Collaborative Project (large-scale integrating project)
Grant Agreement 226273
Theme 6: Environment (including Climate Change)
Duration: March 1st, 2009 – February 29th, 2012

WISER

Water bodies in Europe: Integrative Systems to assess Ecological status and Recovery

DELIVERABLE

Deliverable D4.4-1: Review of fish-based indices to assess ecological quality condition in transitional waters

Lead contractor: **Institute of Estuarine and Coastal Studies, University of Hull**

Contributors: **Rafael Pérez-Domínguez (UHULL), Stefano Maci (USALENTO), Anne Courrat (CEMAGREF), Angel Borja (AZTI), Joao Neto (IMAR), and Mike Elliott (UHULL)**

Due date of deliverable: **Month 6**

Actual submission date: **Month 15**

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013) Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Content

Non-technical summary

Estuaries (areas where rivers meet with the sea) and other coastal areas have been under the damaging influence of human habitation since historical times. Human alteration to once pristine habitats for wildlife has resulted in symptoms of degradation including alteration of watercourses, water quality problems and loss of aquatic fauna such as fish. It is important that these habitats and wildlife are protected from further damage, and that damaged areas are restored through effective management plans. One way to assess habitat conservation status is to analyse a sample of fish living in an estuary. The presence of any fish species indicates that the basic ecological requirements (food, shelter and reproduction) and a minimum water quality or habitat availability are being met. Likewise, finding species with stricter habitat requirements indicates better conservation status and hence less disturbed conditions for that area. Researchers worldwide have used this basic principle to define habitat integrity in monitoring programs. This work reviews sixteen published fish-based indices of estuarine habitat integrity and summarises common development strategies with the aim of improving fish-based monitoring tools in Europe. Most indices are computed from a number of independent fish diversity measures, presence-absence of key species and composition of functional guilds (i.e. group of fish that rely on the same quality attribute). All index developers invest a large amount of effort on the formulation of the reference values, that is the quality or conservation value given to pristine, undisturbed, condition or reference status. Comparatively less effort is invested in the evaluation of the relevance and precision of the assessment. Only half of the indices reviewed attempt any validation and these are limited to simple comparisons between fish-based quality measures and human disturbance level. As yet, there are no fish-based quality measures applicable to all areas in Europe -also known as common metrics. Widening of the geographical relevance will require better precision in the formulation of reference conditions and greater inclusion of functional guild metrics. Improvements are therefore needed in linking human disturbance (or pressure) intensity to new European-wide fish indices and to improve the confidence and robustness of fish-based environmental quality assessment.

1 Introduction

Given that many economic activities and urban areas are concentrated along the coast (Constanza et al. 1997; FAO Statistical Yearbook 2006; Halpern et al. 2008), estuaries and other transitional waters such as coastal lagoons are especially affected by anthropogenic pressures. This has resulted in symptoms of degradation including water quality impairment, salt water intrusion, loss of habitat, biological invasions, harmful phytoplankton blooms and reduction in biodiversity among others (McLusky and Elliott 2004 and references therein). Although environmental protection of aquatic ecosystems is openly acknowledged as a worldwide priority, the precise actions to ensure its conservation are still under a great deal of controversy. National and international initiatives and legislation (Water Quality Act (US Congress, Pub.L. 100-4, 1987) in the USA, OSPAR Convention, Water Framework Directive (WFD; 2000/60/EC) and Marine Strategy Framework Directive (MSFD; 2008/56/EC), Habitats and Species Directive (HD; 1992/43/EC) in the European Union (Apitz et al. 2006), and United Nations Convention on Law of the Sea (UNCLOS, 1982) or Convention of the Biological Diversity (CBD; UNESCO, 2000), among others at international level have come into force to ensure protection of aquatic biodiversity and sustainable use of derived ecosystem goods and services (Borja et al. 2008a). Indeed the most important aim of the management of transitional and coastal waters is to provide economic goods and services while at the same time protect and maintain (and where necessary restore) the ecological functioning of the systems. In order to achieve the fair and effective management plans envisaged at the core of these agreements it is imperative that the conservation and ecological status of aquatic ecosystems can be determined with adequate precision (USEPA 2000, Dale and Beyeler 2001). Similarly, assessments are necessary to implement and guide remedial actions in systems such as transitional waters that are greatly affected by human activities (McLusky and Elliott 2004).

Good ecological status for a given area can be described as the condition where there is a stable presence of the full complement of indigenous species linked by ecological processes in the appropriate physico-chemical environment (Karr 1981, Fairweather 1999). Systems that are closer to this ideal integrity scenario are considered in better conservation status that those that deviate from it. Implicit in the definition of good ecological status is the presence of processes that cause deviation from this idealized ecosystem structure and interfere with the natural ecosystem functioning causing degradation. Assessing the course of degradation in estuarine systems is especially complex because of the large environmental gradients and variability associated with estuarine habitats, hence increasing the ‘noise’ and making the detecting of any ‘signal’ of change difficult (Williams and Zedler 1999; Dale and Beyeler 2001). The difficulty of detecting anthropogenic stress in areas of high natural stress and variability has been termed the estuarine quality paradox (Elliott and Quintino 2007; Dauvin, 2007; Dauvin and Ruellet 2009).

Loss of ecological integrity due to human disturbances is usually assessed on a relative scale where condition or change is compared to the condition expected at undisturbed (or reference) sites where only natural processes operate (Karr 1981, USEPA 2000, Jackson et al. 2000, Borja and Elliott, 2007; Muxika et al. 2007). The WFD requires areas to be compared against a reference condition in which the latter are derived by a physical control, hindcasting, numerical predictive models or, if these are not possible, expert judgement. The necessary reference can seldom be derived from direct assessments as most estuaries in the developed world have been, and still are, largely impacted by human activities. Even where pristine areas within estuaries are still available it is difficult to define expected conditions (i.e. a reference community) for all areas and occasions due to inherent natural variability and uncertainty on biota–environment interactions (Gorshkov et al. 2004, Irz et al. 2008).

In practice, assessments of ecological integrity are normally based on quality measures that are known to correlate with anthropogenic pressures (see Figure 1 for a conceptual diagram). It is therefore assumed in the assessments that intense pressure will lead to stress and impacts on the biological community in a dose response manner (Jordan and Vaas 2000, Gray and Elliott 2009). A priori, any physical, chemical and biological variables (i.e. metrics thereafter) can be used to produce the experimental evidence necessary to assess integrity status (Weisberg et al. 1997; USEPA 2000; Niemi et al. 2004; Gray and Elliott 2009). However, due to the multiple and often unknown mechanisms linking anthropogenic pressures and ecological responses it is difficult to account for all physical and chemical factors leading to the loss of ecological integrity (Karr and Dudley 1981, Fausch et al. 1990, Oberdorff and Hughes 1992; Karr and Chu 1997). Therefore current approaches are increasingly based on measures of biotic integrity, typically community structure (Niemi et al. 2004, Noges et al. 2009). It is of note that most of the measures required by both the EU WFD, in defining good ecological status, and the EU HD, in measuring -Favourable Conservation Status-, are based on structural attributes (i.e. measures of a component amount such as percentage cover or abundance of a species). While it is assumed that these structural elements are a surrogate for functional metrics, by definition involving rate processes, this may not always be the case causing problems in bioassessments (Seegert 2000, de Jonge et al. 2006).

