by a final (temporal) aggregation (see Figure 5.10).



* for each parameter indicator within each area

Figure 5.15 Schematic representation of temporal aggregation of available data ahead of derivation of overall condition assessment.

Temporal component	Assessment	Temporal aggregation

t$_1$

Spatial scale	Assessment process – derivation of:			
	Parameter Indicators	Performance indicators	Performance classification	Overall assessment
Area (i)				
Area (ii)				
Area (iii)				
Area (iv)				
Area (...)				
Member State portion of biogeographic region				**Overall assessment**

t$_2$

Spatial scale	Assessment process – derivation of:			
	Parameter Indicators	Performance indicators	Performance classification	Overall assessment
Area (i)				
Area (ii)				
Area (iii)				
Area (iv)				
Area (...)				
Member State portion of biogeographic region				**Overall assessment**

t$_{...}$

Spatial scale	Assessment process – derivation of:			
	Parameter Indicators	Performance indicators	Performance classification	Overall assessment
Area (i)				
Area (ii)				
Area (iii)				
Area (iv)				
Area (...)				
Member State portion of biogeographic region				**Overall assessment**

t$_n$

Spatial scale	Assessment process – derivation of:			
	Parameter Indicators	Performance indicators	Performance classification	Overall assessment
Area (i)				
Area (ii)				
Area (iii)				
Area (iv)				
Area (...)				
Member State portion of biogeographic region				**Overall assessment**

Final assessment

Figure 5.16 Schematic representation of temporal aggregation of derived condition assessments.

As discussed earlier (Section 5.3) data aggregation has parallel aims: simplifying the available information and producing one or more aggregated metrics that are relatively easily presented and interpreted, whilst maintaining as much useful information as possible within the aggregated outputs.

When there is a choice regarding the order in which spatial and temporal aggregations are undertaken on the same dataset there should be a presumption in maintaining as much information as far through the course of the process as is possible. Following this principle, given a choice of the dimension (spatial or temporal) to which an initial aggregation should be applied, the least variable dimension should be selected. In the relatively stable marine environments typified by UK waters, spatial variability (arising from habitat heterogeneity) is, as a rule-of-thumb, likely to be greater than temporal variability (the marine environment being reasonably well buffered against temporal fluctuations due to temperature, nutrients and pollution). This 'partitioning' of variance was assessed by Anderson & Gribble (1998) who looked at the impact of, *inter alia*, spatial and temporal factors on the distribution of penaeid (prawn) species on the far northern Great Barrier Reef, Australia. They concluded that spatial and spatial-environmental factors accounted for 12.5% and 21.2% of total observed variability respectively, whilst temporal factors accounted for just 2.1%.

Table 5.5 provides an illustration of comparative mean values for an artificial dataset with both a spatial and temporal component: three sites (S_1, S_2 & S_3) over three time periods (T_1, T_2 & T_3). Within Table 5.5 the data at any given site at any given time are assumed to show low variability.

Data for any given site show low temporal variability (i.e. the mean value of the data remains reasonably consistent through time). This example also shows a high spatial variability (the mean value of the data varies between sites).

Table 5.16 Illustration of comparative means and degree of variability of temporally & spatially distributed data.

		Spatial data (across different sites):			
		S_1	S_2	S_3	**Spatially aggregated data**
Temporal data (across different sample dates):	T_1	Mean - high Variability - low	Mean - moderate Variability - low	Mean - low Variability - low	**Mean - moderate Variability - high**
	T_2	Mean - high Variability - low	Mean - moderate Variability - low	Mean - low Variability - low	**Mean - moderate Variability - high**
	T_3	Mean - high Variability - low	Mean - moderate Variability - low	Mean - low Variability - low	**Mean - moderate Variability - high**
	Temporally aggregated data	**Mean - high Variability - low**	**Mean - moderate Variability - low**	**Mean - low Variability - low**	**Mean - moderate Variability - high**

The implication is that, whilst either sequence of aggregation (spatial followed by temporal; or temporal followed by spatial) ultimately provides the same quality data (in this instance a ‘moderate’ mean value, with a ‘high’ variability), initial temporal aggregation produces aggregated data with ‘low’ associated variability. As it is preferable to be able to disaggregate any final output to constituent values with a relatively low variability (as this preserves more of the overall information that was available in the original ‘raw’ constituent data) it is suggested that temporal aggregation should be undertaken ahead of spatial aggregation (i.e. following the process represented in Figure 5.9). However, this conclusion is based on a simple overview and it is recommended that this is an area which should be considered for further investigation.

This approach fits well with the likely availability of data for the scenario under consideration. For any given assessment, where temporal data are available these are first aggregated to provide the requisite indicators for each parameter within a given area. The objective is to derive a set of indicators, from whatever temporal data are available, that represent best estimates over the reporting period that is being assessed for the indicators associated with each of the four parameters.

This may, for different composite areas within a wider biogeographic region, mean that both (temporally) aggregated values and single point values (where there is no repeat, temporal dimension to the data) are being carried forward in the assessment process.

For the scenario under consideration it is likely that, in practice, only the second of the four parameters (species population) will have a temporal component within any given reporting six-year period. Consequently this would be the only parameter for which a temporal aggregation would need to be undertaken, with estimates of the indicators for the other three parameters already representing the reporting period as a whole.

Aggregating temporal data as an initial step in data processing (for example by averaging the data that are available within a given area – such as the data for each of the preceding six years) facilitates subsequent disaggregation. Where new data becomes available in subsequent years it can be added and old data dropped, retaining the most contemporary set of data to derive a rolling six-year average. This approach allows a good level of flexibility and maximises the use of data between different (Type 5) aggregations employed to support major reporting obligations.

5.5.3 European Environment Agency aggregation phase

The aggregation process used by the EEA to determine the conservation status of a given species in a given biogeographical region is explained in ETC/BD (2014), and what follows here is a summary.

Ideally the assessment for each biogeographical region should follow the same method and evaluation matrices as used by the MS (see Appendix G of this report). However, in many cases only the final assessment result is available for one or more of the parameters, rather than the underlying supporting data. Where it is not possible to use the background data provided by the MS directly, the assessments of conservation status for the individual parameters from each country are weighted by the proportion of the overall population that is present in each country and then evaluated against a series of thresholds that are applied sequentially. For species, the preferred weighting is based on population size, with weighting by range being applied where population size data are not available. Where possible the four parameters are evaluated individually and then combined to give a regional assessment using the same method as used by the MS (see Appendix G, Table G.6). Where a weighting is used, the evaluation of conservation status at the biogeographical scale (i.e. aggregating across MS) is undertaken with reference to a series of thresholds, as shown in Figure 5.11.

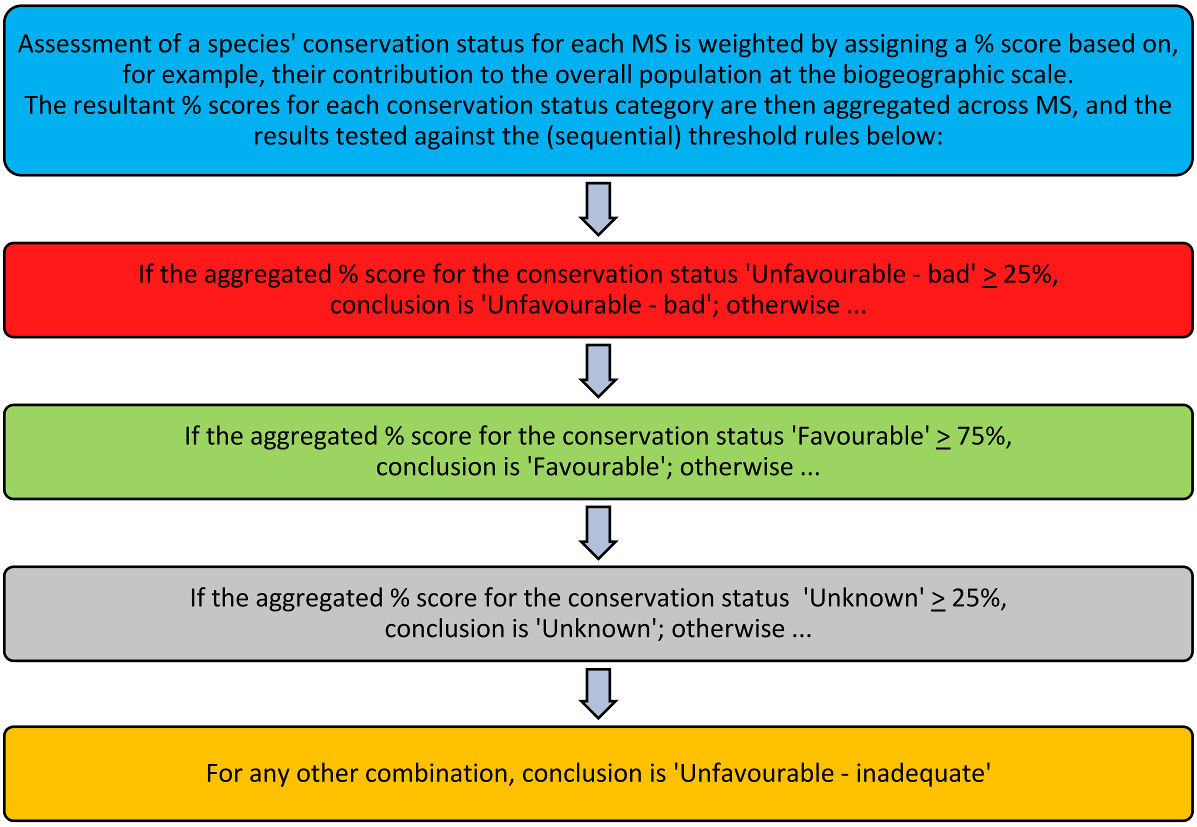


Figure 5.17 Process used by ETC/BD to determine the conservation status of individual parameters at the biogeographic scale (note sequential application of thresholds) – see text for further details.

i. Spatial aggregation

The aggregation process as described here relates to a single biogeographical region, for example the Marine Atlantic. It is necessary to combine data or assessment results for separate areas (e.g. MS territories) to derive a single assessment for reporting at a biogeographic regional scale. In theory this spatial aggregation could be achieved in several ways (see Box 3 – where horizontal red arrows represent stages in the assessment process, and vertical blue arrows represent spatial aggregation within a stage of the assessment process).

Pragmatically, it would seem appropriate to employ the same spatial aggregation option for all four parameters. This pragmatic approach (which aggregates performance indicators from across the different areas) effectively rules out option C.2 as a means of undertaking a spatial aggregation, as only two of the four parameters (species range and population) pass through the ‘Performance indicator’ stage of the assessment process. The situation is further confounded by the fact that a spatial aggregation of a parameter’s values can only be easily considered for the species range and population parameters at the (quantitative) parameter indicator and performance indicator stages of the assessment process. As such, the ETC/BD (2014) recommends varying the spatial aggregation approach by species and by parameter depending on the resolution of available data and information.

In any spatial aggregation across separate MS consideration should be given to the incorporation of a weighting system that accounts for situations where different areas are of different value and contribute to the overall biogeographic region to differing extents (for example in terms of their overall size). It is appropriate to base any such weighting on the available (quantitative) values for the favourable reference range (as this provides an integrated measure of both spatial extent and habitat quality).

However, whilst such weightings can be applied to quantitative values (e.g. range and population) the qualitative nature of the other two parameters (habitat suitability and future prospects) would, in the absence of an agreed, rule-based approach, preclude any attempt at a weighted aggregation.

Box 3 Theoretical options for spatial aggregation across distinct areas to the biogeographic regional scale.

Option C.1: spatial aggregation of parameter indicators for each parameter across Member States, with subsequent derivation of performance indicators and performance classifications for each parameter and overall assessment across all four parameters.

Spatial scale	**Assessment process – derivation of:**			
	Parameter Indicators	Performance indicators	Performance classification	Overall assessment
Member State (i)			-	
Member State (ii)			-	
Member State (iii)			-	
Member State (iv)			-	
Member State (...)			-	
Biogeographical region				**Overall assessment**

Option C.2: calculation of (%) performance indicators for each parameter* at Member State level with subsequent spatial aggregation of performance indicators across Member States, followed by performance classification for each parameter and overall assessment across all four parameters.

Spatial scale	**Assessment process – derivation of:**			
	Parameter Indicators	Performance indicators*	Performance classification	Overall assessment
Member State (i)	→	↓	-	
Member State (ii)	→		-	
Member State (iii)	→		-	
Member State (iv)	→		-	
Member State (...)	→		-	
Biogeographical region	-	→		**Overall assessment**

* Only two parameters (species range and population) pass through this stage of the aggregation process; consequently this option for spatial aggregation only applies to these two parameters.

