ICES WKD1Agg REPORT 2016

ICES ADVISORY COMMITTEE

ICES CM 2016/ACOM:43

REF. ACOM

Report of the Workshop on providing a method to aggregate species within species groups for the assessment of GES for MSFD D1 (WKD1Agg)

29 February–2 March 2016

ICES HQ, Copenhagen, Denmark

ICES CIEM
International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

H. C. Andersens Boulevard 44–46
DK-1553 Copenhagen V
Denmark
Telephone (+45) 33 38 67 00
Telefax (+45) 33 93 42 15
www.ices.dk
info@ices.dk

Recommended format for purposes of citation:

ICES. 2016. Report of the Workshop on providing a method to aggregate species within species groups for the assessment of GES for MSFD D1 (WKD1Agg), 29 February–2 March 2016, ICES HQ, Copenhagen, Denmark. ICES CM 2016/ACOM:43. 53 pp.

Contents

Executive summary

The WKD1Agg workshop was initiated to address the following request: 'In the light of previous advice, provide guidance on the most suitable and defendable approach to aggregate species within species groups (such as birds, mammals, reptiles, fish and cephalopods), for the state assessments of the MSFD'. The final outcome was to be communicated to DG ENV, to further support and facilitate the review of the Commission Decision (2010/477/EU) and the MSFD implementation for D1. The workshop brought together 24 experts from 12 European countries.

The workshop participants considered that the term 'integration' should be used instead of 'aggregation' when referring to combining across criteria and across species or indicators. The group concluded that there were two feasible frameworks for integrating indicators to ecosystem component level: integration of indicators or species within criteria and integration of criteria within species. Within each of these options is the question of how to address subdivision-specific assessments, and integration of indicators or species that are assessed at different spatial scales.

Integration by criteria carries the advantage of a transparent weighting of all criteria, including community aspects. Data for species where only one criterion can be measured can be included without loss of consistency and all criteria receive equal weight in the integration across criteria. Subdivisions can be addressed in different ways depending on whether the aim is to achieve a subdivision or subregional evaluation of GES. Integration of criteria within species carries the advantage of increasing comparability with the Habitats Directive at the species level, though this comparability will decrease if the species and indicators included in the integration at ecosystem component level differ from those addressed by the Habitats Directive. Community considerations are, however, not easily integrated. The group did not consider that species rich and species poor ecosystem components were conceptually different except that integration by species is easier for the latter. In any case, the integration framework selection is a data driven process, where rich data/species integration allows for probabilistic and averaging approaches.

The group considered a range of integration methods, the pros and cons of each, and situations in which each would be applicable. The two integration frameworks are exemplified by worked examples of fish, birds and mammals from all regions, with identification of appropriate integration methods for each level. When integrating criteria within a species, or to species group level, the recommended method is OOAO, and the same recommendation is made when integrating species groups to ecosystem component for both approaches. At other levels (e.g. individual indicators to criteria, species to species group), a mixture of integration methods including OOAO, averages, weighted averages, proportional and probabilistic methods are recommended depending on the specific situation. Retaining and communicating distance to target information was considered useful.

The final choice of integration framework depends on the relative consideration of the wish to facilitate comparability with the Habitats Directive, the wish to use similar methods across all ecosystem components, the wish to give a transparent weight to all criteriavs.giving transparent weight to all species, the wish to incorporate community-level considerations in the integration and finally, the wish to facilitate species or species group assessments incorporating HD criteria assessments. The workshop participants considered that the weighing of these different wishes might vary across assessment efforts, affecting the final choice of integration framework. Furthermore,

the integration approach in high level integration for both schemes can be subject to policy rather than scientific decision, depending on the way the overall GES for each biodiversity component has been determined. The lack of community level indicators was identified as a major gap in the species level assessments during the D1 review process.

Hence, this report lists the pros and cons for both frameworks along with recommended integration methods under each to facilitate informed decision-making.

1 Opening of the meeting

The agenda, aims of the workshop, and expected outcomes were presented at the opening of the meeting. The terms of reference of the group was to 'In the light of previous advice, provide guidance on the most suitable and defendable approach to aggregate species within species groups (such as birds, mammals, reptiles, fish and cephalopods), for the state assessments of the MSFD'. The final outcome will be communicated to DG ENV, to further support and facilitate the review of the Commission Decision (2010/477/EU) and the MSFD implementation for D1. The workshop brought together 24 experts from 12 European countries (Figure1, Annex 1).

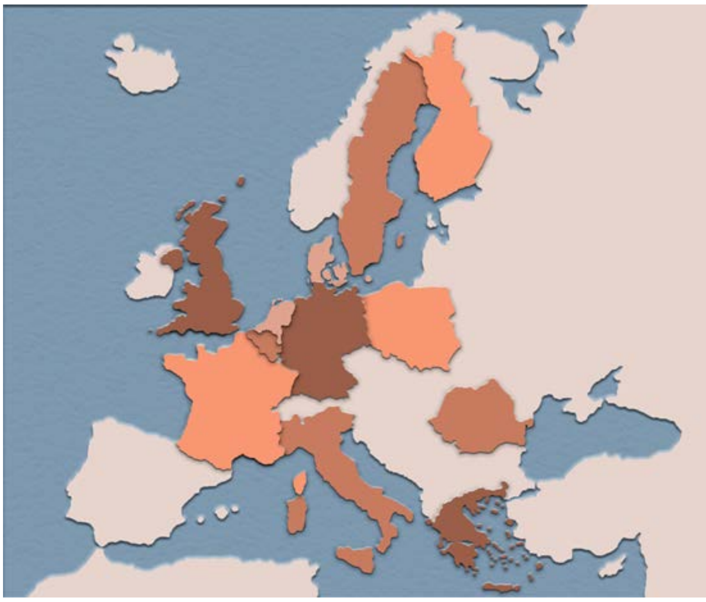

Figure 1. The 12 European countries workshop brought experts together from.

2 Background and initial considerations

The MSFD Regulatory Committee, considering the outcomes of the evaluation of the Member States reporting on the initial assessment, good environmental status (GES) determination and target-setting (COM 2014/97; Palialexis *et al*. 2014), decided to proceed with the review and possible revision of the Commission Decision on criteria and methodological standards on good environmental status of marine waters (GES Decision 2010/477/EU) and of MSFD Annex III. A roadmap for the review of the Commission Decision 2010/477/EU was endorsed by the MSFD Regulatory Committee and includes 3 phases:

- **Phase 1: Technical and scientific review**
- **Phase 2: Consultation and discussion**
- **Phase 3: Decision-making**

For phase 1 of the review, a number of lead experts from the respective MSFD Common Implementation Strategy and Regional Sea Conventions (RSC) groups under the coordination of JRC and ICES prepared the technical and scientific reviews per descriptor that resulted in recommendations for revision and a proposed draft text with changes and the rational for those changes. Additional outstanding issues were identified for D1 that were subsequently discussed in a workshop with the participation of the D1 expert group's members and stakeholders (JRC, Ispra, September 2015). In the workshop, it was highlighted that additional discussion and work on aggregation rules was needed to provide guidance for a coherent and concrete framework for integrating assessments towards the overall assessment of species under D1. DG ENV gave ICES the mandate to organize a dedicated workshop (WKD1Agg) on providing methods to aggregate species within species groups for the assessment of GES for MSFD D1, building on the outcomes of the JRC's workshop, namely the agreed list of biodiversity components/species groups and on the indicative species assessment framework (Table 1 and Figure 2 of the JRC's workshop report (Palialexis *et al*., 2015)).

2.1 Terminology

In this document, we refer to indicator values outside the desired values as the indicator being in poor status, and inside the desired values as the indicator being in good status. The word 'aggregation' carried different notations to participants in the group, and as a result, it was agreed to use the term 'integration' when referring to the process of combining the state of several indicators into one higher-level indicator (e.g. for a species group, criterion or ecosystem component). The term 'integration framework' is used to describe the different options for order of integration (e.g. of species within a criterion, or of criteria within a species). 'Ecosystem component' refers to broad biodiversity groups (Table 3.1) as specified in the JRC's D1 review workshop (Palialexis *et al*., 2015).

2.2 State and pressure indicators

As the request refers only to 'state' the group has only considered integration of state indicators. The group considers that it is necessary to distinguish between pressure and state indicators also in the integration process to facilitate the communication of results.

2.3 Assumptions made

Throughout this document, it is assumed that the indicators suggested have appropriate quality. Further, it is assumed that species groups are only used when deemed relevant and preferably only when supported by a number of indicators. If a species group is represented by only one indicator, it seems more appropriate to investigate the ecosystem component directly without first integrating within species groups.

3 Levels of integration

Four possible integration levels within each ecosystem component were discussed: Species groups, species, criteria and areas. The specific categories in each level are given in table 3.1 and table 3.2. The D1 group recommended using a 'community criterion' which would contain biodiversity indicators at the community level (e.g. fraction of sensitive species increasing).

GES criteria are taken from the 2010 Commission Decision with the exception of criteria 1.7. According to art.12 assessment for art. 8, 9 and 10 (ACOM 2014/97 and Palialexis *et al*. 2014) criterion 1.7 and indicator 1.7.1 were not clearly defined which led member states to omit it or differently interpreted it and report it. In the D1 review workshop it was recommended to omit 1.7 and introduce a new criterion (1.4 mobile species community composition; 1.4.1 Relative abundance of community elements) to fill in the gap of a community-level metric for species. The inclusion of a new community-level criterion was not adapted by DG ENV in their first draft of the revised COM DEC text. The suggested new community criterion is referred to here as criterion 1.X. The group was aware that this criterion may not be adopted, but decided that since this work has not yet concluded, the methods suggested here should be applicable both to criteria 1.1 to 1.3 and the suggested criterion 1.X.

The integration levels considered by the group includes up to the level of ecosystem component. Integrating across ecosystem components (birds, mammals, fish etc.) to descriptor level, and across descriptors, were not considered.

Table 3.1 Specific categories of ecosystem component and species group.

Ecosystem component	Species groups
Birds	Grazing birds
	Wading birds
	Surface-feeding birds
	Pelagic-feeding birds
	Benthic-feeding birds
Mammals	Small toothed cetaceans
	Deep-diving toothed cetaceans
	Baleen whales
	Seals
Reptiles	Turtles
Fish	Coastal fish
	Pelagic shelf fish
	Demersal shelf fish
	Deep-sea fish
Cephalopods	Coastal/shelf cephalopods
	Deep-sea cephalopods

Table 3.2 Criteria, regions, subregions and examples of subdivisions referred to in the text.

Criteria	1.1 Species distribution 1.2 Population size 1.3 Population condition 1.X Ecosystem Structure
Regions	Northeast Atlantic, Baltic, Mediterranean, Black Sea
Subregions	Greater North Sea, Celtic Seas, Bay of Biscay and Iberian coast, wider Atlantic (Macronesia), western Mediterranean Sea, Adriatic Sea, Ionian Sea and central Mediterranan Sea, Aegean-Levantine Sea
Examples of subdivisions	Northwestern North Sea, Northeastern North Sea, Kattegat/Skaggerrak, English Channel, Bothnian Bay, Northern Baltic proper

The methods used for integration at each level should ensure that the final assessment of GES at the ecosystem component level is unbiased by differences in the number of indicators used at each level. For example, two species may be monitored by 5 and 1 indicators, respectively. The selection of integration method should explicitly consider the implication of the different number of indicators, and whether equal weighting should be applied or not. Similar examples can be made with criteria and areas with unequal numbers of indicators associated with them. The group discussed a range of different integration methods, the pros and cons of each, and situations in which they would be appropriate to be applied.

The scientific rationale and consequences of the choice of integration framework and integration methods to use was debated in the group and the conclusions are given below. The decision to use a specific integration method at each level is related to the interpretation of the current revision text and the possible requirement to use integration methods for some species groups which are compatible with requirements under the Habitat and Birds directives. This decision was considered to be in the policy rather than scientific domain. Hence, this report presents pros and cons of choosing methods compatible or not compatible with the current interpretation of related directives.

3.1 Integration within criteria

Under this integration framework, all species group receive a specified weight in the integration regardless of the number of indicators available. The integration follows 4 levels, from species or subdivision to indicator, from indicator to criterion, from criterion to species group and from species group to ecosystem component (figure 2).

3.1.1 Pros and Cons

This type of integration framework is technically possible for indicators under species criteria 1.1–1.3, where each individual criterion is assessed using the available indicators and species. It can easily accommodate assessments on a large number of species. Data for species where only one criterion can be measured can be included without loss of consistency. This framework can furthermore easily accommodate community-level indicators that would be assessed under a community criterion (criterion 1.X), (provided that the community indicator is specific to the species group).

