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

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

DELIVERABLE

Deliverable D4.4-5: Precision and behaviour of fish-based ecological quality metrics in relation to natural and anthropogenic pressure gradients in European estuaries and lagoons

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

Contributors: Rafael Pérez-Domínguez **(UHull)**, Maria C Alvarez **(UHull)**, Angel Borja (**AZTI**), Henrique Cabral (**FCUL**), Anne Courrat (**Irstea**), Mike Elliott **(UHull)**, Vanessa Fonseca (**CO-FCUL**), Anita Franco **(UHull)**, Rita Gamito (**FCUL**), Joxe Mikel Garmendia (**AZTI**), Mario Lepage (**Irstea**), Iñigo Muxika (**AZTI**), Joao M. Neto (**IMAR**), Stephanie Pasquaud (**Irstea/CO-FCUL**), Violin Raykov (**IO-BAS**), and Ainhize Uriarte (**AZTI**)

Due date of deliverable: **Month 36**
Actual submission date: **Month 36**

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

This report summarises the work conducted in Work Package 4.4 – BQE fish in transitional (i.e. estuarine and lagoon) waters (TW) within the project WISER under the sponsorship of the European Commission. It omits most technical details of the analyses given in the four previous Work Package reports, but still provides the necessary information to understand the rationale, approach and underlying assumptions necessary to discuss the results. The focus is therefore to discuss and integrate the results obtained within Work Package 4.4 and with this, make recommendations to improve fish-based ecological assessments in TW, principally estuaries and lagoons. In addition, and to assist with the WFD implementation which is the overarching theme of WISER, the deliverable includes, where appropriate, case studies where we have used multi-metric fish indices currently under development, or already in use for WFD compliance monitoring across Europe. Furthermore, results of the work package have been shared with relevant Geographical Intercalibration Groups (GIGs) supporting the harmonization and equalization process across transitional fish indices in Europe.

Development strategies for fish indices in TW vary but generally include: (1) the calibration of metrics to anthropogenic pressures, (2) the development of reference conditions, (3) the calculation of ecological quality ratios, and (4) the designation of thresholds for Ecological Status (ES) class. New fish indices are developed for a defined geographical area, using specific sampling method and under locally relevant pressure fields. The former two factors, area and sampling methods, define the relevant reference condition in the calculation of Ecological Quality Ratios (EQR) and the latter factor, human pressures, define the index structure and especially the fish metric selection. To assess index relevance across areas, we calculated a suite of transitional fish indices on a standardized WISER dataset and then compared the agreement of the outcomes (using correlation analysis). The application of current indices to areas (or countries) different from the area in which it was originally developed leads to inconclusive or spurious results. The failure to accommodate the different indices to a standardized dataset in this work clearly demonstrates the fundamental reliance of current fish indices on the sampling methods and design of monitoring programmes used in the development of the index. Despite this, for some indices, correlations although weaker are statistically significant, also indicating the possible agreement in successful inter-calibration between these indices. Harmonization of BQE fish methodologies across Europe (common metrics) is unlikely by adapting or creating new fish indices but inter-comparison assessments are possible and valid using a common pressure index to harmonise different indices on a common scale.

We found a negative response of fish quality features to pressure gradients which make BQE fish in TW suitable for greater ecological integration than other BQEs. However, successful assessment of Ecological Status (ES) require a matching combination of fish index, reference values and local dataset gathered with compatible sampling methods. Whole indices provide more consistent overall ES assessments but fish metrics considered individually may be more useful as a means to focus restoration measures. Future work is needed to identify those specific pressures affecting fish assemblages providing targets for minimising the effects of stress in mitigation and restoration plans. In order to achieve this, and although the interpretation of outcomes is still difficult, more recent transitional fish indices are leading in the use of comprehensive appraisal and validation exercises to test the responsiveness of BQEs for the assessment of ES. Here we proposed for the first time a simple sensitivity exercise under realistic scenarios of metric change to explore the expected inertia (i.e. the tendency to buffer ES change after quality alterations), dynamic range (i.e. the ratio between the largest and smallest possible ES values) and most relevant metric components (i.e. the those driving the most likely scenarios leading to ES change) from a multi-metric fish index under relevant human pressure gradients. Overall, the behaviour of multi-metric indices under manipulations of metric scores clearly indicated that metric type, number of metrics used and correlations between metrics are important in determining the index performance, with indices including more and/or uncorrelated metrics or metrics with skewed distribution being less affected by extreme metric manipulations. Results of this analysis may be used to set realistic management targets and also to identify the aspects of the indices that are more likely to affect the outcomes leading to more robust and responsive indices.

Further improvements of fish indices may be attained by reducing the variability confounding biological quality metrics. This variability is undesirable noise in assessments and can be technical (i.e. linked to the method of assessment including sampling effort) or natural (physicochemical and biological). 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) increasing uncertainty in ES assignment. 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. The modelling exercise also demonstrated that unexplained variance remains generally much higher within-systems than between-systems suggesting a higher importance of sources of variability acting at the WB level. Modelling and improved standardization in monitoring campaigns should reduce uncertainty in ES assignment. One important factor that was assessed further was the effect of sampling effort. The results suggest that richness-based metrics require larger sampling efforts although a similar effort-related bias may be an issue for density-based metrics if fish distribution is very patchy (i.e. schooling fish or those aggregated in specific habitats) and insufficient replicates are taken to fully characterise the patchiness in their distribution. It is apparent that to overcome a potential large source of error, the Reference Conditions must be defined according to the level of effort used in the monitoring programme or, conversely, the monitoring must be carried out at the same level of effort used to derive the Reference Condition.

The WP finally explored the use of a predictive linear modelling approach to define reference conditions for fish metrics in transitional waters. The fish response data was modelled together with Corine Land Cover (CLC)-derived pressure proxies (% agricultural, urban and natural land coverage). Based on the obtained models, the expected metric score was 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. Even when significant, the effect of pressures on fish metrics was generally very weak, probably reflecting the use of too-generic pressure indicators (such as land cover data instead of more relevant estuarine proxies such as dredging, port development, waterborne pollutants, etc). The best explanatory models included sampling factors and natural characteristics considered important discriminant features in the definition of water body types. In particular, the present work argues for considering not only estuaries and lagoons as different typologies but also other natural and design characteristic such as the gear type, the sampling season and the salinity class. Furthermore, a relevant reference needs to account for survey design bias, including rare species contribution to assessment datasets, patchiness, choice of pressure proxies or sampling gear. 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 EQRs in Europe where there is a general lack of pristine areas or historical data on fish BQE and it provides an alternative to best professional judgment.

Taking all WP analysis and case studies together, the work conducted has highlighted the following key messages and linked research needs necessary to optimize BQE fish for the quality assessment of transitional waters:

Key Message 01: Harmonization of BQE fish methodologies across Europe (common metrics) is unlikely by adapting or creating new fish indices but inter-comparison assessments are possible and valid using a common pressure index to harmonise different indices on a common scale. Research needs to be focused on more widely-applicable fish indices will require the formulation of completely new indices based on a more flexible use of fish metrics according to system typologies, relevance and, probably, an increased use of functional traits. For current indices, further research on a method of intercalibration is needed.

Key Message 02: BQE Fish in TW respond consistently to human pressure gradients across transitional waters providing the means to assess Ecological Status (ES). Further work will be needed to identify those specific pressures affecting fish assemblages providing targets for minimising the effects of stress in mitigation and restoration plans.

Key Message 03 Although the interpretation of outcomes is still difficult, more recent transitional fish indices are leading in the use of comprehensive appraisal and validation exercises to test the performance of BQEs in the assessment of Ecological Status (ES). Further appraisal of fish indices behaviour is needed to understand the meaning of the quality outcomes, to set realistic management targets and also to identify the aspects of the indices that are more likely to affect the outcomes leading to more robust and responsive indices

Key Message 04 Uncertainty levels associated with metric variability in multi-metric fish indices can be managed to increase the confidence in Ecological Status (ES) class assignment. Further research is needed to include knowledge of habitat partition within systems, to understand metrics behaviour and precision, to test new combination rules allowing metric weighting by robustness and importantly to evaluate more robust sampling tools and methods.

Key Message 05 Reference conditions for BQE fish-based quality assessments can be objectively estimated using predictive modelling. Further refinements will require the use of better pressure proxies, robust metrics amenable to modelling and to account for survey design bias (effort & choice of sampling gear) at the relevant scales used in monitoring programmes.

Introduction

Historical human habitation and resource use in coastal areas have resulted in substantial change of most European estuaries from their original condition (Aubry & Elliott 2006, Zaldivar et al. 2008). Current paradigms consider this change to be detrimental and deviations from the original condition a measure of habitat degradation (Elliott & Whitfield 2011). Estuarine systems are important for ecological functioning in providing ecosystem services which in turn deliver societal benefits (i.e. flood defence, fisheries, water purification, energy) (Lotze et al. 2006). Human pressure on estuarine and coastal areas is likely to continue increasing and, together with global changes, such as sea-level rise, create conflicts between the public, stakeholders and managers. In response to future management needs and to prevent further degradation, the European Union has enacted specific legislation, the Water Framework Directive (WFD), which aims "to establish a framework for the protection of inland surface waters, transitional waters, coastal waters and groundwater" (Article 1, 2000/60/EC). Fish are biological descriptors of quality promoted in this European Directive to indicate estuarine quality status. Fish have a high relevance to the public and managers as there is an intuitive value associated to them. Therefore, fish are often better placed to raise public awareness than for example water chemistry or benthic invertebrates. This argument and the high level ecological quality integration including connectivity with adjacent freshwater and marine areas make fish a desirable biological element of aquatic systems on which to base ecological quality assessment or conservation criteria in management plans.

Despite the important practical advantage of using fishes as a quality element there are technical and ecological difficulties for their incorporation into tools to define estuarine ecological quality (Birk et al. 2012). Some of these difficulties also apply to other biological elements in their link with natural fluctuation of estuarine systems but others are specific to fish such as their inherent mobility, large recruitment variability and/or gear avoidance bias (Karr 1981).

A robust evaluation of ecological quality requires the use of indicator metrics responding in a predictable manner to a predefined gradient of ecological quality (Whitfield & Elliott 2002). There is a large number of these metrics forming indices that are now in use or in development across Europe (Birk et al. 2012). Some rely directly on physicochemical parameters but most, in line with recent developments, centre on the concepts of using biological elements as sentinel of changes in quality (Bain et al. 2000, Niemi et al. 2004, Borja et al. 2009). All require a minimum level of precision and accuracy in the quality evaluation as well as a number of desirable traits to make them practical and fit for use. In an ideal case a biological quality metric/ indicator should:

- produce an accurate and robust indication of change (sensitivity & robustness);
- be responsive to human pressure (calibration);

- have an available reference;
- have a wide biogeographical relevance;
- be technically feasible;
- be of easy interpretation by non-specialists;
- be bound by time and budget constraints.

These criteria are further developed in Appendix 1A and shown to link with the parameters required to monitor natural and anthropogenic change (Elliott 2011).

Sensitivity and robustness of the indication as well as responsiveness to human pressure gradients are features *a priori* necessary in any index. The other characteristics highlighted above are a function both of the monitoring goals and the available science. Fish indices, or any other ecological quality index, are therefore surrogates for monitoring needs and these are ultimately dictated by society (Atkins et al. 2011). In the light of this ultimate linkage with societal needs and expectations is finally necessary to place these indices within the boundaries of what is feasible, essentially a utilitarian framework where indices must perform reasonably well and still be affordable by society. In other words, monitoring programmes need to find a compromise between accuracy and costs. In turn this can be achieved by finding a reasonable compromise between precision and sampling effort (Hering et al. 2006) but also, and more importantly, by using more robust assessment tools (indices) and sampling designs (Niemi et al. 2004).

In line with the work conducted within the WISER fish transitional waters work package (WP4.4) we aim here to discuss these key issues necessary to develop practical estuarine fish tools. Moreover, we aim to promote evidence-based discussion of the underlying ecosystem functioning affected by anthropogenic alterations and the use of fish indices to guide sustainable management of estuarine systems. The specific objectives set were: (1) to review the *state of the art* in estuarine fish-based indices and test and compare available indices using common methods through a dedicated field exercise; (2) to evaluate the sensitivity of fish metrics and indices to different sources of variability (natural, sampling-related, pressure-induced) which might have an effect on the uncertainty of the ecological status assessment, and (3) to propose and where possible test new approaches for the modelling of fish communities to be used in defining reference conditions.

Experimental Approach

A brief description of different approaches and methodologies as well as assumptions implicit in the four previous WP4.4 deliverables is provided below in order to aid with the understanding and interpretation of results. For simplicity we have omitted a number of technical details already presented in these deliverables. To allow the reader to replicate all methodologies and experimental approaches, references to the relevant deliverables are given in the text.

Objective 1 - Review and comparison of current estuarine fish indices

We reviewed seventeen published fish-based indices of estuarine quality worldwide. The comparison was done on common development strategies and assessment methods using cluster analysis and similarity percentage (SIMPER) analysis (for more detailed methods see (Pérez-Domínguez et al. 2010). Six of the indices reviewed (AFI, EFAI, ELFI, TFCI, BHI and Z-EBI) (see Appendix 2 for bibliographical references and Appendix 3A for constituent metrics for the full name and country of development of these indices)were tested and compared by applying them to common datasets obtained during a standardized field sampling campaign conducted in 2009-2010 in 8 different transitional water systems (6 estuaries and 2 lagoons) across Europe (Figure1). The resulting dataset included data from a dedicated WISER transitional fish field survey (5 sites: Varna Bay and Varna Lake, Bulgaria; Mondego estuary, Portugal; Orwell & Stour estuary, UK; Lesina lagoon, Italy) carried out by using standardised sampling protocols (for details see Courrat et al. 2011, Borja et al. 2012, Courrat et al. 2012b). Additional data were collected by the Basque Water Agency (Spain) (3 sites: Nervion, Oiratzun and Bidasoa estuaries) and the Environment Agency (UK) (1 site: Orwell & Stour estuary, integrating the WISER dataset) using comparable sampling protocols.

The resulting extended datasets were combined and collated and fish ecological and feeding guilds, as well as sensitive and reference species lists, were harmonized across sites according to lifecycle theory and bibliographical sources (Appendix 1B). Finally, Ecological Quality Ratios (EQR) were calculated and Ecological Status (ES) classification derived for the different fish indices and water bodies. Analyses were conducted separately for comparable sampling protocols (accounting for sampling method -fyke nets, beam trawls and seine nets-, season and time - day vs. night) and standardised to salinity groups in order to account for differences among water body typologies. To assess the score agreement between indices, Kendall rank correlation coefficients and Kappa values (Fleiss-Cohen weights) between pairs of indices were used.

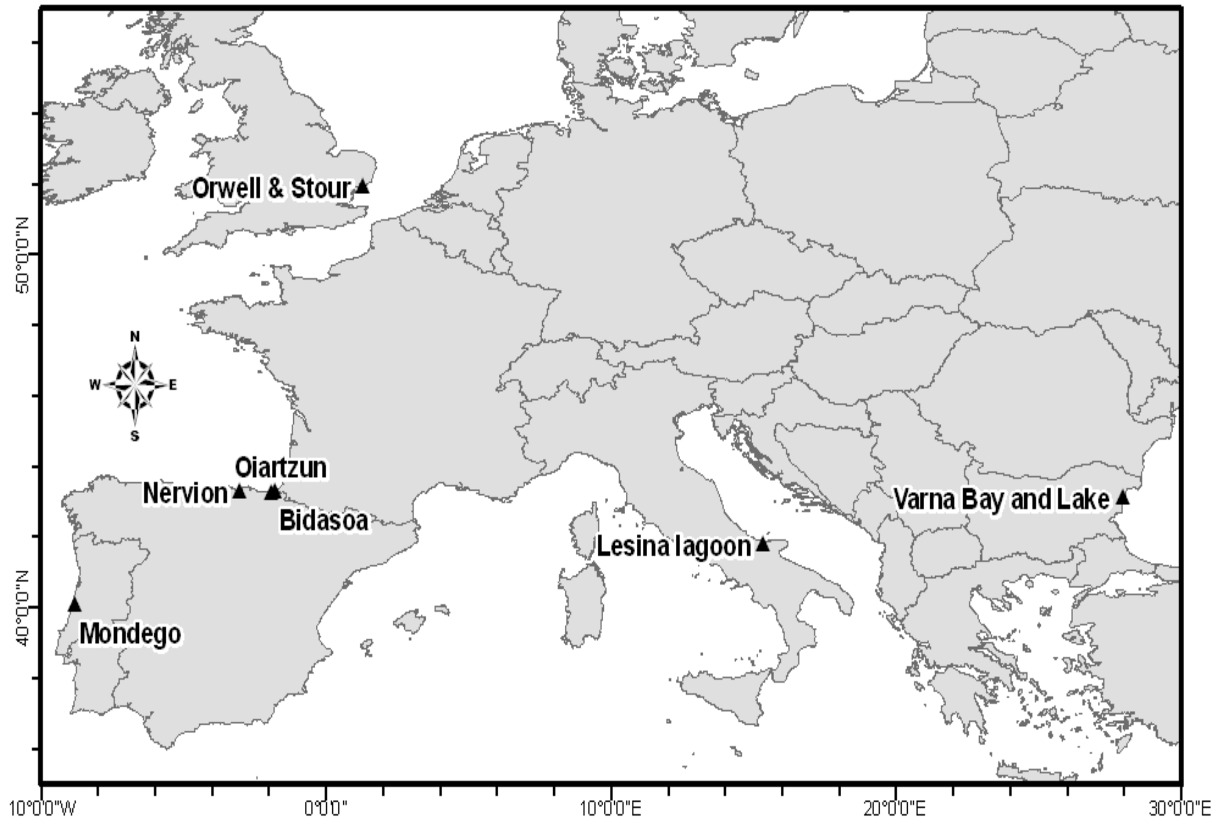


Figure 1. European estuaries sampled for the index performance comparison.

Objective 2 - Sensitivity of metrics and indices possibly affecting robustness of fish-based ecological status assessment

The sensitivity of fish-based indicators (metrics or indices) to different sources of variability was investigated by taking into account both pressures affecting fish assemblages and the effect of natural variability in estuarine/lagoon physico-chemical, hydro-morphological conditions and sampling protocols. Given that the robustness of ecological quality indicators depends on their sensitivity to human pressures, the effect of pressures is regarded as a "signal" whereas the effect of the other (inherent) sources of variability is regarded as "noise". The latter can be introduced by natural seasonal and spatial variability, sampling design and gear-induced variability, laboratory analytical-induced variability, variability due to different numerical tests being used, worker-variability, etc (Gray & Elliott 2009). The higher the signal-to-noise ratio, the more robust a biological indicator is considered. The sensitivity analysis, thus aimed at identifying possible weaknesses within the indices assessment (e.g. metrics or indices highly sensitive to noise and/or with low sensitivity to pressure signal) and allowed to provide recommendations on the ways to reduce the detection of unwanted sources of variability (the 'noise' component) in order to optimise the metric effectiveness as indicator of ecological status. In addition, the knowledge on how the relationships between metrics and pressures can be affected by other system variability (natural 'noise') may be of use in the appropriate definition of reference conditions for the different metrics (see Objective 3).