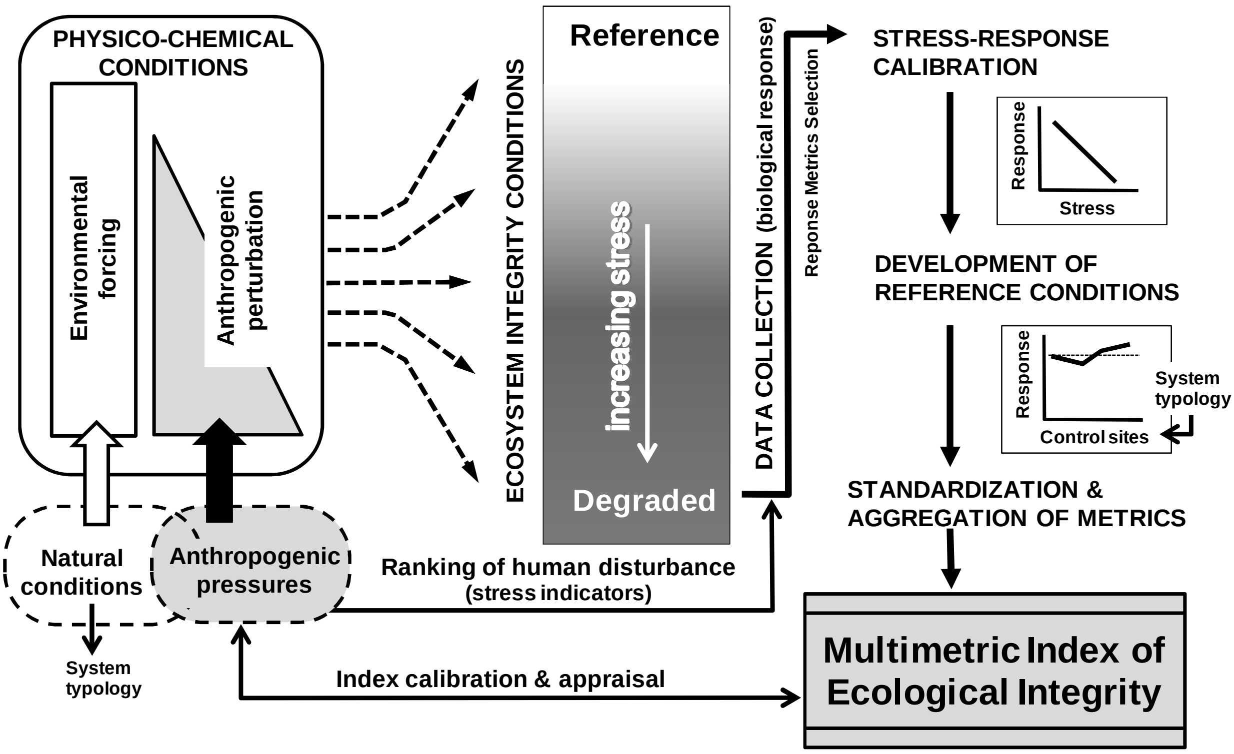

Figure 1. Conceptual diagram illustrating linkages between ecosystem integrity, sources of stress and development of multimetric indices of ecological integrity. The diverging dashed arrows indicate the multiplicity of stressors operating at different scales within estuarine systems.

It is accepted that more holistic assessments are possible when different metrics covering a wide spectrum of responsive ecological and community features are combined (Karr 1981, Niemi et al. 2004, Hering et al. 2006, Borja and Dauer 2008). Despite predicted advantages of indices based on several biological quality elements to convey multiple quality measures in one relevant quality score, their formulation and use is not simple. There are many areas where improvements are necessary such as standardization of sampling and analysis protocols, sensitivity and behaviour of assessment metrics, natural variability of reference communities, relevance of outcomes, and overall validation of indices (Karr and Chu 1997, Fairweather 1999, Dale and Beyeler 2001, Niemi et al. 2004, Hering et al. 2006, Noges et al. 2009).

In the context of the WFD, fish is one of the biological elements included in the quality analysis of Europe's freshwater and transitional water systems (see Borja (2005), for a description of the different elements). The need for robust fish-based bioassessments in transitional waters has resulted in the development of new fish indices specifically tailored to estuarine systems and lagoons. This paper summarizes these developments in fish-based estuarine quality assessments and outlines common approaches in the development of fish indices of biotic integrity. It also aims to review the relevance of current indices and metrics to management needs and to propose future avenues of research to improve assessment of ecological quality status using fish communities in transitional water ecosystems.

2 Fish-based ecological monitoring

When first proposed by Karr (1981) multimetric fish indices pioneering a change in environmental quality assessment from traditional indicators associated with water quality and toxic substances (physico-chemical variables) to biological elements based on community parameters. Since this early work fish communities have been used effectively to convey information of the conservation and ecological quality status of aquatic ecosystems (Roset et al. 2007). The advantages and disadvantages of fish as a biological quality element (BQE, as defined under the WFD) have been extensively discussed in the literature (Karr 1981, Karr and Chu 1997, Whitfield and Elliott 2002, Elliott and Hemingway 2002, Harrison and Whitfield 2004, Breine et al. 2007, 2010). The most useful features of fish are their proven sensitivity to habitats quality loss, their occurrence in all aquatic systems and areas, high level integration of ecosystem functioning, cost effective means of assessment including training on taxonomical competence, their high public value and the direct interpretation of fish community quality condition. The less favourable features include bias associated with sampling gear, uncertain or variable association with pressures and more importantly high mobility and marked seasonal variation (Fausch et al. 1990; Harrison and Whitfield 2004). In the case of transitional waters, fish assemblages are dependent on conditions and pressures both within the riverine catchment and the adjacent marine area as well as those affecting the connectivity in the system leading to complex interactions both in space and time (Elliott and Hemingway 2002). To integrate this wealth of factors and interactions almost all indices now available are based on an aggregated set of metrics generally referred as multimetric fish indices (see Table 1 for the full list of fish indices reviewed).

*Table 1. List of fish indices for transitional waters quality assessment. The number of metrics in the index is given between parentheses. The indices are ranked by year of publication using the earliest appearance in the literature. When indices have been presented in different publications only the more relevant references to the development of the index are presented. * restricted to the ecological quality assessment of estuarine nursery grounds **independent indices for each zone*

Tool name	Abbreviation	Area of use	Type	WFD	References
Index of biotic integrity	IBI[1]	Transitional (Louisiana, USA)	Multimetric (13)	NO	Thompson and Fitzhugh 1986
Community degradation index	CDI	Transitional (South Africa)	Single metric	NO	Ramm, 1988
Index of biotic integrity	IBI[2]	Transitional (Maryland, USA)	Multimetric (9)	NO	Jordan and Vaas 1990, Vaas and Jordan 1991
Biological health index	BHI	Transitional (South Africa)	Single metric	NO	Cooper et al., 1994
Estuarine Biotic	EBI[1]	Transitional (Massachusetts,	Multimetric (12)	NO	Chun et al. 1996, Deegan et al. 1997

Integrity Index		USA)			
Recruitment Index	RI	South Africa	Single metric *	NO	Quinn et al. 1999
Index of biotic integrity	IBI[3]	Transitional (Nagarranset bay, USA)	Multimetric (6)	NO	Meng et al. 2002
AZTI's Fish Index	AFI	Transitional (Basque Country, Spain)	Multimetric (9)	YES	Borja et al. 2004, Uriarte and Borja 2009
Estuarine fish community index	EFCI	Transitional (South Africa)	Multimetric (14)	NO	Harrison and Whitfield, 2004 and 2006
WFD Fish Index for Transitional waters	FITW	Transitional (Holland)	Multimetric (10)	YES	Jager. and Kranenbarg 2004
Fish-based Estuarine Biotic Index	EBI[2]	Transitional (Brackish Scheldt, Belgium)	Multimetric (5)	YES	Breine et al. 2007
Transitional fish classification index	TFCI	Transitional (United Kingdom)	Multimetric (10)	YES	Coates et al., 2007
MJ nursery index	MJNI	Transitional (France)	Non aggregating multimetric (3) *	NO	Courrat et al. 2009
Habitat Fish Index	HFI	Transitional and coastal (Venice Lagoon, Italy)	Multimetric (16)	YES	Franco et al. 2009
Zone-specific Fish-based Estuarine Biotic Index	Z-EBI	Transitional (Brackish and freshwater Scheldt, Belgium)	Multimetric (6) **	YES	Breine et al. 2010
French Multimetric Fish Index	f-MFI (ELFI, *Estuarine and Lagoon Fish Index*)	Transitional (Atlantic and Channel coast (France)	Multimetric (4)	YES	Delpech et al. 2010

Multimetric indices are constructed from an array of fish ecological attributes and therefore considered to be superior to single-metric assessments; in particular they should have wider sensitivity to complex and cumulative pressures and greater relevance to different ecological regions (Karr 1981, Fausch et al. 1990). Despite this, single-metric tools have been proposed for estuaries (Ramm 1988, Cooper et al. 1994). These tools provide a comparative score based on an analysis of similarity between the control community and the actual community. Other multivariate techniques such as ordination or correlation analysis have been also proposed (Fausch et al. 1990). These are simple to compute using available statistical packages, are data driven, can integrate fish functional information and are effective at condensing taxonomic information to a few main ordination axes, or even down to a numerical value of great value at determining recovery or degradation trajectories (Boyle et al. 1984, Fausch et al. 1990, Ramm 1990, Whitfield and Elliott 2002). However, the interpretation of outputs may be difficult to less experienced users and most importantly the outputs may indicate meaningful relationships from random variation when used without ecological knowledge (Fausch et al. 1990). Therefore, these techniques may be more useful as exploratory tools during the evaluation of more complex approaches rather than final assessment tools.

3 Development of multimetric indices

Although not always considered initially, but essential in practice, is the definition of the requirements and performance goals for the final tool. For example, in Europe, indices must be WFD compliant. The WFD explicitly requires assessments based on composition and abundance information of transitional fish fauna. They should be based on current ecological understanding, be biologically meaningful and ideally minimize the uncertainty associated with the classification of ecological status (Noges et al. 2009, Borja et al. 2009a). Moreover, the indices should be readily understandable by biologists, stakeholders, water resource managers and the public.

The procedure to develop indices of fish assemblage integrity follows a more or less complex sequence (Hughes et al. 1998) that starts with an initial appraisal of anthropogenic pressures (Table 2), which is common to any biological quality element (Borja and Dauer, 2008b). Most transitional indices reviewed have been developed using a five step procedure: (i) assessment of the pressure; (ii) fish sampling strategy; (iii) selection of metrics; (iv) formulation of indices, and (v) final appraisal (Table 3).