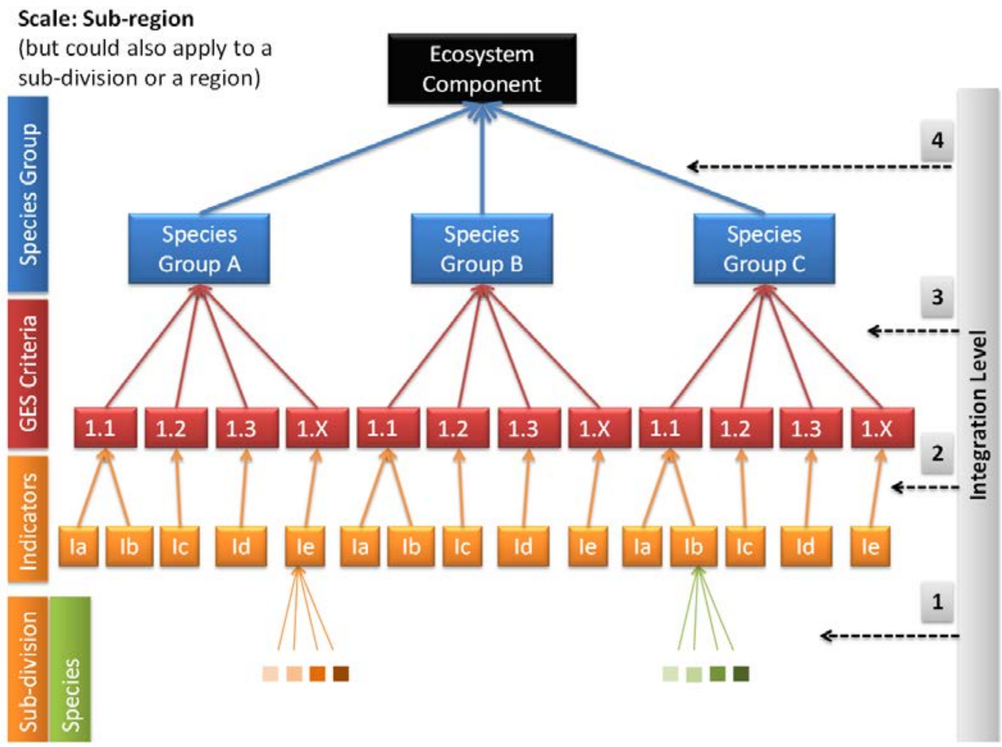

Figure 2: A schema showing integration within criteria. Level 2 integration is within criteria across all species in the species group followed by integration across criteria within species groups, followed by integration to ecosystem component. GES criteria are taken from the Commission Decision. Criterion 1.X refers to a community structure. The indicators and species groups shown are arbitrary examples.

3.2 Integration within species

Under this integration framework, all species receive an equal weight in the integration regardless of the number of indicators available for each species. The integration follows 4 or 5 levels, from subdivision to indicator at subregional scale (if appropriate) from indicator to criterion, from criterion to species, from species to species group and from species group to ecosystem component (Figure3).

3.2.1 Pros and cons

This integration framework is based on an individual species view and hence is technically possible for the species-specific indicators under criteria 1.1–1.3. The integration ensures comparability and streamlining of assessments with the Habitats Directive at the species level, where the status of individual species is assessed using different parameters or indicators, such as range, population, habitat and future prospect. Under the Habitats Directive, 2 of these parameters must be known, otherwise the species is not assessed. Under D1 of the MSFD, no such requirement exists, and species can be assessed even if only one indicator is known for the species, though this will decrease the direct comparability with the Habitats Directive. Furthermore, D1 focuses on species that are representative of the wider ecosystem component and not only the status of species that are identified as threatened or declining under the Habitats Directive. This contributes to possible differences between the assessment results of the two directives for specific ecosystem components. If complete similarity in the assessment of a species group is desired, it may be necessary to exclude data, leading to a more limited data material and possibly a less accurate assessment across all species in the component.

The inclusion of more species will probably provide a more accurate assessment of the ecosystem component but it may prove unfeasible to collect information on several parameters for every species.

This integration framework could allow for, those species that are assessed under the Habitats Directive, the status result for those species to be directly used in MSFD assessments, complemented by species assessments for additional species that are relevant to MSFD but which are not included in the Habitats Directive.

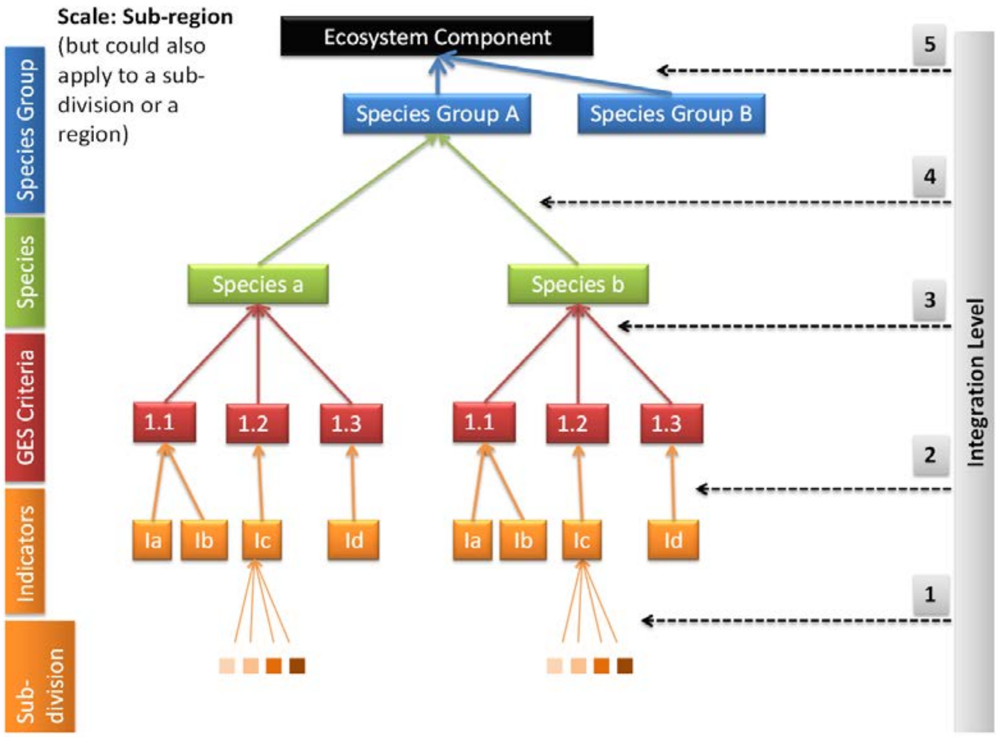

Figure 3: A schema showing integration within species. Level 2 integration is within criteria within each species followed by integration across species within each species group, and finally to ecosystem component. The indicators and species groups shown are generalized examples. This integration framework is similar to that used in Habitats Directive assessments of Favourable Conservation Status.

Community structure criteria (Criterion 1.X) cannot be allocated to species and therefore it must be integrated at a different level, after species have been integrated to species groups. The workshop participants considered that community-level assessments could be integrated into this integration framework as shown in Figure 4. This proposes that the community-level assessment is integrated through an additional step following the integration of species to species groups. This assumes that criterion 1.X is available for the species group. If it is only available at ecosystem component level, it would have to come in at ecosystem component level. The method used for such an integration would need careful consideration. Including the community-level criterion 1.X in this integration framework in this way may result in an overweighting of criterion 1.X (and related assessment uncertainty) compared to the other three criteria (Figure 4).

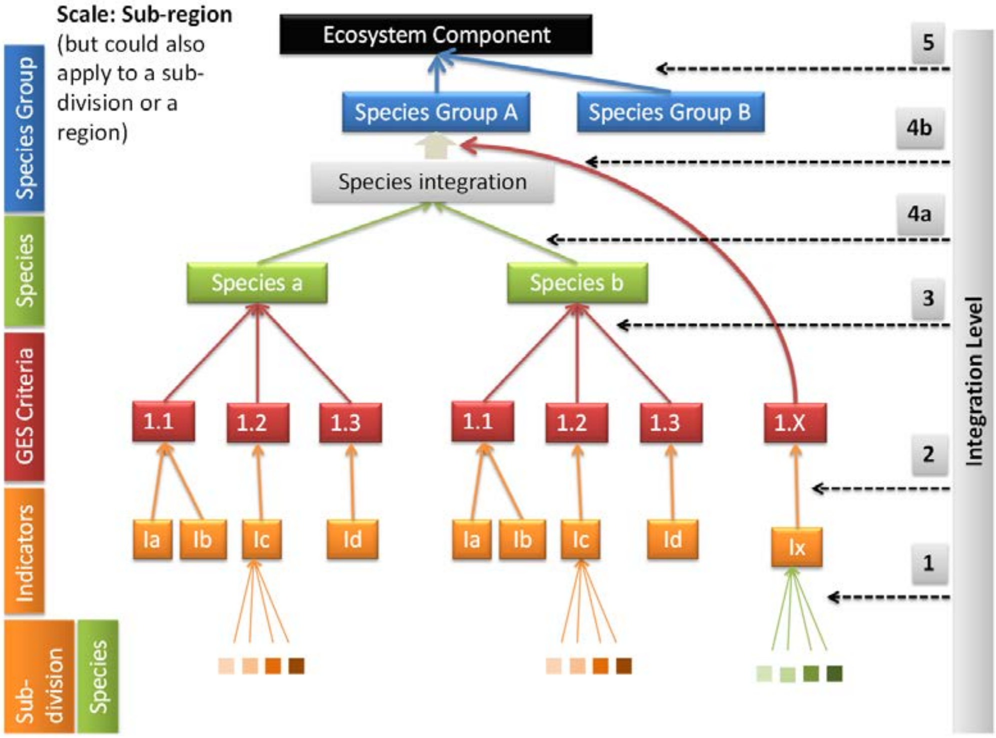

Figure 4: A schema showing integration within species including an integration of community criterion 1.X. Criterion 1.X refers to a community structure criterion.

3.3 Integration of species groups within criteria

3.3.1 Description

This framework is an alternative to that described in section 3.1. Figure 5 shows how assessments of each indicator are first conducted for species within species groups. The assessments of indicators for the same criterion are then integrated within in each species group. The assessments of each species group are subsequently integrated within criteria to gain an assessment for each criterion for the ecosystem component. These criteria assessments are integrated to provide an ecosystem component-level assessment.

3.3.2 Pros and cons

This framework was trialled to investigate an alternative approach. But, as Figure 5 shows, the framework is less intuitive and does not work well with species groups that are unbalanced in terms of numbers of species.

The only advantage of this framework is that it probably works better than the framework in Figure 2 for assessments involving few species that are data-poor. However, this can also be said for the framework in Figure 3 that has further benefits (see section 3.2). Hence this framework offers no additional benefits and is not considered an appropriate integration framework for biodiversity ecosystem, components under Descriptor 1.

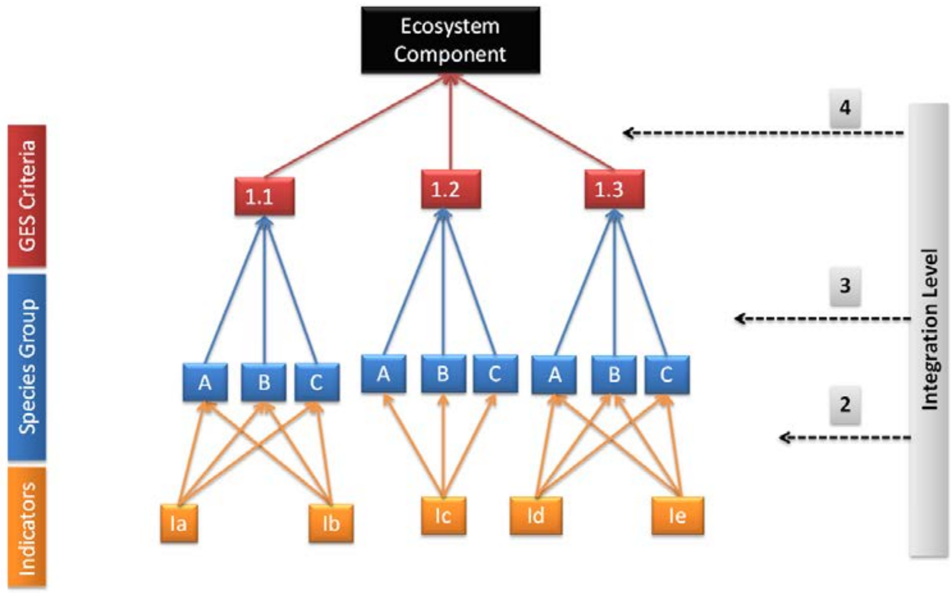

Figure 5: A schema showing integration of species groups within criteria. GES criteria are taken from the Commission Decision. This integration framework is similar to that shown in figure 2 but proposes an opposite approach to integrating across species groups and criteria.