Option C.3: performance indicators and performance classification for each parameter derived for individual Member States, with subsequent spatial aggregation of performance classifications across Member States for each parameter followed by overall assessment across all four parameters.

Spatial scale	**Assessment process – derivation of:**			
	Parameter Indicators	Performance indicators	Performance classification	Overall assessment
Member State (i)	→		↓	-
Member State (ii)	→			-
Member State (iii)	→			-
Member State (iv)	→			-
Member State (...)	→			-
Biogeographical region	-		→	**Overall assessment**

Option C.4: overall assessment undertaken for individual Member States (performance indicators, performance classification and overall assessment across all four parameters) derived for individual Member States, with subsequent spatial aggregation of overall assessment across Member States

Spatial scale	**Assessment process – derivation of:**			
	Parameter Indicators	Performance indicators	Performance classification	Overall assessment
Member State (i)	→			↓
Member State (ii)	→			
Member State (iii)	→			
Member State (iv)	→			
Member State (...)	→			
Biogeographical region	-			**Overall assessment**

This apparent inability to undertake a full spatial aggregation of the data involved in this type of assessment can potentially be addressed in two ways:

- Undertake assessments at the wide (biogeographical) reporting scale, obviating the need to aggregate; or
- develop a rule based approach for spatially aggregating a series of qualitative assessments (whether this is at the stage of the performance classification or the overall assessment).

The first of these approaches might be feasible for those two parameters that cannot be directly aggregated at the parameter indicator stage of overall assessment process (i.e. the habitat suitability and future prospect parameters). If these assessments are undertaken at the wide (biogeographic) reporting scale their performance classifications can be combined with performance classifications for both the range and population parameters based on spatially aggregated (quantitative) parameter indicator data. This has an immediate and significant drawback as regards disaggregation, as the supporting information for the habitat suitability and future prospect parameters cannot be disaggregated for use in other assessments that follow different reporting boundaries.

The second approach, that of aggregating categorical parameter assessments across MS areas to derive an assessment at the scale of the biogeographic region, might itself be undertaken in a number of ways.

Firstly, a straightforward rule-based approach might be employed based on, for example, OOAO, TOAO or a threshold method. Alternatively, a decision tree approach (similar to that already employed to derive the overall assessment at the area level) might be used. However, this latter approach may be somewhat unwieldy as there is no fixed number of areas that would need to be aggregated across.

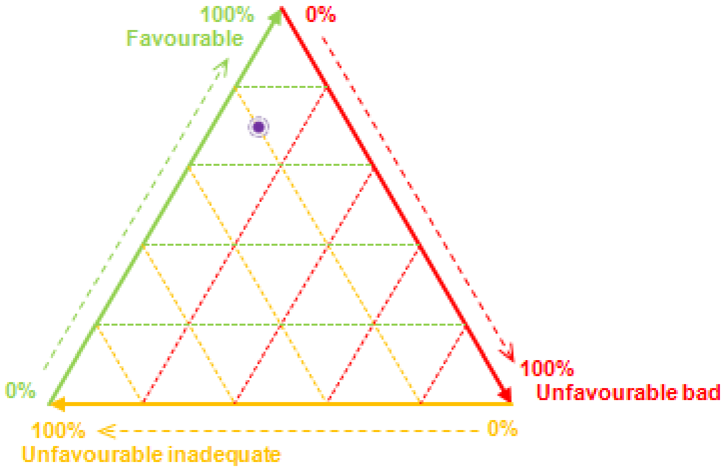
Alternatively, the total area within each status category, expressed as a percentage of the overall (biogeographical) reporting area, can be used as the final assessment metric (see Table 5.6 for an example). This is the approach recommended by the ETC/BD (2014), but it is interesting to note that the FRV are not considered by the ETC/BD (2014). This report recommends that it would be appropriate to use the Favourable Reference Range (FRR) value for area rather than the actual area values, as this would introduce an appropriate and effective weight to account for how relevant each area of assessment was in terms of its habitat suitability and use by the species under consideration. The simple rule-based approach described earlier in Figure 5.11 can then be used to assign a categorical status assessment to a parameter in a biogeographical region (i.e. the reporting area) as a whole.

Table 5.17 Example calculations for deriving overall percentage area values for individual assessment categories.

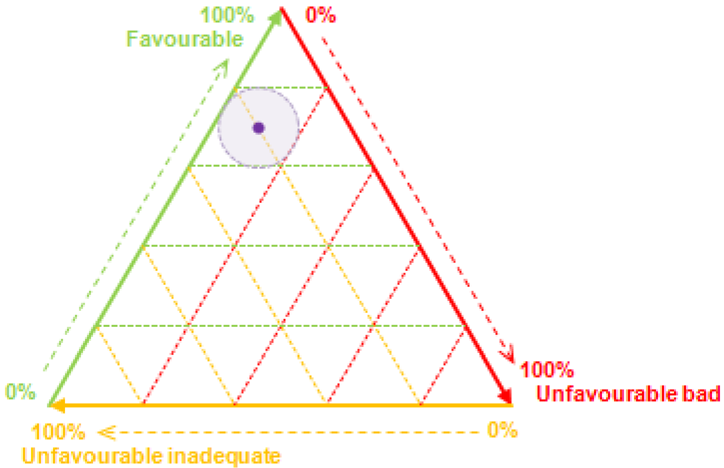
Assessment of conservation status for parameter	Component areas	Favourable reference range: as km^2	as % of total biogeographical (reporting) region	as % of categorised (non-unknown) areas
Favourable	Member State *a*	4,000	38%	40%
	Member State *b*	1,000	10%	10%
	Member State *c*	2,000	19%	20%
	Total	7,000	67%	70%
Unfavourable – inadequate	Member State *d*	2,000	19%	20%
	Total	2,000	19%	20%
Unfavourable – bad	Member State *e*	750	7%	8%
	Member State *f*	250	2%	2%
	Total	1,000	9%	10%
Unknown	Member State *g*	500	5%	-
	Total	500	5%	-
Total biogeographical region		10,500	100%	100%

Additionally, a graphical output might be used to represent the overall assessment. This could be based on a three-way (triangular) plot similar to that used in particle size descriptions. Having derived a (categorical) overall assessment for each component area within the biogeographic region, the sum of the constituent FRR values for each area falling within the three definitive status classes (i.e. 'Favourable', 'Unfavourable - Inadequate', and 'Unfavourable – Bad') would be used to derive the relative percentage values to plot (as per Figure 5.12). As the total (FRR) area that is classified as 'Unknown') would (as a best approximation) be apportioned across the first three categories relative to the areas assigned to each definitive class it would ultimately have no effect on the output of the three-way plot. Instead, the area that is classified as 'Unknown' could be used to adjust the size of the marker on the three-way plot. In this way, the degree of certainty in the output is readily appreciated by the user.

Overall, this second approach has the distinct advantages of allowing for weighting between different constituent areas and of significantly improving the prospect for disaggregation.



Conservation status of component areas	Total (FRR) area km² following component area assessments	% of total region (excluding areas with 'unknown' status)
Favourable	7,000	70 %
Unfavourable - inadequate	2,000	20 %
Unfavourable - bad	1,000	10 %
Unknown	500	-



Conservation status of component areas	Total (FRR) area km² following component area assessments	% of total region (excluding areas with 'unknown' status)
Favourable	7,000	70 %
Unfavourable - inadequate	2,000	20 %
Unfavourable - bad	1,000	10 %
Unknown	5,000	-

Figure 5.18 Example of possible three-way plot outputs with low uncertainty (upper plot) and high uncertainty (lower plot).

5.5.4 Influencing outcomes: predetermination and potential bias at the Member State and biogeographical scales

As discussed above, the derivation of categorical classifications for each of the parameters is likely to be optimal given the number and types of information types that are required for this type of assessment. The final assessment aggregation, whether undertaken at a MS scale or at the wider biogeographical scale, can be undertaken in a number of different ways.

Given four parameters, each assigned to one of four categories, and assuming that all parameters have equal importance, there are 35 potential combinations of categories that need to be considered in terms of an overall (cross-parameter) assessment. The overall assessment outputs for each of these possible combinations are shown in Table 5.7 for the current rule-based (decision tree) approach, together with OOAO, TOAO and a (75%) threshold approach. Where results are confounded by the inclusion of one or more parameters with an 'Unknown' status, the best- and worst-case possible outcomes are given.

Table 5.18 Simulated application of alternative methods for deriving overall assessments from parameter-level assessments.

Number of parameters categorised as:				Overall assessment following different aggregation methods:			
Fav	**Unfav-In**	**Unfav-Bd**	**Ukn**	**Current (decision tree)**	**OOAO**	**TOAO**	**75% threshold***
4	0	0	0	Fav	Fav	Fav	Fav
3	1	0	0	Unfav-In	Unfav-In	Fav	Fav
3	0	1	0	Unfav-Bd	Unfav-Bd	Fav	Fav
3	0	0	1	Fav	Ukn (Fav/Unfav-Bd)	Fav	Fav
2	2	0	0	Unfav-In	Unfav-In	Unfav-In	Unfav-In
2	1	1	0	Unfav-Bd	Unfav-Bd	Unfav-In	Unfav-In
2	1	0	1	Unfav-In	Ukn (Unfav-In/Bd))	Ukn (Fav/Unfav-In)	Unfav-In
2	0	2	0	Unfav-Bd	Unfav-Bd	Unfav-Bd	Unfav-Bd
2	0	1	1	Unfav-Bd	Unfav-Bd	Ukn (Fav/Unfav-Bd)	Unfav-Bd
2	0	0	2	Ukn	Ukn (Fav/Unfav-Bd)	Ukn (Fav/Unfav-Bd)	Ukn (Fav/Unfav-Bd)
1	3	0	0	Unfav-In	Unfav-In	Unfav-In	Unfav-In
1	2	1	0	Unfav-Bd	Unfav-Bd	Unfav-In	Unfav-In
1	2	0	1	Unfav-In	Ukn (Unfav-In/Bd))	Unfav-In	Unfav-In
1	1	2	0	Unfav-Bd	Unfav-Bd	Unfav-Bd	Unfav-Bd
1	1	1	1	Unfav-Bd	Unfav-Bd	Ukn (Unfav-In/Bd))	Ukn (Unfav-In/Bd))
1	1	0	2	Unfav-In	Ukn (Unfav-In/Bd))	Ukn (Fav/Unfav-Bd)	Ukn (Fav/Unfav-Bd)
1	0	3	0	Unfav-Bd	Unfav-Bd	Unfav-Bd	Unfav-Bd
1	0	2	1	Unfav-Bd	Unfav-Bd	Unfav-Bd	Unfav-Bd
1	0	1	2	Unfav-Bd	Unfav-Bd	Ukn (Fav/Unfav-Bd)	Ukn (Fav/Unfav-Bd)
1	0	0	3	Ukn	Ukn (Fav/Unfav-Bd)	Ukn (Fav/Unfav-Bd)	Ukn (Fav/Unfav-Bd)
0	4	0	0	Unfav-In	Unfav-In	Unfav-In	Unfav-In
0	3	1	0	Unfav-Bd	Unfav-Bd	Unfav-In	Unfav-In
0	3	0	1	Unfav-In	Ukn (Unfav-In/Bd))	Unfav-In	Unfav-In
0	2	2	0	Unfav-Bd	Unfav-Bd	Unfav-Bd	Unfav-Bd
0	2	1	1	Unfav-Bd	Unfav-Bd	Unfav-In	Ukn (Unfav-In/Bd))
0	2	0	2	Unfav-In	Ukn (Unfav-In/Bd))	Ukn (Unfav-In/Bd))	Ukn (Unfav-In/Bd))
0	1	3	0	Unfav-Bd	Unfav-Bd	Unfav-Bd	Unfav-Bd
0	1	2	1	Unfav-Bd	Unfav-Bd	Unfav-Bd	Unfav-Bd
0	1	1	2	Unfav-Bd	Unfav-Bd	Ukn (Unfav-In/Bd))	Ukn (Unfav-In/Bd))
0	1	0	3	Unfav-In	Ukn (Unfav-In/Bd))	Ukn (Fav/Unfav-Bd)	Ukn (Fav/Unfav-Bd)
0	0	4	0	Unfav-Bd	Unfav-Bd	Unfav-Bd	Unfav-Bd
0	0	3	1	Unfav-Bd	Unfav-Bd	Unfav-Bd	Unfav-Bd
0	0	2	2	Unfav-Bd	Unfav-Bd	Unfav-Bd	Unfav-Bd
0	0	1	3	Unfav-Bd	Unfav-Bd	Ukn (Fav/Unfav-Bd)	Ukn (Fav/Unfav-Bd)
0	0	0	4	Ukn	Ukn (Fav/Unfav-Bd)	Ukn (Fav/Unfav-Bd)	Ukn (Fav/Unfav-Bd)

* with four parameters, the 75% threshold equates the lowest of the top three categorisations.
NB: categorisations in the above table are given as: Favourable (Fav); Unfavourable-Inadequate (Unfav-In); Unfavourable-Bad (Unfav-Bd); Unknown (Ukn).