*Table 2. Type of common human pressures affecting biological integrity in estuarine system. The table highlights those pressure types with proxies used for pre-classification and scoring of fish metrics. * Invasive species are included as metrics in some indices*

Pressure Class	Pressure type	Included
Hydromorphological	channelling and dredging	YES
	land-reclamation and coastal defence	YES
	port and navigation infrastructure	YES
	flow manipulations (dams, weirs, sluices)	YES
	underwater structures (wrecks, piers, armouring)	NO
Chemical and Physical	nutrient discharge	YES
	waste disposal	YES
	water pollution	YES
	sediment pollution	YES
	noise	NO
	sediment load	NO
	heat exchange	NO
Biological	invasive species	NO*
	resource gathering (fishing, game, natural crops)	NO
	behavioural interference and disturbance	NO

Table 3. Steps and tasks included in the development of multimetric fish indices as discussed in this review.

Developmental sequence (Tasks)
1- Review of pressures and index requirements
Determine pressure field, index scope and quality targets
Selection of ecologically relevant metric types according to relevant pressures
Classify habitat typology and fish functional guilds

2- Selection of sampling methods
    Sampling tools, sampling standardisation and sample analysis
    Indexing period and sampling sites
    Effort level. Precision and accuracy in the assessment
3- Metrics selection and evaluation
    Determine responsiveness to pressures
    Metric redundancy assessment
    Define metric thresholds and scoring system
    Development of reference conditions
    Optimization of sampling methods
4- Index scoring method and ecological status class
    Metric combination rules
    Define ecological quality ratio and thresholds
    Assignment to ecological status class
5- Index calibration and appraisal
    Misclassification rate, sensitivity analysis
    Global uncertainty assessment
    Presentation format and value to end-user

3.1 Review of pressures and index requirements

All estuarine fish indices are built on the assumption of a variable anthropogenic pressure acting upon a normal background of natural variability (Figure 1) (see Table 2 for a summary of pressures). The effects on fish populations should then scale according to the intensity of the disturbance in an approximate dose-response manner and be specific for the pressure type (USEPA 2000). However, this signal is confounded by natural stressors resulting from environmental variability within estuaries in which the natural stress may have the same type of response as the anthropogenic stress (Elliott and Quintino 2007).

In the reviewed literature there are two basic approaches to the definition of biotic quality indices in the context of variable human pressures. The first and simplest approach is to classify undisturbed or, more commonly, least disturbed sites according to the size and diversity of the fish community present across the area being assessed (USEPA 2000, Harrison and Whitfield 2006, Coates et al. 2007). This method does not require previous ranking of survey sites in quality classes (preclassification) or any previous knowledge of the expected reference community. The strength of the approach is its simplicity. It relies in the assumptions that sites are exposed to a varied degree of human pressures (including low or no pressure) and that all sites respond equally to disturbance. When these assumptions are not justified, especially when there is an insufficient representation of reference sites (truly pristine or defined by conservation goals), alternative approaches are needed.

The second common approach uses a preclassification of sites according to hydromorphological, chemical and physical disturbances using a suite of proxies for these anthropogenic pressures (Jordan and Vaas 1990, Deegan et al. 1997, Harrison and Whitfield 2004, Breine et al. 2007, 2010, Delpech et al. 2010). The habitat preclassification is done on a simple rating scale using expert judgment to define expected quality thresholds and is solely intended to provide a reference to evaluate relevant metrics rather than a complete assessment (Breine et al. 2010, Delpech et al., 2010). Water quality, dissolved oxygen, pollutants, channelling, dredging, shore stabilization, intertidal area integrity, land claim, benthic integrity, population density, industrial development have all been used. This approach importantly allows the evaluation of sensitivity of metrics to human pressures and the early elimination of redundant metrics in the indices. However, the choice of pressure proxies and its combination into quality scores can be very subjective and relies heavily on expert judgment and is highly dependant on pressure data availability (Chun et al. 1996, Deegan et al. 1997, Borja et al. 2004, 2009a, Franco et al. 2009, Breine et al. 2010).

For the habitat preclassification of pressures approach to be effective it is necessary to consider the natural make up and variability of the systems. That is it is necessary to classify by habitat typology or whatever unit of assessment is necessary to fulfil the aims of the index (typically the water body as defined in the WFD) (Deegan et al. 1997, Borja et al. 2004, Franco et al. 2009, Breine et al 2010).

Whichever the approach used there is an early need to identify the best and more sensitive fish response metric types according to the pressures acting upon the system. Hering et al. (2006) described a 4-step procedure for the selection of the core metrics for an aggregated index; 1- establishment of a list of possible metrics, 2- metric calculation and elimination of unreliable metrics (those with too narrow range of values or too many outliers, 3- testing the correlation between metrics and some stressor gradient (optional), and 4- removing redundant metrics. Although precise evaluation of the metrics (steps 2 to 4) is normally done later in the index development (see section 3.3.) the review of pressure gradients early in the planning of the index have been used to select a pool of fish metrics whose behaviour in such pressure gradients can be predicted (Hering et al. 2006, Coates et al. 2007, Breine et al. 2010).

3.2 Selection of sampling methods

There are important logistical and cost considerations that affect the method of sampling and the degree of effort in ecological assessments. The fact that any assessment will only be as good as the data used to derive the metrics is recognized explicitly or implicitly in all the indices reviewed. Larger datasets may be better suited for a more robust and general assessment but they require increasing resources. Sample sizes from as little as 36 samples (Meng et al. 2002) to large long-term datasets containing over 1000 sampling events (Coates el at. 2007, Breine et al. 2010). It is important to note that none of the reviewed papers has tested the minimum sampling effort required to get accurate and reliable metric values. Although larger datasets may appear as a safe option, the inherent variability in the samples may compromise any assessment if we can not obtain accurate values of occurrence or densities. Therefore, it is important to know the relevance of a particular metric outcome in the context of the statistical confidence or power associated to a particular sampling effort. The level of effort that is “sufficient” should be considered earlier in the development of future indices as this certainly have important consequences for the selection of robust metrics and assessment of index reliability and uncertainty.

Catch data have been gathered with a suite of gears such as seine nets (Thompson and Fitzhugh 1986, Meng et al. 2002, Harrison and Whitfield 2004, 2006, Coates et al. 2007, Franco et al. 2009), beam trawls (Borja et al. 2004, Coates et al. 2007, Courrat et al. 2009, Uriarte and Borja, 2009, Delpech et al. 2010), otter trawls (Thompson and Fitzhugh 1986, Chun et al. 1996, Deegan et al. 1997), gillnets (Thompson and Fitzhugh 1986, Harrison and Whitfield 2004, 2006, Coates et al. 2007), fyke nets (Coates et al. 2007, Breine et al. 2007, 2010), anchor nets (Jager and Kranenbarg 2004), and visual diver censuses (Angel Perez-Ruzafa pers. communication). Non-capture methods for fish assessment such as acoustic and visual or video-based survey techniques have not been considered yet although there is current research devoted to the development of these techniques (Courrat A & Perez-Dominguez R unpublished data). Properly calibrated, acoustic methods, can provide direct information on functional groups such as pelagic fish which could be a valuable fish metric (Knudsen et al. 2009). Visual scuba diving censuses or video surveillance have a high potential as long as there are favourable conditions (i.e. water clarity). Importantly, the non intrusive survey methods are often well-received as they promote welfare of fish and provide a permanent visual record that can be assessed in several ways.

In addition to the choice of gear, the index time period and the spatial extent of the methods are considered in the design of all indices reviewed. In compliance with the WFD requirements, all newly developed indices aim to develop a tool for the assessment at the water body level, with ecological status judged on a 3-year cycle. However, the development and validation of the indices is often based on different time and spatial scales more appropriate to biological rhythms and ecotypes. Overall, there is good agreement about the index period which is often chosen to coincide with the expected fish diversity and density maxima. In temperate estuaries, with marked seasonal recruitment, this period extends in the northern hemisphere from early summer to autumn with small local deviations due to the composition of the resident fish community and geographical regions. Much less overall information is found on the randomization of sampling sites which is generally not discussed. Most examples, however, use salinity class in the design of the surveys (stratified sampling with a salinity class as a strata) and they attempt to find representative sampling locations based on estuarine typology and expert knowledge.

3.3 Metrics selection & evaluation

USEPA (2000) define a metric as ‘a measurable factor that represents various aspects of biological assemblage, structure, function or other community component’. Metrics and indices based on biological elements such as indicator species, species richness or guild composition are currently favoured among ecological indicators as they provide a direct assessment of ecological integrity as a whole (Bain et al. 2000, McLusky and Elliott 2004). Better assessments are possible when the metrics and indices reflect functional attributes of the ecosystems (Karr 1981). The latter is achievable in fish-based assessment when there is information on the species functional requirement or more commonly termed functional guilds (Elliott et al. 2007, Franco et al. 2008). Essential functional requirements for estuarine fish are grouped at three main guild levels: trophic, habitat and reproduction (Elliott and Dewailly1995; Elliott et al. 2007). This review of indices indicates that the guild approach is followed by all current multimetric indices. By grouping fish in common guilds it is possible to translate taxonomic information directly into proxies of functional features of ecosystems which in turn are valuable in evaluating ecological integrity. Moreover, this approach allows building indicators that do not depend on biogeography of species and hence that can be used on a large geographic scale. Within each guild, metrics are either formulated based on number of individuals, or number of species or relative abundance which together with overall assemblage metrics (diversity, dominance, etc.) comprise a large list of candidate metrics for inclusion in the indices (Noble et al. 2007).