3.4 Reconciling different assessment scales across indicators

3.4.1 Description

When several subdivisions are consistently assessed (i.e. when a large proportion of all indicators are specific to commonly defined subdivisions), an option is to integrate all indicators within each subdivision before integrating across subdivisions to subregion or region scale. Each indicator is assessed for the geographic area where it is valid; for example, a coastal fish indicator could be assessed for several adjacent coastal areas, and not assessed for offshore areas where it is not valid. A pelagic fish indicator could be valid for offshore areas, and an indicator describing a highly mobile species could be valid for the whole regional sea area. An example is shown in Figure. 6.

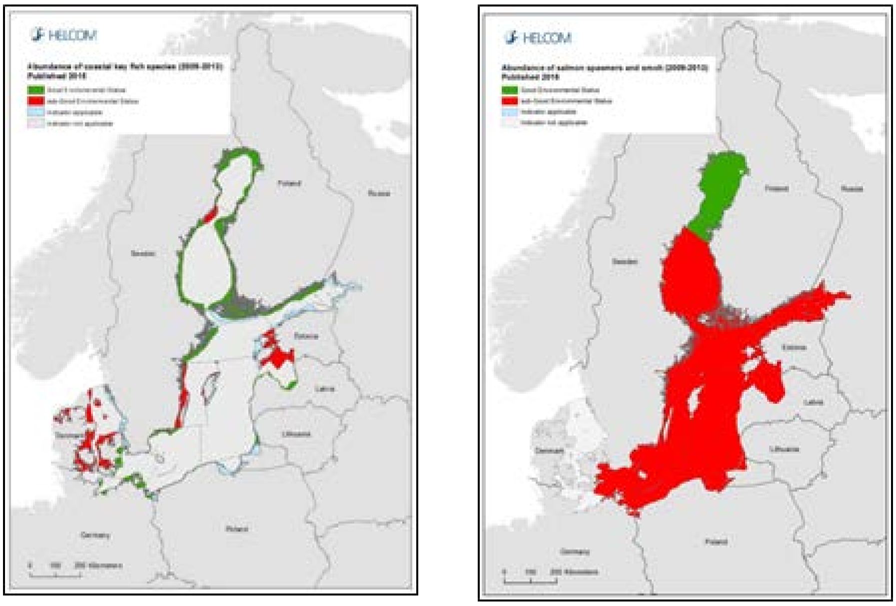

Figure 6. The indicators are defined as applicable or not applicable in specific assessment units, for example coastal fish indicators (left panel) are applicable and assessed in coastal units whereas in the open sea units the indicator is not applicable (striped), whereas migratory salmon (right panel) are applicable in the entire sub-basin assessment unit. Due to differing environmental conditions, different GES boundary values may be defined for different assessment units.

Different parts of the sea could be covered by a varying number of indicators, depending on how many are valid for each area and for which areas sufficient monitoring data are available to carry out an assessment. This type of integration is technically possible for indicators under criteria 1.1 to 1.3 as well as 1.X.

The integration of indicators within subdivisions can potentially follow either of the three other frameworks listed in this section provided that the indicators support the approach. In order to retain as much spatial resolution as possible, the integration can be done at the most detailed of the spatial scales of assessment units that are used in the system. If one indicator covers the entire subregion and only one value is produced for the entire area, this indicator could be applied in subdivisions or smaller assessment areas by assigning the same value of the indicator to each of the smaller assessment units, possibly with a lower weight for the smaller units if it is not considered to be of equal relevance in describing the particular smaller unit compared to indicators assigned as relevant directly to the smaller unit. This would allow integration to be carried out on the spatially most detailed scale of available assessment areas, incorporating the indicator that applies to the wider area, but retaining area-specific details on the other indicators that are assessed at the smaller spatial scale. This approach is used by HELCOM.

Finally, the assessments of the smaller areas can be combined if an overall assessment for the entire subregion or region is required. This final integration step can be done either using one-out-all-out, weighting by surface area, proportion of areas/surface area required to be in GES, or any other rule that is deemed relevant.

3.4.2 Pros and cons

This method has the advantage that subdivision-specific assessments are made which can then be used in management. The integration across areas is however complicated by the need to ensure that indicators which are not specific to one subdivision are not counted several times providing a situation where indicators applying to several subareas receive substantially greater weight. The process to integrate across areas therefore requires care in the way the indicators of different geographic coverage are combined, in order to give each area and indicator the correct weight in the final assessment.

Unless the OOAO method is used at all levels, this may lead to a situation where a good status in an indicator covering several subdivisions could mask poor status of indicators covering only one subdivision. It is possible to address this in the integration process but this may lead to apparent inconsistencies between the subdivision specific integrations and the subregional integration. On the other hand, applying the OOAO approach could make it statistically impossible to reach any defined GES if there are a large number of subdivisions.

4 Integration methods

The group considered various methods for each integration level in the two main integration frameworks. A full list of possible integration methods is provided in Borja *et al.* (2014). Of these, four were considered here in detail as more appropriate for the task i.e. the integration of species into species groups. This section provides a description of each method, its pros and cons, and examples of when it is appropriate to apply.

4.1 One out, all out (OOAO)

One out, all out (OOAO) signifies that all indicators must be in good status for the integrated indicator for the subsequent level to be in a good status (Borja *et al*. 2014, Probst and Lynam 2016). Conversely, if a single indicator is in poor status, the integrated group at the next level is given poor status.

4.1.1 Pros and cons

OOAO has the advantage of being simple to apply, explain and understand. Any poor status rating is visible at higher levels and can be directly traced back to one or more poor status indicators. In this way, there is no 'masking' of poor status indicators by other indicators. The OOAO approach is sensitive to changes in the status of a single indicator and therefore allows a transparent tracking of a deterioration in status. This clarity is obtained at the cost of losing information on progress towards GES (distance to target), and a quantitative value for the indicator, which limits the methods that can be used at subsequent integration levels (e.g. averaging or weighted averaging will not be possible). The presentation of the OOAO assessment result can be complemented by information on the number / percentage of assessed species meeting and failing their indicator values and thereby provide an expression of distance to target and progress towards GES. Further, the method is particularly sensitive to poor status at integrated levels due to random variation in underlying estimated indicators. This risk increases as the number of indicators entering the OOAO integration increases. This is particularly undesirable when an indicator is estimated to just miss the threshold level but the confidence interval of the estimate includes the reference level. There are several sources of uncertainty which can lead to such results arising e.g. measurement error, unclear climate effects and uncertainty in target level and pressure-state response. Thus the OOAO method could easily lead to an overall underestimation of true environmental status. OOAO assigns equal weight to all indicators/criteria, as any one indicator in not good status will result in the higher-level result being not good status. Therefore it should not be applied if indicators/criteria are used which are not well developed or not considered equally important than others.

4.1.2 When is this method considered appropriate?

OOAO is ideally suited to the integration of species groups to ecosystem component level, of criteria to species group level, and of criteria to species or ecosystem component level. It assumes that all indicators to be integrated are equally important elements of GES such that all are required to be in a good state for the component to be in GES, and there is a similar level of certainty in the value of the indicator and in the setting of the threshold value. It is a possible method to integrate species level indicators to species group in the case where all species are considered to be monitored with high precision and it is considered that all species are essential to achieving a good status. The method is not considered appropriate where two indicators may compensate to some degree for each other, such as when indicators of different subdivisions are integrated. The method is not appropriate where there is a large number of indicators to be integrated, and there is a reasonable probability that at any one time, some will be in not good status by chance (natural variability), or due to measurement error and uncertainties in measuring the indicator value and in setting the threshold level. The method is not well suited to retain information on distance to target. It is therefore not appropriate for lower levels of integration, particular where there are a large number of indicators to be integrated.

4.2 Percentage of indicators within limits (proportion)

This is a conditional rule (Borja *et al*. 2014) that specifies a percentage or proportion of assessments (e.g. at species or indicator level) must be in good status for the integrated assessment (i.e. at the subsequent level) to be considered to be in good status. For example, 75% of indicators in good status, or 15 out of 22 species in good status. The percentage or proportion required can be set by expert judgement (Breen *et al*. 2012) or by probabilistic methods (see section 4.2). It is also referred to in this report as a 'proportional' method.

4.2.1 Pros and cons

The use of a percentage as a conditional rule is straightforward and is a simple concept to communicate. Like all conditional rules, it allows expert judgement to be used to combine assessments in a transparent way. The threshold used to define the percentage of assessment results required to be in good status can be more open to challenge if set by expert judgement rather than by more objective probabilistic methods.

4.2.2 When is this method considered appropriate?

This method is appropriate when a larger number of indicator are integrated into one assessment and when it is unlikely for statistical or biological reasons that all the contributing assessments will be in good status simultaneously. The reason for not expecting all assessments to be in good status simultaneously is usually a reflection of the uncertainty around the measurement and assessment of certain indicators and our understanding of what constitutes 'good status'.

For example, the OSPAR EcoQO on seabird population trends used a percentage threshold to assess the status of integrated species-specific indicators: "*Changes in breeding seabird abundance should be within target levels for 75% of species monitored in any of the OSPAR regions or their sub-divisions.*" (ICES 2008). This approach is also proposed for the OSPAR Common Indicator on Marine Bird Abundance (see example in section 5.1.5.2). The approach could be considered appropriate in the context of MSFD if considerable uncertainty exits around the any of the following:

1) uncertainty in the estimation of the indicator metric;
2) uncertainty on what threshold values to use to define an indicator being in 'good' or 'poor' status;
3) uncertainty around the pressure-state relationship, for example whether the indicator is responding only to anthropogenic factors, or the extent to which other drivers such as climate change are impacting on the indicator;
4) uncertainty around the role of other factors as the ecosystem recovers from decades of human pressures. As pressures reduce and the ecosystem recovers, not all species will necessarily improve in status because other 'natural'

factors may affect them negatively e.g. increased predation or competition for resources from other recovering species.

4.3 Probabilistic methods

Probabilistic methods use information of the uncertainty of the indicator status assessment along with the actual assessment result to evaluate the uncertainty related to the final integrated result. The probabilistic method takes account of the status and related uncertainty of all the relevant indicators, to estimate the probability that the system is in good (vs. poor) status. Probabilistic approaches can also be used to evaluate different management measures such as for eutrophication (Barton *et al.*, 2008; Lehikoinen *et al.*, 2013a) and oil spill severity (Lehikoinen *et al.*, 2013b) and to derive the probability of any individual species-specific metric meeting its metric-level target (Greenstreet *et al.* 2012).

This method can be combined with other integration methods, such as weighted averaging, or in the setting of conditional rules. Probabilistic integration can be used to combine indicator assessments that each have an estimate of the probability that their respective target had been met. The resulting integrated assessment outcome would have a known level of uncertainty over whether or not GES had been achieved. Managers would then have to decide how much uncertainty they are willing to tolerate.

4.3.1 Pros and cons

Probabilistic methods allow the explicit inclusion of information about the uncertainty of the indicator values, as well as target levels, while keeping to objective computation formulas (no ad hoc weighing factors needed). They allow the inclusion of indicators with various accuracies, without the possibility that less uncertain indicators will skew the final result. Consequently, the inclusion of more indicators is always an improvement, provided that their uncertainties have been accurately assessed. The understanding of the methods requires specialist modelling skills. Communicating the results and their derivation to non-specialists can be challenging.

4.3.2 When is this method considered appropriate?

This method is typically used when a larger number of assessment results of varying uncertainty are integrated into one assessment. It is appropriate to use when it is unlikely for statistical or biological reasons that all the contributing assessments will be in good status simultaneously. The reason for not expecting all assessments to be in good status simultaneously is usually a reflection of the uncertainty around the measurement and assessment of certain indicators and our understanding of what constitutes 'good status'.

The approach could be considered appropriate in the context of MSFD if there is:

1) variable uncertainty in the estimation of the different indicator metrics;

2) uncertainty on what threshold values to use to define an indicator being in 'good' or 'poor' status;

3) uncertainty around the pressure-state relationship, so that it is not certain that an indicator is responding totally to anthropogenic factors, or the extent to which other drivers such as climate change are impacting on the indicator;

4) uncertainty around the role of other factors as the ecosystem recovers from decades of human pressures. As pressure reduces and the ecosystem recovers, not all species will necessarily improve in status because other 'natural' factors will affect them negatively e.g. increased predation or competition for resources from other recovering species.

