Hydrobiologia (2013) 704:213–229
DOI 10.1007/s10750-012-1276-9

WATER BODIES IN EUROPE

Review Paper

Transitional and coastal waters ecological status assessment: advances and challenges resulting from implementing the European Water Framework Directive

Angel Borja · Mike Elliott · Peter Henriksen · Núria Marbà

Received: 10 May 2012 / Accepted: 4 August 2012 / Published online: 21 September 2012
© Springer Science+Business Media B.V. 2012

Abstract The derivation, performance, sensitivity and inherent uncertainty of ecological quality indicators have become major topics in developing tools for the management of marine, transitional and coastal waters. In reviewing the advances in these waters, related to an ecological status assessment, we show the future challenges to be addressed within the European Water Framework Directive (WFD). Using new analyses carried out under the research project 'Water Bodies in Europe: Integrative Systems to Assess Ecological status and Recovery', we provide a complete set of assessments for the biological quality elements (BQEs) (phytoplankton, macroalgae-seagrasses, macroinvertebrates and fish) to be assessed, as well as the validation of existing indicators and multimetric indices and, in some cases, the development of new assessment indices. We show that these indices respond differently to different human pressures and they each have challenges in defining reference conditions against which future changes are judged. In investigating good ecological potential, as the response to heavily modified water bodies, we show that there are flaws in the Directive, not least in its definitions. Our analyses have also focussed on uncertainty in using the indices and we emphasise the problems of defining ecological class boundaries based on indices which themselves may be combined indices (multimetrics). The analysis shows that some of those multimetrics are redundant and/or are intercorrelated and thus may reduce the sensitivity in defining ecological class boundaries. If this is related to the drivers-pressures-state change-impacts-response approach then there are lessons for management measures aimed at achieving good ecological status and even the potential for legal challenges to decisions based on uncertain indices under the WFD. Hence, we conclude the continued need for advances in assessing pressures and gradients, and defining reference conditions for state change, index development, impact assessment and the validation of indices for each BQE.

Guest editors: C. K. Feld, A. Borja, L. Carvalho & D. Hering / Water bodies in Europe: integrative systems to assess ecological status and recovery

A. Borja (✉)
AZTI-Tecnalia, Marine Research Division, Herrera Kaia, Portualdea s/n, 20110 Pasaia, Spain
e-mail: aborja@azti.es

M. Elliott
Institute of Estuarine & Coastal Studies, University of Hull, Hull HU6 7RX, UK

P. Henriksen
Department of Marine Ecology, National Environmental Research Institute, Aarhus University, 4000 Roskilde, Denmark

N. Marbà
Department of Global Change Research, IMEDEA (CSIC-UIB), Institut Mediterrani d'Estudis Avançats, Miquel Marquès 21, 07190 Esporles (Illes Balears), Spain

Keywords Water framework directive · Methods of quality assessment · Uncertainty · Response to human pressures · Biological quality elements

Introduction

The Water Framework Directive (WFD; 2000/60/EC) establishes a framework for the protection of all European waters, including transitional (estuaries and lagoons) and coastal waters. The WFD requires Member States to assess the ecological status of all water bodies, based especially on the status of the biological elements (phytoplankton, macroalgae, seagrasses, macroinvertebrates and fishes (the latter only in transitional waters rather than in marine waters)) as well as hydromorphological and physico-chemical quality elements. The status requires assessing by comparing monitoring data against proposed reference conditions (Borja et al., 2012). Reference conditions are determined using one of four methods: comparison with an area lacking human pressures or with demonstrated high status, hindcasting to a time prior to significant human pressures, predictive modelling or best expert judgement (Hering et al., 2010).

The adoption of the WFD has produced a plethora of methods to assess the ecological status of the different elements (Birk et al., 2012). However, many challenges in implementing such a complex piece of legislation still need to be addressed (Hering et al., 2010). Some of these challenges, together with others such as the influence of climate change on reference conditions, pressure responses or the recovery of aquatic ecosystems after the removal of pressures, have been addressed in the European research project 'Water Bodies in Europe: Integrative Systems to Assess Ecological status and Recovery' (WISER; www.wiser.eu, completed in February 2012).

This study aimed to provide a complete set of assessments for the transitional and coastal (TraC) biological quality elements (BQEs), which requires the overview of existing indicators and multimetric indices (see Marbà et al., 2012) for seagrasses; Borja et al. (2009a, b) for macroinvertebrates; and Pérez-Domínguez et al. (2012) for fishes), and, in some cases, the development of new assessment indices. The study also aimed (i) to investigate the response of these indices to different human pressures, such as hydromorphological changes, eutrophication, pollution (metals and organic compounds), etc.; (ii) to define the required reference conditions; (iii) to investigate good ecological potential, within heavily modified water bodies (HMWB; see Borja & Elliott,2007); (iv) to estimate uncertainty in assessing the ecological status, including a wide range of geographical regions and types; (v) to determine different ways of combining single metrics into multimetrics or holistic assessment; (vi) to analyse redundancy in component metrics, whether there is duplication or double-counting, and which proportion of the end result is dependent on the individual metrics and (vii) to determine any inter-relationships between the sensitive metrics/indicators.

These objectives are required to link to the drivers-pressures-state change-impacts-response conceptual model (DPSIR) approach (e.g. Elliott, 2002; Atkins et al., 2011) to the assessment index development and validation, response to pressures and setting reference conditions (Fig. 1). The WISER project investigated pressures, disturbance gradients, reference conditions for state change, index development for impact assessment and the validation, sensitivity and uncertainty of the indices.

As such, using information published in WISER project, here we review advances in assessing ecological status in TraC waters to identify future challenges in implementing the WFD. In this way, feedback from the dissemination of the investigations undertaken within WISER project provided valuable lessons for future reviews in terms of stakeholder involvement in assessing the ecological status. Hence, systematic reviews such as this are necessary tools for ecosystem management and formalise the information available based on the weight of evidence, as shown by Stewart et al. (2005).

Study area and data used

In order to address the objectives, existing extensive previous datasets (e.g. Basset et al. (2012a) for macroinvertebrates) were merged with a field sampling survey of the four above-mentioned BQEs, at a series of sites throughout Europe (Table 1), using harmonised sampling and analytical methods (i.e. Borja et al., 2011). This produced data for estimating the statistical uncertainty of the assessment methods.

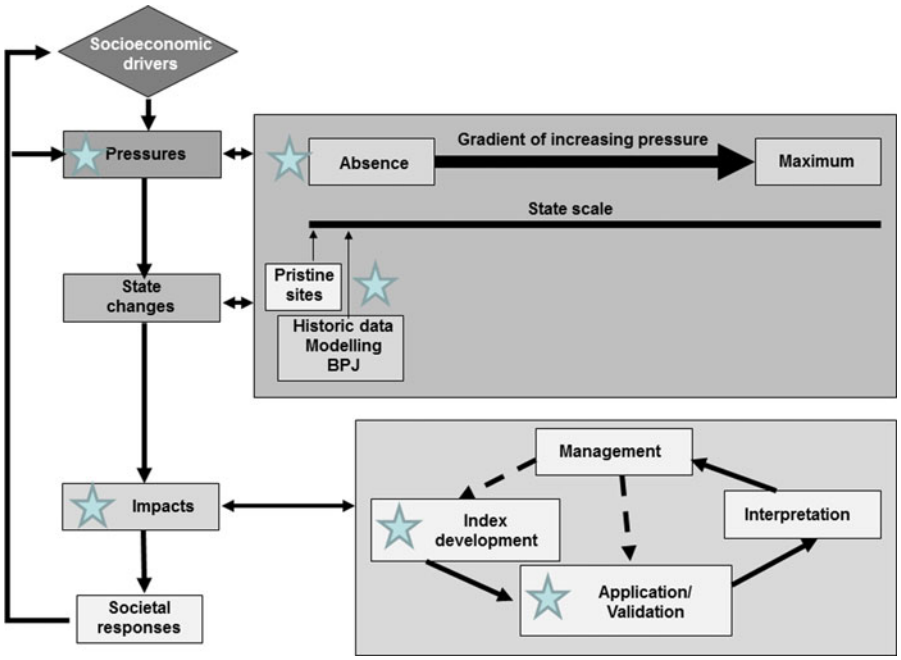

Fig. 1 Conceptual model linking the drivers-pressures-state changes-impacts-responses (DPSIR) approach to the assessment indices development and validation, response to pressures and setting reference conditions, as it was undertaken during the project in coastal and transitional waters. *Stars* indicate the topics addressed during the project. *BPJ* best professional judgment. Adapted from Borja & Dauer (2008) and Borja et al. (2012)

The differences across Europe were assessed for each BQE on a wide range of geographical regions and water types. Hence, the field survey focussed on a Mediterranean lagoon (Lesina lagoon, in Italy), a coastal area in the Mediterranean (for seagrasses, Balearic islands, in Spain), a TraC area within the Black Sea (Varna Bay and lagoon, in Bulgaria), a medium size Atlantic estuary (Mondego, in Portugal), a coastal Atlantic area (the Basque coast, in northern Spain), a Norwegian fjord (Oslofjord) and the Gulf of Finland (Table 1). Data from field surveys were combined with data from existing WFD monitoring programmes, in order to cover the widest range of sites and water body types as possible.