3.3.1 Selection of metrics

Most metrics in use are measures of species diversity (Table 4, species richness and diversity group). Increased diversity is generally assumed to indicate higher quality (Gray 1989). Despite this, it is notable that only approximately half of the metrics in this group summarize biological diversity as a whole. The largest family of metrics in this group involves indicator species which are perceived as important due to known dependencies on precise estuarine quality features (e.g. European smelt, Osmerus eperlanus and well-oxygenated waters) (Jager and Kranenbarg 2004, Breine et al. 2007, 2010). Indicator species metrics using individual taxa are relevant only to the geographical range related to the distribution of the indicator species. For example O. eperlanus is profusely used in North Sea estuaries, but not used in southern or Mediterranean estuaries. This suggests the current prevalence of indices developed within a single estuary or uniform biogeographical zone and indicates the spatially restricted relevance of most current indices.

Habitat use guild (Table 4) is the next relevant group of metrics with most indices containing the estuarine resident guild. This is followed by guilds providing information on habitat preference which are equivalent to indicator species but grouped into a functional guild such as benthic species. This approach gives relevance to the metrics beyond single-species distribution ranges, one of the main advantages of the guild approach. Most of these habitat metrics are expected to decrease together with decreasing estuarine habitat quality although some may have a reverse trend (i.e. benthic habitat destruction may lead to a decrease in benthic metrics but to an increases in pelagic metrics). As the diversity of the system is related to the number of niches available for occupation by species then a more complex or larger transitional water should give higher values for these habitat guilds (Wootton 2008).

With regard to the fish feeding guild (Table 4), the most relevant cluster of metrics is built around number and diversity of predatory (carnivorous and/or piscivorous) fish followed by benthic feeding fish. As in the previous group, the impact of pressures is metric-specific with metrics for specialized feeders generally decreasing with disturbance while metrics associated with omnivores generally increasing with disturbance as long as disturbances remains under a certain level.

Abundance and body condition (i.e. health status) are grouped together as both provide a quantitative measure of fitness (Table 4). Indicator individuals have the largest number of metrics in the group. Abundance is often negatively correlated with impacted systems which have lower abundances overall whereas the number of individuals in poor condition due to disease or malformations is assumed to be correlated with increased disturbance.

The final important group considered in this review includes metrics associated with the use of estuarine habitats by juvenile fish (nursery function, Table 4); these are often marine species that use estuarine habitats on a seasonal basis and when young. The disturbance of the nursery function is considered a direct effect of habitat degradation in estuaries. The relative large number of nursery function metrics found among the indices confirms the overall importance given to this functional guild in most indices (Figure 2).

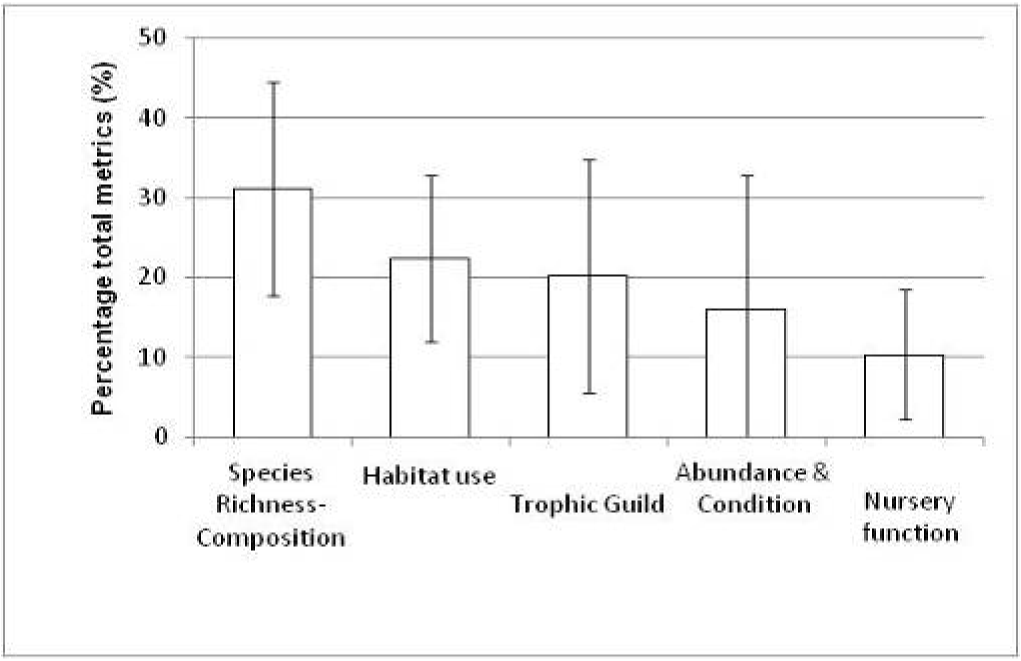

Figure 2. Figure 2. Relative importance of metric distribution across ecological attributes in multimetric indices for transitional waters. Mean (± stdev) MJNI and RI are excluded in the analysis as they are exclusively nursery quality indices.Nursery condition is often identified by Habitat use gulit but it is presented separately due to the recognized value of estuaries as fish nurseried.

3.3.2 Evaluation of metrics

It is assumed that multimetric indices perform better than the individual metrics in predicting habitat quality, dynamic range, precision and robustness. This will only be true when the index contains a balanced and complementary set of metrics that respond to a range of habitat quality degradation parameters (Herring et al. 2006) and are less or not affected by natural variability, i.e. they are specifically sensitive to anthropogenic disturbance. The simplest and more common approach to achieve this goal chooses metrics based on previous successful indices and expected ecological responses to degradation. This method has proved successful and it is the only option when human pressure data are not available to assess the relevance of the metrics (Thompson and Fitzhugh 1986, Meng et al. 2002, Harrison and Whitfield 2004, Borja et al. 2004, Jager and Kranenbarg 2004, Coates et al. 2007, Franco et al. 2009). When pressure data are available and the sampled sites contain a sufficiently large range of pressure, the sensitivity of the metrics can be assessed by box and whisker plots (Jordan and Vaas 1990), regression analysis (Breine et al. 2007, 2010, Delpech et al. 2010), discriminant analysis (Meng et al. 2002, Breine et al. 2007) ANOVA (Deegan et al. 1997), Principal Component Analysis (Breine et al. 2010), or a combination of methods. Stepwise discriminant analysis is used to assess metric relevance and identifies the best combination of metrics eliminating redundancies and metrics with large uncertainty (Jordan and Vaas 1990, Vaas and Jordan 1991, Meng et al. 2002, Breine et at 2007). Furthermore, redundancy screening of metrics is done in some indices after the sensitivity assessment by calculating correlation coefficients between responsive metrics and rejecting one metric from each highly correlated pair (Breine et al. 2010, Delpech et al. 2010).

Methods using formal statistical approaches in the selection process often resulted in more stringent conditions for metric inclusion and some authors use expert judgment to decide on the inclusion of metrics that otherwise will not meet the requirements (Chun et al. 1996, Deegan et al. 1997, Breine et al. 2010). Likewise, the initial pool of metrics included for statistical screening is entirely decided on expert knowledge which requires a sound ecological knowledge of the expected responses. The number and choice of metrics varies with most multimetric indices built around a pool of 9-10 metrics with a maximum of 16 (Franco et al. 2009) and a minimum of 4 (Delpech et al. 2010). There is a tendency for fewer metrics in those indices where metric sensitivity and redundancy has been formally assessed against pressure scores. To overcome some modelling problems in the linkage between pressures and habitat quality the use of fuzzy logic has been trialled in freshwater systems (Ocampo-Duque et al. 2006) and appears to be a viable alternative to add rigour to a decision process based on expert knowledge.

3.3.3 Reference conditions and metric scoring system

In transitional waters there is an increased natural stress and a high inherent variability which introduces natural sources of noise into the assessment. Therefore the reference conditions, and the thresholds between the ecological status, need to be defined using confidence intervals (Borja el al 2009). Increased precision in the assessment will therefore be dependent upon the robustness of the estimations of the reference condition. To reduce the natural noise of spatial and temporal variability, specific reference conditions have been chosen by season, gear type, salinity regime, estuarine system and estuary typology. Therefore the literature distinguishes between metric-, season-, gear-, salinity class-, estuary- and ecotype-specific reference conditions as relevant to the data structure and analysis.

All index developers invest a large amount of effort on the formulation of the reference values. The WFD advocates 4 possible means of defining the reference value (Annex II, 1.3(iii), in the WFD). The more direct approach is to compare with (i) an existing undisturbed site or a site with only very minor disturbance, hence making the further assumption that any changes observed are natural. This is only possible when these exist within the same typology of the system under assessment. None of the indices is able to build a reference condition using this approach in isolation. Instead they explore the alternative 3 approaches which are: (ii) the use of historical data (i.e. hindcasting); (iii) the use of predictive models or, when there is no alternative according to the WFD, (iv) expert judgement. All of these approaches have their advocates in the WFD community but more frequently, a combination of methods is used to calculate by any means whether the measured biological community has deviated from what is expected under good conditions.