4.4 Averages

Averages include methods where the average of indicator values across indicators is used as an integrated indicator. Varieties of this include versions where the indicator is normalized prior to averaging (e.g. to be in the range of 0–1, to be expressed relative to the reference level or to be ordered in categories such as very poor, poor, medium, good and very good). The underlying rationale is that a decrease in one indicator can be compensated by an increase in another indicator.

4.4.1 Pros and cons

A strong advantage is that the methods provide a direct measure of distance to target. This means that information on whether the target is achieved or not as well as information on whether the target is almost achieved, marginally achieved or whether there is still a long way to go, is retained.

When averaging, it is inherently assumed that one or more indicators inside the desired values can compensate for indicators that are outside the desired values. This is both the strength and the weakness of the method. It allows the method to smooth out random variation between indicator levels but in doing so, a single poor status indicator may be masked by a number of good status indicators.

An issue requiring attention when applying averaging is the comparability of the scales used in the indicators that are to be integrated. If the scales are similar the values can be directly averaged, however if the scales differ significantly normalization might be needed whereby different indicator units calculated at different scales are made comparable to each other.

Averaging indicators can be done based on ratios of the indicator measurement value and its target. An issue of the method using the value relative to the reference level is that the range of values taken in each indicator can vary so that an unintended weighting occurs. For example, one indicator may vary between 0.3 and 3 times the reference level whereas another may only vary between 0.8 and 1.4 times the reference level. An incidental low value of the latter indicator will be masked by the variation in the former indicator unless a correction or weighting is made. Averaging indicators can be done for normalized indicator values, which ensures that situations are avoided where indicator scale determines the weight of different indicators in the integration. The method requires the boundary between good and poor indicator status to be defined. Additionally, minimum and maximum values have to be known in order to be able to carry out the normalization method where all indicators take values between 0 and 1.

4.4.2 When is this method considered appropriate?

This method is considered appropriate when good status of one or more indicators can compensate for poor status of other indicators, and when indicators are considered to be measured with observation error, in which case the method decreases the risk of acting on false poor or good status indicators. It is not appropriate when the single indicators are considered to be known without error and the individual indicators do not have the ability to compensate for each other.

4.5 Weighted averages

Weighted averages include methods where the average indicator value is calculated by applying weights to the various indicators in the integration. Weighting may be based on their perceived importance, the area covered by the indicator or the precision and accuracy of the indicator. The underlying rationale is that a decrease in one indicator can be compensated by a specified (but not equal) increase in another indicator.

4.5.1 Pros and cons

The method allows for inclusion of indicators that are correlated with each other but it is still considered appropriate to include them all, or for indicators which are not considered equally reliable or important. Indicators can be weighted to avoid their covariance affecting the weight of individual indicator ensembles in the final assessment result. Further, indicators that are considered crucial to the ecosystem functioning or maintenance of biodiversity can be given a higher weight than other indicators, so that they have a stronger effect on the overall assessment than other indicators. Alternatively, indicators with low confidence (due to few or low-quality data, uncertainty in the GES threshold etc., could be down-weighted to reduce their contribution to the overall assessment result. An advantage of this method is that uncertainties in the assessment are taken into account, though in an ad-hoc manner.

Weighting by precision/certainty has the advantage that the resulting analysis may be more robust as the species/areas with a larger amount of, or more certain, data receive a higher weight in the analysis. However, this method may lead to the undesirable outcome that sensitive or vulnerable species for which less information exists, become less important for the integrated assessment, although these species may be the ones of most concern. In addition, quantity of data are not an automatic guarantee of data quality, as abundant data may be derived from a non-representative area, season, etc. Therefore, automatic computation of weights based on data quantity is not recommended.

Weighting by perceived importance is useful if one wants to take vulnerability or sensitivity of species into account. Species with a critical population size or species status can contribute more to the assessment than the more resilient species by applying weighting. However, the actual weights may need to be determined by expert judgement and thus have a degree of subjectivity.

Weighting by area assumes that the larger the area, the higher is its effect on the outcome of the assessment. However, the importance of an area may not be a function of area size. A large area could contain less species/habitats than a smaller area and may therefore be poorer in terms of biodiversity. Furthermore, a larger area could contain species that are widely distributed while a smaller area could contain very rare and narrowly distributed species, which could be an argument of assigning that smaller area a higher weight.

Choosing the weights for indicators almost inevitably includes a subjective element, and very clear and transparent guidelines are needed in order to apply this method. Further, it has to be noted that weighting only has an effect relative to other indicators in the same average; i.e. if all indicators are considered to have low confidence, weighing them all by the same low value gives the same result as not weighing them in the first place.

4.5.2 When is this method considered appropriate?

The method is appropriate if two or more indicators are correlated with each other due to common drivers or processes, but all of them are however wanted to be included. In that case, down-weighting these might be appropriate. It may also be appropriate when combining indicators which are representative of very different geographical scales; in that case, the indicators could be weighed by their area coverage. Further, the method can be used to reflect that some indicators are considered more important than others (e.g. critically endangered species, key species, icon species etc.).

5 Suggested framework for integration of indicators to determine GES at component level

5.1 Integration within criteria

5.1.1 Integration from species-specific indicators to criterion

The integration from indicator to criterion can use either proportional, probabilistic or averaging (including weighted averaging) methods. All of these methods incorporate the underlying logic that a good status in one indicator can compensate for a lightly poor situation in another indicator. It is necessary to ensure that the indicators were at comparable numerical scales before averaging by e.g. standardizing to values between 0 and 1 or by using other methods with the same aim.

Indicators which take the values 0 and 1 only (e.g. poor or good status) can be used in proportional methods, with the threshold set either by expert judgement or probabilistic methods. In both cases, it is expected that the proportion of allowed indicators at poor status is effectively 0 when the number of indicators is very low (and therefore is equivalent to OOAO) but increases as the number of indicators becomes large. This pattern is seen in indicators such as the sensitive fish indicator and the North Sea seabird example, where the indicator is considered to be in good status if a specified number or proportion of species are in good status in a given year.

Proportional, probabilistic and averaging methods have the common challenge that one or a number of good status indicators may act to hide a poor status indicator. To avoid this, a conditional two-step approach can be used. Under such an approach, unacceptable deviations from the good status would be identified for all indicators and such deviations would automatically elicit a poor status of any higher-level integration.

5.1.2 Integration from criterion to species group

The group agreed that a species group cannot be considered to be at a good status if one or more of the assessed criteria are considered to be in poor status. As a result of this, it is recommended that a one out, all out integration method is used. This integration will mean that all assessed criteria have to be in good status for the species group to be in a good status.

5.1.3 Integration from species group to ecosystem component

The group agreed that an ecosystem component cannot be considered to be at a good status if one or more of the assessed species groups are considered to be in poor status. As a result of this, it is recommended that a one out, all out integration is used. This integration means that all assessed species groups have to be in good status for the ecosystem component to be in a good status.

5.1.4 Worked examples

Below are specific regional worked examples of the application of the approaches favoured by the group (Figures 7–13). The examples use regional sea indicators from OSPAR, HELCOM, the Mediterranean and Black Sea. The integration frameworks and integration methods included in each example below are suggestions made by this group and have not been agreed within any of the Regional Sea Conventions.

5.1.4.1 OSPAR – fish

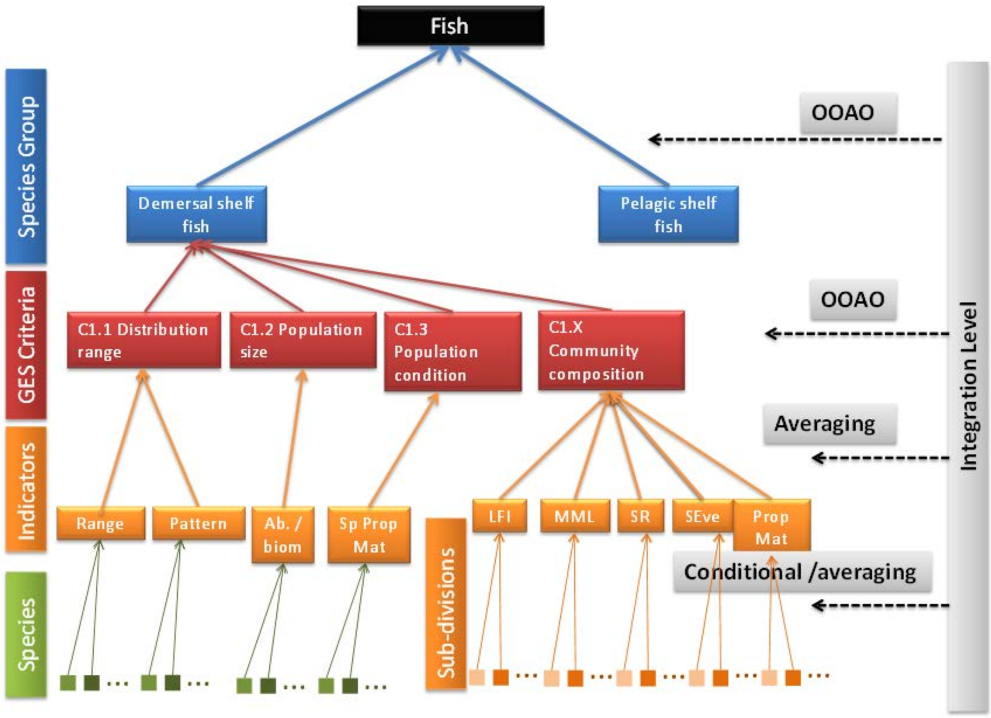

Figure 7: Suggested integration methods at different levels for fish in the OSPAR area if using the integration within criteria framework. Criterion 1.X refers to a community structure criterion. See section 5 above for more detailed descriptions of the different methods.

5.1.4.2 OSPAR –birds

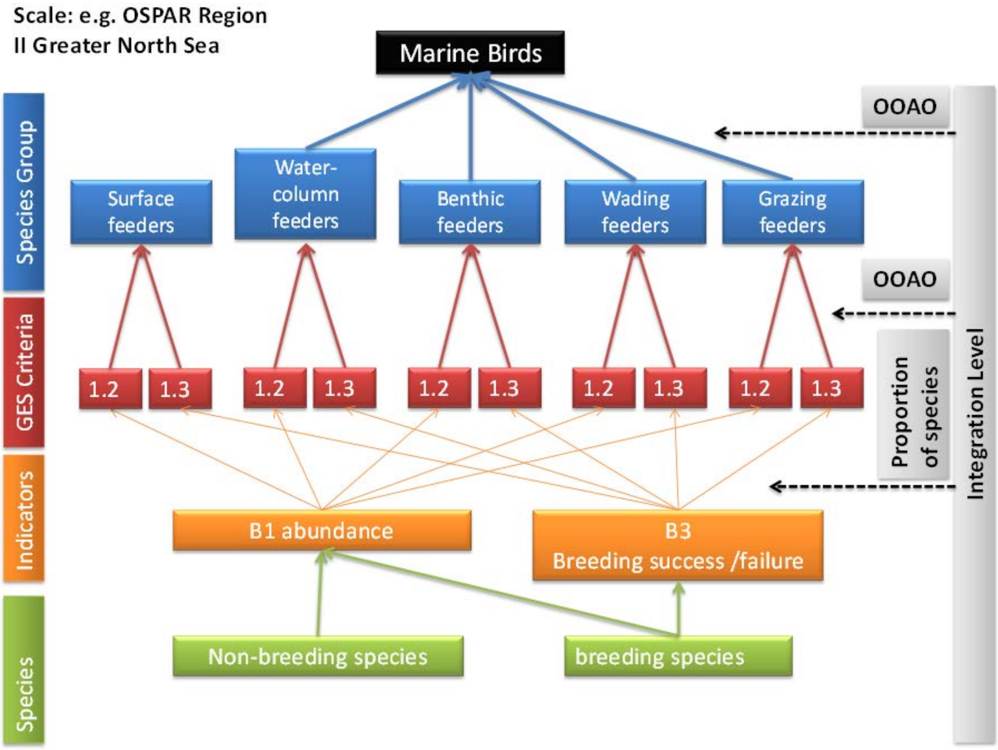

Figure 8: Suggested integration methods at different levels for birds in the OSPAR area if using the integration within criteria framework. See section 5 above for more detailed descriptions of the different methods.