The data presented in Table 5.7 are further summarised as Table 5.8.

Table 5.19 Summary comparison of outputs from different methods for deriving overall condition status assessment from four parameter categorisations.

Overall (cross-parameter) condition status assessment	Distribution of range of possible assessments derived through the application of different aggregation methods: Current (decision tree)	OOAO	TOAO	75% threshold*
Fav	2	1	4	4
Unfav-In	10	4	9	9
Unfav-Bd	20	20	10	11
Ukn	3	0	0	0
Ukn (Fav/Unfav-Bd)	0	4	8	7
Ukn (Fav/Unfav-In)	0	0	1	0
Ukn (Unfav-In/Unfav-Bd)	0	6	3	4

NB: categorisations in the above table are given as: Favourable (Fav); Unfavourable-Inadequate (Unfav-In); Unfavourable-Bad (Unfav-Bd); Unknown (Ukn).

This brief analysis suggests that there are some differences between the different aggregation methods. Unsurprisingly, those methods that apply techniques other than OOAO tend to present more occurrences of overall Favourable condition status assessments (NB OOAO also forms part of the current decision tree approach). The main difference however lies in the fact that the current rule-based approach for combining the categorical outputs from across the four separate parameters has definitive outputs; none of the other methods are able to deal definitively with ‘Unknown’ categorisations and so the best that can be done is to present the result as a range of possible outcomes ranging from best-case to worst-case.

This inability to deal with Unknown categorisations is, coupled with the apparent absence of any significant bias, provides enough justification to recommend that, for this Scenario, the current aggregation decision-tree type approach to parameter aggregation is probably optimal.

5.6 Harmonised marine biodiversity assessments

This section addresses Task 4c and considers the implications of aggregation under each scenario in terms of the potential for disaggregation and the harmonisation of marine biodiversity assessments.

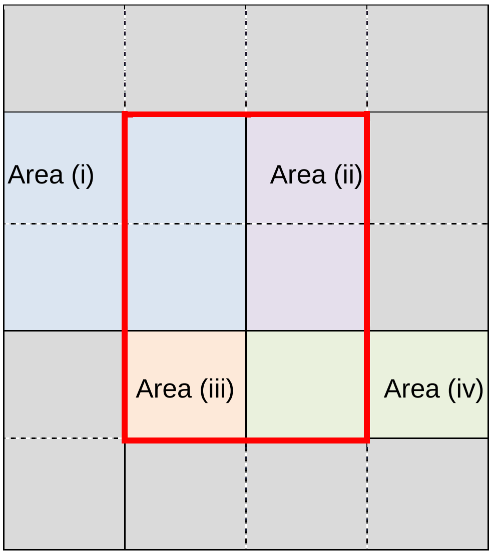
Within Scenario A, although the application of spatial aggregation may take place late (Option A) or early (Option B) in the overall assessment process, the two alternatives both make use of the same fundamental data. In terms of disaggregation potential there is no advantage associated with either of the alternative options described. However, given the likely need to introduce some consideration of weighting for different sub-habitats, it makes operational sense to undertake this operation late in the overall process (in so doing, the weighting process is effectively limited to one set of considerations). In this context, Option A presents a more viable operational alternative.

Scenario B makes use of bird data exclusively, with no ancillary information being required. This reduces the likelihood of conflicts with other assessments (due to, for example, the need for similar data types to be produced in different formats). Species counts for birds within identified colonies across the OSPAR Regions’ subdivisions can be held in a relatively basic form (e.g. a database with just six simple fields: Region, Subdivision, Colony, Species, Date, Count) and can be readily assimilated into an appropriate indicator when required. The same underlying data can be used as 'building blocks' for other assessments, which contributes to achieving the aspiration of harmonised marine biodiversity assessments, although it should be recognised that there is likely to be assumptions made regarding data (spatial) distribution and averaging across areas (see below).

Under Scenario C, the application of spatial aggregation at a high level (Option C.4 from Box 4) has a good disaggregation potential and so better supports a move towards harmonised marine biodiversity assessments than do the other aggregation options.

Where there is a spatial mis-match between biogeographic reporting areas and constituent (monitoring) areas there is the need to apportion information. Taking the stylised example shown as Figure 5.13, the overall biogeographic assessment would need to encompass four separate areas (i – iv). However, these do not each contribute in their entirety. For example, if data are available for the overall population of harbour porpoise in each area, these can be apportioned to the biogeographic reporting area to give an overall population estimate as follows:

$$\text{Population estimate} = \frac{\text{Population}_{\text{Area(i)}}}{2} + \text{Population}_{\text{Area(ii)}} + \text{Population}_{\text{Area(iii)}} + \frac{\text{Population}_{\text{Area(iv)}}}{2}$$



Area (i): with a total area of 4 units; 50% of Area (i) contributes $^{1}/_{3}$ of biogeographic reporting area

Area (ii): with a total area of 2 units; 100% of Area (ii) contributes $^{1}/_{3}$ of biogeographic reporting area

Area (iii): with a total area of 1 unit; 100% of Area (iii) contributes $^{1}/_{6}$ of biogeographic reporting area

Area (iv): with a total area of 2 units; 50% of Area (iv) contributes $^{1}/_{6}$ of biogeographic reporting area

Other adjacent areas

Figure 5.19 A stylised biogeographic reporting area (red rectangle, with a total area of 6 units) composed of contributions from four separate areas i – iv (see key).

By using the data in this way it is possible to build up an overall assessment of a relatively complex biogeographic area using standard 'building blocks' in a nested approach – so facilitating disaggregation and the efficient re-use of data.

Consideration of the relative timing of temporal and spatial aggregation (see section 5.5, above) has drawn the tentative conclusion that aggregation across the dimension that displays the least overall variability (e.g. the temporal dimension) ahead of aggregation across more variable dimensions (e.g. the spatial dimension) provides more useful (less variable) intermediate aggregated elements. This in turn supports more effective disaggregation (overall indicators will be disaggregated to spatially-distributed temporal aggregations with relatively low variability). This has implications for the overall harmonisation of biodiversity assessments as it will be necessary to recalculate temporally aggregated elements on a periodic basis (i.e. as new data become available from ongoing monitoring programmes).

Together, consideration of these three scenarios would suggest that there are no significant barriers to the harmonisation of biodiversity monitoring with subsequent assessment. Data that are collected for use in one reporting obligation can be used to support other reporting obligations, albeit with the possibility of some adjustments to account for spatial discrepancies between monitoring areas and reporting boundaries.

The assessment of harmonisation across marine biodiversity assessments also relates, to some extent, to the synchronicity of reporting schedules (or lack thereof) that is imposed through legislation. This is especially true of six-yearly reporting structures, such as those imposed by the Habitats Directive and the Marine Strategy Framework Directive which may lead to peaks and troughs in the requirement for data acquisition and processing. However, in the case of the scenarios discussed here, it is considered that the data that are required to support reporting can be collected and processed so as to avoid excessive variation in effort over the reporting period.

5.7 Summary and Conclusions

5.7.1 Introduction

Examination of the application of aggregation methods across the three scenarios considered has provided useful insights into some of the subtle, and perhaps hitherto unconsidered, aspects of aggregation as it supports marine biodiversity assessments. It is important to consider that these findings may have relevance to other aggregation applications and are not necessarily restricted to the specific applications discussed.

The following brief paragraphs summarise the main findings and key messages from this consideration of three defined scenarios.

5.7.2 Summaries and learning points

i. Scenario summaries

Scenario A considered the annual assessment and reporting of a single (complex) habitat type (e.g. Annex I reef) at the scale of a single MPA and based on the principles of Common Standards Monitoring (CSM). Potentially, this form of assessment requires the consideration and aggregation of attribute assessments across a number of separate (composite) habitat types or sub-features. Where there is more variability in sub-feature attribute condition apparent across sub-features than across attributes, initial aggregation across attributes should be considered. Conversely, where it is felt that there is more variability in sub-feature attribute condition apparent across attributes than across sub-features, initial aggregation across sub-features should be considered.

The underlying process to support this form of assessment has two distinct aggregation stages: initially aggregating across the condition assessments of attribute or sub-feature combinations (where each combination is assessed as either ‘Favourable’ or ‘Unfavourable’); and subsequent aggregating across the sub-feature status assessments.

The implications of different (qualitative) aggregation approaches or methods for both of these stages of the process were briefly considered. Whilst this highlighted the strong precautionary nature of OOAO, and the more relaxed nature of TOAO, the possible benefits of the adopting an alternative approach - such as a decision tree or threshold approach - should be considered. However, it should be noted that whilst the application of methods such as the threshold approach can provide a much better frequency distribution of potential outcomes, they may not provide the necessary levels of precaution and discrimination that may be required by conservation assessments.

Scenario B considered the assessment of multiple species at the biogeographic (regional sea) scale, reporting every six years (for example under MSFD or OSPAR). It briefly examined two alternative approaches to deriving a spatially aggregated multi-species assessment: aggregating across areas first, and then across species; and aggregating across species first, and then across areas.

The scenario employed targets at a number of stages through the aggregation and assessment process. For example, a 75% target is used in the (spatial) assessment of the number of species passing their ‘trend targets’. It was not possible, within the scope of the project, to assess whether the targets were appropriate, although data modelling might be used in further work to help address this question.

The structure of the aggregation allows for weightings to be introduced; these should be applied to account for a species’ value to, or importance within, each component area of the overall biogeographic region.

The question of how extensive indicator data should be (i.e. how many colonies should be included) in order to ensure that the overall assessment of a given species is valid was also considered. It is suggested that information on typical colony sizes (e.g. mean size and variability), and the numbers and distribution of colonies within the reporting region could be used to identify the number of colonies that should be selected (at random) from within a biogeographic region in order to derive an overall population estimate for a species that deviates from the underlying ‘true’ value by an acceptable amount.

To help present the maximum information from the aggregated assessment, consideration should be given to presenting composite outputs that indicate not just an overall assessment (such as a single GEnS pass/fail statistic) but also the contributions made by individual species (for example by using a ‘fan chart’ type of output).

Scenario C considered the assessment of a single, highly mobile species (harbour porpoise) at the biogeographic scale (employing the approach used for HD reporting). This approach combines four separate parameters (range, population, habitat suitability and future prospects) each of which is based on a discrete area of sea, which may or may not coincide with a delineated reporting area. Data may need to be aggregated or disaggregated to fit with the spatial scope of reporting requirements (Type 1 aggregation), whilst the quantitative information that supports the first two parameters may be available at different temporal resolutions (Type 2 aggregation).

The spatial aggregation involved allows for the incorporation of a weighting system to account for where different areas are of different value and contribute to the overall biogeographic region to differing extents (for example in terms of their overall size).

Such weighting can be based on quantitative values for the favourable reference range which provides an integrated measure of both spatial extent and habitat quality. Although weightings can be applied directly to quantitative values (i.e. range and population) a rule-based approach would be required for the other two (qualitative) parameters (i.e. habitat suitability and future prospects). Under the framework used for HD reporting, condition assessments fall into three groups: favourable, unfavourable-inadequate, and unfavourable-bad. Instead of presenting a single aggregated summary statistic, it is possible to report the area of sea that falls into each of these three classes graphically as in the form of a three-way (triangular) plot similar to that used in particle size descriptions. This approach could also accommodate ‘unknown’ assessments (by using the percentage of area that is classified as ‘Unknown’ to adjust the size of the marker on the three-way plot), enabling the user to readily appreciate the degree of (un)certainty in the assessment output.

This scenario also highlighted the choice that needs to be made in the sequencing of temporal and spatial aggregations. It is recommended that, in general terms, temporal aggregation should be undertaken ahead of spatial aggregation (although it is further recommended that the implications of such options should be investigated).