In practice, the reference community (whether pristine or defined at any other level) is often derived from the assessment dataset itself using the top scoring samples after ranking sites according to human disturbance (pressure-response models) (Jordan and Vaas 1990, Chum et al. 1996, Deegan et al. 1997, Meng et al. 2002, Borja et al. 2004, Uriarte and Borja 2009, Breine et al. 2007, 2010, Delpech et al. 2010) An alternative is to use a simplified model where no pressure data are available by including a large sample size and assuming that less impacted sites are present (Harrison and Whitfield 2004, 2006, Coates et al. 2007. Franco et al. 2009). In both cases the reference community is then derived from the top scoring samples (i.e. 90% percentile) assuming that increased habitat integrity correlates with the distribution of the top scores. Although all methods provide the necessary reference, it is increasingly accepted that there is the need to incorporate some degree of historical (before human impacts) or expert knowledge of the systems (i.e. indicator or conservation species).

3.4 Index scoring method and ecological status class

Once the reference is set, each metric is scored independently depending on where their value lies with respect to the reference. This is done as a relative score in the form of a ratio or directly by setting threshold values that define the quality classes. This produces a single value of ecological quality which is based on a single metric or an aggregated value produced from multimetric indices. All indices use the sum, the average or the weighted product of all individual metric scores. Some indices also report the scores of the individual metrics using radar plots (Jager and Kranenbarg 2004, Breine et al. 2010). Weighting of the metric scores by perceived relevance is only reported in one index (Breine et al. 2010) which uses 1 or 0 weighting depending whether the metric is relevant or not which resulted in de facto formulation of three different indices. Other approaches are possible such as weighting the metrics according to robustness (Courrat A unpublished data).

Several quality scoring systems are available. In the simplest case a sliding scale is used to rate sites with discrete scores (1 to 5) depending on whether their raw value deviates greatly from (score 1), deviates somewhat from (score 3) or is comparable to the reference value (score 5) (Jordan et al. 1990, Thompson and Fitzhugh 1986, Deegan et al. 1997, Jager and Kranenbarg 2004, Harrison and Whitfield 2006, Coates et al. 2007, Uriarte and Borja 2009, Franco et al. 2009). Semi-quantitative scales with intervening scores (2 and 4) have also been used (Coates et al. 2007, Delpech et al. 2010). The number and cut off point for the scores thresholds varies among indices. All WFD compliant indices use a 5-band scoring system although some indices eliminate the top quality bands when these quality statuses are not present in the dataset (Breine et al. 2007, 2010, Delpech et al. 2010).

3.5 Index calibration and appraisal

The relevance and precision of the assessment is obtained by similar modelling exercises as those used to determine the sensitivity of the single metrics. The intention is to gauge whether the new ecological quality score explains known human pressures in estuarine habitats. At present, only about half of the indices attempt any validation and these use correlation analysis between the index ecological quality output and pressure fields to estimate the behaviour of the new index. In practice the calibration uses habitat scores different to those used to define sensitivity of the assessment metrics and use aggregated scores of estuarine condition (Cooper et al. 1994, Meng et al. 2002, Breine et al. 2007, 2010), catch data from a different estuary (Deegan et al. 1997), or pollution proxies (Delpech et al. 2010). The exercise can be as simple as the calculation of a correlation coefficient (Borja et al. 2004, Uriarte and Borja 2009), linear regression (Delpech et al 2010), or box plots and linear regression (Ramm 1988, Cooper et al. 1994, Harrison and Whitfield 2004, Breine et al. 2007, 2010). In some cases, these investigations give the responses of the indices to human pressures (i.e. dredging, engineering works, wastewater discharge), but also to actions taken to remove such pressures (i.e. wastewater treatment) (Uriarte and Borja 2009).

For the WFD implementation it is important to further evaluate the precision and accuracy of the indices and the uncertainty or potential errors in assigning areas to particular ecological classes. Only few examples are given in the literature and most commonly consist of the calculation of the misclassification rate after comparison with a preclassification exercise using habitat quality scores (Harrison and Whitfield 2004, Breine et al. 2007, 2010). Discriminant analysis and residual analysis have also been employed (Meng et al. 2002). Although not a formal analysis, Cooper et al. (1994) discussed the degree of scatter around the average scores as an indication of precision of the index. Rigorous uncertainty analysis providing probability estimates for each ecological status class are lacking.

3.6 Indices comparison and intercalibration

Proposed indices within the WFD should be applicable to the range of types into which the main European eco-regions are divided (European Commission, 2008b). The purpose of defining these types is to enable type-specific reference conditions to be established, making it possible to assess the ecological status for different geographical and habitat conditions (Borja 2005). These type-specific reference conditions are the basis of the classification schemes, and, as such, impact on all subsequent aspects of the implementation of the WFD (including intercalibration of the quality class boundaries assessed by different methodologies, assessment of the quality status of each of the biological elements, and monitoring, assessment and reporting of the water body status) (Borja et al. 2007, 2009b).

In this sense, the comparison and intercalibration of indices is a critical issue within the WFD (Borja et al. 2009b). However, until now, very few investigations have dealt with the fish indices comparison. Hence, some of the estuarine indices available within the WFD were compared by Martinho et al. (2008) and recently some of these methods have been applied to coastal waters under the MSFD (Henriques et al. 2008a and b). It is of note that the MSFD is adopting a more functional based system in which objectives of quality are being tested rather than individual quality elements (Borja 2006, Mee et al. 2008).

The main conclusion of these comparisons is that, despite some variation, all the indices gave consistent results throughout the studied period and the estuary investigated (Mondego, Portugal) (Martinho et al. 2008). Nevertheless, these authors show the high level of mismatch between the selected indices, indicating that there is still a great amount of work to be done in the intercalibration process, and concurrently, further comparisons of different indices for the fish component of transitional waters throughout European member states should be encouraged. Hence, currently, the fish-based indices intercalibration is taking place in all European ecorregions.

4 Discussion

There is much experience and many variations of the original fish-based index of biotic integrity (IBI) (Karr 1981) currently available for the evaluation of freshwater systems. Comparatively much less have been done in estuarine and coastal systems although Whitfield and Elliott (2002) indicated the types of indices available and the guidelines for their use. WFD compliance requires a highly structural analysis, i.e. the ecosystem is divided into a set of biological quality elements, each of which is then assessed according to the ecological structural elements of diversity, species richness and abundance, whereas a functional analysis provides better understanding and more direct insight into processes (de Jonge et al. 2006). The guild approach which allows a direct functional approach to the estuarine assessment is widely adopted in current indices.

A simple analysis uses indicator species or communities and present indices rely heavily on this as indicated by the distribution of metrics in the current indices. A simple taxon-based analysis is combined with a wider guild-based approach where the assessment is done at the functional rather than at the structural (i.e. species composition) level (Karr 1981). Given that there is information on the way fishes use estuarine and adjacent habitats or their reproduction or feeding mode it has been possible to aggregate species within common guilds (Karr 1981, Mathieson et al. 2000, Elliott et al. 2007; Franco et at. 2008) This approach is known to reduce the complexity of aquatic systems (Elliott and Dewailly 1995; Elliott et al. 2007), allows some insights on the functional categories affected by stress (Jager and Kranenbarg 2004) and extend the geographical application of indices beyond single species normal ranges (Elliott and Dewailly 1995, Karr and Chu 1997).

Since the derivation of the Scottish Environment Protection Agency (SEPA) ADRIS classification scheme for estuaries in the 1970s and 1980s it has been recognized the need for indicators which give both the cause of change, such as the levels of contaminants or water parameters, and the effects of that change, such as temporary and permanent habitat loss and community change (McLusky and Elliott 2004). Most fish assessments use multimetric indices. Summing or averaging normalized metric scores produces a balanced integration of independent biotic responses to anthropogenic stress into a single quality ratio value. This aggregation of metric scores into one single value simplifies communication. However, to determine which action is needed to improve the system, single metrics, raw data and expert knowledge should be considered (USEPA 2000, Hering et al. 2006). Indeed, amajor shortcoming of the multimetric approach is the reduction process where a single value representing ecological condition does not identify the cause of impairment. That is, there can be many such scores and these may indicate changes in conservational importance but they may not be immediately obvious in the aggregated score. Since the success of mitigation and restoration plans depends on our ability to minimize the effects of stress, any assessment tool that can both determine conservation status and diagnose damaging pressures can potentially provide cost and time savings for resource managers. Some authors present radar plots where independent metrics scores are presented in a relative scale (Jager and Kranenbarg 2004, Breine et al. 2010). To understand the causes of loss of ecological integrity will require the separation of the index-aggregated scores further into the component metrics and the analysis of stress-component metric relationships. This may be technically possible and relatively simple if the response and sensitivity of the metrics have been carefully validated. However stress-component metric relationships are correlations which do not necessarily correspond to cause-effect relationship. New research leading to better understanding of basic ecological functioning in estuaries is needed to be able to determine the cause of the observed changes. The outcomes should then be directly relevant to management plans and common language for a wide audience (i.e. managers, regulators, policy makers, general public) since the presence of the biota must reflect a required level of environmental quality.