Review of existing indicators and developing new indicators and multimetric indices

The number of ecological quality assessment methods has increased exponentially in recent times (Birk et al., 2012). However, the WISER project evaluated the need to develop new assessment methods, taking into account the lack of methods for some BQEs (e.g. macroalgae) and ecotypes (e.g. hard substratum, lagoons, etc.). It also considered the development of new metrics (e.g. size spectra in phytoplankton and macrobenthos), the use of new methods in sampling or identification [e.g. FlowCAM (Flow Cytometer And Microscope) (Garmendia et al., 2012)] and satellite derived assessment (Novoa et al., 2012), of phytoplankton).

Table 1 Locations sampled for this study, showing the main environmental characteristics

Country	Name	Water type	Ecoregion	Depth (m)	Salinity
Bulgaria	Varna Bay	Coastal bay and lagoon	Black Sea	5–17	Euhaline
Italia	Lesina	Lagoon	Mediterranean Sea	0.6–1.2	Mesohaline
Spain	Balearic Islands	Coast	Mediterranean Sea	0–50	Euhaline
Portugal	Mondego	Estuary	Northeast Atlantic Sea	3–9	Oligo to euhaline
Spain	Basque Country	Coast	Northeast Atlantic Sea	34–100	Euhaline
Norway	Oslofjord	Fjord	Northeast Atlantic Sea	57–452	Euhaline
Finland	Gulf of Finland	Coast	Baltic Sea	0–150	Oligo-mesohaline

Borja & Dauer (2008) and Borja et al. (2009a, b) described the steps to be followed in developing new assessment methods including: (i) selection of candidate metrics; (ii) metric combination; (iii) index validation, using an independent dataset; (iv) index application to different human pressures; (v) index interpretation and (vi) index intercalibration. From the literature, it can be seen that these steps have not always been completed, and, in some cases, there is a lack of calibration and validation of the assessment methods (e.g. Birk et al., 2012). Hence, in this study, the above-mentioned steps have been followed and four new assessment methods have been developed.

Phytoplankton

Morphological-functional traits, as body size and size spectra respond to different types of anthropogenic pressures. A multi-metric Index of Size-spectra Sensitivity (ISS-phyto), which integrates size structure metrics with metrics describing the sensitivity of size classes to anthropogenic disturbance, chlorophyll a and species richness measures were developed, tested and validated (Lugoli et al., 2012). The index was developed using phytoplankton data from 14 Mediterranean and Black Sea coastal lagoons, which were classified as either 'disturbed' or 'undisturbed' ecosystems based on expert quantitative analysis, evaluation of anthropogenic pressures in the catchment area and their current protection and conservation status. The index was effective in discriminating between natural and anthropogenic pressures by presenting significantly higher values at undisturbed than at disturbed sites (Fig. 2).

Macroalgae

Two new assessment methods were developed for macroalgae: one for the Atlantic Iberian coasts (Díez et al., 2012) and the other for Portuguese coasts (Neto et al., 2012). The Rocky Intertidal Community Quality Index (RICQI; Díez et al., 2012) combined into a single value species abundance, morphologically complex algae cover, species richness and faunal cover (herbivore and suspensivore cover, proportion of fauna with respect to the whole assemblage). An independent dataset collected from the Basque coast (N. Spain), before and after the commissioning of a wastewater treatment plant, was used to validate the index. A conceptual model based on these results was proposed to describe successional stages of assemblages along a gradient of increasing environmental disturbance. The performance of this approach was compared with another rocky substratum index (Juanes et al., 2008), currently used as the official method for assessing the ecological status of rocky assemblages along the Spanish Atlantic coast. Both indices responded to changes in community structure, associated with pollution reduction. However, the RICQI index was more sensitive in detecting gradients and changes in disturbance than the other index (see Díez et al., 2012).

The second index (MarMAT: Marine Macroalgae Assessment Tool, Neto et al. (2012)) was developed in Portugal as a multimetric method based on the composition (Chlorophyta, Phaeophyceae and Rhodophyta) and abundance (coverage of opportunists) of marine macroalgae. MarMAT highly and negatively correlated ($P < 0.001$) with the degree of anthropogenic pressures.

Seagrasses

Despite the importance of this component in marine ecosystem functioning, no new method was developed within WISER, although it was necessary to review the current use of seagrass indicators in European monitoring programmes (Marbà et al., 2012). The type and number of indicators used vary across European regions, largely reflecting the regional differences in seagrass flora and plant dynamics. A total of 42 monitoring programmes, aiming at evaluating seagrass health (11 programmes), assessing coastal quality (28 programmes) or both (3 programmes), were identified (Marbà et al., 2012). The programmes span the four European ecoregions and involve the four main European seagrass species (*Zostera nolti*, *Z. marina*, *Posidonia oceanica* and *Cymodocea nodosa*). These programmes use 49 seagrass indicators including a total of 51 seagrass metrics used either on their own or in various combinations of up to 14 metrics per indicator. Mediterranean monitoring programmes include by far the largest diversity of seagrass indicators, followed by those for the North East Atlantic and the Baltic Sea regions, while those of the Black Sea encompass the least diversity of seagrass indicators (Marbà et al., 2012).

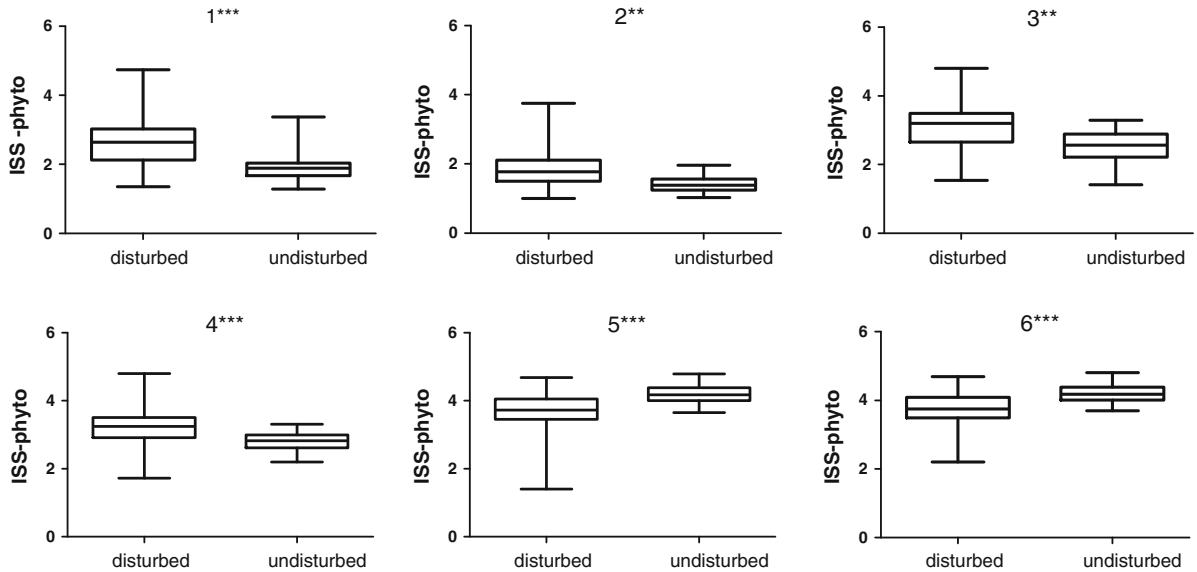

Fig. 2 Significant differences between disturbed and undisturbed Mediterranean lagoons, as assessed by the Index of Size-spectra Sensitivity (ISS-phyto), using six sensitivity models (nos. 1–6) tested on data from WISER phytoplankton sampling: symmetric (1), right asymmetric, i.e. sensitivity increasing with increasing size class; (2–4), and left asymmetric, i.e. sensitivity decreasing with increasing size class (5–6) (modified from Lugoli et al. (2012), for details consult that paper). Statistical comparison (ANOVA test) is shown as: $^{**}P < 0.01$; $^{***}P < 0.001$

Macroinvertebrates

A multimetric Index of Size-spectra Sensitivity (ISS) was developed, tested and validated (Basset et al., 2012a). It integrates size structure with other metrics describing the sensitivity of size classes to anthropogenic disturbance and species richness measures. The ISS was developed using benthic macroinvertebrate data from 12 Mediterranean and Black Sea lagoons. The selected lagoons were classified as either 'disturbed' or 'undisturbed' ecosystems based on expert quantitative analysis, an evaluation of anthropogenic pressures in the catchment area and their current protection and conservation status. Data from another Mediterranean lagoon, characterised by a very strong abiotic stress gradient, were used to validate the index. The ISS clearly discriminated between disturbed and undisturbed sites (Fig. 3).