5.1.4.3 HELCOM – fish

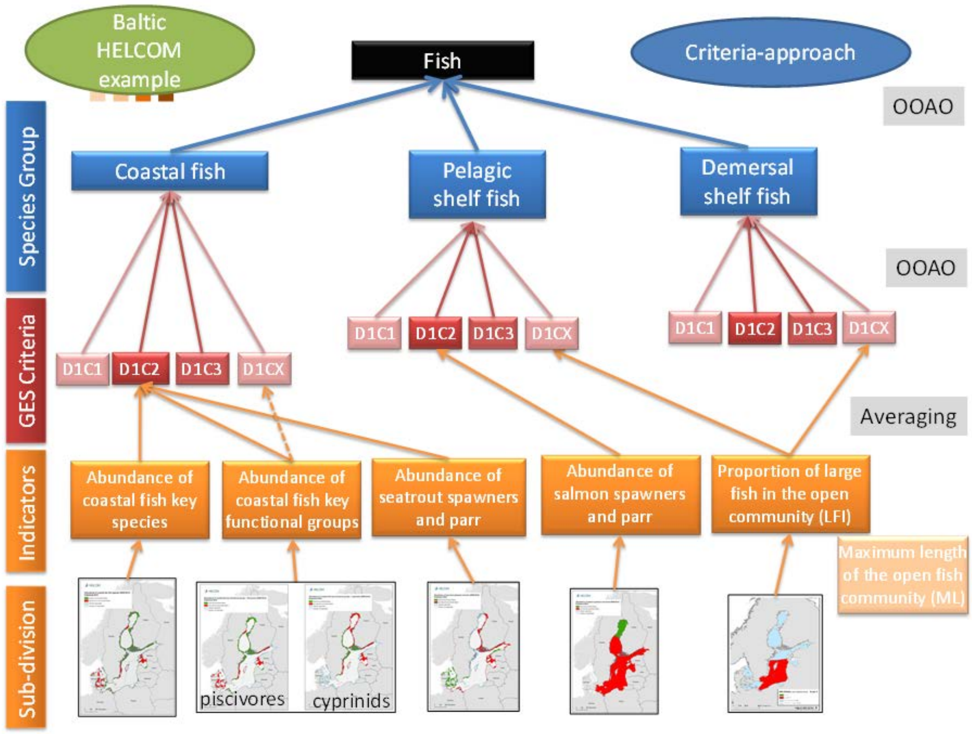

Figure 9: Suggested integration methods at different levels for fish in the HELCOM area if using the integration within criteria framework. Criterion 1.X refers to a community structure. See section 5 above for more detailed descriptions of the different methods.

5.1.4.4 HELCOM - birds

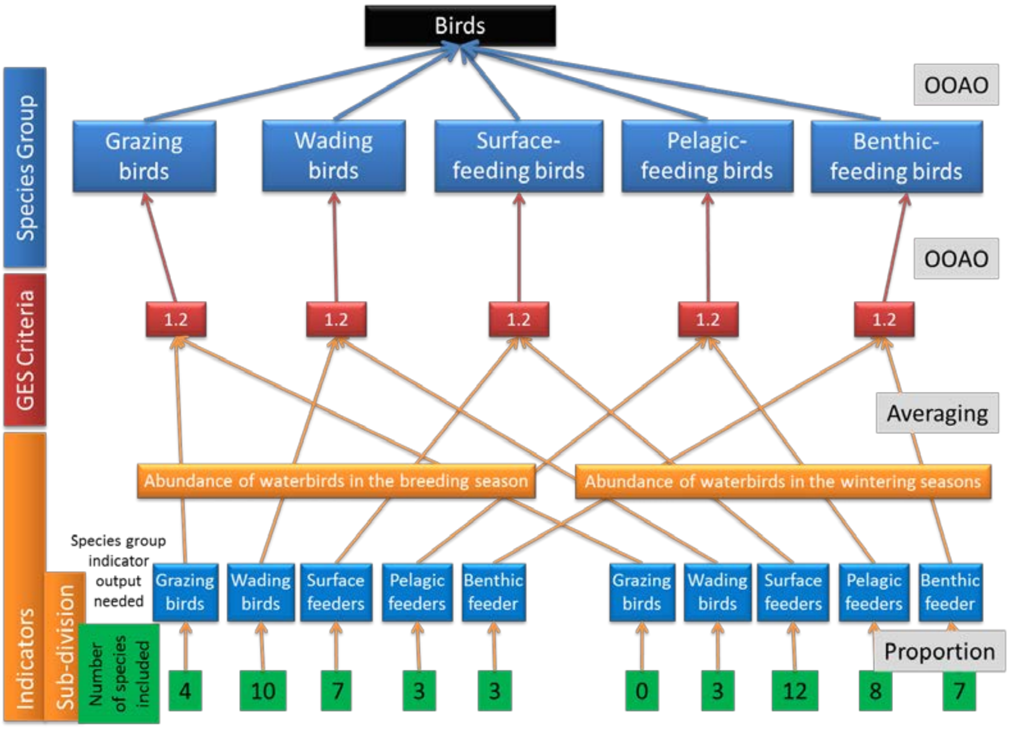

Figure 10: Suggested integration methods at different levels for birds in the HELCOM area if using the integration within criteria framework. See section 5 above for more detailed descriptions of the different methods.

5.1.4.5 HELCOM – Mammals

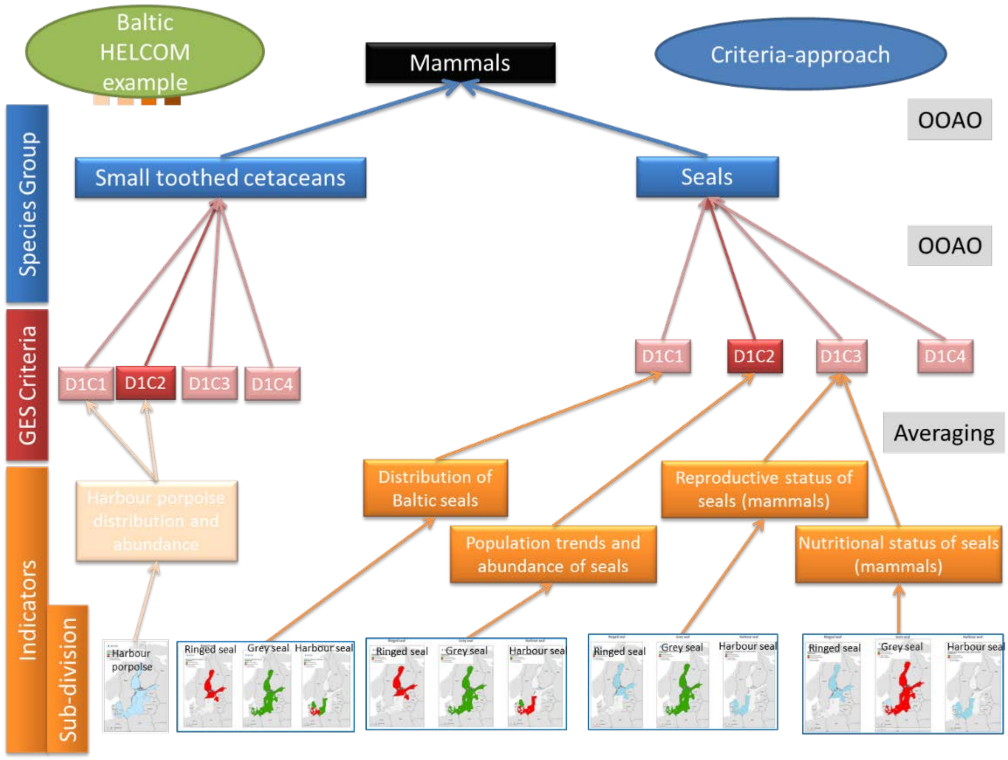

Figure 11: Suggested integration methods at different levels for mammals in the HELCOM area if using the integration within criteria framework. Light colour indicates that the indicator is not adopted or that data for the species is not available. See section 5 above for more detailed descriptions of the different methods.

5.1.4.6 Black Sea - fish

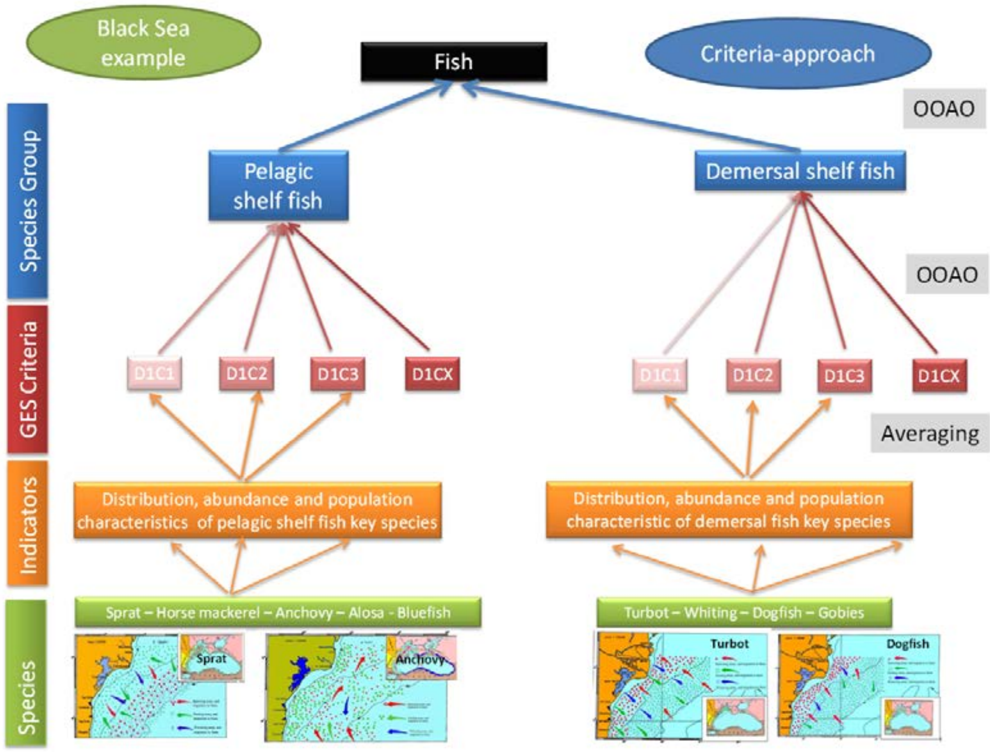

Figure 12: Suggested integration methods at different levels for fish in the Black Sea if using the integration within criteria framework. Criterion 1.X refers to a community structure criterion. See section 5 above for more detailed descriptions of the different methods.

5.1.4.7 Mediterranean Sea – mammals

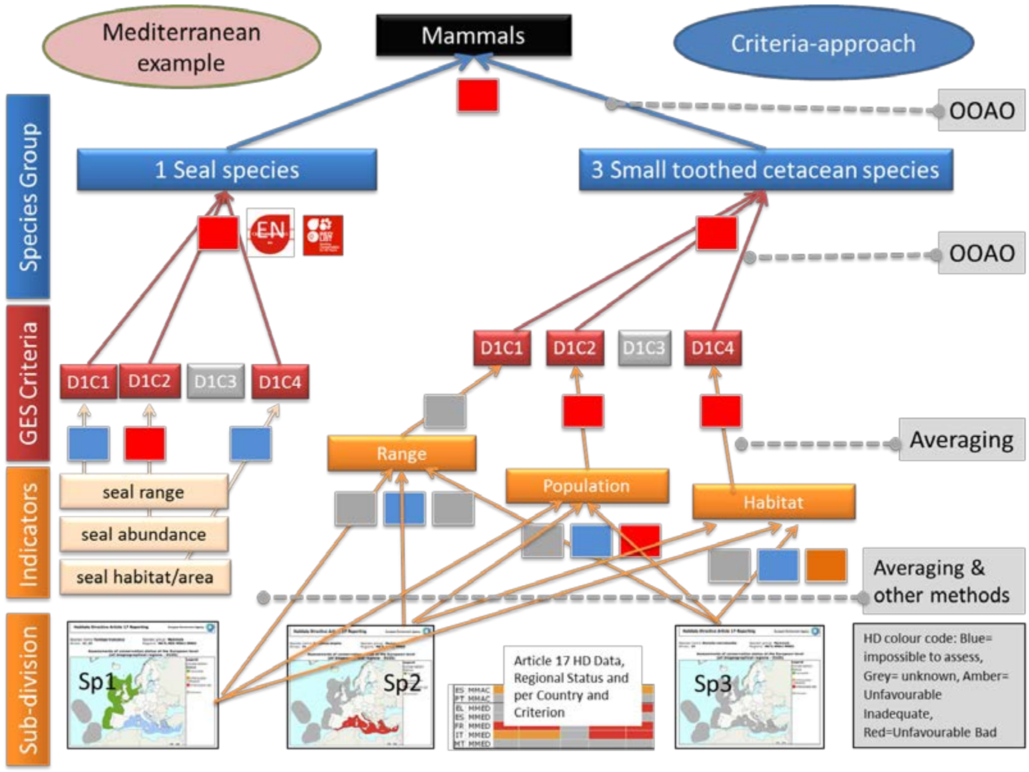

Figure 13: Suggested integration methods at different levels for mammals in the Mediterranean Sea if using the integration within criteria framework. See section 5 above for more detailed descriptions of the different methods. There are three assessed small toothed cetaceans and one seal species in the area.