Karr (1981) first used fish-based multimetric indices integrating characteristics of the community, population and individual organism to assess biological integrity in a relative score. In order to perform their role effectively they rely on a reference condition for comparison. The WFD requires assessments to be made against a baseline in a relatively undisturbed condition and presumably which has associated good quality fish communities. Given historical human pressures within Europe, a baseline based only on current data would be set at a somewhat reduced quality status compared to the original pristine system. The WFD indicates a high or good quality target which therefore should relate to long-term policy actions and restoration programmes for which these multimetric fish indices are being developed. The adjustment of the reference community to reflect good or high status communities and functioning is therefore extremely important as the reference sets the conservation target for a particular water body. Up to now, this is still largely based on expert judgment. In few instances where historical data are available, hindcasting the expected community before human intervention may be a possible method to correct the changed baselines although this has not yet been used in the reviewed indices. In addition, data-driven logistic regression models where metric outputs (model responses of the dependent biological variables) are correlated to environmental and biological factors (i.e. the independent factors and model explanatory variables such as ambient oxygen concentration, temperature, freshwater flow, season, salinity, spawning biomass, etc.) could provide the necessary predictive power to derive statistically-significant models of reference communities (Maes et al. 2007, Delpech et al 2010). It is emphasised there that each of these methods has its disadvantages: high status controls do not exist for most water body types in Europe, hindcasting has the disadvantages in that it is difficult to agree on a year of reference (e.g. before industrialisation) and even if data were available for that time, it is unlikely that conditions could be returned to that status without moving a huge population. Predictive models are currently at an academic level but are not sufficiently developed as management tools. Hence, there may be a demand to rely more heavily on expert judgement as being a more pragmatic and cost-effective manner although environmental managers may be unwilling to have detailed and expensive management measures relying on expert judgement solely.

In order to be effective, an index should be sensitive to cultural (anthropogenic) stressors in a predictable manner but sufficiently robust to be relatively insensitive to natural variability at different spatial and temporal scales (Rice 2003, Noges et al. 2009). The inherent variability and stressful conditions in transitional waters makes this particularly difficult (Elliott and McLusky, 2002). Short term variability (space and time) make extremely difficult to derive a clear reference communities and natural sources of stress such as extreme weather patterns, resource limitation, diseases, etc., have similar effects than anthropogenic stress on fish communities affecting the outcome of the metrics. Assessments therefore need to be based on metrics that are sensitive to the pressures responsible for the loss of integrity and be less affected by natural variation. However, it is difficult to select any set of ecological measures responding to anthropogenic stress and relatively free from background noise. While statistical models or expert knowledge are useful, their relevance and usefulness should be founded in ecological theory and independent appraisal (see Teixeira et al. 2010, for the same issue in benthic communities). Further practical considerations favour those indices that are easily measured using simple tools and show a low variability (Dale and Beyeler 2001, Rice 2003).

All indices found in this review are, to a greater or lesser extent, based on sensitivity, low variability and simplicity premises. However, not all support the metric relevance screening on a formal test using pressure calibration data. Hence, most indices use a conceptual approach where the outcome of pressures on individual metrics is derived from general ecological theory and expert knowledge with little statistical calibration. Instead of representing a fundamental flaw, this approach counters an overly reliance in the calibration dataset to construct the index which may result in dropping important metrics of estuarine function. This review has indicated the reduced number of metrics comprising the indices developed using more rigorous statistical procedures. While the inclusion of many often-correlated metrics may introduce undesirable noise, some degree of metric redundancy is desirable because some metrics may have overlapping sensitivity to multi-pressures acting in transitional waters. This is as a result of variable response lag times, response thresholds, and changes in relative contribution of the metrics across different ecoregions or sampling periods (Fausch et al. 1990, Noges et al. 2009). For example, the richness of sensitive species is likely to be an unresponsive metric in highly degraded areas, and conversely the incidence of diseased or abnormal individuals would only be apparent after substantial degradation, therefore being unresponsive in good quality areas. In an ideal index the complement of metrics would have to include a balanced combination of relevant metrics with a good combined predictive power when validated against all expected conditions possible in a water body. This necessary validation step is seldom present in the reviewed literature and an area where further research is warranted.

The nursery function for marine fish is a main estuarine attribute built into most indices. New recruits have often preferred habitats where they settle within shallow estuarine sites. Enhanced protection increases feeding opportunities and a favourable physicochemical regime is most frequently identified as the underlying evolutionary forces favouring estuarine-dependent life stages. The conceptualization and empirical demonstration of young of the year dependencies on settlement and nursery habitats within estuaries is shown in many studies. The immigration and residency of these life stages is controlled by physical and habitat quality attributes all of which may be affected by human pressures in the estuaries and coastal zones (Maes et al. 2007, Courrat et al. 2009). Hence, it is desirable to determine the habitat needs of each species of functional group in transitional waters and then to relate these to the conservation and management goals for the species (Elliott et al. 2008 HARBASINS ref.).

The need to integrate nursery quality aspects in ecological valuation has prompted the creation of dedicated indices of recruitment (Quinn et al. 1999) and some attempts of the development of fish metrics emphasizing the correlation between the effects of anthropogenic pressures and the nursery function (Amara et al. 2009, Courrrat et al. 2009). It is emphasised that indices should thus be designed with a precise target in mind, with the need for dedicated monitoring and therefore it is necessary to tailor them to capture precise quality aspects affected by specific stressors. This will then require the creation of many different indices. This also requires the need for more efficient ways to integrate the metrics or to weight their relevance according to estuarine typology and the use of more functional guilds may reduce the requirement of indices to only few general ones. This development of common metrics and indices is an area where there is currently a great interest and a valuable field of future research.

Ideally the sampling method should indicate abundance and where possible be quantitative (WFD guidelines). More general assessments are possible when the fishing method is unselective and operates effectively and reproducibly (Breine et al. 2010). However, it is largely recognized that gear efficiency varies with habitat type and species behaviour (Elliott and Hemingway 2002), therefore, all indices reviewed, whether discussed or not, are based on a more or less biased view of the fish assemblage. Gear choice and efficiency is seldom discussed but should be given that the method used will influence the catch identity and size on which the index is based. Similarly the level of effort is crucial to ensure comparability, reproducibility and relevance of reference communities as many metrics are based on diversity estimates (i.e. number of species present).

Karr (1981) discussed the need for an unbiased representation of the fish assemblage in the catch for a meaningful interpretation of the indices. Clearly sampling bias can easily compromise the use of the functional approach based on habitat guilds as fishing devices are often specific for a single type of habitat. For example, the assessment of a pelagic quality attribute (i.e. number of pelagic species) will be wrong if a demersal trawl is used. Likewise, indices based upon abundance metrics may be interpreted wrongly if gear efficiency is low or varies greatly across species or sampling events. Methods usually make no assumptions of gear efficiency or catchability, which may risk basing assessments or developing indices on biased rather than natural community types (Karr 1981, Noges et al. 2009). Coates et al. (2007) provided a parallel assessment using three different gear types (seine nets, otter and beam trawls) by producing scores independently for each gear type. These authors could not combine the different gear output into a ‘multi-method’ reference and presented independent assessment by gear type. However, they stated that by giving a measure of sampling effort to scale the catch data, integration may be possible leading to a more holistic and robust approach. Catch data from different trawl design have already been combined (Delpech et al. 2010), more difficult integration is expected from different gear types (e.g. fyke nets and beam trawls) although common semi quantitative scales may be possible such as probability of capture or recording frequency (Maes et al. 2007, Elliott et al. 2008 HARBASINS). Finally, it is emphasised that the sampling methods are as important as the methods of creating the metrics. Assessment variability due to factors acting at the local scale such as sampling gear, season, salinity, depth, etc are in general much less discussed in the development of the indices than those due to global scale estuarine typology, ecorregion, etc. Similarly, in some cases the sampling design (sampling at the expected fish abundance maxima) suggests current indices are biased towards the highest quality values for an area. In order to increase the confidence on the assessments, local effects will need to be randomized at appropriate scales and the variances and power associated with the metrics or indices assessed (Hughes et al. 1998, Courrat et al 2009). Further research is required with regards of the synergies between sampling methods (gear, replication and randomization) and assessment robustness and uncertainty in quality class assessment.

5 Conclusions and future research areas for improvement

All indices reviewed attempt to summarize the status of fish communities and the environment in a single summary score based on attributes from one to various metrics measures that are sensitive or supposedly sensitive to pressure. It was however evident that there are many types of indices, each created independently and each with different characteristics. Despite this lack of uniformity there is a general agreement in the formulation and structure of the different indices. Most include functional classes or guilds and are multimetric in which the metrics may be combined by simple arithmetic or weighted outputs. There has been a lack of intercomparison between the indices even though they have all been derived to provide a similar quality assessment of estuaries. It was also evident the range of strategies in defining responsive metrics and reference conditions.