The sensitivity of fish-based indices was also analysed with respect to metric changes, in order to understand what was the expected dynamic range of the indices in response to variable pressure gradients. In a multi-metric index it is important to understand the weight that different metrics have on the final index score and thus in the status a WB is given by the assessment. This information is necessary to understand the behaviour of the indices, to facilitate the interpretation of the results and to explore the effect of metrics uncertainty on the robustness of multi-metric indices. In addition, understanding the effect of the different metrics on the index response would allow us to suggest the possible prioritisation of response measures (i.e. towards the improvement of specific metrics) to be undertaken to improve the ecological status of transitional water bodies.

2.1. Fish metrics sensitivity to sources of variability

A sample of seven common fish metrics identified in the initial index review was selected for analysis, namely total density (TD), total number of species (SR), number of estuarine resident species (SR_ER), density of marine migrants (DMM), number of marine migrant species (SR_MM), percentage abundance of omnivorous fishes (RD_O) and percentage abundance of piscivorous fishes (RD_P). Metric values were calculated using the WISER transitional fish data set and additional Basque Water Agency (Spain), IMAR-CMA (Portugal), French Water Agencies (France) and Environment Agency (UK) data collected with comparable sampling protocols in 39 estuaries and 14 lagoons overall between 2003 and 2010. The datasets were organized in a relational database containing a total of 3249 fishing events which included fish data (to derive the metrics), data on sampling protocols (e.g. type of gear, date and time of sampling, etc.) and water quality parameters (e.g. salinity, temperature, depth, etc.). Additional data on the average physicochemical and hydro-morphological characteristics of the studied estuaries/lagoons (e.g. estuarine area, continental shelf width, salinity regime, etc.) as well as Corine Land Cover (CLC)-derived pressure proxies were also included. A conceptual matrix was used to identify *a priori* the most relevant sources of variability and define the expected effect of pressures on the fish metric outcomes. Most relevant variance sources were then quantified using either linear models (LM), linear mixed models (LMM) or generalized linear models (GLM).

In addition to this, a Bayesian probabilistic framework was applied to a sample of French lagoons to combine fish metrics in an objective way, and to provide a rigorous estimate of the uncertainty on the assessments. Briefly the approach relays on the simulation of metrics EQR according to a discrete pressure indicator based on contamination data. GLMs likelihood functions are then used to provide probability densities of the theoretical EQR according to pressure values and to calculate the probability of being in a given ES class. These probabilities were combined using the Bayes theorem to provide the posterior probability of being in each of the five ES classes considered. Under this approach, the probability densities account for the uncertainty associated with the variability of the GLM models (model variance) and the sensitivity of the fish metric to the stressor (model regression parameters) (for further details of the method see Drouineau et al. 2012).

Finally the specific effect of sampling effort on fish metrics was tested on data from four Portuguese estuaries (Ria Aveiro, Tagus, Sado, Mira) collected in May-July 2005/2006. The metrics included in the Portuguese Estuarine Fish Assessment Index (EFAI) were taken into account. Bootstrapping techniques allowed calculating the means and standard deviations of fish metrics for different number of hauls. A cost/bias analysis was also performed in order to provide some evidences on the lowest reliable number of hauls that should be included in monitoring works (for further details see Courrat et al. 2012a).

2.2. Fish indices sensitivity to pressure gradients

In addition, a detailed analysis of fish-based index sensitivity to human pressure gradients was carried out specifically on the AFI and EFAI fish indices, designed for WFD compliance monitoring in Spain and Portugal, respectively. The AFI analysis was carried out on a sample of 12 Spanish estuaries. Initially, 16 pressure proxies and 8 estuarine hydro-morphological descriptors were selected in order to account for both natural and human-induced (pressure) variability. After elimination of highly correlated variables, ordination analysis (Principal Component Analysis) was used to explore the general gradients segregating the studied estuaries. Finally multiple regression analysis was used to assess the sensitivity of AFI index to the more influential pressure proxies, while also taking into account the influence of estuarine hydro-morphological characteristics. The EFAI analysis was done in a smaller sample of 5 estuaries sampled in a single sampling campaign and followed a similar but simplified analytical process (for full details please see Borja et al. 2012).

Further to the empirical modelling exercise, we qualitatively assessed the sensitivity of 8 multi-metric indices (AFI, EFAI, ELFI, Z-EBI, TFCI, EBI, IFCI, IBI, the first five indices being WFD compliant) and their respective metrics to 7 different human pressures and impacts on estuaries (i.e. chemical pollution, eutrophication, loss of habitat, water turbidity, habitat fragmentation, fish mortalities, invasive species, temperature and flow changes). This conceptual analysis was based on a simple scoring method of cause-effect relationship between each pressure and metric/index according to the expected strength of the relationship and response time-lag, as derived from a combination of ecological knowledge, published literature and expert judgement.

The Bayesian method introduced in the previous section (sensitivity to sources of variability) provides pressure-impact statistical models to estimate the probability of being at a certain ES status and anthropogenic pressure level given a pool of candidate fish metrics (Drouineau et al. 2012). This method is relevant to assess sensitivity to pressure proxies and gives an objective criterion for combining multiple fish metrics in a multi-metric fish indicator, taking into account the variability and the sensitivity of the fish metrics as well as the level of certainty given to any evidence included in the analysis.

2.3. Fish indices sensitivity to metric changes

In addition to the behaviour of fish indices against pressures, we investigated the sensitivity of the ELFI and the UK Transitional Fish Classification Index (TFCI) to metric variation under different simulation scenarios, in order to identify the most influential metrics affecting the index variation. The current version of both indices in use for the WFD compliant monitoring was used, and the effect of different scenarios of metric change was assessed based on real WFD monitoring data provided by Irstea/formerly Cemagref (France) and the Environment Agency (EA) in a sample of 68 (ELFI) and 58 (TFCI) water bodies. This sensitivity analysis was done by setting metric values to average level in different upper and lower percentiles (10, 40, 60, 80 percentiles) of the metric distribution and then calculating the percentage change in the final index value with respect to the average value in the studied dataset. In order to account for relationships between metrics (hence their co-variability with the metric driving the scenario), links between metrics were established based on the strength and significance of Spearman rank correlations between metrics, and these were used to allow metrics variation with changes in the driving metrics under the different scenarios (see Borja et al. 2012).

Objective 3 - Modelling reference conditions

Given that there are almost no transitional waters in Europe that can be considered as being in pristine ecological condition and historical data are not available for all transitional waters types, we tested a modelling approach to define type-specific reference conditions for fish assemblages in transitional waters in Europe. The WISER extended dataset together with pressure proxies and physicochemical data introduced above were used. Thirteen fish metrics (including those tested above for sensitivity, see section 2.1) as well as all the metrics composing the French Estuary and Lagoon Fish Index ELFI) were considered. CLC-pressure indicators (% agricultural: Agr, urban: Urb, and natural: Nat land coverage) and modelling methods (LMM, GLMM) used for the modelling of reference conditions were similar to those employed in the metric sensitivity analysis. The analysis was conducted independently for estuaries and lagoons. The variables accounting for sources of natural variability in fish metric included in models for lagoons and estuaries are listed in Appendix 5A. Based on the obtained models, the expected metric score was 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.

Results and Discussion

1. Review and comparison of current estuarine fish indices

1.1. Indices structure and development criteria

Seventeen published fish-based indices of estuarine quality worldwide were reviewed according to their development strategies and assessment methods (Appendix 2) (Pérez-

Domínguez et al. 2012). Most of these indices are multi-metric tools, where a series of fish relevant metrics scores are combined in a summary value. All fish indices found in this review are, to a greater or lesser extent, based on premises of responsiveness (i.e. sensitivity) to human-induced stress, low metric variability and simplicity of measurement, which is in line with general design considerations of other biotic indices (Dale & Beyeler 2001, Rice 2003). Noticeably, most of the published indices refer to estuaries and very few to lagoons or any other transitional water body type.

There is a plethora of numerical techniques for assessing and communicating changes to biological community structure; indeed, Gray & Elliott (Gray & Elliott 2009) give almost 30 families of numerical techniques aimed and used to determine change – many of which have been used to assess anthropogenic change and the effects of pressures. Most metrics in use are measures or derivatives of species richness and diversity (Figure 2). Increased diversity is generally assumed to indicate higher quality (Gray 1989) although this often does not apply in transitional waters subjected to high natural variability irrespective of anthropogenic stressors (Elliott & Whitfield 2011, Basset et al. 2012, Whitfield et al. 2012). Moreover this family of metrics has to be considered with caution because there are very sensitive to sampling effort (Anjos & Zuanon 2007), to introduced species (Ferreira et al. 2007) or to specific disturbances increasing richness (Mayfield 2010). The largest family of metrics in this group involves the number and proportion of indicator species based on their requirement of precise estuarine quality features. Total or relative abundance measures and fish condition or health status are also included as they provide a quantitative measure of fitness at the level of the individual.

Metrics describing the functional structure of fish assemblages are also often included in the reviewed indices. These metrics based on fish guilds, i.e. groups of fish species sharing similar ecological requirements, are in fact a generalization of the fish indicator concept and indicate the availability of functional ecological niches within estuaries (Elliott et al. 2007, Franco et al. 2008a). The guild approach allows a greater understanding of estuarine functioning, by separating estuarine use strategies adopted by fishes. Generally guilds describing the different uses of habitats and feeding resources are used. In particular, the most common habitat use guilds are estuarine residents and marine juvenile migrant species. This choice is justified due to their strong dependence on the estuarine environment for their entire life (estuarine residents) or for sensitive life stages (marine juvenile migrants, using estuaries and lagoons as nursery areas). In particular, the nursery role of transitional waters to juveniles of marine species is often specifically addressed. This, in fact, is recognised as an important function of estuaries and lagoons, which provide shelter and food to juvenile marine fish (Beck et al. 2001, Elliott & Hemingway 2002, Able 2005), thus also supporting marine fishery stocks.

Trophic guilds, the most often used metric family, are centred on trophic guilds, primarily number and diversity of predatory fish (carnivorous and/or piscivorous) followed by benthic feeding fish. In general, the specialised-feeder metric decreases while the omnivore metric increases with increasing level of disturbance, provided that the level of disturbance does not result in the complete collapse of the food resource.

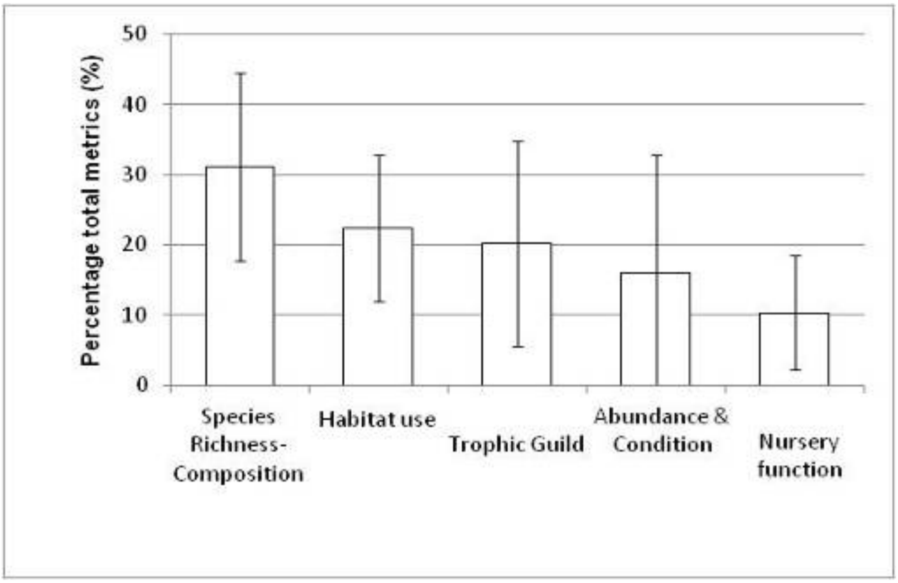


Figure 2. Relative importance of metric distribution across ecological attributes in fish multi-metric indices for transitional waters.

A wide range of approaches for metrics selection is also used, from the simpler (and more common) choice of metrics based on previous successful indices and expected ecological responses to degradation, to more complex statistical methods (e.g. stepwise discriminant analysis, redundancy analysis) often resulting in more stringent conditions for metric inclusion. The resulting number and choice of metrics varies with most multi-metric indices being built around a pool of 9-10 metrics, with a tendency for fewer metrics in those indices where metric sensitivity and redundancy has been formally assessed against pressure scores.

In clear contrast to the overall agreement in metrics, a large variability of methods, protocols and approaches has been employed in the development of fish based indices in transitional waters. A suite of sampling gears such as seine nets, beam trawls, otter trawls, gillnets, fyke nets, and visual diver censuses has been used to generate catch data for the assessments. All these gears and protocols of use are selective to some degree (Pasquaud et al. 2012). The bias introduced by any given gear (due to its selectivity) might be compensated by using complementary methods targeting the different existing niches (Elliott & Hemingway 2002) and adequate sampling effort to characterise a single transitional water body. However, in practice, this compensation would be very difficult to achieve as the combined gear method is unlikely to eliminate all bias and, perhaps more importantly, it will require further assumptions in order to bring the combined catch to a common scale of effort. Alternatively, a more pragmatic approach could be adopted, by using a single, standardised sampling methodology in the knowledge that some components of the assemblages will be missed. This approach is in fact used, leading to a sampling method-specificity built into the indices and therefore dependency on specific gear types and method of use. This is most clearly illustrated by the inclusion of gear specific reference conditions to score fish metrics (Coates et al. 2007, Breine et al. 2010). Furthermore, specific reference conditions are usually chosen according to environmental factors and protocol specificity such as water body type, physiography, season, habitat, and salinity regime, in order to reduce the natural 'noise' of spatial and temporal variability. Therefore, the literature distinguishes between metric-, habitat-, season-, gear-, salinity class-, estuary- and ecotype-specific reference conditions as relevant to the data structure and analysis. In practice, the reference community is derived by either using pressure-response models or by selecting the best 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 where its metric value lies with respect to the reference. 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 scores thresholds varies among indices and estuarine typology and are often calibrated with pressure data if available. As an example, all WFD-compliant indices use a 5-band scoring system.

With regard to the index development, although the analysis of metrics and index sensitivity to pressures and the application of modelling approaches to define reference conditions have increased in recent times, expert opinion is still widely used. Indeed, this follows the suggestion in the WFD in that if other methods of defining change from that which is expected (a physical control, hindcasting or predictive modelling Hering et al. 2010) are not successful or able to be used then best professional judgement will suffice. As indicated above, signals from human pressures may be confounded not only by natural environmental variability (i.e. noise) but also by sampling bias and unsatisfactory sampling effort level resulting in low power assessments. Only about half of the indices attempt any validation and these use correlation analysis between index-computed values and pressure scores to estimate the behaviour of the new index. Rigorous uncertainty analyses providing probability estimates of the robustness and confidence in the ecological status class assignment are commonly lacking.

Overall, the results of the above review highlight the limited applicability of the current fish indices outside specific geographic boundaries (usually not extending further than a single ecoregion and often focusing on one system) and sampling protocol constraints. 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. Whether harmonizing assessment methodologies is not possible a method of intercalibration may be used. Furthermore, in order to increase the confidence on the assessments, local effects will need to be taken into account at appropriate scales and the variances and power associated with the metrics or indices assessed.

Improvements in fish-based estuarine indices of habitat integrity are more urgently needed in four main areas that include: (1) improving the mechanisms of linking anthropogenic pressures and ecological responses; (2) deriving reference conditions, (3) disentangling the effect of natural and anthropogenic stress, and (4) testing the effect of sampling effort and design on indicator outcomes and assessing the uncertainty of outcomes. Once these deficiencies became clear, the initial objectives for the work package were adjusted in order to fill the current knowledge gaps and to allow improvement of the robustness and consistency of fish-based assessment tools in transitional waters. The rest of this document presents analyses and provide case examples illustrating the way fish indices could be improved.

1.2. Consistency of assessment results

Six of the indices reviewed (AFI, EFAI, ELFI, TFCI, BHI and Z-EBI) were tested in order to evaluate their consistency in the ecological status assessment. The metrics composing these indices are listed in Appendix 3A (further details on the characteristics of these indices are provided in (Pérez-Domínguez et al. 2010, Courrat et al. 2011). Given the spatial and methodological constraints highlighted for the different indices during the literature review (Pérez-Domínguez et al. 2010), comparison between the different assessment tools was only possible by applying them to data obtained during a dedicated WISER field sampling campaign conducted in 2009-2010 in 8 different transitional water systems (6 estuaries and 2 lagoons) across Europe (Figure 1). Sampling methods and strategies were standardised in order to obtain comparable datasets while also covering the indices main sampling requirements (Appendix 3B). As the WISER surveys involved multi-gear sampling, each of the gear-specific indices could be tested at least once on the data gathered with the gear for which it was designed.

Seventy-three species and a total of 8687 fish were recorded during the field campaign, with a variable number of taxa recorded in the different systems (Figure 3).

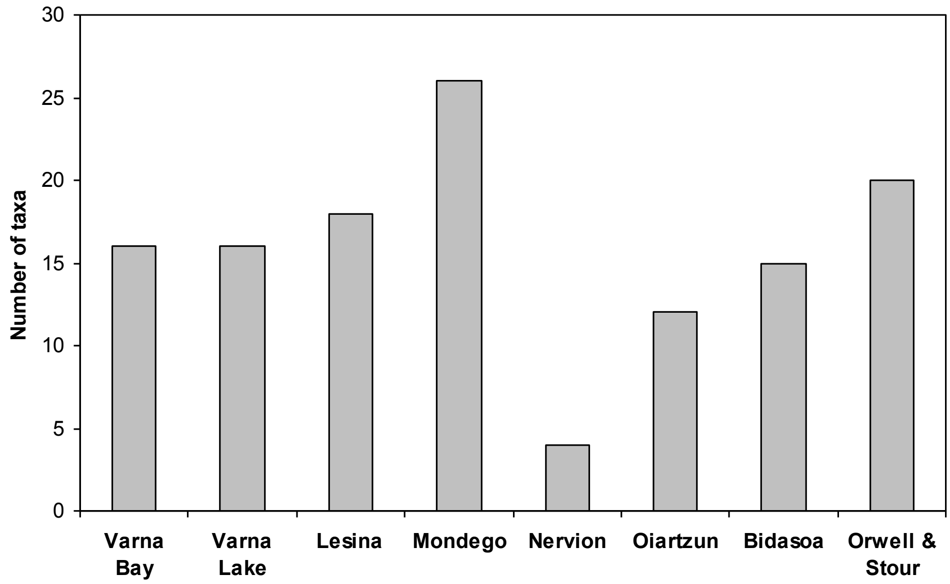


Fig. 3: Number of species caught in the sampled sites. The simple message is the fundamental difference in expectations across study sites for this basic fish metric.

The use of a harmonised dataset resulted in often very different and weakly correlated indices outcomes (both in terms of EQR and ES classification) (Table 1).

Table 1: Overview of indices' EQRs (Ecological Quality Ratios) and ES (Ecological Status) per gear and per day/night or season, for each of the sampled sites. Colours stands for ESs: red: bad; orange: poor; yellow: moderate; green: good; blue: high. White empty cells indicate that the index could not be calculated or no meaningful reference condition could be estimated. The references for fyke nets used for BHI were obtained from a mix of day and night 12 hours fyke nets, thus BHI results are given for combined day and night fyke nets data only.