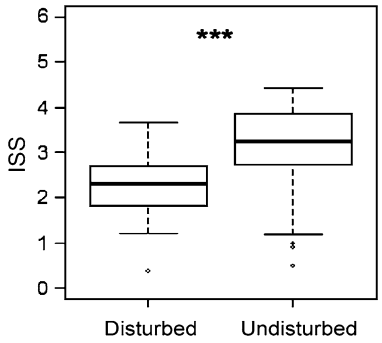

Fig. 3 Comparisons of results of the Index of Size-spectra Sensitivity (ISS) tested for 'undisturbed' and 'disturbed' lagoon sites, within the Mediterranean and Black Seas. Horizontal bars in the box-plot graphs represent the mode of value distribution; box-plot heights represent the 25th and the 75th percentiles, and the error bars represent the maximum non-outlier range. Statistical comparison (Wilcoxon rank test) of undisturbed and disturbed sites is reported in each graph as either *ns* not significant; $^{*}P < 0.05$; $^{**}P < 0.01$; or $^{***}P < 0.001$ (modified from Basset et al., 2012a)

Fish

Although no new indices were developed because of a large number of existing methods, the study reviewed and compared existing fish indices (Pérez-Domínguez et al., 2012). The review included 17 published fish-based indices of the habitat integrity of transitional waters (estuaries, fjords, river mouths, deltas, rías, limans and lagoons) and summarised common development strategies, in different countries worldwide. Most indices are

computed from a number of independent metrics (i.e. they are multimetric indices) and are based on assemblage composition or functional attributes of fish species (e.g. guilds, Elliott et al., 2007). Among metric groups, species richness and composition metrics are the most widely used in current indices, followed by habitat guild (e.g. number or proportion of estuarine species), trophic guild, abundance and condition and finally nursery function metrics. Within these metrics, families, indicator species or guilds associated with estuarine quality features often dominate the indices. Development strategies for the multimetrics vary but generally include (i) selection and calibration of metrics to anthropogenic pressure; (ii) development of reference conditions; (iii) comparison of metric values to reference ones and (iv) designation of thresholds for ecological status class. Only about half of the indices reviewed attempted to validate the index outcomes and these were limited to simple correlation analysis and misclassification rate analysis comparing index ecological status class value and anthropogenic pressure proxies. Currently there are no consistent European-wide fish indices for transitional waters; although under the implementation and wording of the WFD, countries are allowed to adopt their own methods (Hering et al., 2010; Birk et al., 2012). Widening of the geographical relevance will require better precision in formulating of reference conditions and greater inclusion of functional metrics.

All of the above initiatives discussed the nature of existing metrics for BQE and in particular the dependency on metrics describing the assemblage structure. They all demonstrate that a move towards more valuable functional indices are not the least as the proponents take the view that functioning of a system, i.e. the reliance on rate processes, is a more valuable indication of the health of the system than using only structural indices (Borja et al., 2010).

Identification of human pressure-response relationships

Despite the number of assessment methods used in Europe (Birk et al., 2012), studies on the response of assessment indices to human pressures are more scarce, i.e. for phytoplankton (McQuatters-Gollop et al., 2009; HELCOM, 2010; Garmendia et al., 2011), for macroalgae (Guinda et al., 2008; Orfanidis et al., 2011; Sfriso & Facca 2011), for seagrasses (López y Royo et al., 2009), for benthos (Quintino et al., 2006; Chainho et al., 2008; Josefson et al., 2009; Borja et al., 2009a; Neto et al., 2010) and for fishes (Uriarte & Borja, 2009; Delpech et al., 2010; Cardoso et al., 2011).

When studying the response of BQEs to human pressures, the major stressors considered were: (i) hydromorphological pressure, mainly in transitional waters (e.g. structural changes, residence time and flushing rate alterations), including the assessment of the good ecological potential of HMWBs (e.g. in the case of fishes); (ii) eutrophication (restricted to selected BQEs, such as phytoplankton, macroalgae and seagrasses) and (iii) pollution (metals and organic compounds), affecting disturbance-sensitive species, such as in benthic macroinvertebrates. These major stressors have been considered under different pressures (e.g. presence of ports, aquaculture, urban and industrial discharges).

Phytoplankton

The biomass of marine phytoplankton is affected by eutrophication (e.g. Carstensen & Henriksen, 2009). Thus, increasing concentrations of nutrients will support a larger biomass, and indeed the total concentration of chlorophyll a, as a proxy of phytoplankton biomass, was significantly correlated with total nitrogen (TN) across the geographically different sampling localities (Table 1) (Henriksen et al., 2011). However, no clear relationships were found for relative contributions of the different phytoplankton groups across this eutrophication gradient. The composition of phytoplankton communities determined from pigment analysis correlated mainly with salinity and temperature, being less correlated with TN as a measure of eutrophication. This result pinpoints the need for establishing type-specific indicators but generally this is hampered by a lack of a sufficient within-system pressure gradient to establish empirical relationships between the pressure and the indicator.

Seagrasses

Ecological regime shifts affect the response of seagrass indicators to pressures and may delay restoration of seagrass meadows after removing the

pressure (Krause-Jensen et al., 2011). The latter study quantified and compared benthic and pelagic gross primary production (GPP), along spatial and temporal nutrient gradients in a shallow estuary. The estuary experienced a shift from a pristine, seagrass-dominated clear-water regime with high total GPP in the early 20th century to a eutrophic, plankton-dominated regime still with high total GPP in the 1980s when nutrient loadings peaked. Recent reductions in nutrient loadings reduced pelagic GPP as expected, but the water remained turbid and seagrass abundance and GPP did not increase correspondingly. The results suggest that feedback mechanisms, such as increased resuspension of the seafloor and reduced trapping of particles and nutrients, resulting from the loss of seagrasses and their associated ecosystem services then delayed or prevented restoration to a state with seagrass dominance (Krause-Jensen et al., 2011).

Macroinvertebrates

Single metrics (e.g. abundance, number of taxa, and several diversity and sensitivity indices) and multimetric methods (as embedded in 8 of the most common indices used within the WFD) were compared to assess TraC benthic status along human pressure gradients in five distinct environments across Europe (Borja et al., 2011), in Bulgaria, Italy, Portugal, Basque coast (Spain) and Norway (Table 1). Within each system, sampling sites were chosen along an increasing human pressure gradient according to a preliminary classification based on professional judgment. The different indices are largely consistent in their response to pressure gradient, except in some particular cases. Inconsistencies between indicator responses were the most pronounced in transitional waters, highlighting the difficulties of the generic application of indicators to all marine, estuarine and lagoonal environments. However, some of the single metrics and multimetric methods were able to detect such gradients both in TraC environments. Furthermore, the multimetric methods appeared more consistent for detecting such gradients than single indices (Borja et al., 2011). The agreement observed between different methodologies and their ability to detect quality trends across distinct environments show a desirable attribute required for the implementation of the WFD's monitoring plans.

Fish

Using a matching combination of a fish index, reference values and a local dataset, the transitional fish indices (and metrics) can be sensitive to pressure gradients. Pérez-Domínguez et al. (2012) analysed the strength of expected metrics responses to a set of human pressures, and suggested that chemical pollution and loss of habitat were more frequently and more strongly related to fish metrics directly or indirectly reflecting alterations in transitional water fish assemblages. Their analyses provided a conceptual basis for ranking human pressures relative to their expected relevance for fish in transitional waters. In order to further confirm the relationship between fish-quality attributes and pressures, two WFD-compliant indices (from the Spanish Basque Country (AZTI's Fish Index (AFI), Uriarte & Borja (2009)) and Portuguese estuaries [Estuarine Fish Assessment Index (EFAI), Cabral et al., 2012)] were related to a set of pressures acting in these water bodies, while also considering their hydro-morphological descriptors. AFI related to different variables is:

$$\begin{aligned}\text{AFI} = {} & 0.013 + 0.017(\text{average estuary depth}) \\ & - 0.003(\text{global pressure index}) \\ & - 0.001(\text{residence time}) \\ & + 0.028(\text{dredged volume}) \\ & - 0.007(\%\text{ of channelling in ports}) \\ & + 0.009(\%\text{ of channelling out of ports}) \\ & \text{Adjusted } R^2 = 0.859,\ P < 0.05.\end{aligned}$$

Hence, the deeper the estuary, and the shorter the residence time, the pressure index and the channelled ports within the estuary, then the higher the AFI values would be, indicating higher ecological quality. AFI decreases with the increase of pressures. Similar analysis for the EFAI found comparable negative response of the index with increasing pressures. In this case, the EFAI responded to the overall anthropogenic pressure adapted from Aubry & Elliott (2006) (see Pérez-Domínguez et al., 2012).