As yet, there are no European-wide consistent indices but rather each country (or region/ researcher) has created its preferred index. Most indices have a relatively narrow geographical relevance. It was evident that at least in part the widening of the geographical relevance of estuarine fish indices will require better precision in the formulation of reference conditions and greater inclusion of functional metrics. It can be also noted that for instance most of the published indices refers to estuaries and very few to lagoons.

Measures of uncertainty of the indices is often lacking, for example it is not yet known what change in the community is required to produce a step change in the quality class assignment. Also the effects of the gear type and season on the output of the indices have yet to be rigorously interrogated. That is, signals from human pressures may be confounded not only by natural environmental variability (i.e. noise) but also by sampling bias and unsatisfactory effort level resulting in low power assessments.

Improvements in fish-based estuarine indices of habitat integrity are more urgently needed in four main areas that include: 1- Linking stress with response; 2- Derivation of reference conditions, 3- Effect of natural and anthropogenic stress, and 4-Effort and uncertainty.

7 References

Amara, R.., Selleslagh, J.,Billon, G. and Minier, C. 2009. Growth and condition of 0-group European flounder, *Platichthys flesus* as indicator of estuarine habitat quality. Hydrobiologia 627: 87–98.

Apitz, S.E., Elliott, M., Fountain, M. and Galloway, T.S. 2006. European environmental management: moving to an ecosystem approach. Integrated Environmental Assessment and Management 2:80–85.

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

Bain, M.B., Harig, A.L., Loucks, D.P., Goforth, R.R. and Mills, K.E. 2000. Aquatic ecosystem protection and restoration: advances in methods for assessment and evaluation. Environmental Science and Policy 3:S89–S98.

Borja, A. 2005. The European water framework directive: A challenge for nearshore, coastal and continental shelf research. Continental Shelf Research 25: 1768–1783.

Borja, A. 2006. The new European Marine Strategy Directive: Difficulties, opportunities, and challenges. Marine Pollution Bulletin 52: 239–242.

Borja, A. and Elliott, M. 2007. What does `good ecological potential' mean, within the European Water Framework Directive? Marine Pollution Bulletin 54: 1559–1564.

Borja, A., Franco, J., Valencia, V., Bald, J., Muxika, I., Belzunce, M.J. and Solaun, O. 2004. Implementation of the European water framework directive from the Basque country (northern Spain): a methodological approach. Marine Pollution Bulletin 48:209–218.

Borja, A., Josefson, A.B., Miles, A., Muxika, I., Olsgard, F., Phillips, G., Rodríguez, J.G.and Rygg, B. 2007. An approach to the intercalibration of benthic ecological status assessment in the North Atlantic ecoregion, according to the European Water Framework Directive. Marine Pollution Bulletin 55: 42–52.

Borja, A., Bricker, S.B., Dauer, D.M., Demetriades, N.T., Ferreira, J.G., Forbes, A.T., Hutchings, P., Jia, X.P., Kenchington, R., Marques, J.C. and Zhu, C.B. 2008a. Overview of integrative tools and methods in assessing ecological integrity in estuarine and coastal systems worldwide. Marine Pollution Bulletin 56: 1519–1537.

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

Borja, A., Ranasinghe, A. and Weisberg, S.B. 2009a. Assessing ecological integrity in marine waters, using multiple indices and ecosystem components: Challenges for the future. Marine Pollution Bulletin 59:1–4.

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

Boyle, T.P., Sebaugh, J. and Robinson-Wilson, E. 1984. A hierarchial approach to the meassurement of changes in community structure induced by environmental stress. Journal of Testing and Evaluation. 12: 241–245.

Breine, J.J., Maes, J., Quataert, P., Van den Bergh, E., Simoens, I., Van Thuyne, G. and Belpaire, C. 2007. A fish-based assessment tool for the ecological quality of the brackish Schelde estuary in Flanders (Belgium). Hydrobiologia 575: 141–159.

Breine, J.J., Quataert, P., Stevens, M., Ollevier, F., Volckaert, F.A.M., Van den Bergh, E. and Maes, J. 2010. A zone-specific fish-based biotic index as a management tool for the Zeeschelde estuary (Belgium). Marine Pollution Bulletin. In press

Chun, K., Weaver, M.J. and Deegan, L.A. 1996. Assessment of Fish Communities in New England Embayments: Application of the Estuarine Biotic Integrity Index. Biological Bulleting. 191: 320–321.

Coates, S., Waugh, A., Anwar, A. and Robson, M. 2007. Efficacy of a multi-metric fish index as an analysis tool for the transitional fish component of the Water Framework Directive. Marine Pollution Bulletin. l 55: 225–240.

Cooper, J.A.G., Ramm, A.E.L. and Harrison, T.D. 1994. The Estuarine Health Index - a New Approach to Scientific-Information Transfer. Ocean and Coastal Management 25: 103–141.

Costanza, R., d'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, V., Limburg, K., Naeem, S., O'Neill, R.V., Paruelo, J., Raskin, R.G., Sutton, P. and van den Belt, M. 1997. The value of the world's ecosystem services and natural capital. Nature 387: 253–260.

Courrat, A., Lobry, J., Nicolas, D., Laffargue, P., Amara, R., Lepage, M., Girardin, M.and Le Pape, O. 2009. Anthropogenic disturbance on nursery function of estuarine areas for marine species. Estuarine, Coastal and Shelf Science 81: 179–190.

Dale, V.H. and Beyeler, S.C. 2001. Challenges in the development and use of ecological indicators. Ecological Indicators 1:3–10.

Dauvin, J.C. 2007. Paradox of estuarine quality: Benthic indicators and indices, consensus or debate for the future. Marine Pollution Bulletin, 55: 271–281.

Dauvin, J.C. and Ruellet, T. 2009. The estuarine quality paradox: Is it possible to define an ecological quality status for specific modified and naturally stressed estuarine ecosystems? Marine Pollution Bulletin 59: 38–47.

De Jonge, V.N., Elliott, M. and Brauer, V.S. 2006. Marine monitoring: its shortcomings and mismatch with the EU Water Framework Directive's objectives. Marine Pollution Bulletin 53(1-4): 5–19.

Deegan, L.A., Finn, J.T. and Buonaccorsi, J. 1997. Development and validation of an estuarine biotic integrity index. Estuaries 20: 601–617.

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

European Commission, 1992. EU, Parliament and Council Directive 1992/43/EC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora. Official Journal of the European Union, Brussels. L 206:7-50.

European Commission, 2000. EU, Parliament and Council Directive 2000/60/EC of 23rd October 2000. Establishing a Framework for Community Action in the Field of Water Policy, Official Journal of the European Union, Brussels. L327:1-72.

European Commission, 2008a. EU, Parliament and Council Directive 2008/56/EC of 17 June 2008. Establishing a framework for community action in the field of marine environmental policy. Official Journal of the European Union, Brussels. L 164:19-40.

European Commission, 2008b. Commission Decision of 30 October 2008, establishing, pursuant to Directive 2000/60/EC of the European Parliament and of the Council, the values of the Member State monitoring system classifications as a result of the intercalibration exercise (notified under document number C(2008) 6016) (2008/915/EC). Official Journal of the European Union, L332: 20-44.

Elliott, M. and Dewailly, F. 1995. The structure and components of European estuarine fish assemblages. Netherland Journal of Aquatic. Ecology 29 (3-4): 397–417.

Elliott, M. and Hemingway, K.L. 2002. Fishes in Estuaries. Blackwell Science, Oxford, pp636.

Elliott, M. and McLusky, D.S. 2002. The need for definitions in understanding estuaries. Estuarine, Coastal and Shelf Science. 55 (6): 815–827.

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

Elliott, M, Hemingway, K.L., Cutts, N.D., Burdon, D., Pérez-Dominguez, R., Allen, J.H., Thomson, S.H., de Jonge, V.N., Breine, J.J., Van den Bergh, E., Stevens, M., Simoens, I., Jager, Z. and Twisk, F. 2008. HARBASIN report: Work Package 2 Estuarine Ecosystem Functioning, Restoration and Health. ZBB674-F-2008. Available from Interreg North Sea program organization.

Fairweather, P.G. 1999. Determining the 'health' of estuaries: Priorities for ecological research. Australian Journal of Ecology 24:441–451.

FAO. 2006. FAO Statistical Yearbook 2005-2006. 366pp

Fausch, K.D., Lyons, J., Karr, J.R. and Angermeier, 1990. Fish communities as indicator of environmental degradation. American Fisheries Society Symposium 8: 123–144.

Franco, A., Elliott, M., Franzoi, P. and Torricelli, P. 2008. Life strategies of fishes in European estuaries: the functional guild approach. Marine Ecology Progress Series 354: 218–224.

Franco, A, Torricelli, P. and Franzoi, P. 2009. A habitat-specific fish-based approach to assess the ecological status of Mediterranean coastal lagoons. Marine Pollution Bulletin 58:1704–1717.

Gorshkov, V.G., Makarieva, A.M. and Gorshkov, V.V. 2004. Revising the fundamentals of ecological knowledge: the biota–environment interaction, Biological Complexity 1(1):17–36.

Gray, J.S., 1989. The effects of environmental stress on species rich assemblages. Biological Journal of the Linnean Society 37: 19–32.