	Salinity class	Gear	Day / night or season	ELFI (FR)	EFAI (PT)	EFAI (PT) - 5 hauls	AFI (Basque)	TFCI (UK)	Z-EBI (BG)	BHI (S. Africa)
Varna Bay	**2**	**Beam trawl**		0,42	0,66	0,66	0,58	0,61		3,81
Varna lake	**2**	**Beam trawl**		0,33	0,54	0,54	0,42	0,40		1,68
		Fyke net	Night	0,67	0,66	0,66	0,43	0,65	0,1	
Lesina	**2**	**Fyke net Cemagref**		1,00	0,49		0,52	0,69	0,1	9,75
		Fyke net	Day	0,33	0,26	0,26	0,33	0,25	0,1	
			Night	0,33	0,43	0,31	0,41	0,33	0,1	
Mondego	**1**	**Fyke net**	Day		0,67		0,56	0,40	0,1	
			Night		0,67		0,67	0,62	0,1	
		Beam trawl (night only)		0,54	0,67		0,50	0,59		2,99
	2	**Fyke net**	Day		0,71		0,72	0,60	0,1	9,46
			Night		0,66		0,33	0,62	0,1	
		Beam trawl (night only)		0,71	0,66	0,66	0,50	0,47		2,46
	3	**Fyke net**	Day		0,60		0,67	0,43		
			Night		0,71		0,72	0,80		
		Beam trawl (night only)		1,00	0,83	0,77	0,62	0,77		8,17
Nervion	**3**	**Beam trawl**	Spring	0,08	0,66	0,66	0,46	0,35		3,46
			Summer		0,66	0,49	0,42	0,35		
			Autumn	0,17	0,66	0,66	0,29	0,28		2,82
Oiartzun	**3**	**Beam trawl**	Spring	0,00	0,60	0,60	0,40	0,26		4,36
			Summer		0,60	0,60	0,46	0,33		
			Autumn	0,04	0,77	0,71	0,49	0,33		5,21
Bidassoa	**1**	**Beam trawl**	Spring	0,04	0,40		0,39	0,26		1,39
	2		Spring	0,00	0,49		0,33	0,25		1,29
	3		Spring	0,04	0,66	0,66	0,39	0,28		3,26
	3		Summer		0,77	0,77	0,45	0,26		
	3		Autumn	0,13	0,77	0,71	0,54	0,42		4,92
Stour & Orwell	**3**	**Fyke net 24 hours**			0,66		0,38	0,48		8,62
		Beach seine EA		0,75	0,83	0,77	0,52	0,58		
		Beach seine Wiser		0,58	0,71	0,71	0,49	0,47		
		Beam trawl		0,58	0,60	0,54	0,41	0,49		4,17
			Min EQR	0,00	0,26	0,26	0,29	0,25	0,10	1,29
			Max EQR	1,00	0,83	0,77	0,72	0,80	0,10	9,75
			Mean	0,39	0,63	0,62	0,48	0,45	0,10	4,58
			Mediane	0,33	0,66	0,66	0,46	0,43	0,10	3,81

There is in general a low consistency between assessment tools in the evaluation of a transitional water body system. Unlike other biological quality elements, the ecological quality assessment methods based on fish fauna highly rely on specific sampling methodologies. This was clearly identified in the initial review of available indices and discussed in the previous section (Pérez-Domínguez et al. 2012). Constraints imposed by the method of sampling affect the way the monitoring is designed and ultimately the way fish metrics are calculated. For example, some of indices take into account the influence of seasonal variability on fish data, hence defining the index specifically for one or two seasons, usually summer or autumn. Others do not account for this variability and pool data over 1-year period. Similar differences are observed when considering the spatial sources of variability, with some indices being calculated at the fishing event scale and others pooling all data per salinity class or WB. In addition, diversity-based metrics resulted in random/low quality scores when applied outside the area of development of the index due to mismatch in reference conditions. This is especially evident for metrics based on indicator species as using specific taxa limits the spatial applicability of the metric, due to the species geographical range distribution, thus restricting the validity of many indices to a single estuary or biogeographical zone.

Not surprisingly, when used outside of their initial framework, geographical limits or with different sampling methods, there is a greater uncertainty in interpretation if the fish index. Hence, it is clear that the assessment of the ecological status of transitional waters using fish indicators highly depends on the assessment tool (index) used and the corresponding sampling methods. In addition to emphasising the need for the cautious interpretation of fish indices results, the obtained results highlight the importance of intercalibrating the different fish-based assessment tools in order to obtain consistent assessments across wide geographical areas. In European countries, for example, where several indices are in use it is required that good ecological status represents the same level of ecological quality everywhere in Europe despite the tool used (Annex V WFD). Intercalibration (IC) exercises have been undertaken for fish-based indices in the different European ecoregions. These exercises were based upon an approach similar to that one undertaken in this study (i.e. the simultaneous application of multiple standardised sampling and assessment methods to a same water body). However, it is of note that the calculation of each different fish index was done only considering the sampling methodology specifically developed for each particular index. The IC exercise was able to use a common pressure index to which every method can refer to in order to adjust the class boundaries with an harmonization band. The good results of the IC suggests that each fish tool included in the exercise is in fact reacting in a common manner to a same level of human pressures, and providing a good agreement between methods in the diagnostic of ES particularly at the *High*/*Good* and *Good*/*Moderate* boundaries (Lepage M unpublished data). This is the ultimate goal of using fish in ecological assessments and suggests that all indices are relevant and valuable indicators in their own right. This also suggests that the failure to accommodate the different indices to a standardized dataset in this work clearly demonstrates the fundamental reliance of current fish indices on collection methods and design of monitoring programmes used in the development of the index. Therefore adapting indices developed for certain areas and methodologies to other situations remain impractical at best and invalid at worst. More widely applicable fish indices will probably require the formulation of completely new indices based on more flexible use of fish metrics according to system typologies and probably an increased use of functional traits (Breine et al. 2010, Delpech et al. 2010, Drouineau et al. 2012).

2. Sensitivity of metrics and indices possibly affecting robustness of fish-based ecological status assessment

A fish index should be sensitive to anthropogenic stressors ("signal") in a predictable manner but sufficiently robust to be minimally affected by other sources of variability ("noise") at different spatial and temporal scales (Rice 2003, Noges et al. 2009). The difficulty in detecting signals of anthropogenic stress from areas of high natural variability (natural stress), such as estuaries and lagoons, has been acknowledged as the so called "Estuarine Quality Paradox" (Elliott & Quintino 2007, Dauvin & Ruellet 2009, Elliott & Whitfield 2011). The results obtained during the review and comparison of current fish-based indices have highlighted some of the possible confounding sources of variability (noise) which might affect the robustness of these assessment tools. The review highlighted also a wide employment of expert judgement and "common knowledge" in assuming the metrics sensitivity to pressures, whereas more rigorous statistical procedures have been used only in recent indices. In addition, certain subjectivity in the anthropogenic pressure measurement is also present, due to the often limited availability of reliable pressure data or the use of expert judgement to score these. Therefore, the question remains upon the actual sensitivity and robustness of the metrics included in the fish-based indices (and of the indices themselves) and these aspects have been investigated in this project. The following sections summarize different exercises conducted during the lifetime of the project and are used to illustrate a general discussion on the overall robustness of fish-based ecological assessments.

2.1. Fish metric sensitivity to sources of variability

The potential 'noise' factors confounding biological quality metrics can be technical (i.e. those linked to the method of assessment) or natural (Appendix 4A). The resulting best models contained from three (metrics, marine migrant taxa and density of marine migrant taxa- probability of presence, in lagoons) to fourteen explanatory variables (metric estuarine resident taxa, in estuaries) but explained only a relatively small amount of the total variance of fish metrics (maximum 22% for lagoons, 40% for estuaries, Courrat et al. 2012a), suggesting that other factors (not included in the analysis) might contribute to the observed variability.

Nevertheless, the models indicate that metrics showed a significant sensitivity also to a range of technical and natural factors. There was a clear metric dependency in the selection of best explanatory models which indicates that sources of variability (noise) vary according to the metric tested. This is reflected in the different combination of explanatory variables making up the models. The implication for assessments is that different factors might then confound the metric-pressure correlation (signal 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 fish metrics. These parameters must therefore be taken into account in the ecological assessment process. For estuaries, latitude, longitude, source elevation, continental shelf width, size, entrance width, entrance depth, mean annual river discharge, wave exposure at the entrance and intertidal area may affect at least some of the fish metrics tested here (Table 2). Similar factor were highlighted in a similar assessment by Nicolas et al. (Nicolas et al. 2010b). For lagoons, longitude and total cross section of the inlets are the natural parameters explaining some of the between lagoons variability in our dataset. This argues in favour of a typology-based approach in fish-based assessments taking into account these natural parameters. Metrics of relative densities resulted in the highest, unexplained variance (between 99 % and 83 %) and consequently introduce larger uncertainty in the assessments.

Table 2: Best LM and GLM computed separately for lagoons (A) and estuaries (B). Only statistically significant fixed effects (Chi-squared test at 5 % level) are included in the models and only models showing a significant relationship with pressures (Pr) and explaining more than 10% of the metric variance are shown. When significant (Chi-squared test at 5 % level), effect of pressure metrics (regression parameter) is presented. Values given represent the slope. NS: non-significant. Abbreviations for explanatory variables in the models are provided in Appendix 4A. Agr, percentage of agriculture cover; Urb, percentage of urban cover; and Nat. percentage of natural land cover calculated on Corine land cover datasets using a 2km buffer zone around each WB.

Fish metric		**Model**	**Agr**	**Urb**	**Nat**
A) Lagoons					
SR_ER		*Sal class + Season + Sect + Longitude + Pr*	+0.009	NS	NS
DMM	*Probability of presence*	*Sal class + Longitude + Pr*	-0.047	NS	+0.026
	Density when present	*Sal class + Temp + Sect + Pr*	-0.035	+0.043	NS
SR_MM		*Sal class + Longitude + Pr*	-0.042	+0.019	+0.015
RD_O		*Season + Sect + Longitude + Pr*	NS	NS	+0.244
B) Estuaries					
TD		*Sal class + Depth + Season + Lat + Area class + Ent width + Ent depth + Discharge + Wave exposure + Intertidal area + Pr*	+0.015	NS	-0.027
SR		*Sal class + Depth + Season + Lat + Area class + Shelf width + Source elevation + Ent width + Ent depth + Wave exposure + Intertidal area + Pr*	+0.003	NS	-0.005
SR_ER		*Sal class + Season + Lat + Long + Area class + Intertidal area + Pr*	+0.001	+0.002	NS
DMM	Probability of presence	*Sal class + Depth + Lat + Area class + Source elevation + Ent width + Ent depth + Discharge + Wave exposure + Intertidal area + Pr*	+0.016	NS	-0.048
	Density when present	*Sal class + Depth + Season + Lat + Long + Area class + Shelf width + Ent depth + Discharge + Wave exposure + Intertidal area + Pr*	+0.016	NS	-0.046
SR_MM		*Sal class + Depth + Lat + Area class + Shelf width + Ent width + Ent depth + Discharge + Wave exposure + Intertidal area + Pr*	+0.003	-0.003	-0.004

Not all the studied metrics showed a significant sensitivity to human pressures, as measured by land cover data, and only three metrics (number of estuarine resident and of marine migrant taxa, and density of the latter functional group) showed a significant relationship with the land cover pressure proxies in both estuaries and lagoons (Table 2) (see Appendix 4A for the list of fixed factors included in the models). The nature of the relationship (positive or negative) changed with the type of pressure and with the metric, sometimes being contrary to what was expected (e.g. the positive effect of percentage of agricultural land on some fish metrics in estuaries). It is of note, however, that, even though significant, the relationships between metrics and pressures were generally very weak, as indicated by the value of the regression coefficients (Table 2). A number of reasons may be put forward to explain these results but probably the most likely is the type of pressure proxies used in the models. Land use data were used as they provide an objective yet indirect way to summarize human pressures (or lack of pressure, i.e. percentage of natural areas). Unfortunately some pressures affecting fishes in estuaries are poorly or cannot be adequately described using this method (i.e dredging, water abstraction, fishing, navigation, chemical pollution, etc). A more detailed discussion about the suitability of this pressure proxy approach is provided in section 3 on the modelling of reference conditions. Alternatively, the high remaining variability in the models could explain the relatively low sensitivity of fish metrics to the tested pressure indices.

It is of note that, despite the extended dataset used, some factors identified as potentially introducing large variability in the metrics (e.g. habitat type) could not be included in the models due to the lack (or reduced availability) of such data. Moreover, the estimation of interactions between sources of variability was not possible. All these factors which could not be included in the analysis might have a highly relevant effect on the fish metrics. Franco et al. (Franco et al. 2006a) for example, highlighted an important effect of habitat type (e.g. in terms of bottom structure, due, for example, to the presence of seagrass vegetation) on fish assemblages variability within lagoons, leading to significant changes in the abundance and species richness. Although fyke nets used in lagoons allow for a certain integration of fish catches over space (hence over different habitats present in the sampling area) and time compared to other sampling methods (Franco et al. 2012) the overall habitat availability and diversity in different lagoons might affect the total catches and the fish assemblage guild composition (Franco et al. 2008b). In turn, the habitat type-effect might be more relevant within estuaries, given that generally the assessments are conducted by bottom trawling which targets specifically demersal fish assemblages, with a likely higher association with the type of bottom habitat where the haul takes place.

Mixed models using type of WB (estuary or lagoon) as a random factor demonstrated that for all metrics where analysis could be conducted, unexplained variance remains generally much higher within individual systems than between systems suggesting a great importance of sources of variability acting at the within-system level. The within-system variance may be due among other factors to the behaviour of some fish population. Fish species are not uniformly or randomly distributed in nature, but show a variable degree of patchiness. This results in high variance between fishing events within a system and to a lower variance when comparing two systems in a close geographical area. The high within-system variance may be again explained by the lack of data on habitat type in the formulation of the models. Without habitat-characterising data, the variability across samples within a water body due to habitat differences remains unexplained. This within-estuary or -lagoon variability must be accounted for to decrease the uncertainty on the values of fish metrics and thus on the assessment. This may be done simply by increasing the sampling effort or collecting detailed environmental data and habitat characteristics during sampling (i.e. find relevant covariates to account for unexplained variance). This may not be possible due to the multiplicity of possible influential factors. The within system variability includes for example spatial factors, such as the salinity changes within estuaries and lagoons or depth changes within estuaries, and temporal factors, such as the seasonal variability of fish assemblages. Salinity is considered the primary factor affecting fish fauna variability within estuarine and lagoon environments by many authors (Da Sylva 1975, Weinstein et al. 1980, Wagner & Austin 1999, Whitfield 1999, Franco et al. 2006b, Whitfield et al. 2012). It usually influences fish species distribution in relation to their tolerances (Marshall & Elliott 1998) affecting particularly the distribution of those species which are less tolerant to the salinity fluctuations in the estuarine environment (Whitfield 1999, Whitfield et al. 2012). Depth can also affect species and guilds distribution within estuaries, e.g. shallower water habitats are preferably used as nursery areas by several marine species, leading to a general increase in species richness with decreasing depth (Paterson & Whitfield 2000, Smith & Brown 2002). The lack of importance of this factor within lagoons is likely to be due to the general shallowness characterising these systems. Also important temporal patterns of variations within estuaries and lagoons affect the studied metrics, due to the seasonality of both recruitment of estuarine residents and colonization of lagoon environment by marine migrant species (McErlean et al. 1973, Allen & Horn 1975, Hoff & Ibara 1977, Knox 1986, Loneragan & Potter 1990, Potter et al. 1990). Although the within-system variability has proven to be important in affecting the studied metrics, hence their robustness in responding to human pressures, the effect of this source of variability can be reduced not only by removing variability through linked explanatory variables (season, salinity, depth, etc) in models but also by modifying monitoring protocols in order to characterise the metrics in standardised conditions (e.g. during a particular season and with stratified sampling designs taking into account the salinity classes and depth levels within the system). It is of note that these sources of variability are likely to affect the reference conditions for the different metrics, hence some authors have suggested the calibration of fish-based indices separately for specific seasons, different sections of an estuary (accounting for the salinity gradient in it) and habitat types (Deegan et al. 1997, Breine et al. 2007, Coates et al. 2007, Franco et al. 2009).

Between-system variability is generally associated with differences in location of the transitional water systems (longitude for lagoons, latitude for estuaries), and in their characteristics such as morphological and physico-chemical characteristics, total cross section of the inlets in lagoons; source elevation, continental shelf width, size, entrance width, entrance depth, mean annual river discharge, wave exposure at the entrance and intertidal area in estuaries. Geographic location of a water body can be important in affecting the species composition in estuaries and lagoons, due to biogeographic differences in the species distribution ranges (Elliott & Dewailly 1995). Theoretically, species richness decreases with increasing latitude, this effect having been verified in many studies for marine (Poore & Wilson 1993) estuarine (Pease 1999, Harrison & Whitfield 2006) and freshwater fish (Oberdorff et al. 1995). However, it is of note that previous studies on European estuaries (Nicolas et al. 2010a, Nicolas et al. 2010b) did not confirm the significance of this factor in explaining estuarine fish species richness. Also longitude has been observed affecting species richness in Mediterranean lagoons, this effect being related to the presence of areas regarded as sources of colonizing species from adjacent regions along the longitudinal gradient (e.g. the Straits of Gibraltar in the Mediterranean) (Franco et al. 2008b). However, this effect is likely to be reduced when considering functional groups instead of taxonomical identities (Elliott et al. 2007, Franco et al. 2008a). Also, factors affecting the connectivity of estuaries/lagoons with the sea (e.g. estuarine entrance width and depth, cross section of lagoon inlets) or with freshwaters (e.g. river discharge) have been shown to have a significant influence on the functional structure of fish assemblages in transitional waters, particularly affecting those species whose life cycle depend on this connectivity (e.g. marine migrant species) (Pease 1999, Franco et al. 2008b, Nicolas et al. 2010a, Nicolas et al. 2010b). Nicolas et al. (Nicolas et al. 2010a) also showed the importance of continental shelf width in affecting fish species richness in European estuaries, as a wider continental shelf is likely to shelter a greater surface and variety of spawning grounds for different fish species that are likely to enter estuaries as juveniles (Beck et al. 2001, Able 2005). The distinction of estuaries and lagoons into different body types (e.g. according to ecoregion, area, etc.) may partly reduce the metric variability between transitional water systems (as the metrics and their reference conditions would be defined separately for different water body types). Although the inclusion of additional characteristics and factors (e.g. accounting for the connectivity with marine realm) into the definition of water body types might further reduce this variability, this would lead to a high fragmentation of the water bodies of each Member State into multiple types (considering the combination of all the levels of the factors), thus reducing the basis for inter-systems comparisons and hence the power of the analysis which can be applied to the data.

An alternative to the modelling approach, was test trialled in Drouineau et al. (Drouineau et al. 2012). The Bayesian method allows the ability both to select and combine fish metrics taking into account their variability, their sensitivity to pressure or any other relevant feature. For example the method can also be a way to integrate data from expert opinion and it finally gives an assessment of the uncertainty of the diagnostic tool. It was tested on a dataset composed by a sample of 14 French lagoons. The analysis suggests that the quality diagnostics are less variable at the level of the multi-metric indicator than at the level of the fish metrics considered individually (Figure 4). As the uncertainty analyses realised at the fish metric scale in the present work suggested that uncertainty on fish metric may be high, this last result is encouraging and further research on the propagation of uncertainty from fish metric to multi-metric indicator is required, for example it would be of value to test the Bayesian method on other datasets.