In addition to the regression approach, Drouineau et al. (2012) tested an alternative method to establish metric-pressure relationship using a Bayesian approach. This allows for selecting and combining relevant fish metrics, taking into account their sensitivity to pressure, their variability or any other relevant

feature. It was tested on a dataset from 14 French lagoons. The analysis suggested that the quality diagnostics were less variable at the level of the multi-metric indicator than at the level of the fish metrics considered individually. These studies indicated that the BQE fish response to pressures in transitional waters provides a high level of ecological integration to the quality evaluation of those waters.

Definition of reference conditions

Reference conditions are optimally defined/described from data (i) best acquired from multiple sites with similar physical characteristics, within an ecoregion and habitat type; (ii) that ideally represent minimally impaired or undisturbed conditions (i.e. absence or minimal human pressure) and (iii) that provide an estimate of the variability in biological communities and habitat quality due to natural physical and climatic factors (Borja et al., 2012). In essence this can be interpreted as sites which have an *absence* of pressures or a *presence* of high ecological quality. While the former (the identification of pressures) is easier to determine, the latter (the ecological quality) is expensive to determine. However, in Europe there are not many pristine places to be used as reference sites and, consequently, different approaches need to be used in determining reference conditions (i.e. hindcasting, modelling or the best professional judgment).

Phytoplankton

The composition of phytoplankton communities in several water bodies in Europe considered to represent reference conditions was described (Revilla et al., 2010). These water bodies belonged to the Northeast Atlantic and the Mediterranean Sea ecoregions. In addition, data from the non-pristine Baltic Sea were evaluated to provide a characterisation of phytoplankton under good or high ecological status. For the assessment, different methodologies were applied, involving a range of aspects: from the approach for selecting the most suitable datasets, to the laboratory techniques and the numerical analyses employed.

The study in the Baltic Sea ecoregion indicates that by screening 'the best samples' of the present day monitoring data and analysing the literature (Table 2), it may be possible to identify phytoplankton assemblages revealing a presumed good quality. In the Northeast Atlantic ecoregion, the Basque coastal waters present a low risk of eutrophication as indicated by chlorophyll *a* concentration and physico-chemical conditions. This makes this coastal area suitable for establishing phytoplankton reference conditions. In this zone, the dominant taxa were small-sized organisms (2–20 μm) that could belong to different taxonomic groups. Pico-phytoplankton assemblages from the Mediterranean Sea ecoregion were described using Flow Cytometry, an automatic technique. This was found to be more rapid and less laborious than classical epi-fluorescence microscopy, and also allowed increasing the total number of cells counted, which reached thousands of cells observed in a single measurement.

Although the dataset and number of stations studied were not sufficiently representative to draw conclusions for extensive zones, the results obtained on the micro-, nano- and pico-phytoplankton communities can be used for comparison with other coastal waters, allowing the development of composition-based indicators. Estuaries are systems usually more affected by anthropogenic pressure (Elliott & Whitfield, 2011), which, combined with their higher variability, makes it more difficult to establish phytoplankton reference communities.

Benthic vegetation (macroalgae and seagrasses)

The reference conditions in the wide repertoire of methods for water status classification using benthic vegetation encompass historical data (i.e. Eelgrass depth limit, Krause-Jensen & Rasmussen (2009)), composition of vegetation under low-pressure conditions (e.g. Ecological Evaluation Index, Orfanidis et al. (2011)) and reference conditions for each single metric when the index is multimetric (e.g. MarMAT (Neto et al., 2012), *Posidonia oceanica* Multivariate Index (POMI, Romero et al., 2007).

Macroinvertebrates

Surface area, tidal range, confinement and water salinity, which are the drivers of the lagoon typologies proposed in the literature, were significant sources of assessment tool variability (Basset et al., 2012b). This allowed type-specific reference conditions and classification boundaries to be defined, improving the

Table 2 Empirically estimated reference values of chlorophyll a, total phosphorus and total nitrogen and the historical values of Secchi depth in the Baltic Sea

	BB	Q	AS	GF
Chl-a (μg l^{-1})	1.6 (2.4)	1.4 (2.1)	1.4 (2.1)	2.1 (3.2)
Total P (μmol l^{-1})	0.19 (0.29)	0.26 (0.39)	0.45 (0.68)	0.55 (0.84)
Total N (μmol l^{-1})	18.8 (28.3)	17.2 (25.8	17.5 (26.3)	19.3 (28.9)
Secchi depth (m)	6.7 (4.5)	8.6 (5.7)	8.9 (5.9)	5.6 (3.7)

BB outer coastal types of the Bothnian Bay; *Q* the Quark; *AS* the Archipelago Sea; *GF* the eastern Gulf of Finland. The acceptable deviation (50%) from the average reference values is presented in the *brackets* and represents the boundary between good and moderate status

accuracy of ecological status assessment. At the lagoon level, accuracy increased by 100% for the more complex typological schemes and by 83% in a validation test performed on an independent set of highly disturbed sites (expected ecological status from moderate to bad). Nevertheless, a certain degree of uncertainty was still found to affect classification at the study site level, with up to 38% of reference sites classified as moderate to bad. This finding has important implications in the assessment, as some reference sites appear misclassified.

Fish

The modelling approach of fish metrics against the physico-chemical variables has proved useful to derive reference conditions. This is important for the computation of relevant ecological quality ratios (EQRs) in Europe where there is a general lack of pristine areas or historical data on fish BQE and it provides an alternative to the best professional judgment. The literature distinguishes between metric-, habitat-, season-, gear-, salinity class-, estuary- and ecotype-specific reference conditions as relevant to the data structure and analysis (e.g. Pérez-Domínguez et al., 2012). In practice, the reference community is derived by either using pressure-response models or by selecting the highest values (top scoring samples) of the metrics in the dataset, assuming that less-impacted sites are present. Once the reference values are set, each sample is scored independently depending on the metric value in relation to the reference condition. Scoring systems are simple sliding scales rating sites by decreasing degree of deviation from the expected reference. The number and cut-off point for the score thresholds vary among indices and estuarine typology and are often calibrated with pressure data, if available (Pérez-Domínguez et al., 2012).

A predictive linear modelling approach (LM and GLM) has been used to define reference conditions for fish metrics in transitional waters using an extended dataset from European water agencies. The fish response data were modelled together with corine land cover (CLC)-derived pressure proxies (percentage of agricultural, urban, and natural land coverage). The models obtained allowed the expected metric score to be predicted by setting pressure levels either to the lowest observed pressure in the dataset or to zero in order to define the sample and theoretical reference condition, respectively. Extrapolating to the zero value of pressure may be unreliable. Hence, a more conservative approach using the lowest observed pressure values, may give a better prediction (i.e. increases accuracy) but produces a reference condition set at an artificially lower quality level.

Evaluate uncertainty on the use of assessment methods

A central and challenging element in WFD-compliant assessment systems is the estimation of uncertainty (Hering et al., 2010). This reflects the fact that there is no definitive bioassessment and that all results are influenced by several sources of variability and errors, for example variability in sampling, laboratory analysis, and in temporal and spatial variability (Clarke & Hering, 2006; Carstensen, 2007). Given that the WFD requires Member States to take action if a quality status is lower than good status (i.e. moderate, poor or bad), then there are financial repercussions if an area is deemed to fall below the good-moderate boundary. Furthermore, the certainty by which an area is

assigned in a particular class is both very important and may even be legally challenged by an industry financially penalised by the class judgement. Therefore, ecological status classification should always be given in terms of probabilities, even though at present this is used by few assessment systems (Birk et al., 2012).

The underlying statistical principles that are relatively simple and appropriate tools for uncertainty estimation are available (e.g. Clarke & Hering, 2006; Carstensen, 2007) but data are needed, which address the individual sources of error, such as differences between investigators and sampling equipment/analysis, as well as temporal and spatial variation of sampling, affecting the statistical distribution of the assessment results.

The determination of ecological status, and thus the need to invest large amounts of money to remediate problems, is affected by the uncertainty in defining status, especially when metric results are close to the good/moderate class boundary (Hering et al., 2010). Hence, in WISER, uncertainty analysis was a major goal. These analyses included the assessment of different sources of variability (sampling, processing, natural spatial and temporal variation, calculation of metrics and estimation of response curves), as a basis for identifying good indicators (i.e. those sensitive to pressure and of high precision). Combined uncertainty analyses were used to assess the risk of misclassification, in particular across the good/moderate boundary. The sources and magnitude of uncertainty were examined to develop guidance on sampling frequency (temporal variability), number of sampling sites (spatial variability) and analytical methods (harmonised versus non-harmonised). The uncertainty in ecological status classification was estimated using WISERBUGS [WISER Bioassessment Uncertainty Guidance Software (Clarke, 2012)].