The overall process behind this scenario employs a decision tree approach to aggregate across parameters. This approach, which does not seem to be affected by any undue bias, is able to handle situations where there is missing data and condition assessments for specific areas of the marine environment are characterised as unknown.

ii. Aggregation methods employed

All three scenarios (either explicitly or implicitly) employ OOAO as an aggregation method and, within the assessments that have been undertaken here, the precautionary and conservative nature of this approach has been demonstrated through a brief ‘sensitivity’ type analysis. Although these exercises provided a clear demonstration of the potential distribution of resulting condition status assessments when differing aggregation methods were applied to simulated examples, alternative methods (such as TOAO or a threshold method) were found to be similarly blunt. Where it is not possible to apply quantitative methods it may be better to consider the development of decision tree type approaches in order to provide better discrimination between outcomes. In considering possible alternative approaches it is important to note that the primary purpose of any aggregation is to reflect reality and that, consequently, aggregated condition assessments should accurately reflect the reality of the prevailing environmental conditions. Within the social sciences, aggregation results are often validated with independent datasets that have a known common response to the phenomena of interest; based on these validations, the aggregation method or thresholds within the process can be adjusted to better reflect the reality of the situation. Attempts to modify aggregation methods without such an objective supporting process may be misinterpreted as an artificial inflation or deflation of the prevalence of favourable status.

Within its structure, Scenario C employed a decision tree type approach to aggregation. Where alternative aggregation methods were considered the issue of accommodating instances where the condition status assessment for one or more of the parameters involved in the aggregation was ‘unknown’ was flagged. Under such circumstances it was concluded that the decision-tree approach is probably optimal as it allows for ‘unknown’ condition assessments to be handled within the defined aggregation process.

iii. Consideration of summary (Type 5) aggregation

As described, the three scenarios do not include consideration of a Type 5 aggregation. Many of the principles discussed within the three scenarios (influence of aggregation method, importance of weighting, application of temporal aggregation, *etc*.) will apply to Type 5 aggregation. In particular, the application of graphical summary outputs for Type 4 aggregation (as discussed with reference to Scenario C) have a clear relevance to Type 5 aggregation, and should be considered as providing a useful output model for this type of aggregation.

5.7.3 Conclusions and recommendations

Conclusions and recommendations drawn from consideration of the three scenarios are summarised below.

i. Process guidance

Within the constraints of this project, consideration of scenarios that are currently defined (to a greater or lesser extent) within existing guidance documentation reduced the scope for assessing the use of novel (alternative) aggregation methods. The use of established structures, pre-defined suites of attributes and combination rules effectively controls the scope for variability within the aggregation processes. It is recognised that this rigidity is an important requirement for consistent and standardised reporting, especially when such reporting is at the international scale. Overall, the existing aggregation structure within all three scenarios was considered to be fit for purpose. However, despite this, some ‘flexibility’ in the detail of the approaches used to support the three scenarios was identified, and the impact of varying the aggregation structures and aggregation rules for particular scenarios (in terms of overall outputs) could be investigated. Notwithstanding this, as the scenarios that were considered had inherent ‘flexibility’ (in terms of possible aggregation structures) and that this flexibility may give rise to different overall outputs, there is a need for more structured and detailed process descriptions to be produced and adopted.

ii. Alternative models for the aggregation process

Where alternative models for the aggregation process are available, it is suggested that it is more appropriate to consider aggregating across the least variable data first, so maintaining more information within the aggregated outputs.

Consideration of Scenario A highlighted two alternative process models for undertaking an aggregated assessment of a complex habitat based on the principles of Common Standards Monitoring. One model effectively partitioned the data by sub-feature and undertook an initial aggregation across the attribute conditions. The second model partitioned the data by attribute and undertook an initial aggregation across sub-features. The selection of one of these two models should be based on consideration of the sources of variance in the available data. In this case, this would involve a judgement being made regarding the relative variability of attribute condition at the level of sub features. It is suggested that, where there is more variability apparent in sub-feature attribute condition across sub-features than across attributes, initial aggregation should be undertaken across attributes. Conversely, where it is felt that there is more variability apparent in sub-feature attribute condition across attributes than across sub-features, initial aggregation across sub-features should be considered.

Similarly, consideration of Scenario C identified two alternative process models where the data to be aggregated has both spatial and temporal components. The choice between these alternative approaches again follows the principle of aggregating across the least variable dimension first. Accordingly, given the nature of the marine environment, it is tentatively suggested that aggregation should initially be across the temporal dimension rather than spatial dimension.

iii. Temporal data

Where data sets have a temporal component (e.g. Scenario C) it is necessary to derive intermediary statistics for use in aggregation processes. In this context there is an important distinction to be made between cases with temporal replicates of data (which is likely to be the majority of instances for marine biodiversity assessments and reporting) and cases where the underlying data itself has a temporal component (such as continuously recorded data). Scenario C deals with temporally replicated data - and temporal aggregation (whether weighted or unweighted) of such replicates is easily accomplished. However there are other instances of temporal aggregation outside of the scenarios considered that may require summary data to be extracted from time series. Nevertheless this is seen as a data processing, not an aggregation, issue. It is noted however that more work may be required to determine standard methods for deriving suitable summary statistics from time series data.

iv. Output presentation

Consideration of Scenarios B and C highlighted potential opportunities regarding the presentation of outputs. Whilst the underlying objective of aggregation is to reduce the volume of information that is presented, ancillary information can nevertheless be invaluable in providing context for the formal aggregated output. For example, where Type 3 aggregations are being used (providing an aggregation across biodiversity components such as species or habitats) summary outputs might be usefully presented as fan charts (as per Halpern *et al* 2012). Similarly, for Type 4 or 5 aggregations other graphical methods such as three-way plots could be usefully employed. This latter approach can effectively summarise an aggregation of multiple elements (such as across indicators/ indices within a set of assessment criteria, or across assessment criteria of different biodiversity components to deliver an overall assessment) where each element is itself a categorical output taking one of three possible values.

v. Spatial coincidence between monitoring and reporting areas

As far as possible, spatial aggregations need to have information for subdivisions that are spatially coincident with reporting areas. It is apparent from the scenarios considered that this it is not necessarily always the case. This situation is compounded by the fact that reporting boundaries may also be inconsistent between reporting obligations. In cases where boundaries are not coincident it is necessary to apportion data from subdivisions of monitoring or survey areas to derive data representing the areas' contributions to a given reporting area. This can be done on a unit area basis using GIS. It is suggested that standard GIS data layers are identified that describe the spatial extent of all monitoring or survey programs and all reporting areas.

vi. Targets

The need to derive suitable target values was identified within consideration of Scenario B (although it may also apply to a number of other applications). There is currently no clearly defined methodology for this although one possible approach is outlined in the main text.

vii. Weighting

Finally, the application of weightings is repeatedly inferred from the existing process information. The use of weightings improves the realism of the aggregation process, and their influence cannot be downplayed. The role of weighting should be explicitly acknowledged within the process guidance for (aggregated) reporting obligations, including instances where equal weightings are applied as well as where specific weightings are applied to attributes based on (for example) the spatial extent of the reporting area that is represented by the data.

6. Summary Conclusions and Recommendations for Further work

6.1 Summary of work undertaken

The following provides a brief overview of each of the major sections of this report.
Section 2 (in conjunction with Appendix A and Appendix C) reports on the need for aggregation methodologies and the reporting of aggregation in the wider literature. These elements of the work

were undertaken by means of a literature review, and were intended to support the identification of key methodologies or learning points that are applicable to marine biodiversity assessments.

The initial literature review suggested that, with reference to the topic of data aggregation, the majority of references that were identified related to the development of Type 3 aggregations (across biodiversity components). This reflects the relative wealth of literature reporting the production of CIs (composite indicators or indices). Whilst this, in effect, might be considered a trivial form of aggregation, many of the principles that underpin the development of CIs are applicable across other forms of aggregation and it is likely that some sound approaches can be determined from them.

The selection and processing of indicators, and appropriate methods for their normalisation, weighting and aggregation, represent key phases common to the production of all robust composite indicators. The limited review of references from the social sciences that was undertaken highlighted some of the differing methods applied during these phases, and identified a series of learning points, relating to, *inter alia*:

- the transformation of variables to remove bias;
- the application of normalisation techniques (based on minimum and maximum values from baseline data) to rescale inputs;
- the amalgamation of additional indicators or composite indicators into existing assessment indices can be inappropriate if done without evidence of clear and proven relationship between indicators and the ultimate aim of the composite indicator;
- whilst weights can be generated by expert judgement, the use of objective assessments to inform that judgement may be important;
- expert panels can also be employed to undertake validity and quality assessments of the outputs for both CI and AI aggregations;
- the concept of the use of appropriate standardisation methods for indicator variables is important;
- aggregation frameworks may be set up to accept weighting even if suitable weighting values are not available (i.e. an equal weighting can be assumed until improved data becomes available);
- *post hoc* review (validation) of aggregation outputs can be performed using an independent data set (i.e. data not used in the aggregation but which is predicted to show a high correlation with the calculated composite (or assessment) indicators;
- consideration of the use of simple exploratory techniques to identify redundancy across the indicator variables; and
- the use of PCA as a means of deriving weights and of effectively reducing the number of variables that are considered in an aggregation.

The review also provided examples of:

- frameworks for aggregation that can be applied to a wide range of situations, including marine biodiversity assessments
- sound examples of a hierarchical, non-weighted (arithmetic) averaging approach to aggregation;
- use of aggregation techniques to clearly and simply express complex phenomena;
- accommodation of the generation of both CIs and (overarching) AIs within a quantitative, hierarchical weighted aggregation method; and
- the application of weighting at more than one level of an underlying hierarchical structure.

Section 3 (together with Appendix C and Appendix D) reports on the benefits and limitations of selected aggregation methods, meeting the requirements of Objective 2 for the project. To help facilitate this process a quality framework is described, against which the performance of different aggregation methods could be assessed. This quality framework (based on similar frameworks used in the social sciences, and described in detail within Appendix E) covered four main areas (or 'Performance assessment criteria'):

- the **basis for the aggregation** (including whether the approach is methodologically sound; accurate and/or reliable; and allows for weighting);
- the **value of the aggregation** (including whether the approach is easily interpretable, transparent and comprehensible; whether it provides accessible and clear outputs; employs appropriate methods; and produces objective outputs that are able to demonstrate integrity);
- the **aggregation process** itself (including whether it is robust and practical; whether it is accessible; and whether it is able to satisfactorily handle uncertainty and confidence); and
- the subsequent **application of methods** or data (including consideration of repeatability and disaggregation (coherence) potential).

Section 4 considers the information requirements of different aggregation methods, and meets the requirements of Objective 3 for the project. The ability of fundamental aggregation methods (as identified under Section 2) to handle a range of data types (as described by a broad classification system) is assessed. In addition, the availability of suitable data to support aggregation needs is discussed and a number of potential data gaps identified. These gaps included:

- an absence of standardised methods and information sources for setting thresholds, baselines and targets (which can be based on observation of pristine conditions; historical observations; modelled predictions; or best professional (expert) judgement);
- a perceived lack of standardised and transparent conversion methods to convert available quantitative data to categorical or ordinal classes used by some aggregation methods;
- an absence of clearly defined methods for the calculation of weightings for aggregated inputs;
- a shortage of continuous datasets (that could be used to calculate the footprint size of an activity or to monitor the change of a species or habitat range (an issue of point data conversion with spatial analysis methods); and
- a general poor availability of high frequency, long duration temporal studies and high resolution, large extent spatial studies.

Section 5 (together with Annex G) addresses Objective 4 for the project, describing three scenarios where practical use is made of aggregation methods.

Overall, the existing aggregation structures within the scenarios were considered to be fit for purpose and further consideration would suggest that these three selected scenarios do not provide any significant barriers to the harmonisation of biodiversity monitoring. However, one key area that is flagged relates to the need for some possible adjustments that might be required to account for spatial discrepancies between monitoring areas and reporting boundaries.

Although the three scenarios were quite tightly defined, some 'flexibility' in the detail of the approaches used to support aggregation was identified, and the impact of varying the aggregation structures and aggregation rules for particular scenarios (in terms of overall outputs) should be investigated. Notwithstanding this, as the scenarios that were considered had inherent 'flexibility' (in terms of possible aggregation structures), and this flexibility may give rise to different overall outputs, there is a need for more structured and detailed process descriptions to be produced and adopted.