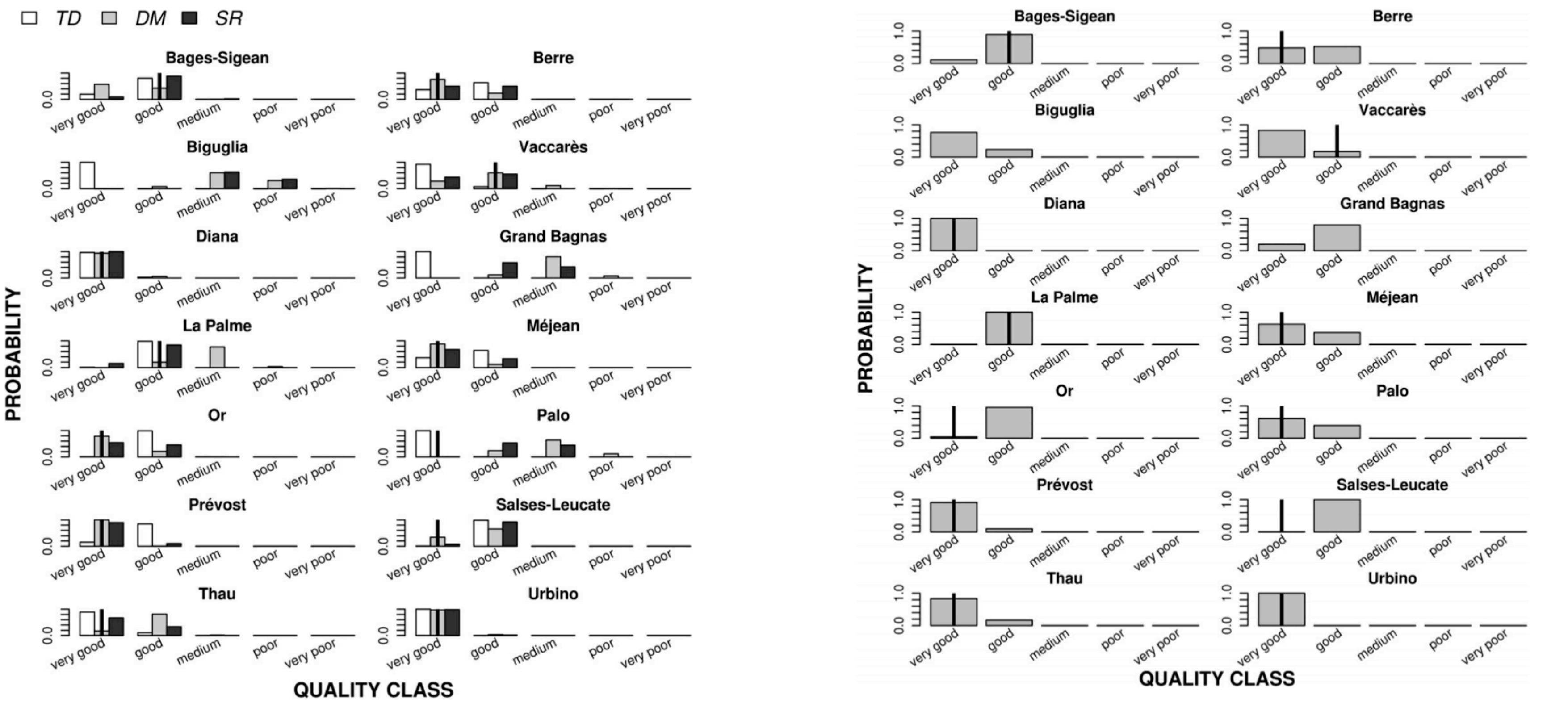


Figure 4: Bayes posterior probability to be in a quality class given; left fish metrics TD (total density), DM (density of marine fishes) and RT (species richness); and right, the corresponding multi-metric index (3 metrics combined). The vertical bold line represents the quality class attributed by a method based on physico-chemical parameters.

It is also important to note that modelling results highly depend on the dataset used. The present work aims to indicate the potential sources of uncertainty affecting fish metrics but the significance of their effect on the tested fish metrics cannot be generalized. As an example, the absence of latitude as a factor affecting lagoons fish metrics is due to its exclusion from the analysis, given the restricted latitudinal distribution of lagoons in the studied dataset. In turn, latitude proved to be the best descriptor for different water body types when considering the functioning of a wider range of Mediterranean lagoons for fish feeding and reproduction (Franco et al., 2008a). Therefore, similar analyses as those carried out in this study should be made on the particular datasets fish indicators are designed for.

The effect of sampling effort on fish metrics could not be assessed in the previous analysis, due to lack of data. Elliott and Dewailly (Elliott & Dewailly 1995) indicated that, for example, sampling effort was the primary influence of recorded estuarine species richness and so any metrics including this would be affected. It is also of note that none of the indices tested before provide any estimation of the risk of error linked to sampling effort (see section “Consistency of assessment results”). However, this factor might have an important effect on the robustness of fish metrics, and the knowledge of its effect might lead to suggestions toward the optimisation of fish monitoring protocols for the implementation of the WFD. Therefore, the effect of sampling effort (measured as number of hauls) on the fish metrics was specifically addressed in a case study analysing the response of metrics from the Portuguese EFAI index to increasing fishing effort (see Appendix 3).

Results showed that sampling effort can be an important source of variability in fish metrics of the EFAI index, especially metrics on number of species, which are common to several other fish-based indices (Figure 5). In turn, metrics based on percentages (relating to 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 (Figure 5). The stabilization of metrics based on species richness changed among 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.

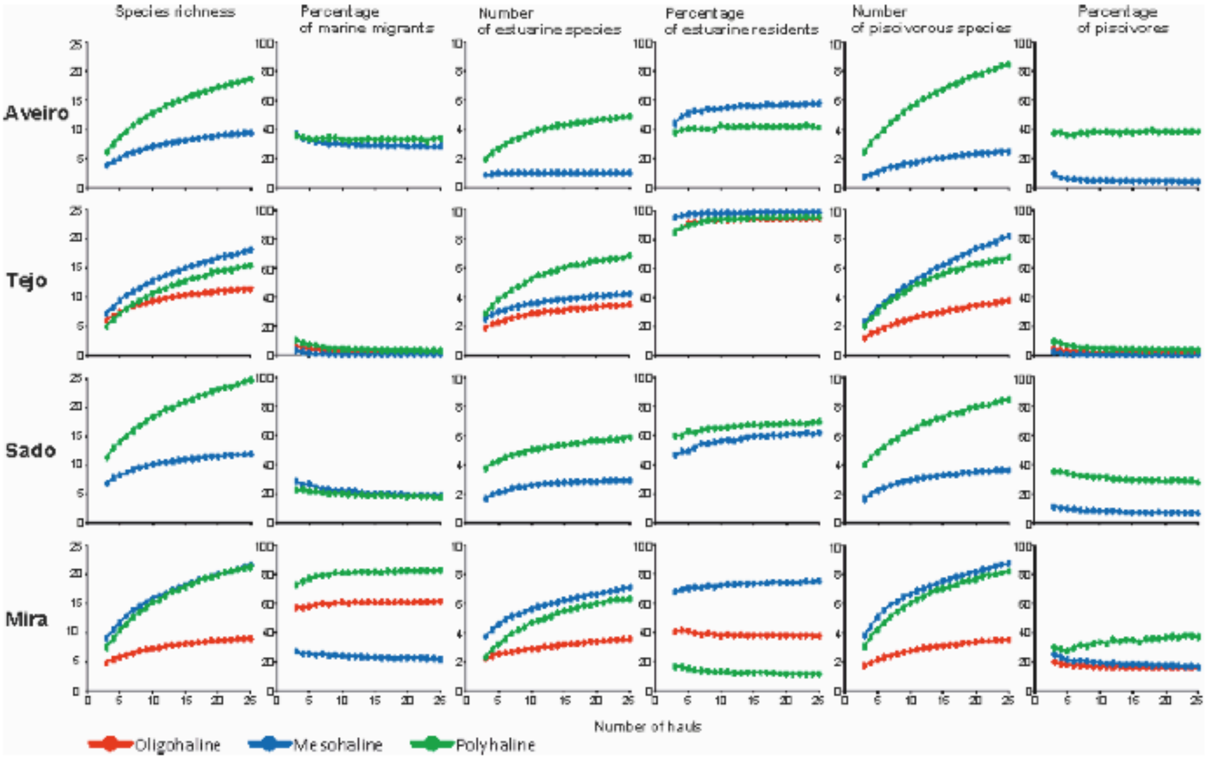


Figure 5: Mean values per salinity zone calculated for different metrics of the EFAI index when a different number of hauls is considered.

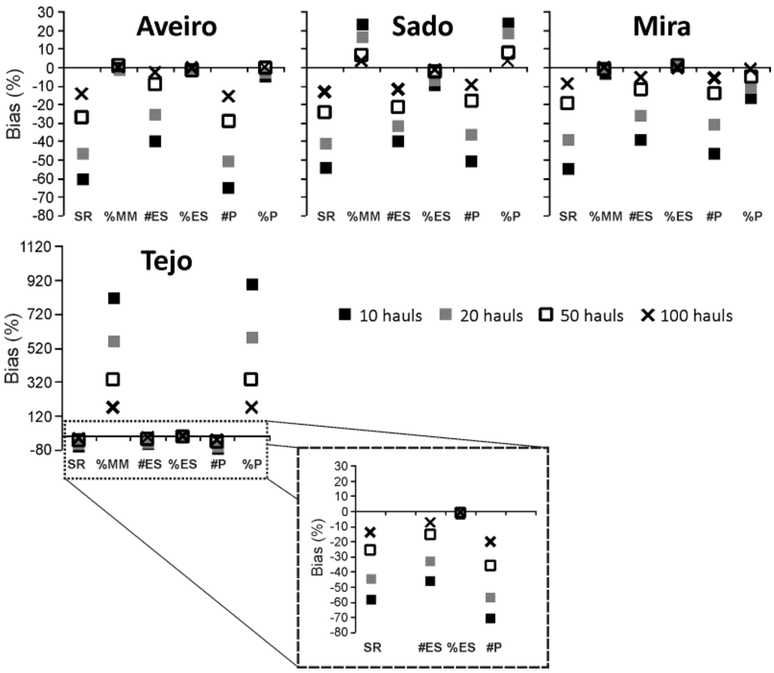


Figure 6: Bias calculated for each metric, per estuary, using 10, 20, 50 and 100 hauls.

A decrease in the metrics bias was generally observed with increasing sampling effort, although it is of note that metrics based on percentages generally showed a smaller bias, even using the smallest number of hauls, than the metrics based on numbers of species (Figure 6). The bias/cost analysis showed that a reduction of sampling bias may be associated to relevant increases (up to 300%) in sampling costs due to the increase of number of hauls and associated processing of the samples. These findings may have implications for the WFD assessment of transitional waters since as they show that sampling effort greatly influences the expectation linked to diversity-based metrics (Hoff & Ibara 1977, Elliott & Dewailly 1995, Harris 1995). This effect is caused by including in the analysis species with an apparent abundance below a certain threshold which prevent the complete characterisation of their presence. These rare species, in some cases a single individual collected on a single occasion, would only be incidentally recorded and therefore add random variability to diversity-based metrics. This in turn affects the relative scores and the outcomes of the assessment. Similar effort related bias may be an issue for density based metrics if fish distribution is very patchy (i.e. schooling fish or aggregated in specific habitats) and not enough replicates are taken to fully characterise the patchiness in their distribution. It is clear that to overcome a potential large source of error, the reference conditions must be defined according to the level of effort used in the monitoring programme or, conversely, the monitoring must be done to the same level of effort used to derive the reference.

The better, more accurate, response of fish metrics, hence possibly of the multi-metric indices, may be obtained simply with more intensive (larger sampling effort and hence greater cost) monitoring programmes, particularly when assessing species richness or spatially or temporally heterogeneous distributions. A higher sampling effort usually allows a higher probability of capturing rare, or less abundant species, hence leading to an increase of the total number of species captured (Krebs 1999, Wootton 2001). This may indirectly reflect also on the functional composition of assemblages, through the influence on those categories, which typically include rare, occasional species, as the case of marine (often piscivorous) stragglers (Franco et al. 2008a). Therefore, to reduce the risk of misclassification and to increase accuracy may be simply a matter of budget, where increasing the number of samples used in the assessment so as to ensure a desirable low uncertainty level might not be economically viable. Improving accuracy without having to increase efforts may be possible by greater use of proportion metrics or the use of less selective gear sets or multi-gear approaches. Alternatively, a more pragmatic decision could be made based on probability of capture, thus considering in the analysis only those aspects for which the sampling method and level of effort allows for a reliable and quantitative estimation. Two possible options are available: (1) weighting of metrics by precision or by species relevance (i.e. species reference lists as in Breine et al. 2010), or (2) pooling samples to get sampling events affording greater habitat or temporal integration (i.e. larger effective samples).

Finally, and although we made the decision to focus firstly on the individual behaviour of the metrics, it is important to consider these not in isolation but together as they are combined in the formulation of multi-metric indices. It is known that metrics are, to some extent, correlated in many indices as some quality attributes reflected in one metric are to some extend included in others (Borja et al. 2012). This redundancy is often dealt with at the early stage of the index development by not using one metric from highly correlated pairs (Breine et al. 2007, Breine et al. 2010, Pérez-Domínguez et al. 2012). However, in other cases where the strength of the correlation is less, this decision is or should be moderated by ecological knowledge. In fact it is necessary to leave some level of redundancy to improve the response of the index across a wider range of quality conditions. That is, a metric may stop responding when the quality reached a certain level but a related metric may complement this lack of sensitivity extending the dynamic range of the index. For example in a severely impacted system, a metric based on sensitive or indicator species may not be at all adequate and probably not responding at the levels where a metric based on trophic composition or diversity is still appropriate. It is possible that at moderate levels of disturbance both type of metrics show some degree of correlation but the combined information captured in the index justify the inclusion of both. No one metric appears to respond to a single pressure only, which is an additional argument in favour to maintain a certain degree of correlation/redundancy in the indices. This level of judgment is very difficult to obtain through statistical modelling of individual metrics and there is the need to determine methods of testing the behaviour of the index under different metric scenarios and explanatory variables. The modelling approach we have used in this section is adequate to explore expected sensitivity for the combined effect into a predetermined index but cannot provide the necessary analytical feedback to improve the metric selection other than trial and error. Choosing the 'correct' set of metrics for the formulation of a useful and robust index is a large task and should, at least in part, be based on those scenarios relevant from an ecological standpoint given the pressure field affecting the systems. An early appraisal of pressures is therefore always needed.

2.2. Fish index sensitivity to pressure gradients

The conceptual analysis carried out on the strength of expected metrics responses to a set of human pressures suggested chemical pollution and loss of habitat as the type of pressures more frequently and more strongly related to fish metrics. These pressures are often regarded as important indirect causes of alterations in transitional water fish assemblages (Vasconcelos et al. 2007, Uriarte & Borja 2009). As a result, a higher sensitivity (in terms of strength of the response) of fish based indices to these two pressures is expected (Appendix 4B). With regard to the time lag in response to human pressures, most fish metrics are expected to have a delayed response to temperature and flow changes, loss of habitat and chemical pollution. The expected results for the fish indices were highly variable, although, on average, temperature and river flow changes were the pressures with longer time lag in the expected response of the fish-based indices (Appendix 4B). This preliminary analysis provided the conceptual basis for the ranking of human pressures in order of expected relevance to fish in transitional waters. In order to illustrate the relationship between fish-quality attributes and pressures further, the AZTI's Fish Index (AFI) measured for the Basque country (Spain) estuaries and the EFAI for the Portuguese estuaries were related to a set of pressures acting in these water bodies, while also considering their hydro-morphological descriptors (see Appendix 4C for a complete list of the variables considered in the analysis).

Multiple regression analysis indentified the following best model (Adjusted R^2= 0.859, p< 0.05) relating AFI index scores to the considered independent variables:

AFI = 0.013 + 0.017(average estuary depth) – 0.003(global pressure index) – 0.001(residence time) + 0.028(dredged volume) – 0.007(percentage of channelling in ports) + 0.009(percentage of channelling out of ports).

It is of note that AFI was specifically designed to be applied using autumn data (Borja *et al.,* 2004), in order to sample the fishing period with a stable community of fishes. Consequently, the regression of AFI and pressures produced a significant model only when considering autumn data (and not spring and summer). This suggests that the application of this index to data from other seasons of the year might be non-valid.

The obtained model indicated that 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. Deeper estuaries are usually larger or have more volume (at comparable area) which are likely to allow for a greater diversity of niches and better migration routes, and these factors likely increase the fish species richness (Elliott & Whitfield 2011, Basset et al. 2012). Deeper estuaries, in fact, often support more resident species and stable populations (França et al. 2009, Uriarte & Borja 2009). Residence time is related to the capacity of the system to retain pollutants and organic matter, in turn influencing the levels of dissolved oxygen, which are important for fishes, as demonstrated in the Basque estuaries (Uriarte & Borja 2009) and others (Elliott & Hemingway 2002). As far as regards human pressures, the AFI clearly responds in a significant way to morphological pressures, which may alter the niches availability, hence leading to significant effects on fish communities (Elliott & Hemingway 2002, Coates et al. 2007). However, it is of note that multiple regression shows also that the more volume dredged and the more channelling out of ports, then the higher the AFI values (higher quality). This may be a spurious correlation, as dredging is only important in some parts of the deeper estuaries, and might be driven by some co-linearity between the two variables. For example, the higher AFI may occur in larger estuaries which support a port and also in more muddy estuaries which require more dredging.

The correlation analysis carried out on the EFAI index values (EQR) assessed in Portuguese estuaries highlighted the negative response to the overall pressure acting in the studied estuaries (as obtained by using two pressure indices integrating the contribution of different pressure indicators; Appendix 4C) (Figure 7). The Portuguese index EFAI responded to the overall anthropogenic pressure level, although limitations in the dataset, in terms of the range of pressures present in the studied estuaries, possibly means that these results cannot be applied in general across all systems. This is especially the case as the pressure levels characterising the Portuguese estuaries considered in the study are generally low. Hence the results cannot provide clear understanding of the behaviour of the tool for higher degradation levels. Therefore, it is recommended that sites covering the full pressure gradient (using scale of Aubry & Elliott 2006) and different specific pressures acting in different types of estuaries should also be considered in further studies.