Phytoplankton

Results from a large scale study quantifying sources of variation in the assessment of phytoplankton communities across European water bodies showed that the main proportion of the variation between pigment measurements (10–68% of variation) was explained by the variation between stations. For measurements of population density recorded as number of cells l^{-1} the main proportion of the variation (35%) was explained by the variation between the sample analysts (Dromph et al., 2012).

Sampling of parameters for characterising the phytoplankton communities, pigments and enumeration of cells, was performed in seven European water bodies. Within-each water body replicate sampling was carried out at several stations. At each station, two to seven water samples were taken, and one sample from each station was further divided into two (for HPLC and chlorophyll *a* analyses) or three (for cell counts) subsamples.

The study showed that increasing the number of stations sampled will have the greatest influence on increasing the precision of pigment concentrations for a specific water body. In contrast, continuous training and inter-calibration of the analysts is the single most important means of increasing the precision of the estimates phytoplankton density recorded as number of cells l^{-1} (Dromph et al., 2012).

Benthic vegetation (macroalgae and seagrasses)

An extensive bio-monitoring dataset, compiled from several macrophyte-based classification methods developed by different Member States, included data addressing spatial, temporal and human-induced sources of variability. This was used to identify, using uncertainty analysis, the major sources of uncertainty for coastal water classification (Mascaró et al., 2012b). The analyses were based on EQR datasets of either official or non-official bio-monitoring programmes of the different indices from which a dataset including sufficient temporal and spatial replication was available. The factors analysed included spatial scales of sampling (variability among zones within a site, among sites within a water body, variability among regions and variability among depths), the temporal scale of sampling (variability among years) and the human-associated source of error (variability between surveyors). These factors represent key sources of variability associated with the design and implementation of a bio-monitoring programme irrespective of BQE, and highlight how specific elements of a sampling design can influence the reliability and robustness of the ecological status classification of coastal water bodies. The study confirmed that the uncertainty analysis associated with the ecological quality classification is necessary to identify and quantify the most important factors that affect the risk

of misclassification (Mascaró et al., 2012b). When applied to macrophyte monitoring programmes, we note that the spatial scales of sampling were the main source of uncertainty, while temporal or human-induced errors seem to be less important. This then influences the design of sampling programmes used in managing ecosystem health. Moreover, the POMI method was assessed for its robustness in classifying the ecological status of Catalan coastal waters (Spain, W Mediterranean) (Bennett et al., 2011). A 7-year dataset, covering 30 sites along 500 km of the Catalan coastline was used to examine which version of POMI (that with 14 or 9 metrics) maximises precision in classifying the ecological status of meadows. Five factors (zones within a site, sites within a water body, depth, years and surveyors) that potentially generate classification uncertainty were examined in detail. Of these, depth was a major source of variability, while all the remaining spatial and temporal factors displayed low variability. POMI 9 matched POMI 14 in all factors, and could effectively replace it in future monitoring programmes (Bennett et al., 2011).

In addition, a dataset from 81 sites distributed throughout 28 water bodies from the coast of Catalonia, Balearic Islands and Croatia was used to determine the uncertainty components of the POMI metrics (Mascaró et al., 2012a), the uncertainty associated with each region and how these change according to the quality status of water bodies. Overall, spatial variability among sites (meadows) within water bodies had the greatest uncertainty that generated the greatest risk of misclassification across the three regions, within which the Balearic Islands had the lowest uncertainty, followed by Croatia and Catalonia. When water bodies classified in good/high quality were separated from those in moderate/poor status classes, it was found that the latter displayed higher levels of uncertainty than the former (Mascaró et al., 2012a).

Macroinvertebrates

Data from the Basque monitoring network were used for an uncertainty analysis of benthic assessment methods. The dataset included M-AMBI (multivariate AZTI's Marine Biotic Index, Muxika et al., 2007) values calculated from soft-bottom macroinvertebrate data from 1995 to 2011 and based on 683 datasets from 48 sampling stations, 4 coastal water bodies and 14 transitional water bodies. Uncertainty associated with spatial and temporal variability was assessed, focussing on between-station variance within water bodies and between-year variance within assessment period (samples were taken annually, but ecological status was assessed every 3 years). The total variance and variance components associated with each factor were estimated for all indices using a linear mixed effects model, treating 'Year' and 'Station' as random factors. Variance components were determined by calculating the proportion of the total variance explained by each individual factor. Then, the uncertainty in ecological status classification was estimated using WISER-BUGS (Clarke, 2012). The results showed that the main source of uncertainty is the between-station variance (97.6%), with interannual variability contributing only 2.4%.

Fish

Technical and monitoring design factors (gear, sampling season and survey protocol including sampling effort), and natural and anthropogenic pressures all affect the variability of fish metrics. The within-system variability is notably larger than the between-system variability (Pérez-Domínguez et al, 2012), and effect is probably due to natural factors and sampling bias. Therefore, the standardization of sampling methods and more robust fish metrics will increase the robustness of the use of the BQE fish in transitional waters.

Potential 'noise' factors (i.e. inherent variability) confounding biological quality metrics can be technical (i.e. those linked to the method of assessment including sampling effort) or natural (physicochemical and biological). Linear models were applied using fish metrics as response variables and a suite of covariates to explain the metric scores and identify the sources of variability affecting them (unpublished data. Available information at WISER Deliverable 4.4.2 part 2: http://www.wiser.eu/download/D4.4-2_part2.pdf). The resulting best models contained from 3 to 14 covariates but explained only a relatively small amount of the total variance (<40% of fish metric variability with a maximum 22% for lagoons and 40% for estuaries). The remaining variability was mainly within-estuary or lagoon and can probably be attributed, at least partly, to both a habitat effect that was not accounted for in the models

and to biological interactions influencing community structure.

Nevertheless, the models indicate that metrics showed a significant sensitivity to a range of technical and natural factors. There was a clear metric dependency in the selection of the best explanatory models which indicates that sources of inherent variability ('noise') vary according to the metric tested. This is reflected in the different combination of factors (covariates or fixed effects) comprising the models. The implication for assessments is that different factors might then confound the metric-pressure correlation (the 'signal' in the signal-to-noise ratio in the assessments) differently. Models showed that salinity class, depth, season, time of fishing (day vs. night) and year of fishing may influence the values of the fish metrics. Mixed models using type of system (estuaries or lagoons) as a random factor demonstrated that unexplained variance remains generally much higher within-systems than between-systems suggesting a higher importance of sources of variability acting at the within-system level.

The effect of sampling effort on fish metrics was not previously analysed but this factor will have an important effect on the variability of fish metrics. The analysis here showed that sampling effort is an important source of variability in fish metrics of the EFAI index, especially metrics dependent on number of species, which are common to several other fish-based indices. In turn, metrics based on percentages (derived from the abundance of marine migrants, estuarine residents, piscivorous species) showed a lower sensitivity to the increase in sampling effort, with values stabilizing after a fewer hauls compared to metrics based on species richness. The stabilization of metrics based on species richness varied between salinity zones, with an increasing number of hauls generally required at higher salinities. In contrast, salinity zone did not have that effect on metrics presented as percentage abundance for different guilds.

General discussion, lessons learnt and conclusions

The analyses described here have provided many lessons, particularly that metrics have to be developed for an area, a type of water body, an analytical method used and/or a BQE. As such they are not easily transposed to other areas and other methods, except in some cases (i.e. macroinvertebrates, Borja et al., 2011). It is of importance, but often neglected and thus emphasised here, that there is the need for the detailed testing of the statistical and ecological behaviour of the metrics. This includes not only their responses to biological changes but also the limits for the estimates and the use to which those estimates are put. In essence, this includes questions such as how valid are those estimates, are the values sufficient for management and will the managers and policy-makers understand the limitations of the indices.

As shown here, we can present the *outputs* of the analyses, i.e. the metric or multimetric values, but emphasise the *outcomes* as the use to which the metrics or multimetrics are put. Increasingly, scientists and managers for whom the indices are designed and used are concerned that, with any quality assessments linked to anthropogenic changes/causes, there may eventually be legal challenges. For example, an industry likely to be held responsible for the deterioration in the quality of an area could mount a legal challenge and even take the view that with such a plethora of indices and approaches (see Birk et al., 2012) they could find another type of assessment to show that their industry was not the cause of the deterioration.

Within all management tools such as the WFD and its counterparts elsewhere such as the Clean Water Act in the US, the assessment of aquatic systems refers to a single pressure-biological response relationship (e.g. eutrophication). However, in estuaries and lagoons anthropogenic activities such as dredging, land reclamation, harbour and industrial development, recreational and tourism development, have produced hydromorphological modifications; furthermore, the water quality of these environments is also affected by complex discharges of pollutants such as domestic and industrial effluents. Biological components have also been subject to human influence through commercial harvesting (e.g. fishing and aquaculture) of certain species as well as the introduction of alien species (either species which compete directly for resources or through the introduction of parasites and disease organisms). Therefore, aquatic systems are affected by multiple activities and human pressures to which they respond in a manner that is yet poorly understood. Such gap in the understanding of response metrics to multiple stressors makes the assessment of the ecological status of marine systems difficult.