A number of insights (described within the main text) were made relating to:
- process guidance;
- the basis of the aggregation methods that are employed;
- alternative models for the aggregation process;
- the use of temporal data;
- output presentation;
- Type 5, summary, aggregation;

- spatial coincidence between monitoring and reporting areas;
- targets; and
- weighting.

6.2 Recommendations for further work

This section addresses Objective 5 for the project, providing recommendations on future work required to improve the availability of information to support aggregation methods, the development of new aggregation methods, and how the concept of harmonised, aggregated marine biodiversity status assessments might be achieved.

6.2.1 Systematic consideration of the structuring of temporal and spatial aggregation

This report has made the pragmatic suggestion that (especially where no associated measure of variability is carried forward) aggregation should be undertaken across the least variable dimension first. In the marine environment it is suggested that this is likely to be the temporal dimension; spatial heterogeneity will tend to give rise to high levels of spatial variability, whilst the buffering capacity of the marine environment will tend to dampen (short-term) temporal variability.

Whilst this approach is intuitive, is easily adopted and, in terms of disaggregation, fits well with data processing, it is suggested that the implications of the relative phasing of temporal and spatial aggregation be tested using representative datasets and statistical modelling approaches.

6.2.2 Documentation of the reporting process

The use of categorical approaches to support (qualitative) data aggregations is well established in marine biodiversity assessment reporting. However, review of available guidance to support the development of the scenarios considered as part of this report suggests that the process that underpins some reporting methods in the UK (e.g. CSM) could be better defined. In particular, the role played by aggregation in each reporting process should be clearly highlighted and described. Where ambiguities or uncertainties exist, these should be highlighted and work commissioned to clarify the practical options available, and to provide detailed supporting process guidance.

6.2.3 Category definitions and targets

Allied to the review of process documentation, the application of targets (e.g. to assign indicators to pass/fail or to high/good/moderate/poor/bad categories) is identified within existing guidance. However, what is missing is a robust and transparent narrative on the provenance of such targets. Currently there is the risk that some targets may be viewed as opportunistic or arbitrary.

It is suggested that the methods used for setting targets should be reviewed to ensure that this fundamental stage in the various defined aggregation processes is supported by appropriate targets which are fit for purpose and defined by methodologically sound principles.

In addition, it is to see how (categorical) changes at a low level of the aggregation process due to the application of different targets or category boundaries can result in changes (in assessment conclusions) at a higher level of the aggregation.

6.2.4 Allowing for the differential contribution made to an aggregation by data from different sources

The aggregation of spatial data (such as population abundance estimates) potentially suffers where the boundaries or spatial extent of monitoring or survey units do not fully align with reporting areas (see Figure 5.13). This introduces the question of where indicator data should be sourced from and what contribution to the overall aggregation it should make.

The pragmatic approach, and one which maximises the use and re-use of collated (and archived) data (i.e. disaggregation potential), is to simply scale the indicator data from each separate monitoring area that contribute to the overall reporting area (Areas (i) to (iv) in Figure 5.13). This can be done simply by using the spatial extent of monitoring area that contributes to the reporting area expressed as a proportion of the total extent of the monitoring area.

Again using the scenario outlined in this would mean that indicator data from each monitoring area would be scaled as follows (Table 6.1):

Table 6.20 Example of indicator data scaling based on spatial contribution to reporting area.

Monitoring area	Abundance estimates (total individuals)	Total extent of monitoring area (notional units)	Extent of monitoring area contributing to reporting area (notional units)	Scaling factors for indicator data	Scaled abundance estimates used in spatial aggregation process
Area (i)	7,000	4	2	0.5	3,500
Area (ii)	5,000	2	2	1.0	5,000
Area (iii)	2,000	1	1	1.0	2,000
Area (iv)	4,000	2	1	0.5	2,000

It is recommended that, as a minimum, a standard set of GIS layers describing the spatial extent of all monitoring or survey units and reporting areas (e.g. biogeographic seas) that lie within the wider UK reporting area should be developed. These layers can then be used to extract the required series of statistics indicating the % area of each monitoring unit that contributes to a given reporting area.

In addition, consideration should be given to the redesign of monitoring units to be more nested or ‘complementary’ (and to tessellate more completely, matching reporting area or biogeographic boundaries more closely).

6.2.5 Weighting

Weighting has been identified as being key to the process of aggregation (Nardo *et al* 2005). The use of weighting within the three scenarios discussed in this report (whether implicit or explicitly described in relevant reporting guidance documents) is centred around the use of ‘area’ as a weighting factor. In the discussions relating to the scenarios the spatial extent of the area(s) represented by the contributory data or indicators has been taken as the principle source of weighting. This is, of course, just one of the available options; other values that might be used could include the amount of data (e.g. data replication or density of coverage – with greater weight being given to, for example, replicated data or data derived from areas with intensive sampling) or the variability or noise that is inherent in the data (with greater weight being given to data that are associated with greater accuracy or lower variability). In situations where area is retained as a weighting factor, the value of the weighting might be improved by considering the value or 'quality' of the area that is represented. For example, contributory indicator data from a large area of sub-optimal habitat might merit an equal weighting to a smaller area of optimal habitat (this concept was introduced under discussions of Scenario C where it was suggested that Favourable Reference Range values might be used as the basis for weighting as this would effectively embody both the 'quantity' and 'quality' dimensions).

The use of different sources of information to derive weights (especially the choice between the alternatives based around the source of the data and the quality of the data) should be reviewed. It is recommended that guidance on deriving and applying appropriate weightings should be produced (to sit within or alongside improved process guidance).

Weighting is not only related to fully quantitative empirical assessments, but can also be implicit in some existing processes where expert judgement is used (e.g. CSM). The production of appropriate weightings from expert judgement can be formalised through the application of methods such as AHP (see Barnard & Boyes 2013) and it is recommended that consideration be given to embedding such approaches within defined aggregation and reporting processes.

6.2.6 Output interpretation and presentation

A key role of aggregation is to take a set of complex information (for example spatially distributed species monitoring, habitat and physico-chemical data) and derive a reduced number of data (often a single 'assessment index') that can be easily presented and reported. At the same time, these aggregated 'assessment indices' need to be understood and easily interpreted.

Significant progress towards these joint objectives can be made by presenting data from the lower levels of aggregation (i.e. composite indicators) together with the overall assessment index. For example, where the population size of several species of seabird are being aggregated to derive a representation of overall seabird abundance, data can be presented both as a high level summary (single numeric value) and a visual indication of the relative performance of each contributory species. Figure 6.1 provides an example of such a graphic where the overall (aggregated) assessment is shown as the central figure, and the contribution made by different species, for example the percentage attainment of the relevant population target (shown below as a red dashed line), is represented by the radius of separate segments of a fan chart. This type of informational graphic (which is adopted from a graphic presented by Halpern *et al* 2012) can be further enhanced by using the width of each segment as an indicator of relative weighting for each contributory indicator.

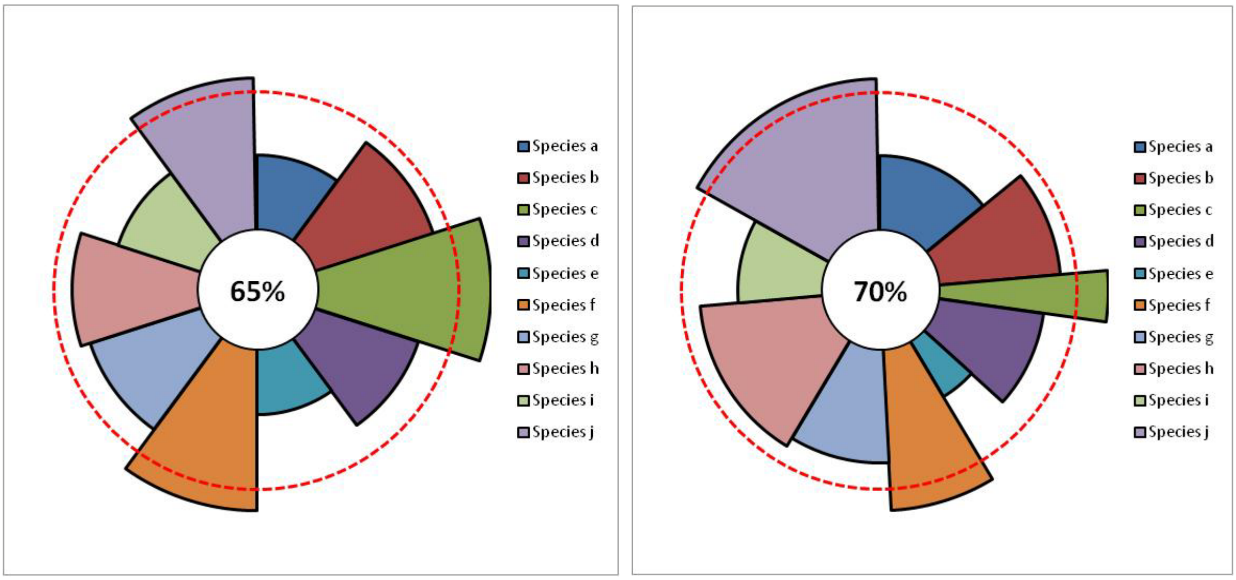


Figure 6.20 Example fan chart outputs for aggregated seabird abundance data using equal (left) and variable (right) weighting (see text for details).

This particular presentation approach is versatile and can be adopted for other assessments including, for example, Type 5 aggregations such as the ‘Index of global ocean health and benefits’ reported by Halpern *et al* 2012. It could also feasibly be used to present assessments of descriptors under MSFD.

In situations where aggregations are based not on quantitative data but on qualitative (categorical) data, aggregated outputs can be very poor in terms of the information they convey. This is the case, for example, for Article 17 type reporting where data are categorised to one of three mutually exclusive condition classes and a rule based approach used to spatially aggregate across what could, in actuality, be a range of classifications. In such cases, where the derived data are presented in terms of their association with three mutually exclusive classes, data can be presented visually as a three way ‘triangle ‘plot’. The percentage of sites that are categorised to each of three categories can be plotted on three independent axes. For example, in Figure 6.2, 60% of sites within a reporting area are designated ‘Favourable’, 25% are ‘Unfavourable-inadequate’, and 15% are ‘Unfavourable-bad’. In deriving the statistics for such a plot the overall level of membership within each category may be weighted (e.g. by site area). Also, uncertainty in the data (for example the relative contribution made by sites where the performance categorisation is unknown) can be represented by the size of the data marker used (a larger marker indicating more uncertainty). Although the overall assessment for the reporting area (based on, for example, a decision tree approach) – which could be reflected in the marker colour that is used in the triangle plot - might be ‘Unfavourable-inadequate’ the additional information that is conveyed by a triangle plot provides a useful aid to interpretation that is simply unavailable using the more commonly used categorical reporting methods.

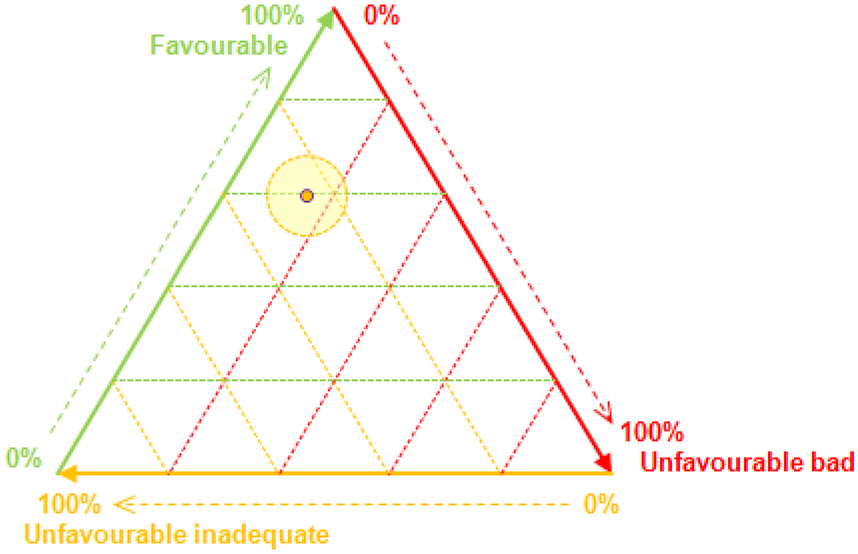


Figure 6.21 Example triangle plot.