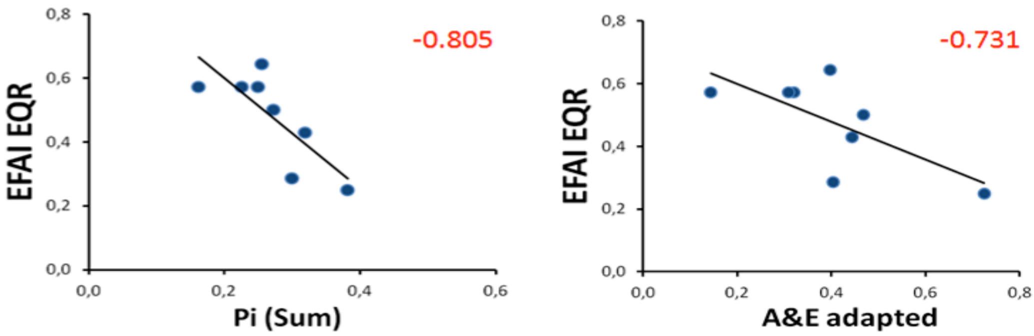


Figure 7: Response of EFAI against the overall pressure (measured as Pi (Sum) – sum of pressures, and as A&E adapted – pressure index adapted from Aubry & Elliott 2006 (see Appendix 4B). The regression coefficient is given in the figure.

These results obtained with the AFI and EFAI along with the evidence of the IC exercise suggest that, using the appropriate matching combination of fish index, reference values and local dataset, multi-metric fish indices can be considered sensitive to pressure gradients. Further confirmation was obtained with the

With regards the results obtained for the model relating the AFI index to human pressures in the Basque country estuaries, it is of note that a lower uncertainty (degree of unexplained variability) was recorded compared to the results obtained in the analysis of metrics sensitivity to pressures. In addition to the advantage of applying the index to the types of estuaries in the geographical area for which the assessment tool was designed, other factors might determine the better response of the model relating the fish based index to pressures obtained for the Basque and Portuguese case study. On the one hand, there is a relatively high similarity in terms of morphological features in this set of water bodies, hence possibly contributing to reduce the uncertainty in the index response to pressures. On the other hand, it should be noted that more specific pressure indicators have been considered in this analysis rather than the more generic land cover data considered before. The pressure indicators used for the Basque country, and also for the Portuguese case study, have been selected based on their specific possible effect on fish fauna in transitional waters (Deegan et al. 1997, Meng et al. 2002, Aubry & Elliott 2006, Borja 2006, Breine et al. 2007, Vasconcelos et al. 2007, Uriarte & Borja 2009), hence a higher sensitivity of fish based indices to such types of pressure indicators is expected. In turn, land cover data, being more generic proxies for pressures in transitional waters, might not be as good as indicators of actual pressures affecting fish assemblages (hence the response of assessment tools based on this biological quality element) in these systems. As introduced earlier, dredging, water pollution and many other pressures are not included in the land cover proxies. The advantage of measuring specific pressures relies also on the need for indicators which give detailed information on the cause of change together with the effects of the change observed (McLusky & Elliott 2004). Since the success of mitigation and restoration plans depends on our ability to minimise the effects of stress, any assessment tool that can both determine conservation status and diagnose specific damaging pressures can potentially provide cost and time savings for resource managers. The use of too-generic pressure indicators (such as land cover data) would increase the difficulty of deciding how to intervene to improve the ecological status (as indicated by the fish index response) by reducing the pressures. However, it must be acknowledged that data on detailed pressure indicators as considered above may not be available for many transitional water bodies, leading to the impossibility to test their relationship with fish assessment tools on a wider context (and, as shown above, there are some limitations on the generalisation of the above results, due to the specific dataset and indices tested). The advantage of using the CORINE land cover data as proxies for pressures is that they, although showing a weaker relationship with fish based metrics (hence possibly with indices), allow for a standardised characterisation of pressures over a wider range of transitional water bodies. Therefore, it brings to the assessment the necessary standardization across all transitional systems included in these analysis.

Up to date the sensitivity of indices to gradients of degradation in transitional waters has been tested using regression analysis. This same approach has been used here to test EFAI and AFI response to pressure proxies. The Bayesian approach presented in the previous section (Fish metric sensitivity to sources of variability) is equally relevant to assess metric-pressure relationship (Drouineau et al. 2012). The method relies on modelling the response of fish metrics using pressure as response variables which is identical in concept to the approach used to model reference conditions (see section 3 in the discussion). The use of the approach in this capacity, i.e. sensitivity assessment, allows for the selection of relevant metrics or to combine them in multi-metric indices. There was a good agreement between ES class assignments and pressure proxy (contamination data in this exercise, *a priori* more relevant to fish than land cover proxies) which suggest a sensitive index (Figure 4). The important addition of the Bayesian approach to ecological quality assessments is that provides a measure of confidence in the ES class assignment which is a necessary feature in any management plan.

2.3. Fish index sensitivity to metric changes

Sensitivity analysis in this context is the analysis of the response of multi-metric indices to the manipulation of metric scores to values higher or lower than its average under realistic scenarios of change; this therefore identifies the most influential metrics in determining the index response. To our knowledge, this sensitivity study has never been used before in ecological assessment. The sensitivity analysis is described further in Borja et al. (Borja et al. 2012) but in essence it consisted in the shift of average metric values to extreme percentiles according to their covariation with the other metrics included in the multi-metric index under assessment. This creates realistic scenarios of change driven by a metric that can be explored in different ways. For example it could provide indication of the magnitude of metric improvement (or worsening) to produce a quality status change and whether reaching the threshold for the change is realist. Part of the usefulness of this exercise is its computational simplicity and the graphing technique used, the ‘tornado’ diagram (Figure 8). This novel visual tool provides an effective way to convey the influence of each component metric on the determined ES using the fish BQE. The methods is used in economy to help understand the implication and risks associated with economic models and to prioritize investment decisions. Analogous use is here proposed for the first time to characterize the expected behaviour of multi-metric fish indices and inform management decisions. For example, results of this analysis may be used to guide the implementation of management/conservation plans, e.g. by prioritising the restoration and/or conservation of underlying fish quality futures underpinning these influential metrics. This is important since, for example, the WFD requires all water bodies to reach the *Good* status.

The analysis indicated that the TFCI index is especially sensitive to manipulations of M1 (species composition), M4 (number of taxa that contributes 90% of the abundance), M5 (number of estuarine resident taxa), M6 (number of estuarine-resident marine taxa), M8 (number of benthic invertebrate feeding taxa) and M9 (number of piscivorous taxa), whereas the ELFI index showed a higher sensitivity (under all scenarios) to DT (total fish density), DB (density of benthic species) and RT (total species richness) (Figure 8) (Appendix 4D).

In considering the UK waterbodies included in the analysis, the results for the TFCI index showed that the minimal changes required to bring the overall water body classification to *Good* will be obtained by increasing M4 (number of taxa that comprises 90% of the abundance) and M6 (number of estuarine-resident marine taxa) to the average of the top 40 percentile of the sample population (Figure 8). This has been estimated to be an increase to score value of 5 for M4 and 4 for M6 from their current average value score of 3 for both (the actual increase in the original metric value being dependent on the water body type assessed, hence on the type-specific reference conditions defined for it). It is of note that efforts to improve M4 (e.g. to top 10 percentile average) would be of little value, given that they would not lead to any relevant additional improvement in the final status assessment, contrary to that observed for metric M6. This information can be used to provide targets for management purposes that may be bound within what it is possible to achieve given the current status but can be used to set more ambitious management goals when conditions improve. With regard to the ELFI index and the French water bodies included in the analysis, the results indicate that a higher effort (hence higher costs) would be required to reach a *Good* status compared to the TFCI results. With ELFI, metric improvement is effective in influencing quality classification only when increasing the total density of species and especially the density of benthic species (metrics DT and DB) to their highest possible values (at least to the top 10 percentile average value) (Figure 8). Hence, the scores of these metrics need to improve from their average score value of 1.5 and 2 respectively to a score of 4 for both.

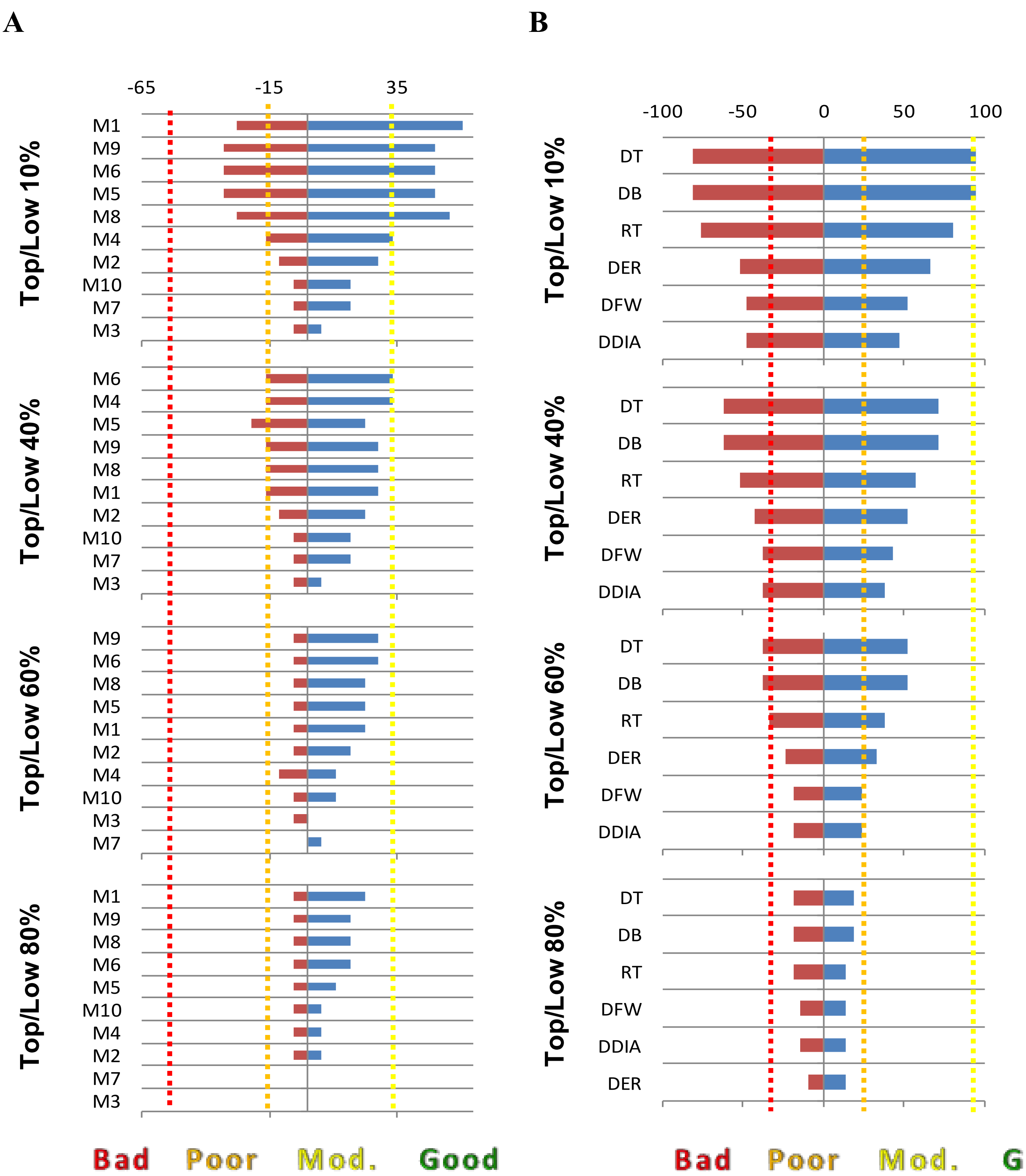


Figure 8: Tornado diagram showing the percentage change of the TFCI (A) and ELFI (B) index from its average value (central axis) under the different scenarios of metric variation. Low (red bars) and top (blue bars) scenarios for each of the 10, 40, 60 and 80 percentiles of metrics distribution in the studied dataset are shown (the greatest effect is therefore expected for the Top/Low 10% and the lowest for the Top/Low 80% which will result in the bars closer to the central axis). Vertical coloured lines indicating the threshold for the different WFD classifications are also shown. See Appendix 2A for metrics abbreviations.

It is of note that the above results are highly dependent on the distribution of the metrics scores in the range defined by the data set used (Appendix 4E). For example, tornado diagrams for the TFCI resulted in a stronger effect towards the upper percentiles compared to the lower percentiles. This is due to a skewed distribution of several metric scores throughout the data set. Nevertheless, based on the actual WFD data, the distribution of the metrics is determined by the actual scores recorded and represents a realistic appraisal of the range and expected frequency of the different scores. It can be also said that given the sample size we expect to have a range of quality scores including the best and worst metric values that can be expected. Furthermore and since the indices have proven to show a response to human pressures (i.e. IC exercise, Lepage & Coates 2011), the observed metric scenarios could be assumed to reflect the dynamic range expected from the fish tool under relevant human pressure gradients. If this is the case, we could use the boundaries of the different scenarios to set realistic management targets and also to identify the aspects of the indices that are more likely to affect the outcomes leading to more robust and responsive indices.

Overall, the behaviour of multi-metric indices under manipulations of metric scores clearly indicated that metric type, number of metrics used and correlations between metrics are important in determining the index performance, with indices including more and/or uncorrelated metrics or metrics with skewed distribution being less affected by extreme metric manipulations. It is still not clear whether this lower sensitivity indicates an increased robustness of the indices to spurious metric scores or a lack of sensitivity to pressure gradients. Comparisons with pressure proxies will be needed to allow further interpretations of the behaviour of the multi-metrics and their component metrics. However, considering that the indices, if successfully calibrated towards anthropogenic pressure gradients, are designed to be a proxy of pressure, we suggest this analysis presents a realistic appraisal of the expected inertia (i.e. tendency to buffer ES change after quality alterations) and dynamic range (i.e. ratio between the largest and smallest possible values) of a multi-metric fish index. If the conditions deviate strongly, improve or deteriorate, from the average (i.e. the current situation) the index outcomes will probably still be bound by the limits identified in this exercise. The relative simplicity of this approach makes it very easy to understand by non–specialists and the tornado plots a simple way to highlight quality aspects on which the restoration efforts will produce the most desirable effects. However, as this is just a hypothetical exercise, it does not provide a true indication of change at the level of individual water bodies, the assessment should be taken only to provide a general interrogation of the multi-metric index behaviour.

Finally, the analysis has offered also a simple means of comparing the behaviour of two of the currently available WFD fish indices in the resulting ecological status classification of water bodies. The average classification of water bodies by the French ELFI index falls in the *Poor* boundary while the UK TFCI average classification falls in the *Moderate* boundary. This indicates that remediation processes to bring water bodies to the desired *Good* status will need to be stronger in France than in the UK. However, it is also expected the ELFI to show a greater reaction which may compensate at least in part the wider gap to reach the *Good* status assuming the ES classification obtained by these indices is comparable. Taken together with indices sensitivity to pressure, the information provided by this relatively simple technique is extremely valuable to evaluate the likelihood to achieve a predefined quality status, to link expected index change to specific pressure types and to assess the usefulness of individual metrics on conservation targets in management plans. The challenge remains as how to integrate all these lines of evidence to further improve and refine current methodologies.

3. Modelling reference conditions

The WFD proposes four approaches to built type-specific biological reference conditions: (i) the spatially-based identification of good quality sites, i.e. existing undisturbed sites or sites with only very minor disturbance; (ii) modelling by using hindcasting methods (based on historical data and information); (iii) predictive modelling of the community against the physico-chemical variables, or, (iv) expert judgement (as a last resort according to the wording of the WFD). A combination of these methods can also be used although as shown by (Hering et al. 2010), each of these poses difficulties. Given historical human pressures within most estuaries worldwide, a reference computed only on current data would be set at a somewhat reduced quality status compared to the pristine system (Hering et al. 2006). The adjustment of the reference community to reflect extant good or high status communities and functioning is therefore extremely important and, until now, this is still largely based on expert judgment. Even in the current situation where pristine systems are generally lacking, predicting the expected community before human intervention may be a possible method to correct the changed baselines where historical data are available (Borja et al. 2010, Breine et al. 2010). In addition, data-driven logistic regression models where metric outputs are correlated to environmental and biological factors could provide the necessary predictive power to derive statistically-significant models of original reference communities (Maes et al. 2007). In this context, the aim of the present modelling work is to test a predictive linear modelling approach (LMM and GLMM, Bolker et al. 2009) to define reference conditions for fish metrics in transitional waters in Europe.

In particular, there is the need to predict the changes in the fish community or elements of it (the dependent variables) using the independent variables of the physico-chemical environment. It is acknowledged that the latter go some way to explaining the former by indicating the relationships between the environmental tolerances of the species and the number of niches available for colonisation. However, given that the resultant biological community is then the result of the physico-chemical environment over which is superimposed the biological interactions (such as competition, predator-prey cycles, recruitment processes both inside the estuary and elsewhere) then such models will only ever have the ability to predict part of the variability in the ecological data (Gray & Elliott 2009). Further development of these models to include the biological interactions and the influences in the adjacent parts of the system (the catchment and marine areas) would therefore be valuable.

The identification of the fixed factors in the LMM and GLMM models gives a first indication of features driving the fish metric and hence the community present in transitional waters. The idea behind this exercise is then use response models to derive a prediction for the expected value of the reference for each type of transitional water body. Significant mixed models (LMM and GLMM) relating metric response to pressures (as assessed by land cover data) were obtained only for three metrics, namely the number of marine migrant species (SR_MM) and density of benthic invertebrate feeder fishes (DIB) assessed in lagoons, and percentage of omnivorous fishes (RD_O) in estuaries (Table 3). The variability components ascribed to the non-pressure factors included in the models were considered in the modelling of reference conditions, e.g. by calculating separately reference conditions for different salinity classes and season. The effects of these factors are generally very influential in defining the reference value, thus highlighting the importance of standardizing any monitoring programme to ensure an accurate quality evaluation of transitional waters using fish BQE.

Table 3: Best model selected after mixed model analysis computed separately for lagoons (A) and estuaries (B). Only statistically significant fixed effects (Chi-squared test at 5 % level) are included in the models and only models showing a significant relationship with pressures (Pr) are shown. When significant (Chi-squared test at 5 % level), effect of pressure metrics included in each mixed model is presented (values given represent the slope.). - indicates that the pressure is not included in the specific mixed model (see (Courrat et al. 2012b) for details of the analysis). Abbreviations for metrics and covariates in the models are provided in Appendix 3A and 5A, respectively.

Fish metric	Model	Agr	Urb	Nat
A) Lagoons				
SR_MM	*Sal class + Pr*	-0.043	-	+0.013
DIB	*Sal class + Season + Sect + Area + Pr*	-	-0.06	-
B) Estuaries				
RD_O	*Sal class + Season + Discharge + Intertidal area + Pr*	+0.041	-	-

As observed in the metric sensitivity analysis, even when significant, the effect of pressures on fish metrics was generally very weak, as indicated by the low regression coefficients. This might indicate a generally low sensitivity of the modelled fish metrics to human pressures. However, it must be considered that the results highly depend also on the choice of the pressure indicator, and that more specific indicators (in term of their ability to indicate pressures likely affecting fish assemblages in transitional waters) might provide a more accurate assessment, in addition to the limitations in data availability (as discussed in section 2.2 regarding land cover proxies limitation in estuarine fish assessments)

Ideally the reference value could be set independently from pressure proxies if there is sufficiently good information of species responses across natural environmental gradients at pristine sites. As indicated above and elsewhere (Hering et al. 2010), the reference condition can be defined as either the *presence* of good ecology or the *absence* of pressures. Although the latter is more costly to demonstrate and requires a large and rigorous sampling procedure, in this case, the reference value would be indicating good biology rather than absence of pressure. However, this information is seldom available for transitional waters. Instead pressure-fish response predictive modelling represents an alternative way to forecast the expected theoretical reference at a zero level of pressure. Nevertheless, as there is no transitional water body in such conditions, results obtained this way may be inaccurate as they require an extrapolation outside the limits of the models. A compromise is to set the reference to the level of the least impacted sites. This increases accuracy but produced a reference condition set at an artificially diminished quality status which may be far from the true reference condition. Both of these predictive approaches (zero or minimum pressure) were applied here to establish reference conditions for the main fish metrics included in transitional water.