Gray, J.S. and Elliott M. 2009. The ecology of marine sediments. Second edition, Oxford University Press, Oxford. 256pp

Halpern, B.S., Walbridge, S., Selkoe, K.A., Kappel, C.V., Micheli, F., D'Agrosa, C., Bruno, J.F., Casey, K.S., Ebert, C., Fox, H.E., Fujita, R., Heinemann, D., Lenihan, H.S., Madin, E.M.P., Perry, M.T., Selig, E.R., Spalding, M., Steneck, R., and Watson, R. 2008. A Global Map of Human Impact on Marine Ecosystems. Science, 319: 948–952.

Harrison, T.D. and Whitfield, A.K. 2004. A multi-metric fish index to assess the environmental condition of estuaries. Journal of Fish Biology 65: 683–710.

Harrison, T.D. and Whitfield, A.K. 2006. Application of a multimetric fish index to assess the environmental condition of South African estuaries. Estuaries and Coasts 29: 1108–1120.

Henriques, S., Pais, M.P., Costa, M.J. and Cabral, H. 2008a. Development of a fish-based multimetric index to assess the ecological quality of marine habitats: the Marine Fish Community Index. Marine Pollution Bulleting 56: 1913–1934.

Henriques, S., Pais, M.P., Costa, M.J. and Cabral, H. 2008b. Efficacy of adapted estuarine fish-based multimetric indices as tools for evaluating ecological status of the marine environment. Marine Pollution Bulleting 56:1696–1713.

Hering, D., Feld, C.K., Moog, O. and Ofenböck, T. 2006. Cook book for the development of a Multimetric Index for biological condition of aquatic ecosystems: experiences from the European AQEM and STAR projects and related initiatives. Hydrobiologia 566: 311–324.

Hughes, R.M., Kaufmann, P.R., Herlihy, A.T., Kincaid, T.M., Reynolds, L. and D.P. Larsen, 1998. A process for developing and evaluating indices of fish assemblage integrity. Canadian Journal of Fisheries and Aquatic Sciences 55: 1618–1631.

Irz, P., De Bortoli, J., Michonneau, F., Whittier, T.R., Oberdorff, T. and Argillier, C. 2008. Controlling for natural variability in assessing the response of fish metrics to human pressures for lakes in north-east USA. Aquatic Conservation-Marine and Freshwater Ecosystems 18: 633–646.

Jackson, L.E., Kurtz, J.C., Fisher, W.S. 2000. Evaluation Guidelines for Ecological Indicators. EPA/620/R-99/005. U.S. Environmental Protection Agency, Office of Research and Development, Research Triangle Park, NC. 107 pp.

Jager, Z. and Kranenbarg, J. 2004. Development of a WFD. Fish Index for transitional waters in the Netherlands. Report. RIKZ/2004/606w.

Jordan, S.J. and Vaas. P.A. 1990. An index of ecosystem integrity for Northern Chesapeake Bay. Environmental Science and Policy. 3: S59–S88.

Karr, J.R. and Dudley, D.R. 1981. Ecological perspective on water quality goals.Environmental Management 5: 55–68.

Karr, J.R. 1981. Assessment of Biotic Integrity Using Fish Communities. Fisheries 6: 21–27.

Karr, J.R. and Chu, E.W. 1997. Biological monitoring and assessment: using multimetric indexes effectively. Report No. EPA 235-R97-001, University of Washington, Seattle, Washington

Knudsen, F.R., Hawkins, A, McAllen, R. and Sand O. 2009. Diel interactions between sprat and mackerel in a marine lough and their effects upon acoustic measurements of fish abundance. Fisheries Research 100: 140-147.

Maes, J., Stevens, M., Breine, J.J. 2007. Modelling the migration opportunities of diadromous fish species along a gradient of dissolved oxygen concentration in a European tidal watershed Estuarine, Coastal and Shelf Science 75: 151–162.

Martinho, F., Viegas, I., Dolbeth, M., Leitao, R., Cabral, H.N. and Pardal, M.A. 2008. Assessing estuarine environmental quality using fish-based indices: Performance evaluation under climatic instability. Marine Pollution Bulletin 56: 1834–1843.

Mathieson, S., Cattrijsse, A., Costa, M.J., Drake, P., Elliott, M., Gardner J. and Marchand, J. 2000. Fish assemblages of European tidal marshes: a functional guilds-based comparison. Marine Ecology Progress Series 204: 225–242.

McLusky, D.S., and Elliott, M. 2004. The Estuarine Ecosystem: Ecology, Threats and Management, Oxford University Press, Oxford. 226pp.

Mee, L.D., Jefferson, R.L., Laffoley,m Dd'A. and Elliott, M. 2008. How good is good? Human values and Europe's proposed Marine Strategy Directive. Marine Pollution Bulletin 56: 187–204.

Meng, L., Orphanides, C.D. and Powell, J.C. 2002. Use of a Fish Index to Assess Habitat Quality in Narragansett Bay, Rhode Island. Transactions of the American Fisheries Society 131: 731–742.

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

Niemi, G., Wardrop, D., Brooks, R., Anderson, S., Brady, V., Paerl, H., Rakocinski, C., Brouwer, M., Levinson, B. and McDonald, M. 2004. Rationale for a New Generation of Indicators for Coastal Waters. Environmental Health Perspectives 112(9): 979–986.

Noble, R.A.A., Cowx, I.G., Goffaux, D. and Kestemont, P. 2007. Assessing the health of European rivers using functional ecological guilds of fish communities: standardising species classification and approaches to metric selection. Fisheries Management and Ecology 14: 381–392.

Noges, P., van de Bund, W., Cardoso, A.C., Solimini, A.G. and Heiskanen, A.S. 2009. Assessment of the ecological status of European surface. Hydrobiologia 633: 197–211.

Oberdorff, T. and Hughes, R.M. 1992. Modification of an Index of Biotic Integrity Based on Fish Assemblages to Characterize Rivers of the Seine Basin, France. Hydrobiologia 228:117–130.

Ocampo-Duque,W., Ferré-Huguet, N., Domingo, J.L. and Schuhmacher, M. 2006. Assessing water quality in rivers with fuzzy inference systems: A case study. Environment International 32(6): 733–742.

Quinn, N.W., Breen, C.M., Whitfield, A.K. and Hearne, J.W. 1999. An index for the management of South African estuaries for juvenile fish recruitment from the marine environment. Fisheries Management and Ecology 6: 421–436.

Quintino, V., Elliott, M. an dRodrigues, M. 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.

Ramm, A.E.L. 1988. The Community Degradation Index - a New Method for Assessing the Deterioration of Aquatic Habitats Water Research, p293-301.

Ramm, A.E.L. 1990. Application of the Community Degradation Index to South-African Estuaries. Water Research 24: 383–389.

Rice, J. 2003. Environmental Health Indicators. Ocean and Coastal Management 46: 235–259.

Roset, N., Grenouillet, G., Goffaux, D., Pont, D. and Kestemont, P. 2007. A review of existing fish assemblage indicators and methodologies. Fisheries Management and Ecology 14: 393–405.

Seegert, G. 2000. The development, use, and misuse of biocriteria with an emphasis on the index of biotic integrity. Environmental Science and Policy 3: S51–S58.

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

Thompson, B. A., and Fitzhugh, R.G. 1986. A use attainability study: An evaluation of fish and macroinvertebrate assemblages of the lower Calcasieu River, Louisiana. Louisiana Dep. Environ. Qual., Office of Water Reso. Rep

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

UNESCO, 2000. Solving the Puzzle: The Ecosystem Approach and Biosphere Reserves. UNESCO, Paris.

Uriarte, A. and Borja, A. 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.

USEPA. 2000. Estuarine and Coastal Marine Waters: Bioassessment and Biocriteria Technical Guidance. Report No. EPA-822-B-00-024, U.S. Environmental Protection Agency. Office of Water, Washington, D.C. 20460.

Vaas, P.A. and Jordan, S.J. 1991. Long term trends in abundance indices for 19 species of Chesapeake Bay Fishes: Reflection in trends in the Bay ecosystem. In: New Perspectives in the Chesapeake System: a Research and Management Partnership. Proceedings of a Conference. Mihursky JA, Chaney A, eds. Chesapeake Research Consortium Publ. No. 137. Solomons, MD, p.539-546.

Weisberg, S.B., Ranasinghe, J.A., Dauer, D.M., Schaffner, L.C., Díaz, R.J. and Frithsen, J.B. 1997. An estuarine benthic index of biotic integrity (B-BI) for Chesapeake Bay. Estuaries, 20: 149–158.

Whitfield, A.K. and Elliott M. 2002. Fishes as indicators of environmental and ecological changes within estuaries: a review of progress and some suggestions for the future. Journal of Fish Biology 61: 229–250.

Williams, G.D. and Zedler, J.B. 1999. Fish assemblage composition in constructed and natural tidal marshes of San Diego Bay: Relative influence of channel morphology and restoration history. Estuaries 22: 702–716.

Wootton, J.T. 2008. Causes of species diversity differences: a comparative analysis of Markov models. Ecology Letters 4(1): 46–56.