This consideration of supporting (or contributory) information was highlighted by Prins *et al* (2013) who suggested the development of other reporting metrics for aggregations. The ultimate expression of environmental status as being ‘pass’ or ‘fail’ (i.e. GEnS achieved or not) effectively leaves out a lot of ancillary information, especially in situations where an extremely parsimonious rule based approach such as OOAO has been used. It is much more informative to relate the ultimate assessment to the number of spatial units, percentage areas or components that pass or fail a given target. It is recommended that the development and adoption of appropriate reporting formats to deliver these benefits is taken forward as further work. As described, both fan charts and triangle plots provide flexible tools for helping to achieve this objective.

Finally, it is noted that there is, within UK marine biodiversity assessments, a strong leaning towards the assessment of (spatial) aggregations based on qualitative data. Although more demanding in terms of data requirements, quantitative aggregation methods arguably provide a greater degree of flexibility and allow more detailed insights into system performance to be derived. It would be useful to understand the barriers that exist to prevent the adoption of more quantitative methods. Whilst it is likely that data availability (or provision) and policy drivers are, together, responsible in large measure for the current qualitative/quantitative balance it is recommended that a review is undertaken to identify the range of changes (e.g. to monitoring, organisational structure, or policy) that would be required to facilitate the adoption of more quantitative aggregation methods for reporting marine biodiversity assessments.

6.2.7 Sense checking and ground-truthing

As EU Member States move toward full implementation of the MSFD, aggregation approaches used to derive reporting metrics may tend to include more elements (species, habitats, physico-chemical metrics, *etc*.) and become increasingly more complex. It is recommended that an over-arching framework be developed (as recommended by Nardo *et al* 2005) to help place the wide (and potentially expanding) range of marine biodiversity assessments and their associated aggregations into context and to provide a framework for considering how the aggregations work (for example, are they meaningful; do they respond in a predictable and stable manner; do they accurately represent the particular phenomena of interest).

Such a framework would also serve to provide a background for the comparison of different (Type 5) aggregations, allowing for cross-referencing between aggregated indicators. In this context, comparisons should be made between aggregations that purport to indicate the same general phenomena (e.g. GEcS under WFD and GEnS under MSFD) but which do not, ideally, use the same contributory indicators. Such sense checking between the different Type 5 aggregations should be undertaken routinely and reported on as this serves to increase transparency of the overall process and boost acceptance of the outputs.

6.2.8 Overview of recommendations

The following summarises the recommendations for further work:

(i) The implications of the relative phasing of temporal and spatial aggregation should be tested using representative datasets and statistical modelling approaches;

(ii) the role played by aggregation in each reporting process should be clearly highlighted and described. Where ambiguities or uncertainties exist, these should be highlighted and work commissioned to clarify the practical options available, and to provide detailed supporting process guidance;

(iii) methods that are currently used for setting targets should be reviewed to ensure that defined aggregation processes are supported by appropriate targets which are fit for purpose and which are defined by methodologically sound principles;

(iv) a representative data set should be used to assess the consequences (in terms of overall assessment) of assigning quantitative data to different categorical groups. Specifically, a ‘sensitivity’ type analysis on categorised assignments should be undertaken;

(v) UK SNCBs should, as a minimum, ensure that they have a standard set of GIS layers describing the spatial extent of all monitoring or survey units and reporting areas (e.g. biogeographic seas), and consideration should be given to the redesign of monitoring units to ensure that they are more nested or ‘complementary’;

(vi) the use of different sources of information to derive weights (especially the choice between the alternatives based around the source of the data, the quality of the data and, for area-based weights, the value or quality of spatial areas represented by data) should be reviewed. It is recommended that guidance on deriving and applying appropriate weightings should be produced (to sit within or alongside improved process guidance). In addition to considering approaches relating to quantitative data, the production of weightings from qualitative sources such as expert judgement can be formalised (for example through the application of methods such as AHP) and it is recommended that consideration be given to embedding such approaches within defined aggregation and reporting processes;

(vii) it is recommended that the development and adoption of appropriate reporting formats to deliver the benefits of presenting supporting or contributory (indicator or composite indicator) information alongside high-level aggregated outputs is taken forward as a discrete project;

(viii) it is recommended that a review is undertaken to identify the range of changes (e.g. to monitoring, organisational structure, or policy) that would be required to facilitate the adoption of more quantitative aggregation methods for reporting marine biodiversity assessments;

(ix) it is recommended that an over-arching framework be developed to help place the wide (and potentially expanding) range of marine biodiversity assessments and their associated aggregations into context and to provide a framework for considering how the aggregations work. Such a framework should facilitate routine sense checking (cross referencing) between the different Type 5 aggregations; it is recommended that such cross-referencing be routinely undertaken and reported.

7. References

ALDEN, R.W., DAUER, D.M. RANASINGHE, J.A., SCOTT, L.C. & LLANSÓ, R.J. 2002. Statistical verification of the Chesapeake Bay benthic index of biotic integrity. *Environmetrics*, **13**: 473-498.

ALDER, J., CULLIS-SUZUKI, S., KARPOUZI, V., KASCHNER, K., MONDOUX, S., SWARTZ, W., TRUJILLO, P., WATSON, R. & PAULY, D. 2010. Aggregate performance in managing marine ecosystems of 53 maritime countries. *Marine Policy*, **34**: 468-476.

ANDERSON, M.J. & GRIBBLE, N.A. 1998. Partitioning the variation among spatial, temporal and environmental components in a multivariate data set. *Australian Journal of Ecology*, **23**: 158-167.

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**: 175-185.

BACARO, G., BARAGATTIA, E. & CHIARUCCI, A. 2009. Using taxonomic data to assess and monitor biodiversity: are the tribes still fighting? *Journal of Environmental Monitoring*, **11**: 798-801.

BARBOUR, M.T., GERRITSEN, J., GRIFFITH, G.E., FRYDENBORG, R., McCARRON, E., WHITE, J.S. & BASTIAN, M.L. 1996. A framework for biological criteria for Florida streams using benthic macroinvertebrates. *Journal of the North American Benthological Society*, **15(2)**: 185-211.

BARNARD, S. & BOYES, S.J. 2013. *Review of Case Studies and Recommendations for the Inclusion of Expert Judgement in Marine Biodiversity Status Assessments* - A report for the Joint Nature Conservation Committee, JNCC Report No.490 by the Institute of Estuarine and Coastal Studies, University of Hull.

BASSET, A., BARBONE, E., BORJA, A., BRUCET, S., PINNA, M., QUINTANA, X.D., REIZOPOULOU, S., ROSATI, I. & SIMBOURA, N. 2012. A benthic macroinvertebrate size spectra index for implementing the Water Framework Directive in coastal lagoons in Mediterranean and Black Sea ecoregions. *Ecological Indicators*, **12**: 72-83.

BEKETOV, M.A., KATTWINKEL, M. & LIESS, M. 2013. Statistics matter: data aggregation improves identification of community-level effects compared to a commonly used multivariate method. *Ecotoxicology*, **22**: 1516-1525.

BELIAEFF, B. & PELLETIER, D. 2011. A general framework for indicator design and use with application to the assessment of coastal water quality and marine protected area management. *Ocean & Coastal Management*, **54**: 84-92.

BELLUCCI, C.J., BECKER, M.E., BEAUCHENE, M. & DUNBAR, L. 2013. Classifying the health of Connecticut streams using benthic macroinvertebrates with implications for water management. *Environmental Management*, **51**: 1274-1283.

van den BERGH, J.C.J.M. & van VEEN-GROOT, D.B. 1999. *Constructing Aggregate Environmental-Economic Indicators: A Comparison of 12 OECD Countries*. Tinbergen Institute Discussion Papers 99-064/3, Tinbergen Institute: Amsterdam, the Netherlands.

BIAN, L. & BUTLER, R. 1999. Comparing effects of aggregation methods on statistical and spatial properties of simulated spatial data. *Photogrammetric Engineering & Remote Sensing*, **65(1)**: 73-84.

BORJA, A. & DAUER, D.M. 2008. Assessing the environmental quality status in estuarine and coastal systems: Comparing methodologies and indices. *Ecological Indicators*, **8**: 331-337.

BORJA, A. & RODRÍGUEZ, J.G. 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., BALD, J., FRANCO, J., LARRETA, J., MUXIKA, I., REVILLA, M., RODRÍGUEZ, J.G., SOLAUN, O., URIARTE, A. & VALENCIA, V. 2009a. Using multiple ecosystem components, in assessing ecological status in Spanish (Basque Country) Atlantic marine waters. *Marine Pollution Bulletin*, **59**: 54-64.

BORJA, A., BARBONE, E., BASSET, A., BORGERSEN, G., BRKLJACIC, M., ELLIOTT, M., GARMENDIA, J.M., MARQUES, J.C., MAZIK, J., MUXIKA, I., NETO, J.M., NORLING, K., RODRÍGUEZ, J.G., ROSATI, I., RYGG, B., TEIXEIRA, H, & TRAYANOVA, A. 2011. 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., BRICKER, S.B., DAUER, D.M., DEMETRIADES, N.T., FERREIRA, J.G., FORBES, A.T., HUTCHINGS, P., JIA, X., KENCHINGTON, R., MARQUES, J.C. & ZHU, C. 2008. Overview of integrative tools and methods in assessing ecological integrity in estuarine and coastal systems worldwide. *Marine Pollution Bulletin*, **56**: 1519-1537.

BORJA, A., ELLIOTT, M., ANDERSEN, J.H., CARDOSO, A.C., CARSTENSEN, J., FERREIRA, J.G., HEISKANEN, A.-S. & ZAMPOUKAS, N. 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., RANASINGHE, A. & WEISBERG, S.B. 2009b Assessing ecological integrity in marine waters, using multiple indices and ecosystem components: Challenges for the future. *Marine Pollution Bulletin*, **59**: 1-4.

BROOKS, R., McKENNEY-EASTERLING, M., BRINSON, M., RHEINHARDT, R., HAVENS, K., O'BRIEN, D., BISHOP, J., RUBBO, J., ARMSTRONG, B. & HITE, J. 2009. A Stream–Wetland–Riparian (SWR) index for assessing condition of aquatic ecosystems in small watersheds along the Atlantic slope of the eastern U.S. *Environmental Monitoring and Assessment*, **150**: 101-117.

CAIRNS, J., MCCORMICK, P.V. & NIEDERLEHNER, B.R. 1983. A proposed framework for developing indicators of ecosystem health. *Hydrobiologia*, **263**: 1-44.

CARONI, R., van de BUND, W., CLARKE, R.T. & JOHNSON, R.K. 2013. Combination of multiple biological quality elements into waterbody assessment of surface waters. *Hydrobiologia*, **704**: 437-451.

CEFAS 2002. *Development of indicators on the quality of the marine environment: proceedings of the Seabed Disturbance Indicators Workshop* (J.Jones, Ed.). Five Lakes Hotel, Tolleshunt Knights, Essex, UK, 6&7 February 2002.

CERTAIN, G., SKARPAAS, O., BJERKE, J-W., FRAMSTAD, E., LINDHOLM, M., NILSEN, J-E., NORDERHAUG, A., OUG, E., PEDERSEN, H-C., SCHARTAU, A-K., VAN DER MEEREN, G.I., ASLAKSEN, I., ENGEN, S., GARNÅSJORDET, P-A., KVALØY, P., LILLEGÅRD, M., YOCCOZ, N.G. & NYBØ, S. 2011. The Nature Index: A general framework for synthesizing knowledge on the state of biodiversity. *PLoS ONE*, **6(4)**: e18930. doi:10.1371/journal.pone.0018930. Available at: http://www.plosone.org/article/info:doi/10.1371/journal.pone.0018930 [accessed October 2012].

CZÚCZ, B., MOLNÁR, Z., HORVÁTH, F., NAGY, G.G., BOTTA-DUKÁT, Z. & TÖRÖK, K. 2012. Using the natural capital index framework as a scalable aggregation methodology for regional biodiversity indicators. *Journal for Nature Conservation*, **20**: 144-152.

DALTON, T., FORRESTER, G. & POLLNAC, R. 2012. Participation, process quality, and performance of marine protected areas in the wider Caribbean. *Environmental Management*, **49**: 1224-1237.

DeWAN A.A. & ZIPKIN, E.F. 2010. An Integrated Sampling and Analysis Approach for Improved Biodiversity Monitoring. *Environmental Management*, **45**: 1223-1230.

DIAZ, R.J., SOLAN, M. & VALENTE, R.M. 2004. A review of approaches for classifying benthic habitats and evaluating habitat quality. *Journal of Environmental Management*, **73**: 165-181.

DOBBIE, M.J. & DAIL, D. 2013. Robustness and sensitivity of weighting and aggregation in constructing composite indices. *Ecological Indicators*, **29**: 270-277.