Hence, a challenge in TraC waters is to underpin decision making, risk assessment and management of these systems under complex multiple stress background (e.g. combination of different pollutants, hydrodynamic alteration, invasive species, etc.). Research should not only enhance the understanding of multiple stressor interactions and accumulation (Crain et al., 2008; Thrush et al., 2008; Ban et al., 2010), but also pay special attention to species-stressor relationships and impacts on the ecological functioning, stability and resilience of these ecosystems. Hence, we emphasise that further work is required to assess cause-effect relationships (Adams, 2005) and use holistic approaches or tools to diagnose changes in the ecological status of the marine systems, in relation to multiple stress conditions. It is of note that some BQEs could be more sensitive to some of the pressures than to others (e.g. phytoplankton and macroalgae to eutrophication; seagrasses and fishes to hydromorphological changes or habitat loss; etc.). However, these responses could be masked depending on the dynamics of the water bodies (i.e. different residence time, flow regime, etc.), and, in management terms, by an inconsistent definition of types across Europe.

Moreover, the metrics used to assess the status must be able to discriminate between natural and anthropogenic changes; otherwise false conclusions maybe reached regarding the status being as the result of human pressures. Hence, the metric and multimetrics need a response linked to the magnitude of pressures and the signal–noise relationships; the latter indicating the anthropogenic signal and the inherent variability. Indicators and monitoring methods in which these are used are required to fulfil many attributes (Elliott, 2011) of which a consistent and reliable pressure-response relationship is the most important.

Some assessment methods, used within the WFD for different BQEs, have demonstrated their ability to differentiate anthropogenic stress from natural variability (see Section 'Identification of human pressure-response relationships' above). In general, we assume that human activities (drivers) lead to pressures, which cause impacts (DPSIR approach, Elliott (2002), Atkins et al. (2011)). However, if a water body has an inbuilt resilience to withstand and recover from stress, then we cannot make such an assumption. Hence, we need more investigations to allow discriminating the response of indices to human and natural stress within TraC waters, allowing for the inherent resilience of these systems.

The WFD may be regarded as having a relatively straightforward basis: *what ecology is in an area, what should be there without human pressures, if these two aspects differ then put in a programme of measures as long at these are not prohibitively expensive*. Despite this, the complexity of implementation (Hering et al., 2010) has resulted from the ecological complexity, the uncertainty in the system, the variability of areas and water body types and the desire by Member States to have their own methods of implementation (Birk et al., 2012).

The analyses here have also shown the importance of determining thresholds to pressure changes and boundaries across the scale (from bad to high) of ecological status. Of greatest importance, because it triggers the need for action (and again thus the need for expensive mediation and mitigation), is the *Good-Moderate Status* boundary. Hence, the sensitivity of the metric and multimetric indices to change and data manipulation and the degree of uncertainty in placing an area within a status class are important consideration for policy and management. The work described here is among the first to indicate, for the BQEs for TraC waters, the sensitivity of the indices.

The value of the analysis relies on the ability of metrics to integrate changes within the water body but also to separate hydromorphological changes from other stressors such as pollution by nutrients or persistent chemicals. While the latter may easily be remedied, through industrial or domestic sewage treatment or changing farming practices, the hydromorphology changes maybe the result of infrastructure (barrages, weirs, etc.), coastline modification or abstraction (Aubry and Elliott, 2006), the removal of the latter becoming more expensive (if at all possible) because of human occupation. Although the WFD requires Good Ecological Status to be achieved after a programme of measures, if the area is hydromorphologically modified then only Good Ecological Potential (GEP) has to be achieved (Borja & Elliott, 2007). WISER aimed to assess the meaning of GEP in this context but concluded that while GES was an entity to be achieved and measured using metrics, GEP was a misnomer in that logically an area can 'have a potential' if it cannot 'be in a potential' (K. Mazik, IECS, unpubl. Available information at WISER Deliverable 4.3.3: http://www.wiser.eu/results/deliverables/).

It is accepted that GEP refers to the ecology of an area once the hydromorphological pressures have been removed. Hence, we conclude that either the metrics tested as described here are not appropriate despite the wording of the WFD or that GEP is synonymous with the methods of defining a reference condition.

The present analysis has attempted to demonstrate the meaning of, and problems of defining, reference conditions, i.e. by finding a physical control area, hindcasting, predictive modelling or the use of the best professional judgement (BPJ) (e.g. Hering, et al., 2010; Teixeira et al., 2010; Borja et al., 2012). Although the use of pristine areas or the least disturbed areas is the preferred method in setting reference conditions, it is recognised that pristine transitional habitats are rare in Europe. An alternative to this method could be the use of historical reference conditions (hindcasting). However, the biggest problem here is determining what is the baseline date and to acknowledge that the baseline may shift in time (Duarte et al., 2009) especially with climate changes. There may also be temporal changes in indicator values due, for example, to climatic oscillations. Hence, hindcasting can be highly misleading unless: (i) the causes of cyclical oscillations are well established; (ii) reference conditions are available for the whole oscillation cycle and (iii) the positioning of both the reference condition and the considered year are strictly identical relative to the oscillation cycle (Borja et al., 2012). Modelling and BPJ can be used in setting reference conditions when adequate information is missing. However, although there are attempts at modelling the community structure of some BQEs (Valesini et al., 2010; Quataert et al., 2011), none of these are suitable in decision-support systems. This leaves BPJ as the most practical tool for setting reference conditions. The WFD wording indicates that BPJ should be used as a last resort when, in fact, the problems with the other methods of defining reference conditions suggest that it should be the first route. BPJ has been demonstrated to be very valuable in assessing the status of many geographical areas within USA and Europe (Teixeira et al., 2010), with a common set of criteria among different experts (Borja et al., 2012). Hence, we encourage exploring the use of these alternative options in setting reference conditions in aquatic systems. This can be done through a large panel of experts, with the aim to extract a common set of criteria for reference conditions within different European marine types.

The analysis of the indices here in relation to the pressure gradients also illustrates a questionable part of the WFD in that its indicators as required are dominated by structural aspects such as the number of species, the diversity or the percentage cover. We have argued elsewhere (Borja et al., 2010) that this is a 'deconstructuring structural approach', where the ecosystem is divided into its ecological parts (the BQE), these are tested individually for their status and then these assessments are combined in a 'one-out-all-out' approach (OOAO) to indicate the overall status. This raises concerns that firstly, as combining elements in an inappropriate way might not give an indication of overall status, and secondly, because of the OOAO principle, site assessment is determined by the quality of the lowest scoring element which, in some cases, may be the least suitable (Borja & Rodríguez, 2010). For example, by its nature the phytoplankton of a turbid estuary and the macrobenthos of a low-salinity area both will be naturally poor irrespective of the pressures. Furthermore, we emphasise that it is the functioning of the system, rather than only the structure, that should be important in reaching conclusions about the well-being of an area.

Hence, the nature of any water body, irrespective of the human pressures, influences the nature of the indices and their behaviour. This is particularly the case with regard to estuaries where, for example, the use of any index of anthropogenic stress is similar to that for natural stress. Indeed the latter even gives the same pattern as the former, what has been called the *Estuarine Quality Paradox* (Elliott & Quintino, 2007). This influences the ability of an index to indicate the levels of stress. Similarly, in the case of fishes in transitional waters and perhaps more than any of the other BQE, the nature of the community is heavily influenced by the characteristics of the catchment and the populations at sea (Elliott & Whitfield, 2011), hence being less dependent on the magnitude of pressures in the transitional water body.

The sensitivity and uncertainty analyses here have shown the importance of the amount of data and indeed rely on having an adequate number of replicates, thus defining the inherent variability (Clarke, 2012). The determination of sensitivity and uncertainty requires the users of the indices to know the inherent variability of the indices within the test area, for example the variance of the metric values around the average. Of course, having this information then

relies on the effort expended in determining the metric values (and in turn related to the costs of the assessment and monitoring). This indicates five other problems for the TraC areas: (i) in the inherent variability especially when trying to characterise a large water body with a few spot samples; (ii) the fact that pressures in these water bodies may act on a small area of the large water body rather than the whole water body (e.g. a sewage discharge into a large estuary); (iii) the large natural spatial variability of the water bodies; (iv) the fact that these areas perhaps have a larger range of pressures (operating both inside and outside the areas (Elliott & Whitfield, 2011) than other water bodies and (v) the cost of monitoring the TraC areas. Each of these has a large effect on the determination of the metrics and the way in which the metrics are used in water body management.

Acknowledgments This paper is a result of the project WISER (Water bodies in Europe: Integrative Systems to assess Ecological status and Recovery) funded by the European Union under the 7th Framework Programme, Theme 6 (Environment including Climate Change) (contract No. 226273), www.wiser.eu. We thank all WISER participants in Module 4 ('Indicators for marine and transitional waters') for their contributions. This is contribution number 592 of the Marine Research Division (AZTI-Tecnalia).