The two theoretical reference conditions predicted for each metric using the above models (by setting pressure levels to the lowest observed pressure in the dataset or to zero) are shown in Appendix 5B. As expected, the values of the metrics generally increase at reference conditions compared to the average metric values in the dataset. The theoretical zero pressure reference has, generally, the largest predicted values although an opposite response is observed in some metrics and land cover pressure proxies. The largest difference between the two reference thresholds in the model were found in the species richness metric (SR_MM), therefore the choice of the reference condition (either based on the absence of pressure or on the least impacted site) is likely to have a relevant impact on the metric EQR, hence possibly affecting the overall assessment in fish indices using that metric. In turn, this effect seems minor when considering the other two metrics assessed (based on density), as the difference between the two reference conditions is relatively smaller.

Density-based metrics then appear to be a more robust choice in explaining and reflecting the nature of the community but in some cases these metrics have shown average values in the dataset higher (or lower, depending on the particular metric response) than the reference condition itself. This effect was due to the presence of extreme values of the metrics in the dataset (e.g. very high abundances of a certain species recorded occasionally) which highly affected the calculation of average metric values. This influence of extreme observations is likely reduced in the predictive models for reference conditions due to the transformation applied to metric data. Modelling density values based on count data are especially sensitive to observed patchiness in species distribution. Fish schooling behaviour, aggregation in specific habitats and seasonal cycles of recruitment in estuaries often induce samples with large deviations from the mean abundance when the sampling effort is limited. The presence of these outlying values makes modelling main tendencies in the dataset difficult resulting in large residuals and uncertain predictions. Therefore, the type of metric, and sampling method and effort have probably influenced this result, indicating a high sensitivity of density metrics to outliers, leading to possible difficulties in interpreting ES when theoretical reference conditions are worse than the observed conditions in the dataset.

The approach used in this work consisting of modelling reference conditions separately for each fish metric may lead to application problems as not all the metrics included in multi-metric index are suitable for this treatment. There are possible alternative options to obtain reference conditions in use and these can be applied in combination to overcome this limitation. Alternatively, fish metrics may be modelled directly using multi-metric techniques used in community ecology such as redundancy analysis (RDA) or canonical correlation analysis (CCA) (Jongman et al. 1995, Lepš & Šmilauer 2003). These are in essence canonical ordination methods and thus a generalization of the PCA that was used on the analysis of pressure proxies. They model species gradients according to explanatory variables under the assumption that there is an underlying structure driving the metric scores. With such a gradient analysis it may be possible to detect patterns of variation (i.e. a response model) in the species data (or fish metrics) that can be explained by pressure proxies and natural characteristics of the systems. Under this approach the pressure proxies can be used as covariates to remove anthropogenic effects from these gradients and then produce a model with only natural characteristics of the systems as explanatory variables (Jongman et al. 1995). Using forward selection or other model optimization protocols it would then be possible to define best models which could then be investigated further to define reference values. It is important to recognize that such an approach will have also disadvantages and will equally reflect non-normal metric response curves, inappropriate pressure proxies and/or the unavailability of explanatory variables for the definition of water body typologies.

The modelling approach attempted here has proven useful despite the difficulties and has provided good insight into ecological processes and design factors that affect the derivation of reference conditions and the deviation from those reference conditions. Rare species contribution to assessment datasets, patchiness, choice of pressure proxies or sampling effort are few examples where further work is needed.

Key messages, evidence and recommendations for future work:

From the range of analysis and case studies presented in this study, several lessons have been learned about the use of fish in the assessment of ecological quality in estuaries. First we have briefly highlighted main aspects of the conclusions of the study as 'Key Messages'. In addition to these we then summarise our interpretation of results and finally provide recommendations for future research:

KEY MESSAGES 4.4

- Available indices
  - current fish indices structure require specific sampling methodologies
  - they are mostly locally relevant even though the interpretation of outcomes is still difficult
  - more recent fish indices are leading in the use of appraisal and validation exercises to test the performance of BQE in ecological quality assessments
- Uncertainty
  - The use of statistical models (LMM, LM, GLM…) enable to disentangle effect due to natural variability and protocol (sampling) from effects due to anthropogenic disturbance
  - Information on fish habitat during survey could probably explain a large part of the within estuary/lagoon variability remaining in the models and could lead to a reduction of uncertainty
  - Further work is needed to better develop the Mixed Model approach and WISERBUG to assess the propagation of uncertainty from the metric level to the multi-metric indices
  - Bayesian approach provides confidence intervals and a rigorous estimate of uncertainty on the assessment at the scale of individual metrics or the index
- Sensitivity
  - There is a proven response to pressure gradients in fish indices.
  - The response of multi-metric indices is better than single metrics
  - The index sensitivity analysis to change in metrics was very useful for no specialist to understand the behaviour of indices
- Reference conditions
  - The modelling approach of fish metrics against the physico-chemical variables has proved useful to derive Reference Conditions
  - However the problems with the score distribution of some metrics remain unsolved
  - Modelling will probably need to be used in combination with other methods
  - Modelling to include the biological interactions as well as the environment influence on the fish community is required
- Sampling effort & standardization
  - better information on habitat and physical characteristics will reduce uncertainty at the metric level leading to more robust assessments
  - a minimum effort is required to minimize misclassification (i.e. prevent wrong quality class assignment)
- Intercalibration
  - harmonization of methodologies across Europe is unlikely by adapting or creating new fish indices (common metrics).
  - But possible using a common pressure index to bring different indices to a common scale

Conclusions and Suggestions for further work

CONCLUSIONS D4.4-1 (Review of current fish indices)

- Development strategies vary but generally include (1) the calibration of metrics to anthropogenic pressures, (2) development of reference conditions, (3) calculation of ecological quality ratios, and (4) designation of thresholds for ecological status class.
- Only a few indices include any evaluation of the index behaviour and performance, detailed description of sampling methodologies, sensitivity to pressure and value to end-user.
- Future work is needed to expand the geographical relevance of indices, improve index sensitivity to pressure and metric behaviour, derive reference conditions, assess uncertainty and improve precision.

CONCLUSIONS D4.4-2 part1 (Indices intercomparison)

- Current fish indices can be used in biological monitoring only if there is compatibility between index structure, sampling method, reference condition and geographical area.
- Even after standardising and optimising sampling protocols the interpretation of the indices outcomes when applied to different geographical areas is unreliable.
- Future work is needed to expand geographical relevance; however, it is unlikely that a common European tool may be developed. Flexible combination of metrics (weighting by relevance) and functional guild are options that should be investigated.

CONCLUSIONS D4.4-2 part2 (Uncertainty)

- Technical and monitoring design factors -gear, sampling season, survey protocol (including sampling effort), natural and anthropogenic pressures all affect variability of fish metrics.
- Modelling metric variance for the uncertainty was limited to subsets of the dataset; sampling methodology factors (gear, season, etc) were easy to model but to identify natural variability and pressure gradients effects, sufficient dataset reflecting natural condition and understanding of pressure effect are needed.
- With the available dataset, the best models explained less than 40 % of fish metric variability. The remaining variability was mainly within estuary or lagoon variability and can probably be attributed, at least in part, to both a habitat effect that was not accounted for in the models and to the influence of biological interactions in influencing community structure.

- This high degree of within-system variability may hamper the detection of pressure effect on fish metrics. Increasing sampling effort could improve assessments but this increases costs.
- Further work is needed to test how the residual variability can be accounted for, explained and/or decreased without increasing the number of replicates (effort). This may be achievable but will require a better knowledge of habitat partition within systems, understanding of metrics behaviour and precision, testing new combination rules allowing metric weighting by robustness and importantly research on new and more robust sampling tools (i.e. acoustic and video methods).
- Further work is needed to understand the propagation of the uncertainty from the fish metrics to the multi-metric indicator.
- A Bayesian framework was proposed to combine objectively fish metrics in a multi-metric fish indicator, taking into account the sensitivity and the variability of the fish metrics.
- Further work is needed to test the Bayesian approach on other datasets as well as to include expert opinion (as informative prior) in the assessment. Further testing of responses to pressure fields is also needed and to link these to restoration scenarios.

CONCLUSIONS D4.4-3 (Responsiveness and Sensitivity to pressure)

- Using a matching combination of fish index, reference values, and local dataset, the transitional fish index (& metrics) can be sensitive to pressure gradients.
- Strength and time lag of the expected index response varied across all indices tested indicating sensitivity differences probably associated with constituent metrics.
- Metric type, correlation between metrics and the number of metrics included in multi-metric tools all influence index sensitivity to change.
- The magnitude and scale of index responses to pressure gradients is influenced by the structure of the indices, especially number of metrics, the correlation between metrics and distribution of metric scores.
- Sensitivity analysis to change in metrics very useful to understand the behaviour of indices and evaluate expectations for management purposes.
- Further testing of responses to pressure fields is also needed and to link these to restoration scenarios.

CONCLUSIONS D4.4-4 (Reference conditions)

- The modelling approach proposed here to define type-specific reference conditions for fish assemblages in transitional waters in Europe appears promising.

- Extrapolating to zero value of pressure may be unreliable. Hence a more conservative approach using the lowest pressure values, may give a better prediction (i.e. increases accuracy) but produces a reference condition set at an artificially diminished quality level.
- The best explanatory models included sampling factors and natural characteristics of estuaries and lagoons. These are considered important discriminant features in the definition of water body types. In particular, the present work argues for considering not only estuaries and lagoons as different typologies but also other natural and design (logistic) characteristic such as the gear type, the sampling season and the salinity class.
- Several improvements are required for the method. In particular:
    - the creation of relevant standardised pressure data for fish at a European scale;
    - analyses are required at scales relevant to specific monitoring programmes;
    - other methods to select the fixed effects in the models must be tested.
- Further work needs to concentrate on how to combine single metric reference conditions into multi-metric indices references, and for the particular implementation of the WFD further work is required to find ways to calculate reference conditions at scales relevant to the waterbodies.

Acknowledgments:

The authors would like to thank the help of European Water agencies for the provision of data and help with the interpretation of their datasets. Thanks are specially due to the Basque Water Agency (URA) for fish data on the Basque estuaries, the IMAR -Institute of Marine Research (C.M.A.), University of Coimbra, Portugal for data on the Mondego estuary, the United Kingdom Environment Agency for fish data in the Orwell and Stour estuary and the ONEMA and the French Water Agencies for fish data on the French estuaries and lagoons. The authors are also grateful to all the people who participated in the WISER field samplings. K. Markham, S Coates and A. Waugh assisted in the Stour and Orwell sampling campaign. The sampling in the Mondego estuary was supported by IMAR. N Castro provided help with the Mondego field work. Finally, S. Maci in Lesina lagoon provided assistance in the field and coordinated the support of the Department of Biological and Environmental Sciences and Technologies, University of Salento, Lecce. S Coates and J Breine helped with the assignation of fish guilds and calculation of fish indices. We also thank Mike Dunbar for his help on the use of mixed models. Final thanks are also given to the coordinators of WISER, Christian Feld and Daniel Hering for their support and coordination work.

References

Able KW (2005) A re-examination of fish estuarine dependence: Evidence for connectivity between estuarine and ocean habitats. Estuarine Coastal and Shelf Science 64:5-17

Allen LG, Horn MH (1975) Abundance, diversity and seasonality of fishes in Colorado lagoon, Alamitos bay, California. Estuarine and Coastal Marine Science 3:371-380

Anjos MBd, Zuanon J (2007) Sampling effort and fish species richness in small terra firme forest streams of central Amazonia, Brazil. Neotropical Ichthyology 5:45-52

Atkins JP, Burdon D, Elliott M, Gregory AJ (2011) Management of the marine environment: Integrating ecosystem services and societal benefits with the DPSIR framework in a systems approach. Mar Pollut Bull 62:215-226

Aubry A, Elliott M (2006) The use of environmental integrative indicators to assess seabed disturbance in estuaries and coasts: Application to the Humber Estuary, UK. Mar Pollut Bull 53:175-185

Bain MB, Harig AL, Loucks DP, Goforth RR, Mills KE (2000) Aquatic ecosystem protection and restoration: Advances in methods for assessment and evaluation. Environmental Science and Policy 3:S89-S98

Basset A, Barbone E, Elliott M, Li B-L, Jorgensen SE, Lucena-Moya P, Pardo I, Mouillot D (2012) A unifying approach to understanding transitional waters: fundamental properties emerging from ecotone ecosystems. . Estuarine, Coastal & Shelf Science in press.

Beck MW, Heck KL, Able KW, Childers DL, Eggleston DB, Gillanders BM, Halpern B, Hays CG, Hoshino K, Minello TJ, Orth RJ, Sheridan PF, Weinstein MR (2001) The identification, conservation, and management of estuarine and marine nurseries for fish and invertebrates. Bioscience 51:633-641

Birk S, Bonne W, Borja A, Brucet S, Courrat A, Poikane S, Solimini A, van de Bund W, Zampoukas N, Hering D (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

Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, White J-SS (2009) Generalized linear mixed models: a practical guide for ecology and evolution. Trends in ecology & evolution (Personal edition) 24:127-135

Borja A (2006) The new European Marine Strategy Directive: Difficulties, opportunities, and challenges. Mar Pollut Bull 52:239-242

Borja A, Dauer DM, Elliott M, Simenstad CA (2010) Medium- and Long-term Recovery of Estuarine and Coastal Ecosystems: Patterns, Rates and Restoration Effectiveness. Estuaries and Coasts 33:1249–1260

Borja A, Ranasinghe A, Weisberg SB (2009) Assessing ecological integrity in marine waters, using multiple indices and ecosystem components: Challenges for the future. Mar Pollut Bull 59:1-4

Borja A, Uriarte A, Muxika I, M. GJ, Uyarra MC, Courrat A, Lepage M, Elliott M, Pérez-Domínguez R, Alvarez MC, Franco A, Cabral H, Pasquaud S, Fonseca V, Neto JM (2012) Report detailing Multimetric fish-based indices sensitivity to anthropogenic and natural pressures, and to metrics’ variation range. In: WISER Deliverable D44-3

Breine J, Quataert P, Stevens M, Ollevier F, Volckaert FAM, Van den Bergh E, Maes J (2010) A zone-specific fish-based biotic index as a management tool for the Zeeschelde estuary (Belgium). Mar Pollut Bull 60:1099-1112

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

Coates S, Waugh A, Anwar A, Robson M (2007) Efficacy of a multi-metric fish index as an analysis tool for the transitional fish component of the Water Framework Directive. Mar Pollut Bull 55:225-240

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

Courrat A, Lepage M, Pérez-Domínguez R, Uriarte A, Fonseca V, Raykov V, Borja A, Neto JM, Cabral H, Alvarez MC (2011) Testing the adaptability and behaviour of 6 fish indices on a common dataset composed of multiple gears samplings from 8 estuaries and lagoons. In: WISER Deliverable D44-2 (part 1)

Courrat A, Lepage M, Alvarez MC, Borja A, Cabral H, Elliott M, Franco A, Gamito R, Neto JM, Pérez-Domínguez R, Raykov V, Uriarte A (2012a) The ecological status assessment of transitional waters: an uncertainty analysis for the most commonly used fish metrics in Europe. In: WISER Deliverable D44-2 (part 2)

Courrat A, Lepage M, Alvarez MC, Borja A, Cabral H, Elliott M, Neto JM, Pérez-Domínguez R, Raykov V, Uriarte A (2012b) Report summarising the definitions of reference conditions using predictive models for ecological endpoints for fish in transitional waters. In: WISER Deliverable D44-4

Da Sylva DP (1975) Nektonic food webs in estuaries. In: Cronin L (ed) Estuarine Research, Book 1. Academic Press, New York

Dale VH, Beyeler SC (2001) Challenges in the development and use of ecological indicators. Ecological Indicators 1:3-10

Dauvin JC, Ruellet T (2009) The estuarine quality paradox: Is it possible to define an ecological quality status for specific modified and naturally stressed estuarine ecosystems? Mar Pollut Bull 59:38-47

Deegan LA, Finn JT, 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, Boet P, Girardin M, Lepage M (2010) Development of a fish-based index to assess the ecological quality of transitional waters: The case of French estuaries. Mar Pollut Bull 60:908-918

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

Elliott A, Whitfield AK (2011) Challenging paradigms in estuarine ecology and management. Estuarine, Coastal & Shelf Science 94:306-314

Elliott M (2011) Marine science and management means tackling exogenic unmanaged pressures and endogenic managed pressures – a numbered guide. Mar Pollut Bull 62:651-655

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

Elliott M, Hemingway KL (2002) Fishes in Estuaries, Vol. Blackwell Science, Oxford

Elliott M, Quintino V (2007) The estuarine quality paradox, environmental homeostasis and the difficulty of detecting anthropogenic stress in naturally stressed areas. Mar Pollut Bull 54:640-645

Elliott M, Whitfield AK, Potter IC, Blaber SJM, Cyrus DP, Nordlie FG, Harrison TD (2007) The guild approach to categorizing estuarine fish assemblages: a global review. Fish and Fisheries 8:241-268

Ferreira T, Oliveira J, Caiola N, De Sostoa A, Casals F, Cortes R, Economou A, Zogaris S, Garcia-Jalon D, Ileu M, Martinez-Capel F, Pont D, Rogers C, Prenda J (2007) Ecological traits of fish assemblages from Mediterranean Europe and their responses to human disturbance. Fisheries Management and Ecology 14:473-481

França S, Costa MJ, Cabral HN (2009) Assessing habitat specific fish assemblages in estuaries along the Portuguese coast. Estuarine, Coastal and Shelf Science 83:1-12

Franco A, Elliott M, Franzoil P, Torricellil P (2008a) Life strategies of fishes in European estuaries: the functional guild approach. Marine Ecology-Progress Series 354:219-228

Franco A, Franzoi P, Malavasi S, Riccato F, Torricelli P, Mainardi D (2006a) Use of shallow water habitats by fish assemblages in a Mediterranean coastal lagoon. Estuarine Coastal and Shelf Science 66:67-83

Franco A, Franzoi P, Torricelli P (2008b) Structure and functioning of Mediterranean lagoon fish assemblages: A key for the identification of water body types. Estuarine Coastal and Shelf Science 79:549-558

Franco A, Malavasi S, Zucchetta M, Franzoi P, Torricelli P (2006b) Environmental influences on fish assemblage in the Venice Lagoon, Italy. Chem Ecol 22:105-118

Franco A, Torricelli P, Franzoi P (2009) A habitat-specific fish-based approach to assess the ecological status of Mediterranean coastal lagoons. Mar Pollut Bull 58:1704-1717

Franco A, Pérez-Ruzafa A, Drouineau H, Franzoi P, Koutrakis ET, Lepage M, Verdiell-Cubedo D, Bouchoucha M, López-Capel A, Riccato F, Sapounidis A, Marcos C, Oliva-Paterna FJ, Torralva-Forero M, Torricelli P (2012) Fish assemblages in coastal lagoon habitats: effect of sampling method. Estuarine, Coastal and Shelf Science, (in press)