DODGSON, J.S., SPACKMAN, M., PEARMAN, A. & PHILLIPS, L.D. 2009. *Multi-criteria analysis: a manual*. Department for Communities and Local Government: London.

DROUINEAU, H., LOBRY, J., DELPECH, C., BOUCHOUCHA, M., MAHÉVAS, S., COURRAT, A., PASQUAUD, S. & LEPAGE, M. 2012. A Bayesian framework to objectively combine metrics when developing stressor specific multimetric indicator. *Ecological Indicators*, **13**: 314-321.

EC 2005. *Overall approach to the classification of ecological status and ecological potential*. WFD CIS Guidance Document, Luxembourg, 13, 48 pp.

ETC/BD 2011. *Assessment and reporting under Article 17 of the Habitats Directive: Explanatory notes & guidelines for the period 2007-2012 (final version)*. Available at: https://circabc.europa.eu/sd/a/2c12cea2-f827-4bdb-bb56-3731c9fd8b40/Art17%20-%20Guidelines-final.pdf [accessed July 2014]

ETC/BD 2014. *Article 17 Reporting - Assessments of conservation status at the EU biogeographical level - Public consultation*. ETC/BD Technical paper 3/2014, Paris. Available at: http://bd.eionet.europa.eu/activities/Reporting/Article_17/Documents/ART17%20public%20consultation%20guide.pdf [accessed July 2014]

FALCONE, J.A., CARLISLE, D.M. & WEBER, L.C. 2013. Quantifying human disturbance in watersheds: Variable selection and performance of a GIS-based disturbance index for predicting the biological condition of perennial streams. *Ecological Indicators*, **10**: 264-273.

FULAZZAKY, M.A. 2013. Assessing the suitability of stream water for five different uses and its aquatic environment. *Environmental Monitoring and Assessment*, **185**: 523-535.

GLEASON, M., FOX, E., ASHCRAFT, S., VASQUES, J., WHITEMANE, E., SERPA, P., SAARMAN, E., CALDWELL, M., FRIMODIG, A., MILLER-HENSON, M., KIRLIN, J., OTA, B., POPE, E., WEBER, M. & WISEMAN, K. 2013. Designing a network of marine protected areas in California: Achievements, costs, lessons learned, and challenges ahead. *Ocean & Coastal Management*, **74**: 90-101.

GOTWAY, C.A. & YOUNG, L.J. 2002. Combining incompatible spatial data. *Journal of the American Statistical Association*, **97**: 632-648.

GOTWAY, C.A. & YOUNG, L.J. 2007. A geostatistical approach to linking geographically aggregated data from different sources. *Journal of Computational and Graphical Statistics*, **16(1)**: 115-135.

HALE, S.S. & HOLLISTER, J.W. 2009. Beyond data management: how ecoinformatics can benefit environmental monitoring programs. *Environmental Monitoring and Assessment*, **150**: 227-235.

HALPERN, B.S., LONGO, C., HARDY, D., MCLEOD, K.L., SAMHOURI, J.F., KATONA, S.K., KLEISNER, K., LESTER, S.E., OLEARY, J.A., RANELLETTI, M., ROSENBERG, A.A., SCARBOROUGH, C., SELIG, E.R., BEST, B.D., BRUMBAUGH, D.R., CHAPIN, F.S., CROWDER, L.B., DALY, K.L., DONEY, S.C., ELFES, C., FOGARTY, M.J., GAINES, S.D., JACOBSEN, K.I., KARRER, L.B., LESLIE, H.M., NEELEY, E., PAULY, D., POLASKY, S., RIS, B., ST MARTIN, K., STONE, G.S., RASHID SUMAILA, U. & ZELLER, D. 2012. An index to assess the health and benefits of the global ocean. *Nature*, **488**: 615-620.

HAMMOND, A., ADRIAANSE, A. & RODENBURG, E. 1995. *Environmental indicators: a systematic approach to measuring and reporting on environmental policy performance in the context of sustainable development*. World Resources Institute, 58pp

HARRISON, T.D. & KELLY, F.L. 2013. Development of an estuarine multi-metric fish index and its application to Irish transitional waters. *Ecological Indicators*, **34**: 494-506.

HECK, N., DEARDEN, P., MCDONALD, A. & CARVER, S. 2011. Developing MPA performance indicators with local stakeholders' input in the Pacific Rim National Park Reserve, Canada. *Biodiversity and Conservation*, **20**: 895-911.

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

HESS, G.R., BARTEL, R.A., LEIDNER, A.K., ROSENFELD, K.M., RUBINO, M.J., SNIDER, S.B. & RICKETTS, T.H. 2006. Effectiveness of biodiversity indicators varies with extent, grain, and region. *Biological Conservation*, **132**: 448-457.

HILL, J.M., EARNSHAW, S., BURKE, C. & GALLYOT, J. 2012. *Reviewing and recommending methods for determining reference conditions for marine benthic habitats in the North-East Atlantic Region*. - A report for the Joint Nature Conservation Committee, JNCC Report No.464 by Marine Ecological Surveys Ltd.

HINCHEN, H. 2014. *Review of marine biodiversity assessment obligations in the UK. Part I: A summary of the marine biodiversity assessment obligations stipulated within national and international legislative and policy instruments*. JNCC Report 497, JNCC, Peterborough. 123pp.

HOLLERT, H., HEISE, S., PUDENZ, S., BRÜGGEMANN, R., AHLF, W. & BRAUNBECK, T. 2002. Application of a sediment quality triad and different statistical approaches (Hasse diagrams and fuzzy logic) for the comparative evaluation of small streams. *Ecotoxicology*, **11**: 311-321.

JACKSON, D.L. & McLEOD, C.R. (Editors), 2000, 2002. *Handbook on the UK status of EC Habitats Directive interest features: provisional data on the UK distribution and extent of Annex I habitats and the UK distribution and population size of Annex II species*, Revised 2002, JNCC Report 312, 180 pp.

JNCC 2004a. *Common Standards Monitoring guidance - generic introduction for marine feature guidance*. JNCC. 9pp. Available online at: http://jncc.defra.gov.uk/PDF/CSM_marine_introduction.pdf [accessed July 2014]

JNCC 2004b. *Common Standards Monitoring guidance - Littoral rock and inshore sublittoral rock habitats*. JNCC. 27pp. Available at: http://jncc.defra.gov.uk/PDF/CSM_marine_rock.pdf [accessed July 2014]

JOLLANDS, N. 2006. How to aggregate sustainable development indicators: a proposed framework and its application. *International Journal of Agricultural Resources, Governance and Ecology (IJARGE)*, **5(1)**: 18-34.

JOLLANDS, N., LERMIT, J. & PATTERSON, M. 2003. The usefulness of aggregate indicators in policy making and evaluation: a discussion with application to eco-efficiency indicators in New Zealand. Economics and Environment Network, Australian National University.

JOLLANDS, N., LERMIT, J. & PATTERSON, M. 2004. Aggregate eco-efficiency indices for New Zealand - a principal components analysis. *Journal of Environmental Management*, **73**: 293-305.

KONTULA, T. & RAUNIO, A. 2009. New method and criteria for national assessments of threatened habitat types. *Biodiversity Conservation*, **18**: 3861-3876.

LEE, K-M., LEE, S.Y. & CONNOLLY, R.M. 2012. Combining process indices from network analysis with structural population measures to indicate response of estuarine trophodynamics to pulse organic enrichment. *Ecological Indicators*, **18**: 652-658.

LLANSÓ, R.J., SCOTT, L.C., HYLAND, J.L., DAUER, D.M., RUSSELL, D.E. & KUTZ, F.W. 2002. An estuarine benthic Index of Biotic Integrity for the Mid-Atlantic region of the United States: II Index Development. *Estuaries*, **25(6A)**: 1231-1242.

LLANSÓ, R.J., VØLSTAD, J.H., DAUER, D.M. & DEW, J.R. 2009. Assessing benthic community condition in Chesapeake Bay: does the use of different benthic indices matter? *Environmental Monitoring and Assessment*, **150**: 119-127.

MARTíNEZ-CREGO, B., ALCOVERRO, T. & ROMERO, J. 2010. Biotic indices for assessing the status of coastal waters: a review of strengths and weaknesses. Jo*urnal of Environmental Monitoring*, **12**: 1013-1028.

MAVRIC, B., URBANIC, G., LIPEJ, L. & SIMBOURA, N. 2013. Influence of sample size on ecological status assessment using marine benthic invertebrate-based indices. *Marine Ecology*, **34**: 72-79.

McMAHON, B.J., ANDERSON, A., CARNUS, T., HELDEN, A.J., KELLY-QUINN, M., MAKI, A., SHERIDAN, H. & PURVIS, G. 2012. Different bioindicators measured at different spatial scales vary in their response to agricultural intensity. *Ecological Indicators*, **18**: 676-683.

MOLDAN, B., JANOUŠKOVÁ, S. & HÁK, T. 2012. How to understand and measure environmental sustainability: Indicators and targets. *Ecological Indicators*, **174**: 4-13.

MÖLLMANN, C., LINDEGREN, M., BLENCKNER, T., BERGSTRÖM, L., CASINI, M., DIEKMANN, R., FLINKMAN, J., MÜLLER-KARULIS, B., NEUENFELDT, S., SCHMIDT, J. O., TOMCZAK, M., VOSS, R. & GÅRDMARK, A. 2013. Implementing ecosystem-based fisheries management: from single-species to integrated ecosystem assessment and advice for Baltic Sea fish stocks. *ICES Journal of Marine Science*, doi:10.1093/icesjms/fst123.

MPA MONITORING ENTERPRISE. 2011. *South Coast MPA Monitoring Plan*. California Ocean Science Trust, California, USA.

MUXIKA, I., BORJA, A. & BALD, J. 2007. Using historical data, expert judgement and multivariate analysis in assessing reference conditions and benthic ecological status, according to the European Water Framework Directive, *Marine Pollution Bulletin*, **55**, 16–29.

NARDO, M., SAISANA, M., SALTELLI, A., TARANTOLA, S., HOFFMAN, A. & GIOVANNINI, E. 2005. *Handbook on constructing composite indicators: Methodology and User Guide*. OECD Statistics Working Paper, 2005/03. OECD Publishing.

NESTLERODE, J.A., ENGLE, V.D., BOURGEOIS, P., HEITMULLER, P.T., MACAULEY, J.M. & ALLEN, Y.C. 2009. An integrated approach to assess broad-scale condition of coastal wetlands—

the Gulf of Mexico Coastal Wetlands pilot survey. *Environmental Monitoring and Assessment*, **150**: 21-29.

NEUMAYER, E. 2001. The human development index and sustainability – a constructive proposal. *Ecological Economics*, **39**, 101-114.

NORMANDER, B., LEVIN, G., AUVINEN, A-P., BRATLI, H., STABBETORP, O., HEDBLOM, M., GLIMSKÄR, A. & GUDMUNDSSON, G.A. 2012. Indicator framework for measuring quantity and quality of biodiversity - Exemplified in the Nordic countries. *Ecological Indicators*, **13**: 104–116.

OECD & JRC 2008. *Handbook on constructing Composite Indicators: methodology and user guide*. OECD, Paris.

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(4)**: e19231. doi:10.1371/journal.pone.0019231.

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

OSPAR 2009. *CEMP assessment report: 2008/2009 - Assessment of trends and concentrations of selected hazardous substances in sediments and biota*. Monitoring and assessment series, publication number 390/2009. London, OSPAR, 79pp.

OSPAR 2014. *Bird 1: Technical Specification*. Information Document presented by Netherlands, as convenor of ICG-COBAM to the Meeting of the Biodiversity Committee, The Hague: 18-22 February 2014. BDC 14/2/Info.2-E, London, OSPAR.

PARACCHINI, M. L., PACINI, C., JONES, M. L. M. & PEREZ-SOBA, M. 2011. An aggregation framework to link indicators associated with multifunctional land use to the stakeholder evaluation of policy options. *Ecological Indicators*, **11**: 71-80.

PARACCHINI, M.L., PACINI, C., CALVO, S. & VOGT, J. 2008. Weighting and aggregation of indicators for sustainability impact assessment in the IP SENSOR context. In: Helming, K., Tabbush, P., Perez Soba, M. (Eds.), *Sustainability Impact Assessment of Land Use Changes*. Springer-Verlag, 349-372.