References

Adams, S. M., 2005. Assessing cause and effect of multiple stressors on marine systems. Marine Pollution Bulletin 51: 649–657.

Atkins, J. P., D. Burdon, M. Elliott & A. J. Gregory, 2011. Management of the marine environment: integrating ecosystem services and societal benefits with the DPSIR framework in a systems approach. Marine Pollution Bulletin 62(2): 215–226.

Aubry, A. & M. Elliott, 2006. The use of environmental integrative indicators to assess seabed disturbance in estuaries and coasts: application to the Humber Estuary, UK. Marine Pollution Bulletin 53: 175–185.

Ban, N. C., H. M. Alidina & J. A. Ardron, 2010. Cumulative impact mapping: advances, relevance and limitations to marine management and conservation, using Canada's Pacific waters as a case study. Marine Policy 34: 876–886.

Basset, A., E. Barbone, A. Borja, S. Brucet, M. Pinna, X. D. Quintana, S. Reizopoulou, I. Rosati & N. Simboura, 2012a. A benthic macroinvertebrate size spectra index for implementing the Water Framework Directive in coastal lagoons in Mediterranean and Black Sea ecoregions. Ecological Indicators 12: 72–83.

Basset, A., E. Barbone, A. Borja, M. Elliott, G. J. Lasinio, J. C. Marques, K. Mazik, I. Muxika, J. M. Neto, S. Reizopoulou, I. Rosati & H. Teixeira, 2012b. Natural variability and reference conditions: setting type-specific classification boundaries for lagoon macroinvertebrates in the Mediterranean and Black Seas. Hydrobiologia. doi:10.1007/s10750-012-1273-z.

Bennett, S., G. Roca, J. Romero & T. Alcoverro, 2011. Ecological status of seagrass ecosystems: An uncertainty analysis of the meadow classification based on the *Posidonia oceanica* multivariate index (POMI). Marine Pollution Bulletin 62: 1616–1621.

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

Borja, A. & D. M. Dauer, 2008. Assessing the environmental quality status in estuarine and coastal systems: comparing methodologies and indices. Ecological Indicators 8: 331–337.

Borja, A. & M. Elliott, 2007. What does 'Good Ecological Potential' mean, within the European Water Framework Directive? Marine Pollution Bulletin 54: 1559–1564.

Borja, A. & J. G. Rodríguez, 2010. Problems associated with the 'one-out, all-out' principle, when using multiple ecosystem components in assessing the ecological status of marine waters. Marine Pollution Bulletin 60(8): 1143–1146.

Borja, A., I. Muxika & J. G. Rodríguez, 2009a. Paradigmatic responses of marine benthic communities to different anthropogenic pressures, using M-AMBI, within the European Water Framework Directive. Marine Ecology 30: 214–227.

Borja, A., A. Miles, A. Occhipinti-Ambrogi & T. Berg, 2009b. Current status of macroinvertebrate methods used for assessing the quality of European marine waters: implementing the Water Framework Directive. Hydrobiologia 633: 181–196.

Borja, Á., M. Elliott, J. Carstensen, A.-S. Heiskanen & W. van de Bund, 2010. Marine management – towards an integrated implementation of the European Marine Strategy Framework and the Water Framework Directives. Marine Pollution Bulletin 60: 2175–2186.

Borja, A., E. Barbone, A. Basset, G. Borgersen, M. Brkljacic, M. Elliott, J. M. Garmendia, J. C. Marques, K. Mazik, I. Muxika, J. M. Neto, K. Norling, J. G. Rodríguez, I. Rosati, B. Rygg, H. Teixeira & A. Trayanova, 2011. Response of single benthic metrics and multi-metric methods to anthropogenic pressure gradients, in five distinct European coastal and transitional ecosystems. Marine Pollution Bulletin 62: 499–513.

Borja, Á., D. M. Dauer & A. Grémare, 2012. The importance of setting targets and reference conditions in assessing marine ecosystem quality. Ecological Indicators 12: 1–7.

Cabral, H. N., V. F. Fonseca, R. Gamito, C. I. Gonçalves, J. L. Costa, K. Erzini, J. Gonçalves, J. Martins, L. Leite, J. P. Andrade, S. Ramos, A. Bordalo, E. Amorim, J. M. Neto, J. C. Marques, J. E. Rebelo, C. Silva, N. Castro, P. R. Almeida, I. Domingos, L. S. Gordo & M. J. Costa, 2012. Ecological quality assessment of transitional waters based on fish assemblages in Portuguese estuaries: the Estuarine Fish Assessment Index (EFAI). Ecological Indicators 19: 144–153.

Cardoso, I., M. P. Pais, S. Henriques, L. C. da Fonseca & H. N. Cabral, 2011. Ecological quality assessment of small estuaries from the Portuguese coast based on fish assemblages indices. Marine Pollution Bulletin 62: 992–1001.

Carstensen, J., 2007. Statistical principles for ecological status classification of Water Framework Directive monitoring data. Marine Pollution Bulletin 55: 3–15.

Carstensen, J. & P. Henriksen, 2009. Phytoplankton biomass response to nitrogen inputs: a method for WFD boundary setting applied to Danish coastal waters. Hydrobiologia 633: 137–149.

Chainho, P., M. L. Chaves, J. L. Costa, M. J. Costa & D. M. Dauer, 2008. Use of multimetric indices to classify estuaries with different hydromorphological characteristics and different levels of human pressure. Marine Pollution Bulletin 56: 1128–1137.

Clarke, R. T., 2012. Estimating confidence of European WFD ecological status class and WISERBUGS (WISER Bioassessment Uncertainty Guidance Software). Hydrobiologia. doi:10.1007/s10750-012-1245-3

Clarke, R. T. & D. Hering, 2006. Errors and uncertainty in bioassessment methods – major results and conclusions from the STAR project and their application using STARBUGS. Hydrobiologia 566: 433–439.

Crain, C. M., K. Kroeker & B. S. Halpern, 2008. Interactive and cumulative effects of multiple human stressors in marine systems. Ecology Letters 11: 1304–1315.

Delpech, C., A. Courrat, S. Pasquaud, J. Lobry, O. Le Pape, D. Nicolas, P. Boët, M. Girardin & M. Lepage, 2010. Development of a fish-based index to assess the ecological quality of transitional waters: The case of French estuaries. Marine Pollution Bulletin 60: 908–918.

Díez, I., M. Bustamante, A. Santolaria, J. Tajadura, N. Muguerza, A. Borja, I. Muxika, J. I. Saiz-Salinas & J. M. Gorostiaga, 2012. Development of a tool for assessing the ecological quality status of intertidal coastal rocky assemblages, within Atlantic Iberian coasts. Ecological Indicators 12: 58–71.

Dromph, K. M., S. Agusti, A. Basset, J. Franco, P. Henriksen, J. Icely, S. Lehtinen, S. Moncheva, M. Revilla, L. Roselli & K. Sørensen, 2012. Sources of uncertainty in assessment of marine phytoplankton communities. Hydrobiologia, this issue.

Drouineau, H., J. Lobry, C. Delpech, M. Bouchoucha, S. Mahevas, A. Courrat, S. Pasquaud & M. Lepage, 2012. A Bayesian framework to objectively combine metrics when developing stressor specific multimetric indicator. Ecological Indicators 13: 314–321.

Duarte, C. M., D. J. Conley, J. Carstensen & M. Sánchez-Camacho, 2009. Return to *Neverland*: shifting baselines affect eutrophication restoration targets. Estuaries and Coasts 32: 29–36.

Elliott, M., 2002. The role of the DPSIR approach and conceptual models in marine environmental management: an example for offshore wind power. Marine Pollution Bulletin 44: iii–vii.

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

Elliott, M. & V. Quintino, 2007. The estuarine quality paradox, environmental homeostasis and the difficulty of detecting anthropogenic stress in naturally stressed areas. Marine Pollution Bulletin 54: 640–645.

Elliott, M. & A. Whitfield, 2011. Challenging paradigms in estuarine ecology and management. Estuarine, Coastal & Shelf Science 94: 306–314.

Elliott, M., A. K. Whitfield, I. C. Potter, S. J. M. Blaber, D. P. Cyrus, F. G. Nordlie & T. D. Harrison, 2007. The guild approach to categorizing estuarine fish assemblages: a global review. Fish and Fisheries 8: 241–268.

Garmendia, M., M. Revilla, J. Bald, J. Franco, A. Laza-Martínez, E. Orive, S. Seoane, V. Valencia & Á. Borja, 2011. Phytoplankton communities and biomass size structure (fractionated chlorophyll "a"), along trophic gradients of the Basque coast (northern Spain). Biogeochemistry 106: 243–263.