5.2 Integration within species

5.2.1 Integration from criterion to species

The group agreed that a species cannot be considered to be at a good status if one or more of the assessed criteria are considered to be in poor status. As a result of this, it is recommended that a one out, all out integration is used. This integration will mean that all assessed criteria have to be in good status for the species to be in a good status.

5.2.2 Integration from species to species group

The integration from species to species group can be either OOAO, proportional, probabilistic or through averaging across indicators. All of these methods except OOAO incorporate the underlying logic that a good status in one indicator can compensate for a lightly poor situation in another indicator or that indicators are inherently measured with observation error and hence may provide false poor or good status ratings. An example could be where a large number of species are measured. The status of each of these species could then be assessed and a proportional integration method, with the threshold value set through expert judgement or probabilistic methods. It is expected that the proportion of allowed indicators at poor status is effectively 0 (OOAO) when the number of indicators is very low but increases as the number of indicators becomes large. This pattern is similar to those seen in indicators such as the sensitive fish indicator and the OSPAR seabird indicator where a specified proportion of the indicator species are allowed to be in poor status in a given year.

The proportional method has the challenge that it may allow a very poor status indicator to be integrated into a higher-level 'good' status assessment. To avoid this case, a conditional two-step approach can be used. Under such an approach, unacceptable deviations from the good status would be identified for all indicators and such deviations would automatically elicit a poor status of any higher-level integration. An example could be an endemic and potentially endangered or threatened species.

5.2.3 Integration from species group to ecosystem component

The group agreed that an ecosystem component cannot be considered to be at a good status if one or more of the assessed species groups are considered to be in poor status. As a result of this, it is recommended that a one out, all out integration is used. This integration will mean that all assessed species groups have to be in good status for the ecosystem component to be in a good status.

5.2.4 Worked examples

Below are shown specific regional worked examples of the application of the approach (from OSPAR, HELCOM, and the Mediterranean Sea, Figure 14–16).

5.2.4.1 OSPAR – mammals

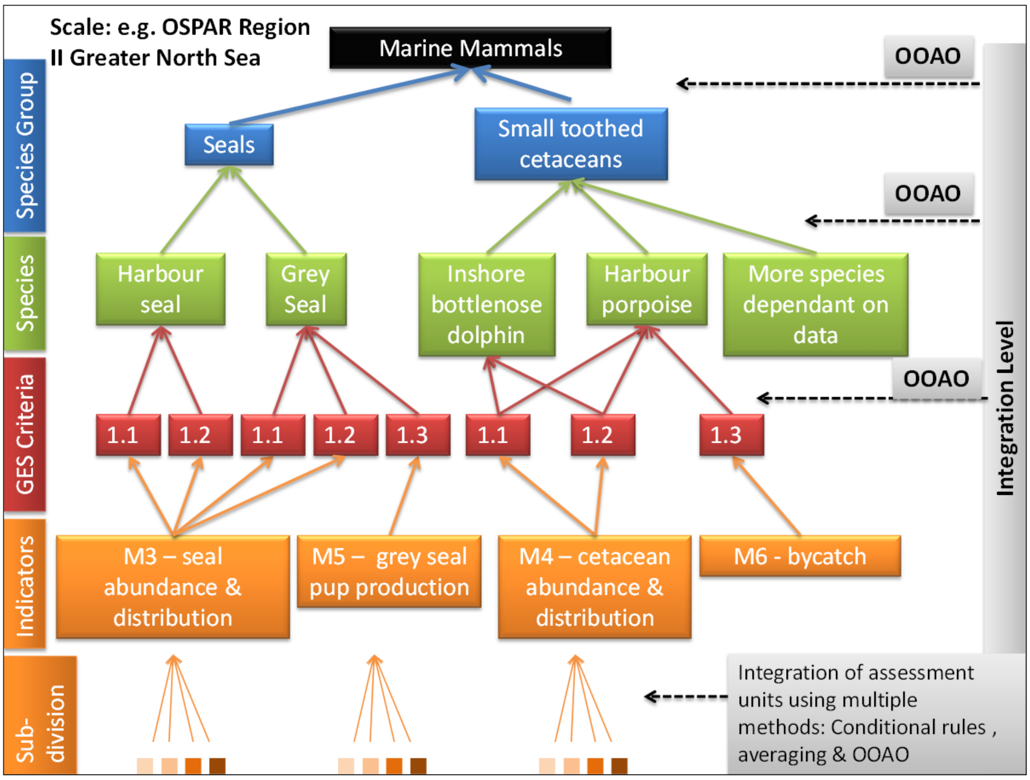

Figure 14: Suggested integration methods at different levels for mammals in the OSPAR area if choosing the integration within species framework. See section 5 above for more detailed descriptions of the different methods. As there are few species in each species group, OOAO is used in the integration from species to species group.

5.2.4.2 Mediterranean – mammals

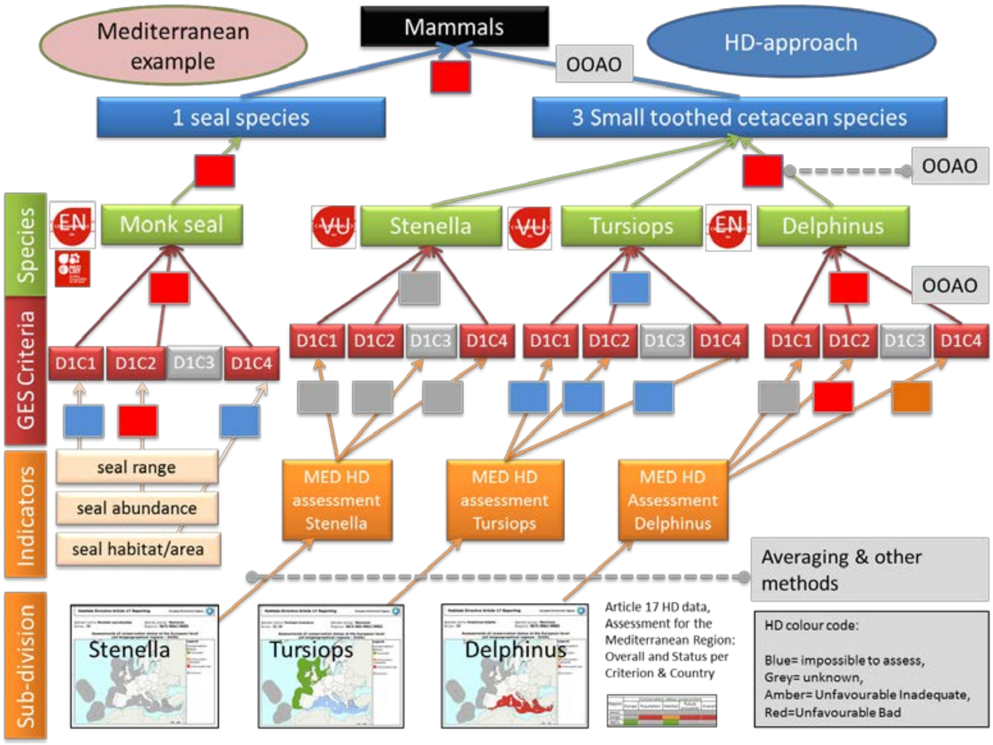

Figure 15: Suggested integration methods at different levels for mammals in the Mediterranean Sea if choosing the integration within species framework. As there is only 1 spp in the 'seals' species group, no integration method is needed, and the outcome for the monk seal is effectively the outcome for the species group. As there are few indicators for each species and few species in each species group, OOAO is used in the integration. See section 5 above for more detailed descriptions of the different methods.

5.2.4.3 HELCOM mammals

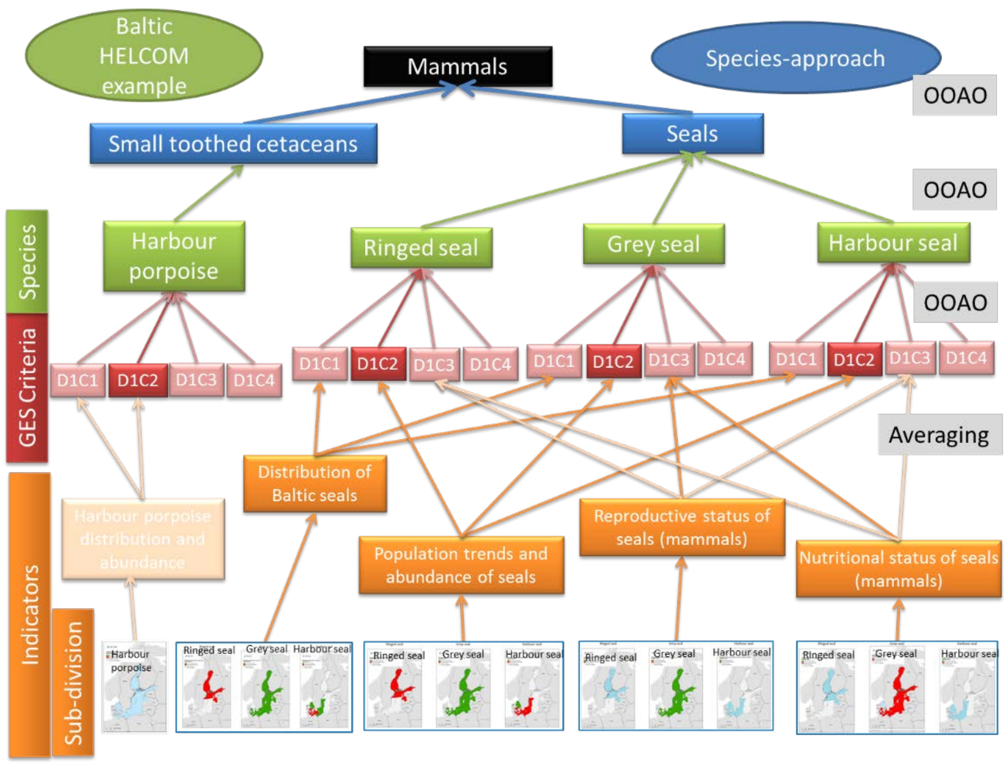

Figure 16: Suggested integration of indicators at different levels for mammals in the HELCOM area if choosing the species-approach. Criterion 1.X refers to a community structure criterion. As there is only one species in the 'small toothed cetaceans' group, the outcome for harbour porpoise is effectively the outcome for the species group. Light colour signifies that the indicator is currently not adopted or that data for the species is not available. See section 5 above for more detailed descriptions of the different methods.

5.3 Integration by area

Subdivisions can be addressed in two different ways depending on whether the aim is to achieve a subdivision or subregional evaluation of GES. The two methods are conceptually similar and are both treated in this section. For the subregional integration, all levels are included but for subdivision integration, the integration across subdivisions is omitted and all subdivisions assessed separately.

5.3.1 Integration from species or subdivision to subregional indicator

The integration from species or subdivision to subregional indicator is recommended to be either proportional, probabilistic or through averaging over species/subdivisions. All of these methods incorporate the underlying logic that a good status in one species/subdivision can compensate for a lightly poor situation in another species/subdivision. An example could be where multiple colonies of a species are surveyed. Indicators which take the values 0 and 1 only (e.g. poor or good status) can be used in proportional and probabilistic methods. In both cases, it is expected that the proportion of allowed species/subdivisions at poor status is effectively 0 when the number of species/subdivisions is very low but increases as the number of species/subdivisions becomes large.

Proportional, probabilistic and averaging methods have the common challenge that one very good status species/subdivision status may act to hide a very poor species/subdivision status. To avoid this case, a conditional two step approach can be used instead. Under such an approach, unacceptable deviations from the good status would be identified for all species/subdivisions and such deviations would automatically elicit a poor status of any higher level integration. Such an approach would also allow a process where certain species/subdivisions were considered of primary importance and hence could not be compensated for by other species/subdivisions at good status. An example could be an endemic and potentially endangered or threatened species.

6 Conclusions

The workshop participants considered that there were two feasible integration frameworks for integrating indicators to ecosystem component level: the integration within criteria and integration within species. Within each of these options is the question of how to address indicators that are assessed at different spatial scales, and how to aggregate across assessment areas or subdivisions to produce subregional or regional assessments.

Integration within species carries the advantage of increasing comparability with the Habitats Directive at the species level. Community considerations are, however, not easily integrated and when different species have information for different numbers of criteria, this is masked in the higher-level integration. This method is exemplified by marine mammal examples in this report as these are covered by the Habitats Directive and the number of species is often low. However, the group did not consider that mammals were conceptually different from other ecosystem components other than in this respect.