PAUL, J.F., SCOTT, K.J., CAMPBELL, D.E., GENTILE, J.H., STROBEL, C.S., VALENTE, R.M., WEISBERG, S.B., HOLLAND, A.F. & RANASINGHE, J.A. 2001. Developing and applying a benthic index of estuarine condition for the Virginian biogeographic province. *Ecological Indicators*, **1**: 83-99.

PÉREZ-DOMÍNGUEZ, R., MACI, S., COURRAT, A., LEPAGE, M., BORJA, A., URIARTE, A., NETO, J.M., CABRAL, H., ST.RAYKOV, V., FRANCO, A., ALVAREZ, M.C. & ELLIOTT, M. 2012. Current developments on fish-based indices to assess ecological-quality status of estuaries and lagoons. *Ecological Indicators*, **23**: 34-45.

PIET, G.J., ALBELLA, A.J., ARO, E., FARRUGIO, H., LLEONART, J., LORDAN, C., MESNIL, B., PETRAKIS, G., PUSCH, C., RADU, G. & RÄTZ, H-J. 2010.. *Marine Strategy Framework Directive – Task Group 3 Report Commercially exploited fish and shellfish*. EUR – Scientific and Technical Research series, Luxembourg, JRC/ICES, EUR 24316 EN, 82 pp.

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

PRIMPAS, I., TSIRTSIS, G., KARYDIS, M. & KOKKORIS, G.D. 2010. Principal component analysis: Development of a multivariate index for assessing eutrophication according to the European water framework directive. *Ecological Indicators*, **10**: 178-183.

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*. Report number 1207879-000. Deltares; Delft, NL. 61pp

RICE, J., ARVANITIDIS, C., BORJA, A., FRID, C., HIDDINK, J., KRAUSE, J., LORANCE, P., RAGNARSSON, S.Á., SKÖLD, M. & TRABUCCO, B. 2010. *Marine Strategy Framework Directive – Task Group 6 Report Seafloor integrity*. EUR – Scientific and Technical Research series, Luxembourg, JRC/ICES, EUR 24334 EN, 73 pp.

RODRIGUEZ-RODRIGUEZ, D. & MARTINEZ-VEGA, J. 2012. Proposal of a system for the integrated and comparative assessment of protected areas. *Ecological Indicators*, **23**: 566–572.

RODRÍGUEZ-RODRÍGUEZ, D. & MARTÍNEZ-VEGA, J. 2013. Results of the implementation of the System for the Integrated Assessment of Protected Areas (SIAPA) to the protected areas of the Autonomous Region of Madrid (Spain). *Ecological Indicators*, **34**: 210-220.

ROMBOUTS, I., BEAUGRAND, G., ARTIGAS, L.F., DAUVIN, J.-C., GEVAERT, F., GOBERVILLE, E., KOPP, D., LEFEBVRE, S., LUCZAK, C., SPILMONT, N., TRAVERS-TROLET, M., VILLANUEVA, M.C. & KIRBY, R.R. 2013. Evaluating marine ecosystem health: Case studies of indicators using direct observations and modelling methods. *Ecological Indicators*, **24**: 353-365.

ROSENBERG, R., BLOMQVIST, M., NILSSON, H.C., CEDERWALL, H. & DIMMING, A. 2004. Marine quality assessment by use of benthic species-abundance distributions: a proposed new protocol within the European Union Water Framework Directive. *Marine Pollution Bulletin*, **49**: 728-739.

SALAS, F., MARCOS, C., NETO, J.M., PATRÍCIO, J., PÉREZ-RUZAFA, A. & MARQUES, J.C. 2006. User-friendly guide for using benthic ecological indicators in coastal and marine quality assessment. *Ocean & Coastal Management*, **49**: 308-331.

SAMY, M., SÁNCHEZ-LIZASO, J. M. & FORCADA, A. 2011. Status of marine protected areas in Egypt. *Animal Biodiversity and Conservation*, **34(1)**: 165-177.

SCHOLES, R. J. & BIGGS, R. 2005. A biodiversity intactness index. *Nature*, **434**: 45-49.

SCHOOLMASTER, D.R., GRACE, J.B., SCHWEIGER, E.W., GUNTENSPERGEN, G.R., MITCHELL, B.R., MILLER, K.M. & LITTLE, A.M. 2013. An algorithmic and information-theoretic approach to multimetric index construction. *Ecological Indicators*, **26**: 14-23.

SHARPE, A. 2004. *Literature Review of Frameworks for Macro-indicators*. Centre for the Study of Living Standards, Ottawa, Canada.

SHERBININ, D. A., REUBEN, A., LEVY, M. & JOHNSON, L. 2013. *Indicators in Practice: How Environmental Indicators are Being Used in Policy and Management Contexts*. New Haven and New York: Yale and Columbia Universities.

SIMBOURA, N., PANAYOTIDIS, P. & PAPATHANASSIOU, E. 2005. A synthesis of the biological quality elements for the implementation of the European Water Framework Directive in the Mediterranean ecoregion: The case of Saronikos Gulf. *Ecological Indicators*, **5**: 253-266.

SIMBOURA, N., ZENETOS, A., PANCUCCI-PAPADOPOULOU, M.A., 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(1)**: 140-152.

SIMONS, J.D. & SMITH, C.R. 2009. Texas National Coastal Assessment (2000–2004): challenges, solutions, lessons learned and future directions. *Environmental Monitoring and Assessment*, **150**: 167-179.

SINGH, R.K., MURTY, H.R., GUPTA, S.K. & DIKSHIT, A.K. 2012. An overview of sustainability assessment methodologies. *Ecological Indicators*, **15**: 281-299.

SKARPAAS, O., CERTAIN, G. & NYBØ, S. 2012. The Norwegian Nature Index – conceptual framework and methodology. *Norsk Geografisk Tidsskrift (Norwegian Journal of Geography)*, **66(5)**: 250-256.

STARK, J.D. 1998. SQMCI: A biotic index for freshwater macroinvertebrate coded abundance data. *New Zealand Journal of Marine and Freshwater Research*, **32(1)**, 55-66.

STODDARD, J.L., HERLIHY, A.T., PECK, D.V., HUGHES, R.M., WHITTIER, T.R. & TARQUINIO, E. 2008. A process for creating multimetric indices for large-scale aquatic surveys. *Journal of the North American Benthological Society*, **27(4)**: 878-891.

SYRBE, R.U., MICHEL, E. & WALZ, U. 2013. Structural indicators for the assessment of biodiversity and their connection to the richness of avifauna. *Ecological Indicators*, **31**: 89-98.

TETT, P., GOWEN, R., MILLS, D., FERNANDES, T., GILPIN, L., HUXHAM, M., KENNINGTON, K., READ, P., SERVICE, M., WILKINSON, M. & MALCOLM, S. 2007. Defining and detecting undesirable disturbance in the context of marine eutrophication. *Marine Pollution Bulletin*, **55**: 282-297.

TOGNELLI, M.F., FERNÁNDEZ, M & MARQUET, P.A. 2009. Assessing the performance of the existing and proposed network of marine protected areas to conserve marine biodiversity in Chile. *Biological Conservation*, **142**: 3147-3153.

TRENKEL, V.M. & ROCHET, M.J. 2010. Combining time trends in multiple metrics for identifying persistent changes in population processes or environmental stressors. *Journal of Applied Ecology*, **47**: 751-758.

TUEROS, I., BORJA, A., LARRETA, J., RODRÍGUEZ, J.G., 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.

VAČKÁŘ, D., BRINK, B., LOH, J., BAILLIE, J.E.M. & REYERS, B. 2012. Review of multispecies indices for monitoring human impacts on biodiversity. *Ecological Indicators*, **17**: 58-67.

van den BERGH, J. C. & van VEEN-GROOT, D. B. 1999.. *Constructing aggregate environmental-economic indicators: a comparison of 12 OECD countries* (Report number 99-064/3). Tinbergen Institute.

VOLLENWEIDER, R.A., GIOVA NARDI, F., MONTANARI, G. & RINALDI, A. 1998. Characterization of the trophic conditions of marine coastal waters with special reference to the NW Adriatic Sea: Proposal for a trophic scale, turbidity and generalized water quality Index. *Environmetrics*, **9**: 329-357.

WARD, T.J. 2011. *SOE 2011: National marine condition assessment – decision model and workshops*. Report prepared for the Australian Government Department of Sustainability, Environment, Water, Population and Communities on behalf of the State of the Environment 2011 Committee. Canberra: DSEWPaC, 2011.

WEISBERG, S.B., RANASINGHE, J.A., LINDA C. SCHAFFNER, L.C., DIAZ, R.J., DAUER, D.M. & FRITHSEN J.B. 1997. An estuarine Benthic Index of Biotic Integrity (B-IBI) for Chesapeake Bay. *Estuaries*, **20(1)**: 149-158.

WIKLE, C.K. & BERLINER, L.M. 2005. Combining information across spatial scales. *Technometrics*, **47(1)**: 80-91.

YODER, C.O. & BARBOUR, M.T. 2009. Critical technical elements of state bioassessment programs: a process to evaluate program rigor and comparability. *Environmental Monitoring and Assessment*, **150**: 31-42.

ZHOU, P. & ANG, B.W. 2009. Comparing MCDA aggregation methods in constructing Composite Indicators using the Shannon-Spearman measure. *Social Indicators Research*, **94**: 83-96.

ZHOU, P., ANG, B.W. & POH, K.L. 2006. Comparing aggregating methods for constructing the composite environmental index: an objective measure. *Ecological Economics*, **59**: 305-311.

ZHOU, P., FAN, L-W. & ZHOU, D-Q. 2010. Data aggregation in constructing composite indicators: a perspective of information loss. *Expert Systems with Applications*, **37**: 360-365.

8. Glossary

AA	Averaging Approach
AI	Assessment Index
AHP	Analytic Hierarchy Process
AMBI	AZTI Marine Biotic Index
BAT	Benthic Assessment Tool
BD	Birds Directive
BEQI	Benthic Ecosystem Quality Index
B-IBI	Benthic Index of Biotic Integrity
BII	Biodiversity Intactness Index
BQE	Biological Quality Element
BQI	Benthic Quality Index
CI	Composite Indicator
CBD	Convention on Biological Diversity
CMS	Convention on the Conservation of Migratory Species of Wild Animals
CR	Conditional Rule
CSA	Conservation of Seals Act
CSM	Common Standards Monitoring
DEVOTES	DEVelopment Of innovative Tools for understanding marine biodiversity and assessing good Environmental Status
EEA	European Environment Agency
EEI	Environmental Integrative Indicator
EEZ	Exclusive Economic Zone
EQR	Ecological Quality Ratio
ETC/BD	European Topic Centre on Biological Diversity
EUBS	European Union Biodiversity Strategy
FRP	Favourable Reference Population

FRR	Favourable Reference Range
FRV	Favourable Reference Value
GEcS	Good Ecological Status
GEnS	Good Environmental Status
GNP	Gross National Product
HBDS	Healthy and Biologically Diverse Seas
HD	Habitats Directive
HDI	Human Development Index
HELCOM	Baltic Marine Environment Protection Commission (Helsinki Commission)
HLMO	High Level Marine Objectives
I	Indicator
IMF	International Monetary Fund
INFOMAR	Integrated Mapping for the Sustainable Development of Ireland's Marine Resource
ISS	Index of Size-spectra Sensitivity
lnRR	Log-response ratio
LUF	Land-use function
OECD	Organisation for Economic Co-operation and Development
OOAO	One Out, All Out
OSPAR	Oslo and Paris Convention for the protection of the marine environment of the North East Atlantic
M-AMBI	Multivariate AZTI Marine Biotic Index
MAIA	Mid-Atlantic Integrated Assessment
MARLIN	Marine Life Information Network
MCAA	Marine and Coastal Access Act
MCZ	Marine Conservation Zone
MPA	Marine Protected Area
MPS	Marine Policy Statement
MS	Member State
MSA	Marine Scotland Act
MSFD	Marine Strategy Framework Directive
MMI	Multi-Metric Index
NAA	Non-Averaging Approach
NCI	Natural Capital Index
NI	Nature Index
NQI	Norwegian Quality Index
PA	Protected Area
PCA	Principal Components Analysis
QSR	Quality Status Report
R&D	Research & Development
SAW	Simple Additive Weighting
SSSI	Special Site of Scientific Interest
SWR	Stream–Wetland–Riparian
TOAO	Two Out All Out
TRIX	Triple Exponential (technical analysis)

UNCLOS	United Nations Convention on the Law of the Sea
WCA	Wildlife and Countryside Act
WDI	World Development Indicator
WP	Weighted Product
WFD	Water Framework Directive