Ecological Indicators 61 (2016) 871–881

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Short Communication

Integrated assessment results depend on aggregation method and framework structure – A case study within the European Marine Strategy Framework Directive

W. Nikolaus Probst[a,*], Christopher P. Lynam[b]

[a] Johann Heinrich von Thünen-Institute of Sea Fisheries, Palmaille 9, 22767 Hamburg, Germany
[b] Centre for Environment, Fisheries & Aquaculture Science, Lowestoft Laboratory, Pakefield Road, Lowestoft, Suffolk NR33 0HT, UK

ARTICLE INFO

Article history:
Received 6 July 2015
Received in revised form 6 October 2015
Accepted 14 October 2015
Available online 10 November 2015

Keywords:
Integration
Aggregation
Indicator
Criterion
Descriptor
GES
Principal of equal probability

ABSTRACT

The ecosystem approach (EA) to environmental management is commonly associated with the holistic, integrative assessment of ecosystem status, where assessments of single elements are aggregated across one or multiple levels. Such an integrative assessment is required by the European Marine Strategy Framework Directive (MSFD). Member states of the European Union must assess the environmental status of their marine waters every six years. For this purpose the MSFD is structured into eleven descriptors of good environmental status (GES), which in turn are subdivided into 29 criteria containing a total of 56 different ecosystem indicators. These 56 indicators are recommended to be used by the member states to assess the status of biodiversity, invasive species, exploited fisheries resources, food webs, seafloor integrity, hydrological conditions as well as the impacts of eutrophication, contamination, litter and anthropogenic noise. The nested structure of the indicators within the commission decision on criteria for GES provides a hierarchy of information, for which aggregation at different levels may be necessary, namely within criteria, within descriptors and across descriptors. However, to date no aggregation rules have been provided by the European Commission. This study explores the implications of five commonly used aggregation methods, once applied with the assessment structure outlined for the MSFD, on the aggregated assessment outcomes at the level of criteria, descriptors and overall GES. Assessing the 56 indicators within the nested structure of the MSFD led to different outcomes between the different methods. Furthermore, all five methods were sensitive to the number of aggregated elements, with higher numbers of assessed elements being associated with lower probability of reaching GES. To overcome this drawback, two new aggregation methods were developed with the aim to ensure that the probability of achieving GES was equal for each indicator, criteria, descriptor and the overall environmental status. This aim was termed as principal of equal probability (PEP). In practice, only one out of the two aggregation methods developed succeeded in maintaining PEP across all hierarchical levels. Whether PEP is imperative for multi-element ecosystem assessments remains to be debated, however, both scientists and managers should be aware of PEP and the implications of its violation.

1. Introduction

Integrated environmental assessments (IEA) of pressures and states within marine ecosystems have become important objectives in marine policies due to the implementation of ecosystem based management approaches throughout the world (Borja et al., 2008; Diekmann and Möllmann, 2010; Levin et al., 2009, 2014; Link and Browman, 2014; Toth and Hizsnyik, 1998). Generally, IEA are based on environmental or ecological indicators, which are assessed against a threshold separating an acceptable from a non-acceptable status (Borja et al., 2011; Gascuel et al., 2014; Greenstreet et al., 2012; Probst et al., 2013; Probst and Stelzenmüller, 2015). The outcome of the single indicator assessments can then be integrated to achieve assessment scores at higher levels of information in which several indicators are combined (Borja et al., 2004; Halpern et al., 2012; HELCOM, 2010; Korpinen et al., 2012).

In Europe, the implementation of IEA for marine ecosystems has been formalised within the Marine Strategy Framework Directive (MSFD). The MSFD aims to achieve good environmental status (GES)

* Corresponding author. Tel.: +49 40 38905202; fax: +49 40 38905263.
E-mail address: nikolaus.probst@ti.bund.de (W.N. Probst).

http://dx.doi.org/10.1016/j.ecolind.2015.10.040
1470-160X/

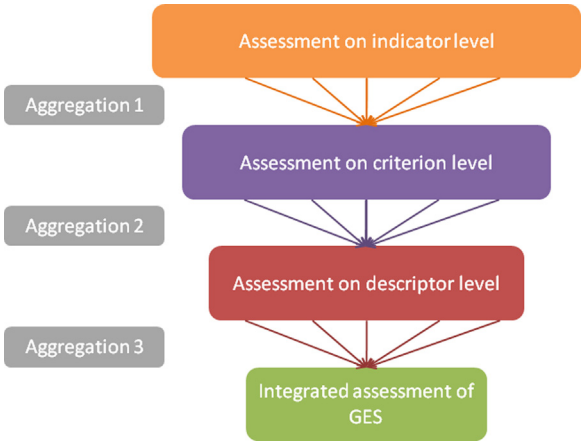

Fig. 1. Possible aggregation levels for the integrated assessment of marine regions within the EU-Marine Strategy Framework Directive (after Claussen et al., 2011).

in all European Marine waters by 2020. The progress towards GES shall be monitored by the EU member states across several ecosystem components (including marine mammals, birds, fish, benthos, plankton and macrophytes). Based on quantitative data IEA shall be performed every six years with the first IEA having been undertaken in 2012 (EU-COM, 2008).