Garmendia, M., M. Revilla & L. Zarauz, 2012. Testing the usefulness of a simple automatic method for particles abundance and size determination to derive cost-effective phytoplankton indicators, in large monitoring networks. Hydrobiologia, this issue.

Guinda, X., J. A. Juanes, A. Puente & J. A. Revilla, 2008. Comparison of two methods for quality assessment of macroalgae assemblages, under different pollution types. Ecological Indicators 8: 743–753.

HELCOM, 2010. Towards a tool for quantifying anthropogenic pressures and potential impacts on the Baltic Sea marine environment: a background document on the method, data and testing of the Baltic Sea pressure and impact indices. Baltic Sea Environmental Proceedings, 125: 72 pp.

Henriksen, P., M. Revilla, S. Lehtinen, P. Kauppila, S. Kaitala, S. Agusti, J. Icely, A. Basset, S. Moncheva & K. Sørensen, 2011. Assessment of pigment data potential for multi-species and assemblage indices. WISER Deliverable 4.1-2 (http://www.wiser.eu/download/D4.1-2.pdf).

Hering, D., A. Borja, J. Carstensen, L. Carvalho, M. Elliott, C. K. Feld, A. S. Heiskanen, R. K. Johnson, J. Moe, D. Pont, A. L. Solheim & W. van de Bund, 2010. The European Water Framework Directive at the age of 10: A critical review of the achievements with recommendations for the future. Science of The Total Environment 408(19): 4007–4019.

Josefson, A. B., M. Blomqvist, J. L. S. Hansen, R. Rosenberg & B. Rygg, 2009. Assessment of marine benthic quality change in gradients of disturbance: Comparison of different Scandinavian multi-metric indices. Marine Pollution Bulletin 58: 1263–1277.

Juanes, J. A., X. Guinda, A. Puente & J. A. Revilla, 2008. Macroalgae, a suitable indicator of the ecological status of coastal rocky communities in the NE Atlantic. Ecological Indicators 8: 351–359.

Krause-Jensen, D. & M. B. Rasmussen, 2009. Historisk udbredelse af ålegræs i danske kystområder. In Danish with English summary. Danmarks Miljøundersøgelser – Scientific report from NERI no. 755, Aarhus University, Denmark: 38. http://www.dmu.dk/Pub/FR755.pdf

Krause-Jensen, D., S. Markager & T. Dalsgaard, 2011. Benthic and pelagic primary production in different nutrient regimes. Estuaries and Coasts. doi:10.1007/s12237-011-9443-1

López y Royo, C., C. Silvestri, M. Salivas-Decaux, G. Pergent & G. Casazza, 2009. Application of an angiosperm-based classification system (BiPo) to Mediterranean coastal waters: using spatial analysis and data on metal contamination of plants in identifying sources of pressure. Hydrobiologia 633: 169–179.

Lugoli, F., M. Garmendia, S. Lehtinen, P. Kauppila, S. Moncheva, M. Revilla, L. Roselli, N. Slabakova, V. Valencia, K. M. Dromph & A. Basset, 2012. Application of a new multi-metric phytoplankton index to the assessment of ecological status in marine and transitional waters. Ecological Indicators 23: 338–355.

Marbà, N., D. Krause-Jensen, T. Alcoverro, S. Birk, A. Pedersen, J. M. Neto, S. Orfanidis, J. M. Garmendia, I. Muxika, A. Borja, K. Dencheva & C. M. Duarte, 2012. Diversity of European seagrass indicators: Patterns within and across regions. Hydrobiologia, this issue.

Mascaró, O., S. Bennett, N. Marbà, V. Nikolic, J. Romero, C. M. Duarte & T. Alcoverro, 2012a. Uncertainty analysis along the ecological quality status of water bodies: The response of the *Posidonia oceanica* multivariate index (POMI) in three Mediterranean regions. Marine Pollution Bulletin 64: 926–931.

Mascaró, O., T. Alcoverro, K. Dencheva, I. Díez, J. M. Gorostiaga, D. Krause-Jensen, T. J. S. Balsby, N. Marbà, I. Muxika, J. M. Neto, V. Nikolic, S. Orfanidis, A. Pedersen, M. Perez, J. Romero, 2012b. Exploring the robustness of macrophyte-based classification methods to assess the ecological status of coastal and transitional ecosystems under the Water Framework Directive. Hydrobiologia, this issue.

McQuatters-Gollop, A., A. J. Gilbert, L. D. Mee, J. E. Vermaat, Y. Artioli, C. Humborg & F. Wulff, 2009. How well do ecosystem indicators communicate the effects of anthropogenic eutrophication? Estuarine, Coastal and Shelf Science 82: 583–596.

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

Neto, J. M., H. Teixeira, J. Patrício, A. Baeta, H. Veríssimo, R. Pinto & J. C. Marques, 2010. The response of estuarine macrobenthic communities to natural- and human-induced changes: dynamics and ecological quality. Estuaries and Coasts 33(6): 1327–1339.

Neto, J. M., R. Gaspar, L. Pereira & J. C. Marques, 2012. Marine Macroalgae Assessment Tool (MarMAT) for intertidal rocky shores. Quality assessment under the scope of the European Water Framework Directive. Ecological Indicators 19: 39–47.

Novoa, S., G. Chust, J. M. Froidefond, C. Petus, J. Franco, E. Orive, S. Seoane & A. Borja, 2012. Water quality monitoring in Basque coastal areas using local chlorophyll-a algorithm and MERIS images. Journal of Applied Remote Sensing 6: 1–27.

Orfanidis, S., P. Panayotidis & K. P. Ugland, 2011. Ecological Evaluation Index continuous formula (EEI-c) application: a step forward for functional groups, the formula and reference condition values. Mediterranean Marine Science 12: 199–231.

Pérez-Domínguez, R., S. Maci, A. Courrat, M. Lepage, A. Borja, A. Uriarte, J. M. Neto, H. Cabral, V. St.Raykov, A. Franco, M. C. Alvarez & M. Elliott, 2012. Fish-based indices to assess the ecological quality condition of estuaries and lagoons. Ecological Indicators 23: 34–45.

Quataert, P., P. Verschelde, J. Breine, G. Verbeke, E. Goetghebeur & F. Ollevier, 2011. A diagnostic modelling framework to construct indices of biotic integrity: A case study of fish in the Zeeschelde estuary (Belgium). Estuarine, Coastal and Shelf Science 94: 222–233.

Quintino, V., M. Elliott & A. M. Rodrigues, 2006. The derivation, performance and role of univariate and multivariate indicators of benthic change: case studies at differing spatial scales. Journal of Experimental Marine Biology and Ecology 330: 368–382.

Revilla, M., S. Agustí, M. Garmendia, P. Kauppila, S. Kaitala & S. Lehtinen, 2010 Report on identification of type-specific phytoplankton assemblages for three ecoregions. WISER Deliverable 4.1-1 (http://www.wiser.eu/download/D4.1-1.pdf).

Romero, J., B. Martínez-Crego, T. Alcoverro & M. Pérez, 2007. A multivariate index based on the seagrass *Posidonia oceanica* (POMI) to assess ecological status of coastal waters under the Water Framework Directive (WFD). Marine Pollution Bulletin 55: 196–204.

Sfriso, A. & C. Facca, 2011. Macrophytes in the anthropic constructions of the Venice littorals and their ecological assessment by an integration of the "CARLIT" index. Ecological Indicators 11: 772–781.

Stewart, G. B., C. F. Coles & A. S. Pullin, 2005. Applying evidence-based practice in conservation management: lessons from the first systematic review and dissemination projects. Biological Conservation 126: 270–278.

Teixeira, H., Á. Borja, S. B. Weisberg, J. Ananda Ranasinghe, D. B. Cadien, D. M. Dauer, J.-C. Dauvin, S. Degraer, R. J. Diaz, A. Grémare, I. Karakassis, R. J. Llansó, L. L. Lovell, J. C. Marques, D. E. Montagne, A. Occhipinti-Ambrogi, R. Rosenberg, R. Sardá, L. C. Schaffner & R. G. Velarde, 2010. Assessing coastal benthic macrofauna community condition using best professional judgement – Developing consensus across North America and Europe. Marine Pollution Bulletin 60: 589–600.

Thrush, S. F., J. E. Hewitt, C. W. Hickey & S. Kelly, 2008. Multiple stressor effects identified from species abundance distributions: Interactions between urban contaminants and species habitat relationships. Journal of Experimental Marine Biology and Ecology 366: 160–168.

Uriarte, A. & A. Borja, 2009. Assessing fish quality status in transitional waters, within the European Water Framework Directive: setting boundary classes and responding to anthropogenic pressures. Estuarine, Coastal and Shelf Science 82: 214–224.

Valesini, F. J., M. Hourston, M. D. Wildsmith, N. J. Coen & I. C. Potter, 2010. New quantitative approaches for classifying and predicting local-scale habitats in estuaries. Estuarine, Coastal and Shelf Science 86: 645–664.