Gray JS (1989) EFFECTS OF ENVIRONMENTAL-STRESS ON SPECIES RICH ASSEMBLAGES. Biological Journal of the Linnean Society 37:19-32

Gray JS, Elliott M (2009) Ecology of Marine Sediments: science to management, Vol. Oxford University Press, Oxford

Harris JH (1995) The Use of Fish in Ecological Assessments. Australian Journal of Ecology 20:65-80

Harrison TD, Whitfield AK (2006) Estuarine typology and the structuring of fish communities in South Africa. Environmental Biology of Fishes 75:269-293

Hering D, Borja A, Carstensen J, Carvalho L, Elliott M, Feld CK, Heiskanen AS, Johnson RK, Moe J, Pont D, Solheim AL, van de Bund W (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 408:4007-4019

Hering D, Feld CK, Moog O, Ofenbock 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

Hoff JG, Ibara RM (1977) FACTORS AFFECTING SEASONAL ABUNDANCE, COMPOSITION AND DIVERSITY OF FISHES IN A SOUTHEASTERN NEW-ENGLAND ESTUARY. Estuarine and Coastal Marine Science 5:665-678

Jongman RHG, Ter Braak CJF, Van Tongeren OFR (1995) Data Analysis in Community and Landscape Ecology, Vol. Cambridge University Press, Cambridge

Karr JR (1981) Assessment of Biotic Integrity Using Fish Communities. Fisheries 6:21-27

Knox GA (1986) Estuarine Ecosystems: A System Approach, Vol 1. CRC Press, Boca Raton, USA

Krebs CJ (1999) Ecological Methodology, Vol. Addison Wesley, California, USA

Lepage M, Coates S (2011) Water Framework Directive Intercalibration Phase 2 : Milestone 6 report. In. European Commission, Directorate General JRC, Joint Research Centre, Institute of Environment and Sustainability

Lepš J, Šmilauer P (2003) Multivariate Analysis of Ecological Data using CANOCO Vol. Cambridge University Press, Cambridge

Loneragan NR, Potter IC (1990) Factors influencing community structure and distribution of different life-cycle categories of fishes in shallow waters of a large Australian estuary. Mar Biol 106:25-37

Long ER, Macdonald DD, Smith SL, Calder FD (1995) Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments. Environmental Management 19:81-97

Lotze HK, Lenihan HS, Bourque BJ, Bradbury RH, Cooke RG, Kay MC, Kidwell SM, Kirby MX, Peterson CH, Jackson JBC (2006) Depletion, degradation, and recovery potential of estuaries and coastal seas. Science 312:1806-1809

Maes J, Stevens M, Breine 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

Marshall S, Elliott M (1998) Environmental influences on the fish assemblage of the Humber estuary, UK. Estuarine Coastal and Shelf Science 46:175-184

Mayfield MM, Bonser, S. P., Morgan, J. W., Aubin, I., McNamara, S. & Vesk, P. A. (2010) What does species richness tell us about functional trait diversity? Predictions and evidence for responses of species and functional trait diversity to land-use change. Global Ecology and Biogeography 19:423-431

McErlean AJ, O'Connor SG, Mihursky JA, Gibson CI (1973) Abundance, diversity and seasonal patterns of estuarine fish populations. Estuarine and Coastal Marine Science 1:19-36

McLusky DS, Elliott M (2004) The Estuarine Ecosystem: Ecology, Threats and Management, Vol. Oxford University Press, Oxford

Meng L, Orphanides CD, Powell JC (2002) Use of a fish index to assess habitat quality in Narragansett Bay, Rhode Island. Transactions of the American Fisheries Society 131:731-742

Nicolas D, Lobry J, Le Pape O, Boet P (2010a) Functional diversity in European estuaries: Relating the composition of fish assemblages to the abiotic environment. Estuarine Coastal and Shelf Science 88:329-338

Nicolas D, Lobry J, Lepage M, Sautour B, Le Pape O, Cabral H, Uriarte A, Boet P (2010b) Fish under influence: A macroecological analysis of relations between fish species richness and environmental gradients among European tidal estuaries. Estuarine Coastal and Shelf Science 86:137-147

Niemi G, Wardrop D, Brooks R, Anderson S, Brady V, Paerl H, Rakocinski C, Brouwer M, Levinson B, McDonald M (2004) Rationale for a new generation of indicators for coastal waters. Environmental Health Perspectives 112:979-986

Noges P, van de Bund W, Cardoso AC, Solimini AG, Heiskanen AS (2009) Assessment of the ecological status of European surface waters: a work in progress. Hydrobiologia 633:197-211

Oberdorff T, Guegan JF, Hugueny B (1995) Global scale patterns of fish species richness in rivers. Ecography 18:345-352

Pasquaud S, Brind'Amour A, Berthele O, Girardin M, Elie P, Boet P, Lepage M (2012) Impact of the sampling protocol in assessing ecological trends in an estuarine ecosystem: The empirical example of the Gironde estuary. Ecological Indicators 15:18-29

Paterson AW, Whitfield AK (2000) Do shallow-water habitats function as refugia for juvenile fishes? Estuarine Coastal and Shelf Science 51:359-364

Pease BC (1999) A spatially oriented analysis of estuaries and their associated commercial fisheries in New South Wales, Australia. Fisheries Research 42:67-86

Pérez-Domínguez R, Maci S, Courrat A, Borja A, Neto J, Elliott M (2010) Review of fish-based indices to assess ecological quality condition in transitional waters. In: WISER Deliverable D44-1

Pérez-Domínguez R, Maci S, Courrat A, Lepage M, Borja A, Uriarte A, Neto JM, Cabral H, Raykov V, Franco A, Alvarez MC, Elliott M (2012) Current developments on fish-based indices to assess ecological-quality status of estuaries and lagoons. Ecological Indicators In press

Poore GCB, Wilson GDF (1993) MARINE SPECIES RICHNESS. Nature 361:597-598

Potter IC, Claridge PN, Warwick RM (1990) Consistency of seasonal changes in an estuarine fish assemblage. Marine Ecology Progress Series 32:217-228

Rice J (2003) Environmental health indicators. Ocean & Coastal Management 46:235-259

Smith KF, Brown JH (2002) Patterns of diversity, depth range and body size among pelagic fishes along a gradient of depth. Global Ecology and Biogeography 11:313-322

Uriarte A, 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

Vasconcelos RP, Reis-Santos P, Fonseca V, Maia A, Ruano M, Franca S, Vinagre C, Costa MJ, Cabral H (2007) Assessing anthropogenic pressures on estuarine fish nurseries along the Portuguese coast: A multi-metric index and conceptual approach. Science of the Total Environment 374:199-215

Wagner CM, Austin HM (1999) Correspondence between environmental gradients and summer littoral fish assemblages in low salinity reaches of the Chesapeake Bay, USA. Marine Ecology-Progress Series 177:197-212

Weinstein MP, Weiss SL, Walters MF (1980) Multiple determinants of community structure in shallow marsh habitats, Cape Fear river estuary, North-Carolina, USA. Mar Biol 58:227-243

Whitfield AK (1999) Ichthyofaunal assemblages in estuaries: A South African case study. Reviews in Fish Biology and Fisheries 9:151-186

Whitfield AK, Elliott A (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

Whitfield AK, Elliott M, Basset A, Blaber SJM, West RJ (2012) Paradigms in estuarine ecology – the Remane diagram with a suggested revised model for estuaries: a review. Estuarine, Coastal & Shelf Science 97:78-90

Wootton JT (2001) Causes of species diversity differences: a comparative analysis of Markov models. Ecology Letters 4:46-56

Zaldivar JM, Cardoso AC, Viaroli P, Newton A, de Wit R, Ibanez C, Reizopoulou S, Somma F, Razinkovas A, Basset A, Holmer M, Murray N (2008) Eutrophication in transitional waters: an overview. Transitional Waters Monographs 1:1-78

Appendix 1: Indicator properties and Fish guilds assignments

A. Conceptual table with the required properties of indicators and monitoring parameters for successful marine management (from Elliott 2011 based on sources therein)

Property	Explanation
Anticipatory	Sufficient to allow the defence of the precautionary principle, as an early warning of change, capable of indicating deviation from that expected before irreversible damage occurs.
Biologically important	Focuses on species, biotopes, communities, etc. important in maintaining a fully functioning ecological community.
Broadly applicable and integrative over space and time	Usable at many sites and over different time periods to give an holistic assessment which provides and summarises information from many environmental and biotic aspects; to allow comparisons with previous data to estimate variability and to define trends and breaches with guidelines or standards.
Concrete / results focussed	We require indicators for directly observable and measurable properties rather than those which can only be estimated indirectly; concrete indicators are more readily interpretable by diverse stakeholders who contribute to management decision-making.
Continuity over time and space	Capable of being measured over appropriate ecological and human time and space scales to show recovery and restoration.
Cost-effective	Indicators and measurements should be cost-effective (financially non-prohibitive) given limited monitoring resources, i.e. with an ease/economy of monitoring. Monitoring should provide the greatest and quickest benefits to scientific understanding and interpretation, to society and sustainable development. This should produce an optimum and defensible sampling strategy and the most information possible.
Grounded in theory / relevant and appropriate	Indicators should reflect features of ecosystems and human impacts that are relevant to achieving operational objectives; they should be scientifically sound and defensible and based on well-defined and validated theory. They should be relevant and appropriate to management initiatives and understood by managers.
Interpretable	Indicators should reflect the concerns of, and be understood by stakeholders. Their understanding should be easy and equate to their technical meanings, especially for non-scientists and other users; some should have a general applicability and be capable of distinguishing acceptable from unacceptable conditions in a scientifically and legally defensive way.
Low redundancy	The indicators and monitoring should provides unique information compared to other measures.
Measurable	Indicators should be easily measurable in practice using existing instruments, monitoring programmes and analytical tools available in the relevant areas, to the required accuracy and precision, and on the time-scales needed to support management. They should have minimum or known bias (error), and the desired signal should be distinguishable from noise or at least the noise (inherent variability in the data) should be quantified and explained, i.e. have a high signal to noise ratio. They need

	to be capable of being updated regularly, being operationally defined and measured, with accepted methods and Analytical/Quality Control/Quality Assurance and with defined detection limits.
Non-destructive	Methods used should cause minimal and acceptable damage to the ecosystem and should be legally permissible.
Realistic / attainable (achievable)	Indicators should be realistic in their structure and measurement and should provide information on a 'need-to-know' basis rather than a 'nice-to-know' basis. They should be attainable (achievable) within the management framework.
Responsive feedback to management	Indicators should be responsive to effective management action and regulation and provide rapid and reliable feedback on the findings. Such feedback loops should be determined and defined prior to using the indicator.
Sensitive to a known stressor or stressors	The trends in the indicators should be sensitive to changes in the ecosystem properties or impacts, to a stressor or stressors which the indicator is intended to measure and also sensitive to a manageable human activity; they should be based on an underlying conceptual model, without an all-or-none response to extreme or natural variability, hence potential for use in a diagnostic capacity.
Socially relevant	Understandable to stakeholders and the wider society or at least predictive of, or a surrogate for, a change important to society.
Specific	Indicators should respond to the properties they are intended to measure rather than to other factors, and/or it should be possible to disentangle the effects of other factors from the observed response (hence having a high reliability/specificity of response and relevance to the endpoint).
Time-bounded	The date of attaining a threshold/standard should be indicated in advance. They are likely to be based on existing time-series data to help set objectives and also based on readily available data and those showing temporal trends.
Timely	The indicators should be appropriate to management decisions relating to human activities and therefore they should be linked to that activity; thus providing real-time information for feedback into management giving remedial action to prevent further deterioration and to indicate the results of or need for any change in strategy.

B. Fish species caught and corresponding commonly agreed ecological guilds

Species caught and corresponding guilds for which a common assignment was reached. Ecological guilds: ER: Estuarine Resident species; DIA: Diadromous species; FW: Freshwater species; MJ: Marine Juvenile species; MA: Marine Adventitious species; MS: Marine Seasonal species. Position guilds: P: Pelagic; B: Benthic; D: Demersal. Trophic guilds: F: Piscivorous (exclusively); Z: Zooplankton feeder; IS: Supra benthic Invertebrate feeder; IB: Benthic Invertebrate feeder; O: Omnivorous. Blank: no data.

Species	Ecological guild	Position guild	Trophic guild
Alosa	DIA	P	Z
Ammodytes tobianus	ER	B	Z
Anguilla anguilla	DIA	D	O
Aphanius fasciatus	ER	D	IB
Aphia minuta	ER	P	Z
Arnoglossus kessleri	ER	B	IB
Arnoglossus laterna	MA	B	IB
Arnoglossus thori	MA	B	IB
Atherina boyeri	ER	P	Z
Atherina pontica	MJ	P	Z
Atherina presbyter	ER	P	Z
Barbus barbus	FW	D	IB
Buglossidium luteum	MA	B	IB
Callionymus lyra	ER	B	IB
Callionymus risso	ER	B	IB
Chelidonichthys lucernus	MJ	B	IS
Chelon labrosus	DIA	D	O
Ciliata mustela	ER	B	O
Ciliata septentrionalis	MA	D	IS
Clupea harengus	MJ	P	Z
Conger conger	MA	D	F
Dicentrarchus labrax	MJ	D	IS
Diplodus annularis	MA	D	IS
Diplodus sargus	MJ	D	IS
Diplodus vulgaris	MJ	D	IS
Engraulis encrasicolus	MS	P	Z
Gambusia holbrooki	ER	P	IS
Gasterosteus aculeatus	ER	D	IB
Gobiidae	ER		
Gobius niger	ER	B	IB
Gobius paganellus	ER	B	IB
Hippocampus guttulatus	ER	B	Z
Hippocampus hippocampus	ER	B	Z
Knipowitschia panizzae	ER		
Labrus merula	MA	D	IB
Liza aurata	DIA	D	O
Liza ramada	DIA	D	O
Liza saliens	DIA	D	O
Micropterus salmoides	FW	P	F
Mugil cephalus	DIA	D	O
Mugilidae	DIA	D	O
Mullus barbatus ponticus	ER	B	IB
Mullus surmuletus	MJ	B	IB
Neogobius	ER		IB
Neogobius cephalargoides	ER	D	IB
Neogobius gymnotrachelus	ER	B	IB
Neogobius melanostomus	ER	B	IB
Oreochromis niloticus niloticus	ER	D	O
Osmerus eperlanus	DIA	P	IS

Parablennius tentacularis	ER	B	O
Platichthys flesus	DIA	B	IB
Pleuronectes platessa	MJ	B	IB
Pomatoschistus	ER	B	
Pomatoschistus marmoratus	ER	B	IB
Pomatoschistus microps	ER	B	IB
Pomatoschistus minutus	ER	B	IB
Proterorhinus marmoratus	ER	B	IB
Raja undulata	MA	B	IB
Salaria pavo	ER	B	O
Sardina pilchardus	MJ	P	Z
Scophthalmus rhombus	MJ	B	IB
Scorpaena notata	MA	B	IS
Scorpaena porcus	MA	D	IS
Solea senegalensis	MJ	B	IB
Solea solea	MJ	B	IB
Sprattus sprattus	MJ	P	Z
Symphodus bailloni	ER	D	O
Symphodus cinereus	ER	D	IS
Symphodus roissali	ER	D	IS
Syngnathus abaster	ER	D	Z
Syngnathus acus	ER	D	Z
Syngnathus rostellatus	ER	P	Z
Syngnathus typhle	ER	D	F
Trachurus mediterraneus	MJ	P	F
Zebrus zebrus	ER	B	IS
Zeugopterus punctatus	MA	B	IS
Zoarces viviparus	ER	B	O
Zosterisessor ophiocephalus	ER	B	F

Note: given ontogenetic changes in species functional characteristics (e.g. feeding preferences), the species allocation to each guild reflected the dominant life stages present in the estuaries/lagoons, as inferred from the dominant size-range of fish species in the samples.

Appendix 2: Reviewed fish indices for transitional waters quality assessment

Tool name	Abbreviation	Area of use	Type (no. Metrics)	WFD	References
Index of biotic integrity	IBI[1]	Transitional (Louisiana, USA)	Multi-metric (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)	Multi-metric (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 Integrity Index	EBI[1]	Transitional (Massachusetts, USA)	Multi-metric (12)	NO	Chun et al. 1996, Deegan et al. 1997
Recruitment Index	RI	South Africa	Single metric *	NO	Quinn et al. 1999
Index of biotic integrity	IBI[3]	Transitional (Nagarranset bay, USA)	Multi-metric (6)	NO	Meng et al. 2002
AZTI's Fish Index	AFI	Transitional (Basque Country, Spain)	Multi-metric (9)	YES	Borja et al. 2004, Uriarte and Borja 2009
Estuarine fish community index	EFCI	Transitional (South Africa)	Multi-metric (14)	NO	Harrison and Whitfield, 2004 and 2006
WFD Fish Index for Transitional waters	FITW	Transitional (Holland)	Multi-metric (10)	YES	Jager. and Kranenbarg 2004
Fish-based Estuarine Biotic Index	EBI[2]	Transitional (Brackish Scheldt, Belgium)	Multi-metric (5)	YES	Breine et al. 2007
Transitional fish classification index	TFCI	Transitional (United Kingdom)	Multi-metric (10)	YES	Coates et al., 2007
MJ nursery index	MJNI	Transitional (France)	Non aggregating multi-metric (3) *	NO	Courrat et al. 2009
Habitat Fish Index	HFI	Transitional and coastal (Venice Lagoon, Italy)	Multi-metric (16)	YES	Franco et al. 2009
Zone-specific Fish-based Estuarine Biotic Index	Z-EBI	Transitional (Brackish and freshwater Scheldt, Belgium)	Multi-metric (6) **	YES	Breine et al. 2010
French Multi-metric Fish Index *(Estuarine and Lagoon Fish Index)*	f-MFI *(ELFI)*	Transitional (Atlantic and Channel coast (France)	Multi-metric (4)	YES	Delpech et al. 2010
Estuarine Fish Assessment Index	EFAI	Transitional (Portugal)	Multi-metric (7)	YES	Cabral et al. 2011