At its lowest level, the MSFD contains 56 indicators, which are clustered into 29 criteria. These 29 criteria in turn are grouped into 11 descriptors of GES (EU-COM, 2010) (Table S1), which could be aggregated[1] into an overall status assessment (Borja et al., 2011; Halpern et al., 2012). Indicators, criteria and descriptors therefore reflect a hierarchy of information (Borja et al., 2014). The actual rules for aggregation within the MSFD have not to date been further specified by the EU (Borja et al., 2013), though several proposals exist within the scientific literature for various descriptors (Cardoso et al., 2010; Froese et al., 2015; Probst et al., 2013) and for the overall status (Borja et al., 2011). The EU member states performed their initial assessments in 2012, but in many cases their assessment and aggregation methods were not harmonized. Many member states aggregated the assessment results of single MSFD-elements without applying rigid aggregation procedures, mostly based on expert opinion (EU-COM, 2014). Therefore some inconsistencies and uncertainties exist in how and to which hierarchical level aggregation should be performed for the next environmental status assessment in 2018 (Borja et al., 2013). Nonetheless during the first stage of implementation of the MSFD there seemed to be consensus that aggregation should occur in three steps across indicators within criteria, across criteria within descriptors and across descriptors (Cardoso et al., 2010; Claussen et al., 2011) (Fig. 1).

To date few studies have investigated the consequences of different aggregation methods for IEA (Barnard and Strong, 2014; Moe et al., 2015; Ojaveer and Eero, 2011). This study contrasts different aggregation approaches within the indicator framework of the MSFD and compares the assessment outcomes at each level in the hierarchy i.e. criterion, descriptor and overall GES.

Assessment results are usually associated with uncertainty due to natural variability or observational error (Moe et al., 2015; Porszt et al., 2012). For these reasons the probability of obtaining at least one negative assessment result will increase with an increasing number of aggregated elements (Borja and Rodriguez, 2010; Moss et al., 2003; Noges et al., 2009). Because the MSFD descriptor and criteria clusters differ in the number of elements they contain, it can be assumed that clusters with many elements will have a lower probability of achieving GES than clusters which only contain few elements. Therefore the aggregated assessment outcome will be influenced by the number of elements per aggregation and hence by the nested, imbalanced structure of the MSFD. The second objective of this study was to develop an aggregation method which accounts for the design of the MSFD-indicator framework allowing each indicator, criterion and descriptor as well as overall GES to be achievable with equal probability, such that no element in the assessment is given unintentional precedence. Thirdly, a generic formula was empirically derived to transfer the principles for determining assessment benchmarks to other indicator frameworks.

2. Materials and methods

2.1. The reference indicator data set

To compare the influence of different aggregation methods on the aggregated assessment results a reference data set containing 1000 random Monte Carlo runs (MCR) for each MSFD-indicator was created. Each MCR created a random value between 0 and 1 resulting in a random-uniform distribution of indicator status (see Fig. 2A as example for a random-uniform distribution). The indicator data set was used to apply the different aggregation methods described below 1000 times to obtain distributions of the status of each criterion, descriptor and overall GES.

2.2. Defining the threshold for good environmental status (GES)

The status of each indicator was standardised to values between 0 and 1. This range was chosen to simulate ecological quality ratios (EQR) (Borja et al., 2007). EQR are used within the European Water Framework Directive (WFD) to allow a between-indicator comparison for metrics with different units and value ranges (Borja et al., 2004, 2011; Noges et al., 2009). EQR are based on the ratio between the environmental target (or reference condition within the context of the WFD) and the current state. EQR are grouped into five categories ('high', 'good', 'moderate',' poor', 'bad'). The boundaries between the five categories can be set somewhere between 0 and 1 e.g. between 'high' and 'good' at 0.8, between 'good' and 'moderate' at 0.6, between 'moderate' and 'poor' at 0.4 and between 'poor' and 'bad' at 0.2 (Borja et al., 2007). The first two categories ('high' and 'good') are considered to be consistent with GES of the MSFD (Claussen et al., 2011) and hence in this study the threshold for GES was set between 'good' and 'moderate' (Borja et al., 2011). Accordingly, an indicator was considered to reach GES when its score was >0.6.

2.3. The initial aggregation methods

Within the MSFD, the status of single indicators may have to be aggregated within criteria, these in turn may have to be aggregated within descriptors and finally the overall status of the ecosystem may have to be determined across all eleven descriptors (Fig. 1). This study started out by analyzing the implications of five different aggregation methods, which were applied to the reference indicator data set. These five methods are described in Table 1 and were:

- the 'one-out-all-out'-rule (OOAO)
- arithmetic mean (AA)
- median (ME)

[1] Some authors make a distinction between integration and aggregation. According to Borja et al. (2014) aggregation occurs between several elements which are similar, while integration occurs between elements which differ e.g. in spatial or temporal scales. Because I did not consider such structural differences in indicators and treated them as similar entities, the term 'aggregation' is used synonymously to the term 'integration' throughout the study.