Integration within criteria carries the advantage of a transparent weighting of all criteria, including community aspects when carrying out the next level of integration to ecosystem component level. Data for species where only one criterion can be measured can be included without loss of consistency and all criteria receive equal weight in the integration across criteria. Community criteria have been suggested for fish and birds and hence this method is exemplified here for fish, birds and mammals in all regions. This framework is more appropriate where there is a larger number of species, due to the increasing chance of some species missing information on one or more criteria with an increasing number of species.

Subdivisions can be addressed in different ways depending on whether the aim is to achieve a subdivision or subregional evaluation of GES. The area integration can be performed for both integration frameworks. Down-scaling of indicator assessments from larger spatial areas to smaller areas can be carried out to allow the integration of indicators with differing spatial scales, prior to combining areas to a regional or subregional assessment.

The recommended integration method differs between integration levels under both of the two frameworks. At the level of integrating criteria (whether to species or ecosystem component), the recommended method is OOAO, and the same recommendation is made when integrating species groups to ecosystem component for both frameworks. At other levels, a mixture of averages, weighted averages, proportional and probabilistic methods are recommended depending on the specific situation.

The final choice of integration framework depends on the relative consideration of the wish to facilitate comparability with the Habitats Directive (integration within species), the wish to use similar methods across all ecosystem components, the wish to give a transparent weight to all criteria (integration within criteria)vs.giving transparent weight to all species (integration within species) and finally, the wish to incorporate community-level considerations in the integration in a simple way. The workshop participants considered that the weighing of these different wishes was a policy decision rather than a scientific decision. Hence, this report lists the pros and cons for both frameworks along with recommended integration methods for different integration levels under each framework, to facilitate informed decision-making.

7 References

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

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

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

Breen, P., Robinson, L. A., Rogers, S. I., Knights, A. M., Piet, G., Churilova, T., Margonski , P., Papadopoulou N. , Akoglu E. , Eriksson A., Finenko Z., Fleming-Lehtinen V., Galil B., Goodsir F., Goren M., Kryvenko O., Leppanen J.M., Markantonatou V., Moncheva S., Oguz T., Paltriguera L., Stefanova K., Timofte F., F. Thomsen, F., (2012). An environmental assessment of risk in achieving good environmental status to support regional prioritization of management in Europe. Mar. Policy 36, 1033–1043. doi: 10.1016/j.marpol.2012.02.003

Greenstreet, S. P. R., Rossberg, A. G., Fox, C. J., Le Quesne, W. J. F., Blasdale, T., Boulcott, P., Mitchell, I., Millar, C., and Moffat, C. F. 2012. Demersal fish biodiversity: species-level indicators and trends-based targets for the Marine Strategy Framework Directive. ICES Journal of Marine Science, 69: 1789–1801.

ICES. 2008. Report of the Workshop on Seabird Ecological Quality Indicator, 8-9 March 2008, Lisbon, Portugal. ICES CM 2008/LRC:06. 60 pp.

Jennings, S., and Dulvy, N. K. 2005. Reference points and reference directions for size-based indicators of community structure. ICES Journal of Marine Science, 62: 397–404.

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

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

Link, J. S. 2005. Translating ecosystem indicators into decision criteria. ICES Journal of Marine Science, 62: 569–576.

Palialexis A., Tornero A. V., Barbone E., Gonzalez D., Hanke G., Cardoso A. C., Hoepffner N., Katsanevakis S., Somma F., Zampoukas N., 2014. In-Depth Assessment of the EU Member States' Submissions for the Marine Strategy Framework Directive under articles 8, 9 and 10. EUR – Scientific and Technical Research series. Luxembourg: Publications Office of the European Union. EUR 26473 EN, 149 pp. doi: 10.2788/64014.

Palialexis A., A. C. Cardoso, L. Avellan, J. Batsleer, B. Campos, D. Connor, S. Greenstreet, L. Guérin, N. Häubner, J. Hawkridge, P. Heslenfeld, K. Coull, J. Krause, B. Mavri, L. Miossec, M. Orlando-Bonaca, , J. Reker, C. Romao, H. Teixeira and K. Tsiamis; 2015. Report of the JRC's Descriptor 1 workshop to support the review of the Commission Decision 2010/477/EU concerning MSFD criteria for assessing Good Environmental Status; EUR 27715; doi:10.2788/444834.

Probst WN, Lynam CP (2016) Integrated assessment results depend on aggregation method and framework structure – A case study within the European Marine Strategy Framework Directive. Ecological Indicators 61: 871-881 http://dx.doi.org/10.1016/j.ecolind.2015.10.040

Shin, Y-J., Rochet, M-J., Jennings, S., Field, J. G., and Gislason, H. 2005. Using size-based indicators to evaluate the ecosystem effects of fishing. ICES Journal of Marine Science,62: 348–396.

Annex 1: List of participants

Name	Address	Phone/Fax	E-mail
Andrea Rau	Thünen Institute Institute of Baltic Sea Fisheries Alter Hafen Süd 2 18069 Rostock Germany	+49 3818116 138	andrea.rau@thuenen.de
Andreas Palialexis	DG Joint Research Centre Via Enrico Fermi 2749 21027 Ispra (VA) Italy		andreas.palialexis@jrc.ec.europa.eu
Anik Brindamour	Ifremer Nantes Centre Rue de l'île d'Yeu 44311 Nantes Cédex 03 France	+45 35 88 33 78	Anik.Brindamour@ifremer.fr
Anna Rindorf (Chair)	DTU Aqua - National Institute of Aquatic Resources Jægersborg Allé 1 2920 Charlottenlund Denmark		ar@aqua.dtu.dk
Bruna Campos	BirdLife Europe and Central Asia Avenue de la Toison d'Or 67 Brussels 1060 Belgium		Bruna.Campos@birdlife.org
Christopher Lynam	Cefas Lowestoft Laboratory Pakefield Road Lowestoft Suffolk NR33 0HT UK		chris.lynam@cefas.co.uk
Ciarán Murray	Aarhus University Nordre Ringgade 1 8000 Aarhus C Denmark		cim@bios.au.dk

Name	Address	Phone/Fax	E-mail
Håkan Wennhage	Swedish University of Agricultural Sciences Department of Aquatic Resources Institute of Marine Research 453 30 Lysekil Sweden	+46 761 33 4455	hakan.wennhage@slu.se
Ian Mitchell	Joint Nature Conservation Committee Inverdee House Baxter Street Aberdeen AB11 9QA UK	+44 1224 266573	ian.mitchell@jncc.gov.uk
Jacob Carstensen	Aarhus University Nordre Ringgade 1 8000 Aarhus C Denmark		jac@bios.au.dk
Jarosław Krogulec	The Polish Society for the Protection of Birds OTOP/BirdLife Poland Odrowąża 24 Marki 05-270 Poland		jaroslaw.krogulec@otop.org.pl
Jochen Krause	Federal Agency for Nature INA Vilm D-18581 Putbus Germany	+49 38301 86127	Jochen.krause@bfn-vilm.de
Johnny Reker	European Environment Agency Kongens Nytorv 6 DK-1050 Copenhagen K Denmark	+45 3343 5955	Johnny.Reker@eea.europa.eu
Jurgen Bastleer	VisNed 8320 AB URK Netherlands	+31 6 12 89 36 84	jbatsleer@visned.nl
Laura Boicenco	National Institute for Marine Research and Development Grigore Antipa 300 Mamia Bld. Constanta 900 581 Romania		laura_boicenco@yahoo.com

Name	Address	Phone/Fax	E-mail
Laura Uusitalo	Finnish Environment Institute (SYKE) Mechelininkatu 34a 00251 Helsinki Finland		laura.uusitalo@ymparisto.fi
Lena Avellan	HELCOM Helsinki Commission (Baltic Marine Environment Protection Commission) Katajanokanlaituri 6 B FI-00160 Helsinki Finland	+358 40 162 2054	lena.avellan@helcom.fi
Mark Dickey-Collas	ICES International Council for the Exploration of the Sea H. C. Andersens Boulevard 44-46 1553 Copenhagen V Denmark	+4533386759	mark.dickey-collas@ices.dk
Nadia Papadopoulou	Hellenic Centre for Marine Research Anavyssos 19013 Anavissos 190 13 Attica Greece		nadiapap@her.hcmr.gr
Simon P.R. Greenstreet	Marine Scotland Science Marine Laboratory 375 Victoria Road Aberdeen AB11 9DB UK	+44 1224 295417	Simon.Greenstreet@scotland.gsi.gov.uk
Suzannah Walmsley	ABPMer Quayside Suite Medina Chambers Town Quay Southampton SO14 2AQ UK		swalmsley@abpmer.co.uk

Name	Address	Phone/Fax	E-mail
Victoria Magath	University of Hamburg Centre of Natural History Martin-Luther-King-Platz 3 20146 Hamburg, Germany		victoria.magath@uni-hamburg.de

Annex 2: Agenda

Workshop on providing a method to aggregate species within species groups for the assessment of GES for MSFD D1 (WKD1Agg)

Biscay Room, ICES H.Q. Copenhagen, Denmark

10:00, 29 February–16:00, 2 March 2016

Chair: Anna Rindorf, Denmark

1. Introductions and welcome

Participants will be welcomed to the workshop.

2. Aim of the workshop

The agenda, aims of the workshop, and expected outcomes will be reviewed. Participants will be invited to provide initial feedback on the proposed agenda and process.

Please be aware that the term of reference is related to methods to integrate within species groups under Descriptor 1 and potentially with additional information from other descriptors, and hence is not addressing how to integrate across species groups.

3. List potential methods to integrate across indicators within species groups

Participants are invited to briefly present potential weighting methods (i.e. not software, but the underlying conceptual model) together with a technical documentation.

Presentations from HELCOM, DEVOTES and Marine Scotland (15 minutes each). If there are others who wish to present methods, please let Anna know.

4. Review of pros and cons of potential methods to integrate across indicators within species groups

Following the presentation in plenary, we divide into subgroups to review potential weighting methods. Subgroups report to plenary and a combined list is constructed. Pros and cons of each method is listed.

5. Conclusion

The main conclusions of the workshop will be reviewed and summarized.

Annex 3: Workshop Terms of Reference

The **Workshop on providing a method to aggregate species within species groups for the assessment of GES for MSFD D1** (WKD1Agg), chaired by Anna Rindorf, Denmark, will be established and will meet at ICES HQ, Copenhagen, Denmark, 22–24 February 2016 to:

a) In the light of previous advice, provide guidance on the most suitable and defendable approach to aggregate species within species groups (such as birds, mammals, reptiles, fish and cephalopods), for the state assessments of the MSFD.

WKD1Agg will report by 24 March 2016 for the attention of the ACOM Committee.

Supporting information

Priority	High, in response to a special request from GDENV on the Common Implementation (CIS) of the MSFD. The advice will feed into ongoing efforts to provide guidance on the operational implementation of the MSFD.
Scientific justification	Guidance on the most appropriate method to aggregate species within species groups for the assessment of GES for D1. This guidance can allow for a limited set of differing methods for different species groups, if required. The EC is requesting guidance on the most suitable and defendable approach to aggregate species within species groups (birds, mammals, reptiles, fish and cephalopods), for the state assessments of the MSFD. The request is driven by the need to consider how to aggregate species within the appropriate species groups for D1. This should focus on assessments of state (species groups). An open workshop will examine potential approaches for aggregation and will also incorporate the recent work by the ICES review of MSFD descriptor foodweb (D4), WKFOOWI and WGECO. The report from this workshop will then be examined by ICES WGECO (April 2016). An advice drafting group will take place in May, with a release in late May.
Resource requirements	ICES secretariat and advice process.
Participants	Workshop with researchers and RSCs investigators
Secretariat facilities	Yes.
Financial	Covered by DGENV special request.
Linkages to advisory committees	Run through ACOM.
Linkages to other committees or groups	Links to CSGMSFD and SCICOM.
Linkages to other organizations	Links to RSCs and EC.

Annex 4: Technical minutes from the WGECO

- Method to aggregate species within species group Review Group
- Dates: 6–13 April 2016
- Reviewed by the Working Group on the Ecosystem Effects of Fishing Activities
- Working Group: WKD1Agg

Consider methods to integrate indicators in support of integrated assessment of GES at the MSFD descriptor level (in collaboration with the DEVOTES project and building on work from WGBIOV) (tor f) and the request 'In the light of previous advice, review the guidance on the most suitable and defendable approach to aggregate species within species groups (such as birds, mammals, reptiles, fish and cephalopods), for the state assessments of the MSFD that was developed by the Workshop on providing a method to aggregate species within species groups for the assessment of GES for MSFD D1 (WKD1Agg).'

> It is recognized that the advice request was for D1 only, however, WGECO considered that the integration approaches should also be appropriate for other descriptors, especially D3, and probably also D4 and D6. WGECO have elaborated on the links to D3 in the following section on ToR F. Therefore, WGECO both reviewed the WKD1Agg report and then proceeded to consider the recommendations on integration given by WKD1Agg and WKGESFish together with WGECO considerations.

Review the guidance developed by the Workshop on providing a method to aggregate species within species groups for the assessment of GES for MSFD D1 (WKD1Agg).

WGECO reviewed the draft report of WKD1Agg principally in terms of the value of the report in providing advice for MSFD managers and of its consistency with previous recommendations from WGECO on integrating MSFD indicators. WGECO did not have access to the revised MSFD.

The report was considered as very useful, and covered the main elements needed to combine indicators.

Before drafting the advice in response to the request a number of important issues need to be considered:

- The section about integration methods is incomplete, for example it lacks the fuzzy operators which formalize a wide flexibility in integration methods (Silvert 1997, 2000). The section about averages also fails to mention geometric means, whereas it has been shown to be generally more relevant than arithmetic mean for indicator integration (Ebert and Welsh, 2004). The section about integration methods should be structured according to the methods which integrate quantitative indicators vs those which integrate assessments. Quantitative indicators may be aggregated by e.g. averages and weighted averages, which are essentially the same (averages being weighted averages with equal weights). By contrast, the percentage of indicators within limits (of which the "one out all out" rule is just a special case, with the required proportion 100%) integrates assessments resulting from combinations of indicators with the associated reference point, i.e. a qualitative statement such as "Good environmental status" or "Outside safe biological limits". The qualitative assessments require the availability of thresholds established in a consistent way for all the indicators to be integrated.

- A corollary is that averaging cannot be used after an assessment has been made at a lower level; which disqualifies the example in Section 5.1.4.4, Figure 10, where proportions are used at the lower level, preventing any averaging to be used at the next level.
- In a number of examples e.g. 4.4 and 4.5 (averages and weighted averages), the text mentions the idea of indicators "compensating" for each other. This terminology is confusing. At a simple level it could mean that a bad GES in one indicator, could be "compensated" by a good one in another. However, the text also suggests that indicators may or may not have the "ability to compensate for each other". This implies a judgement, and would be very difficult to substantiate. We would suggest the use of the term "cancel out" and make it explicit that this means that having one good, and one bad indicator in an integration step could lead to them "cancelling" each other out.
- It was quite difficult to identify the "cons" in the "pros and cons" sections. The general practice appeared to be to give the advantages, and then suggest a variety of approaches that might mitigate the "cons" that were not detailed clearly. We would suggest that the advice could include bulleted lists for the pros and cons, and then possibly suggest the mitigation of the cons specific to each of these. A table of method and pros and cons like the one given in Section 8.2 of this report might be useful.
- In many cases, the text discussed the problems of producing a single result from the integration at any level might "cover up" GES issues, at the lower level. This would apply to any of the averaging approaches. The text also differentiates between indicators that are below but close to a "good" status, from those very far from being "good". In all such cases we would recommend that, whatever the outcome of the integration process, the reporting should also detail where such cases had arisen. This would be vital information to target remediation or research.
- At the higher levels of integration the text implies that after an OOAO process, the components or components causing that "out" should be identified. As with the averaging discussed in the previous bullet, these should include how bad the components causing the "out" evaluation actually were.
- On page 7 and thereafter the text uses the term "areas" as distinct from "regions" and sub-regions" detailed in table 3. We understand these "areas" to be specific to the management areas in the Baltic Sea where a slightly different approach was proposed. However, this should be made clear, and it should be questioned whether a Baltic specific step in the integration is useful.
- The report explicitly avoids considering "pressure indicators", based on the specific request for advice. WGECO believes that any useful integration process should be able to encompass both pressure and state indicators at a disaggregated level.

- Many indicators so far proposed can be considered as "surveillance", rather than "operational", indicators. It should be made clear that this advice refers to operational indicators only, and probably only those where reference levels or some other good/bad state threshold can be identified.
- WGECO noted that there was some attempt to harmonise the D1 indicator approach with that used for the Habitats Directive. While this is not a problem in itself, it was felt that harmonisation should probably be emphasised first within the MSFD and only then across directives.
- The diagrams were felt to be valuable and helped in explaining both the principles and how it might work in action. However, if these are to be used in the advice there are some quality and consistency issues that need to be addressed. For instance there is virtually no difference between Figures 2 and 3 whereas they are supposed to illustrate two contrasted aggregation processes; in Figures 8, 9, 11 and 12 there was no indication of an integration process at the lowest level.

Minor points

- Section 2. The information referred to from Palialexis *et al*. should probably be included.
- Section 3 paragraph 1. End of the sentence should probably include "in abundance."
- Figure 6 is very poor quality, and impossible to see the elements indicated in the legend.
- Section 4.5. Paragraph 2. Line 2. This seems to be conflating more data with better data and mixing up quality and quantity of data. It is probably about quality of data.
- Section 5.1. Final paragraph. It is not clear what an "unacceptable deviation from the good status means" unless it simply refers to good vs bad??

There have been several ICES workshops that considered the issue of aggregation/integration in relation to an assessment of status of an MSFD Descriptor:

- For D1 the WKD1Agg the ToR was "In the light of previous advice, provide guidance on the most suitable and defendable approach to aggregate species within species groups (such as birds, mammals, reptiles, fish and cephalopods), for the state assessments of the MSFD."
- For D3 the WKGESFish the ToR was: "Conduct the assessment of criterion 3.1, 3.2 and 3.3, to evaluate the GES status of selected stocks (as examples)" which specifically involved an exploration of the methods for integrating indicator assessment results within stocks across Criteria 3.1 and 3.2 as well as the aggregation of stocks within criteria.

WGECO considers that the methods applied to assess the status of MSFD descriptors should not be considered in isolation for each of the descriptors separately but also across descriptors as this is likely to increase consistency in the assessments and thus comparability of the outcome of the assessments which is an advantage if trade-offs between the descriptors need to be considered.

Consider methods to integrate indicators in support of integrated assessment of GES at the MSFD descriptor level – comparing WKD1Agg, WKGESFish and WGECO approaches.

The WKD1Agg concluded that there were two feasible integration frameworks relevant for the assessment of D1: the integration within criteria and integration within species. This was also found by WKGESFish which, as an aid to compare and discuss these different frameworks, proposed to distinguish between the words 'aggregation' and 'integration' which are often used synonymously but can carry different connotations i.e.

- 'aggregation' refers to the combination of several elements which are similar (e.g. the aggregation of the same indicator across species or stocks or aggregation of species within the same criterion), whereas
- 'integration' refers to the synthesis of several elements, which are not comparable (e.g. the integration of several indicators or criteria within a stock, species or species group) (Borja *et al*., 2014).

Below we discuss the conclusions from WKD1Agg together with those coming from WKGESFish and considerations by WGECO. It should be noted that the two methods are identical if One Out All Out methods are used for integration.

WGECO also questioned the [usefulness][relevance for management] of highly aggregated or integrated evaluations that are effectively averages of a wide and varied set of indicators. Averaging up to an ecosystem component or to a species group, and then evaluating that component/group as being in or out of GES would tend to disguise the main issues. For components/groups that are in GES, managers and stakeholders would presumably need to identify those elements that were still outside GES and to target remedial action. The same would be true if the overall evaluation were to be outside GES; the managers would still need to know which specific indicators were driving this. It would also be useful to know by how much any given indicator was outside GES. Arguably, managers may want to focus on those closest to reaching their targets, although this may be "picking low hanging fruit". A better approach might be to carry out an ecological risk assessment (e.g. Hobday *et al*., 2011), and identify the most important elements outside GES. The key point is that these high level integration results may have some political value, but be of little use for targeting attention or action where it would be needed.

Integration within species		
D1 (WKD1Agg)	D3 (WKGESFish)	D1, D3, D4 and D6 (WGECO)
(+) Increases comparability with Habitats Directive	(-) Decreases comparability with Common Fisheries Policy	(-) decreases comparability with D1-habitat, D4 and D6, neither of which are species based
(-) Community considerations are not easily integrated		(-) Assessment of D1-habitat, D4 and D6 is not likely to depend on species specific indicators but rather indicators involving several species into a single indicator (e.g. biomass or production of a trophic guild)
(-) When different species have information for different numbers of criteria, this is masked in the higher-level integration.	(-) There are many stocks with information on only one of the criteria. If the requirement is that only stocks with both criteria can be included a lot of information is lost. Alternatively, allowing also stocks with only one criterion could be an incentive to provide only one (best) criterion for each stock. In both cases the outcome of the assessment will be affected.	
	(-) Proposes to retain the focus on stocks rather than species, as this promotes coherence with the CFP (which considers stocks) this would therefore require having to aggregate stocks into species.	

Aggregation within criteria		
D1: WKD1Agg	D3: WKGESFish	WGECO
(-) Decreases comparability with Habitats Directive	(+) Increases comparability with Common Fisheries Policy	
(+) Data for species where only one criterion can be measured can be included without loss of consistency and all criteria receive equal weight in the integration across criteria.	(+) Avoids the problems associated with combining indicators with different evidence base and levels of confidence (i.e. primary vs. secondary indicators), as well as the conceptual problem of combining pressure-related indicators (3.1) with state-related indicators (3.2).	(+) Retains the equal importance of all criteria, emphasizing that no one criterion is considered more important than any other criterion.
	(+) Proposes to retain the focus on stocks rather than species, as this promotes coherence with the CFP. This therefore avoids having to aggregate stocks into species.	(+) Combining stocks into species or criteria can be done either assuming stocks to be equally important or by assuming stocks to differ in importance. The current implementation of the CFP uses the former method but using an agreed weighting, such as biomass or areal coverage, may be more appropriate to e.g. foodweb considerations.
(+) Transparent weighting of all criteria, including community aspects when carrying out the next level of integration/aggregation.		(+) Transparent weighting of all criteria when carrying out the next level of integration/aggregation.
(+) More appropriate where there is a larger number of species, due to the increasing chance of some species missing information on one or more criteria with an increasing number of species	(+) Because D3 assessment usually involves large number of fish species, this is considered most appropriate method	(+) More appropriate where there is a larger number of indicators in accordance with the advice for D4 to monitor at least three trophic guilds.

Integration/Aggregation methods		
D1 (WKD1Agg)	D3 (WKGESFish)	D1, D3, D4 and D6 (WGECO)
The recommended integration method differs between integration levels under both of the two frameworks. At the level of integrating criteria, the recommended method is OOAO, and the same recommendation applies when aggregating species groups. For aggregation across species a mixture of averages, weighted averages, proportional and probabilistic methods are recommended depending on the specific situation.	At the level of integrating criteria, the recommended method is OOAO. The focus was mainly on two aggregations methods, the One-Out-All-Out approach (OOAO) and averaging methods. Probabilistic methods were only briefly discussed during the meeting, but were deemed to hold high potential for the determination of GES threshold levels for the aggregation of stocks within criteria.	At the level of integrating criteria, the recommended method is OOAO. For aggregation across indicators within a criteria a mixture of averages, weighted averages, proportional and probabilistic methods are recommended depending on the specific situation.

Conclusions		
D1: WKD1Agg	D3: WKGESFish	WGECO
The final choice of integration framework depends on the relative consideration of the wish to facilitate comparability with the Habitats Directive (integration within species), the wish to use similar methods across all ecosystem components, the wish to give a transparent weight to all criteria (integration within criteria) vs. giving transparent weight to all species (integration within species) and finally, the wish to incorporate community-level considerations in the integration in a simple way. The workshop par-ticipants considered that the weighing of these different wishes was a policy decision rather than a scientific decision.	The aggregation of stocks within criteria was considered as preferable by most workshop participants, because the aggregation of stocks within criteria would avoid the problems associated with combining indicators with differing evidence base and levels of confidence (primary vs. secondary indicators), as well as the conceptual problem of combining pressure-related indicators (3.1) with state-related indicators (3.2). The aggregation of stocks within criteria would also have the advantage that the available information would be optimally used and that results match those for the CFP.	Comparability across methods makes the final assessment at descriptor level more transparent and facilitates communication. There does not appear to be a good argument for using species approach in one descriptor but criteria approach in another.

References

Hobday, A. J., A. D. M. Smith, *et al*. 2011. Ecological risk assessment for the effects of fishing. Fisheries Research 108(2–3): 372–384.

Silvert, W. 1997. Ecological impact classification with fuzzy sets. Ecological Modelling 96:1–10.

Silvert, W. 2000. Fuzzy indices of environmental conditions. Ecological Modelling 130:111–119.

Ebert U., Welsch H., « Meaningful environmental indices: a social choice approach », Journal of environmental economics and management, 47, 270–283, 2004.