* restricted to the ecological quality assessment of estuarine nursery grounds

**independent indices for each zone

Appendix 3: Comparison of fish assessment tools

A: List of metrics composing the tested fish indices

Index	Country	Metrics
ELFI *(Estuary and Lagoon fish index)* **Estuaries**	France	1. Total density (TD) 2. Density of Diadromous species (DDIA) 3. Density of Marine Juvenile Migrants (DMJ) (meso- & polyhaline zones only) 4. Density of Benthic species (DB) 5. Density of estuarine resident (DER) 6. Total species richness (SR) 7. Density of freshwater species (DFW) (oligohaline zone only)
ELFI **Lagoons**		1. Density of Benthic Invertebrate feeder species (DIB) 2. Density of Zooplankton feeders (DZ) 3. Density of Diadromous species (DDIA)
AFI *(AZTI Fish Index)*	Basque country	1. Richness (number of species) 2. Pollution indicator species (% individuals) 3. Introduced species (% individuals) 4. Fish health (damage,diseases) (% affection) 5. Flat fish presence (% individuals) 6. Trophic composition (% omnivorous) 7. Trophic composition (% piscivorous) 8. Estuarine resident (number of species) 9. Resident species (% individuals)
Z-EBI *(Zone-specific Estuarine index of Biotic Integrity)*	Belgium	1. Total number of piscivorous species (MnsPis) Oligohaline 2. Total number of pollution intolerant species (MnsInt) Oligohaline 3. Total number of diadromous species (MnsDia) Oligo- & Mesohaline 4. Total number of individuals (MnsInd) Oligohaline 5. Total number of marine migrating species (MnsMms) Oligo- & Mesohaline 6. Total number of estuarine species (MnsErs) Oligohaline 7. Total number of species (MnsTot) Mesohaline 8. Total number of specialised spawners (MnsSpa) Mesohaline 9. Total number of habitat sensitive species (MnsHab) Mesohaline 10. Percentage of pollution intolerant individuals (MpiInt) Mesohaline **Note: Freshwater zone metrics are not included in the table**
TFCI *(Transitional Fish Classification Index)*	United Kingdom	1. Species composition 2. Presence of Indicator species 3. Species relative abundance 4. Number of taxa that make up 90% of the abundance 5. Number of estuarine resident taxa 6. Number of estuarine-dependant marine taxa 7. Functional guild composition 8. Number of benthic invertebrate feeding taxa 9. Number of piscivorous taxa 10. Feeding guild composition
BHI *(Biological Health Index)*	South Africa	BHI = $10(J)[\ln(P)/\ln(P_{max})]$ where J is the number of species in the system / the number of species in the reference community; P is the potential species richness (number of species) of each reference community and P_{max} is the maximum potential species richness from all the reference communities.
EFAI *(Estuarine Fish Community Index)*	Portugal	1. Species richness (SR) 2. Percentage of marine migrants (%MM) 3. Estuarine resident species (ES): 3a.*Percentage of individuals, 3b. Number of species* 4. Piscivorous species (P): *Percentage of individuals, Number of species* 5. Diadromous species (D) 6. Introduced species (I) 7. Disturbance sensitive species (S)

B: Sampling protocols applied in the study

*Overview of the samplings considered in the present work, and number of fishing events per estuary or lagoon, salinity class (1 - oligohaline (0 – 5), 2 - mesohaline (5 – 18) and 3 - polyhaline/euhaline (>18)), season and gear that were used to compute the fish indices.; * Varna lake is here considered as a lagoon though in Bulgaria it is usually considered as a "liman lake", which is a lake formed at the mouth of a river where flow is blocked by a bar of sediments (Violin Raykov, pers. com[1]). Hence, it is a particular type of lagoon.*

Site	Estuary / lagoon	Data source	Salinity class	Season	Gear	Number of fishing events
Varna Bay	Estuary	WISER survey	2	autumn	Beam trawl	10
Varna Lake	Lagoon*		2	autumn	Beam trawl	7
					Fyke net	6
Lesina	Lagoon		2	autumn	Fyke net Cemagref	3
					Fyke net	18
Mondego	Estuary		1	autumn	Beam trawl	3
					Fyke net	4
			2		Beam trawl	6
					Fyke net	6
			3		Beam trawl	6
					Fyke net	5
Nervion	Estuary	AZTI / Basque Water Agency	3	autumn	Beam trawl	9
				spring	Beam trawl	9
				summer	Beam trawl	9
Oiartzun	Estuary		3	autumn	Beam trawl	12
				spring	Beam trawl	12
				summer	Beam trawl	12
Bidasoa	Estuary		1	spring	Beam trawl	3
			2	spring	Beam trawl	3
			3	autumn	Beam trawl	12
				spring	Beam trawl	6
				summer	Beam trawl	12
Orwell & Stour	Estuary	WISER survey & UK Environment Agency	3	spring	Beam trawl	9
					Fyke net 24 hours	3
					Beach seine EA	15
					Beach seine Wiser	12

[1] Dr. Violin Raykov, IO-BAS, Institute of Oceanology, "Parvi May" Str. 40, P.O. Box 152, Varna 9000, Bulgaria

Appendix 4: Sensitivity of fish-based indices

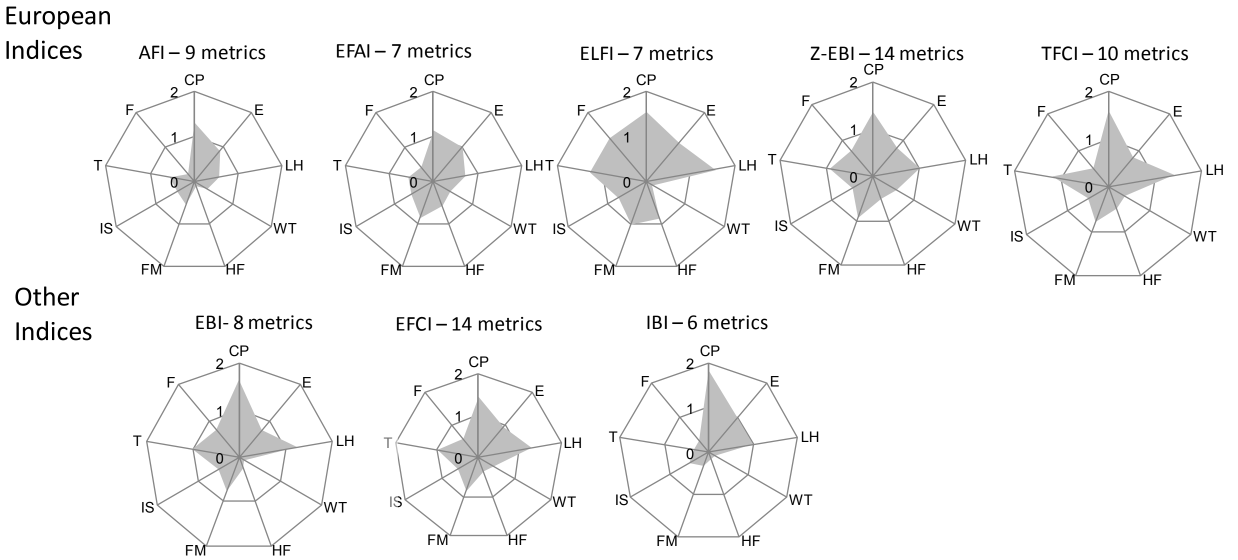
A: Potential noise factors of variability affecting fish metrics in transitional waters systems. These were identified using both expert opinion and bibliographic sources following methodologies outlined in Courrat et al. (2009), Delpech et al. (2010) and Nicolas et al. (2010b).

[Y: likely effect; N: no effect; ?: unknown]		Fish metrics						
*[Due to data availability constraints or to colinearity, only some factors (indicated by *) have been included in the sensitivity analysis]*		Total density	Total no. of sp.	No. estuarine resident sp.	Density of marine migrants	No. of marine migrant sp.	% of omnivorous ind.	% of piscivorous ind.
Potential sources of variability: -within TW systems		**TD**	**SR**	**SR_ER**	**DMM**	**SR_MM**	**RD_O**	**RD_P**
Spatial variability	Depth *(*)*	Y	Y	Y	Y	Y	Y	Y
	Bottom temperature *(Temp) (*)*	Y	Y	?	Y	Y	?	?
	Salinity *(Sal class) (*)*	Y	Y	Y	Y	Y	?	?
	Turbidity	Y	Y	Y	Y	Y	Y	Y
	Oxygen	Y	Y	Y	Y	Y	Y	Y
	Habitat	Y	Y	Y	Y	Y	Y	Y
	Bottom structure	Y	Y	Y	Y	Y	Y	?
Temporal variability	Day *vs.* night *(*)*	Y	Y	Y	Y	Y	Y	Y
	Season *(*)*	Y	Y	N	Y	Y	?	?
	Date (ex. early vs. late spring)	Y	Y	N	Y	Y	Y	Y
	Interannual *(*)*	Y	Y	Y	Y	Y	Y	Y
Sampling method	Gear (type and characteristics)	Y	Y	Y	Y	Y	Y	Y
	Speed & Duration	Y	Y	Y	Y	Y	Y	Y
	Tide	Y	Y	Y	Y	Y	Y	Y
	With or against current	Y	Y	Y	Y	Y	Y	Y
	Operator	Y	Y	Y	Y	Y	Y	Y
	Sampling effort	Y	Y	Y	Y	Y	Y	Y
	Method to chose sampling sites	Y	Y	Y	Y	Y	Y	Y
Sample processing	Errors in species identification	N	Y	Y	Y	Y	Y	Y
	Taxonomical resolution	N	Y	Y	Y	Y	Y	Y
	Subsampling	Y	Y	Y	Y	Y	Y	Y
-between TW systems								
Estuary / lagoon *(*)*		Y	Y	Y	Y	Y	Y	Y
Size of estuary / lagoon *(Area class) (*)*		?	Y	Y	?	Y	?	?
Latitude *(Lat) (*, only for estuaries)*		Y	Y	Y	Y	Y	?	?
Longitude *(Long) (*)*		?	?	?	?	?	?	?
Ecoregion *(*)*		?	Y	?	Y	Y	?	?
Source elevation *(*)*		?	?	?	?	?	?	?
Catchment area *(*,only for lagoons)*		?	?	?	?	?	?	?
Mean annual river discharge *(Discharge) (*)*		?	?	?	?	?	?	?
Estuary entrance width *(Ent width) (*)*		?	Y	?	?	Y	?	?
Estuary entrance depth *(Ent depth) (*)*		?	?	?	?	?	?	?
Tidal range *(*)*		?	?	?	?	?	?	?
Intertidal area (in class) *(*)*		Y	?	?	?	?	?	?
Wave exposure *(*)*		?	?	?	?	?	?	?
Continental shelf width *(Shelf width) (*)*		?	Y	?	?	?	?	?
Section of inlets for lagoons (*Sect*) *(*)*		?	?	?	?	?	?	?
Littoral substrate		?	?	?	?	?	?	?
Fishes not identified at the species level		N	Y	Y	Y	Y	Y	Y

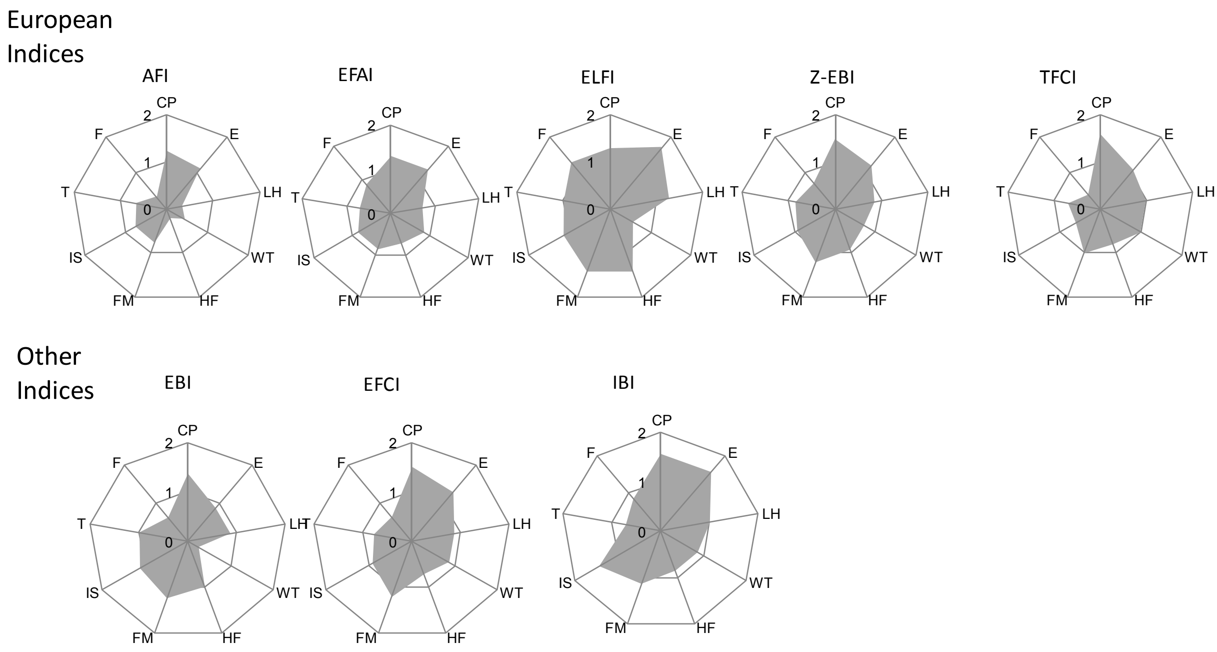
B: Expected strength and time lag in the response of multi-metric fish indices to human pressures

Pressures abbreviations: chemical pollution (CP), eutrophication (E), loss of habitat (LH), water turbidity (WT), habitat fragmentation (HF), fish mortalities (FM), invasive species (IS), temperature (T) and flow (F) changes. Index abbreviations are detailed in the Appendix 2.

Strength of response *(axes of the radar plots represent the average strength assigned to the metrics for each pressure; scores range between 0 (no relationship) and 2 (strong strength). Top row shows WFD-compliant indices.*



Time lag of response *(axes of the radar plots represent the average of the time lags detected by the metrics for each pressure; sores range between 0 (no response) and 2 (response in a short time lag))*



C: Variables included in the analyses of AFI and EFAI indices sensitivity to pressures

Variables, including the form (transformation) in which they have been used in the analysis of AFI index sensitivity to pressures (Borja et al. 2012). Variables highlighted in bold face were found to explain the biological response variables (BEST analysis). Variables underscored best characterized the species assemblage (BEST analysis). Finally, highlighted in grey are the variables found to best explain AFI scores (BACKWARD multiple regression analysis).

Variables	Variable type	Name	Units/measure	Transformation
Biological	Response	Number of taxa	N	
		Abundance	N	√√
		Diversity	Shannon	
		Equitability	Pielou	
Human		Population	hab km^{-2}	
Pressures	Explanatory	Industrial plants	n	log (1+x)
		Ports	n	
		Port area	km^2	log (1+x)
		Berths	n	
		Dredged volume	m^3 $year^{-1}$	
		Farms in the catchment	n	log (1+x)
		Human Pressures present	n	log (1+x)
		Human Pressures present	n km^{-2}	
		Human Pressures present	n km^{-1}	
		Total pressure index, see (Uriarte & Borja 2009)		
		Global pressure index (as used in NEA-GIG intercalibration group)		
		Water pollution index	%	
		Sediment pollution index	%	
		Channeling in ports	%	
		Channeling out of ports	%	
		Loss of intertidal area	%	
		Nutrient loadings	N kg day^{-1} km^{-2}	
Hydromorpho-		Estuary length	km	log (1+x)
logical	Explanatory	Average estuary depth	M	
		Estuary volume	Hm^3	log (1+x)
		Estuary subtidal volume	Hm^3	log (1+x)
		Floodplain surface	Ha	log (1+x) → removed
		Subtidal surface	%	
		Intertidal surface	%	→removed
		Average tidal prism	km^2	log (1+x) → removed
		Catchment area	km^2	log (1+x)
		River flow	m^3 s^{-1}	log (1+x)
		Flushing time	Hr	
		Residence time period	days	
		Continental shelf width	km	log (1+x)
		Distance to the estuary mouth	km	log (1+x)
		Orientation of the estuary	degrees	log (1+x)

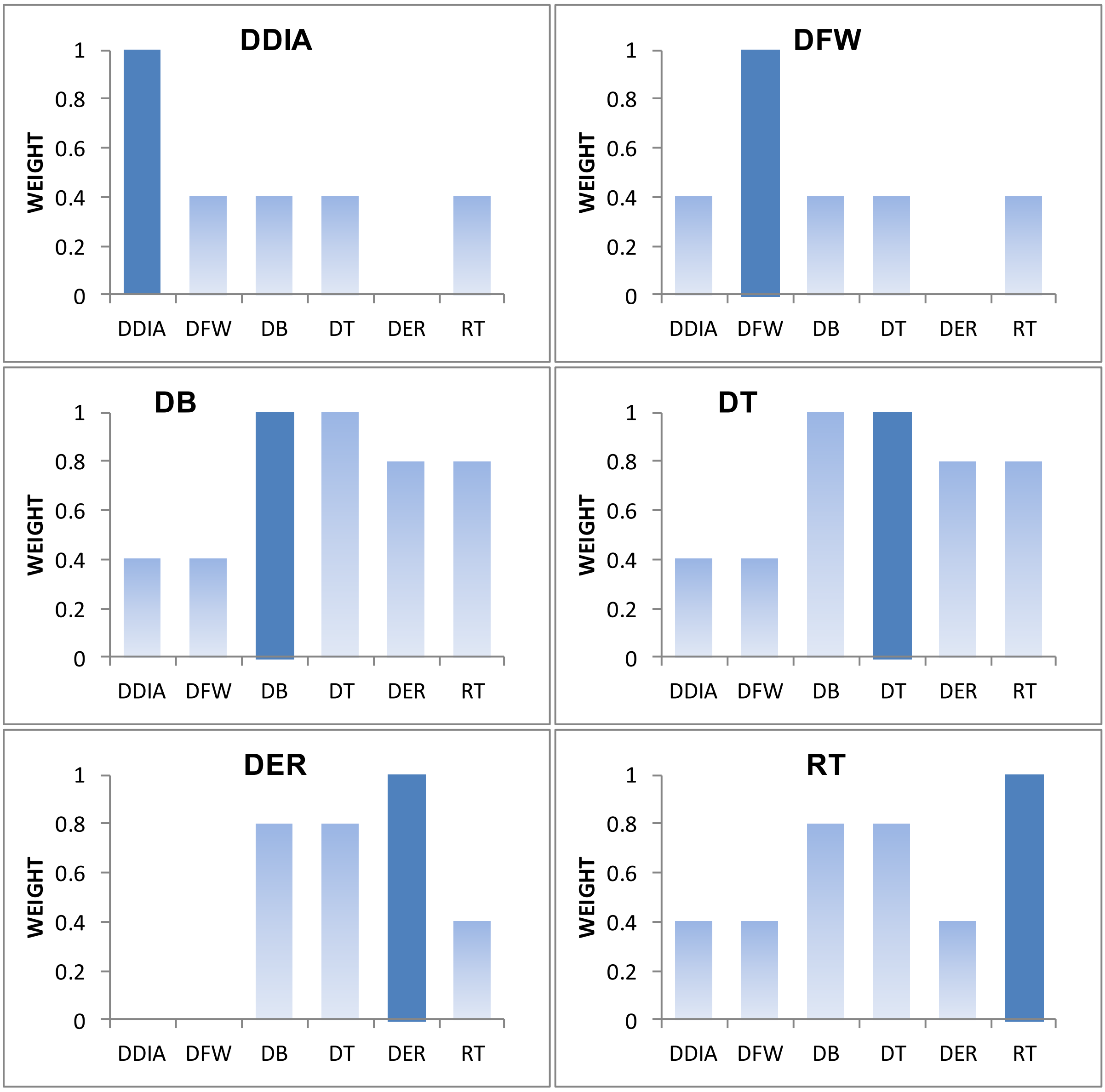
Pressure indicators used to quantify the total pressure present on each site in Portuguese estuaries. Type of data used and the source of information used to collect the data. The different specific pressure indicators were combined into two overall pressure indices (Pi (sum) and EII – environmental integrative indicators) which were then related to the EFAI index scores (Borja et al. 2012).*(ERL – effects range low; ERM – effects range medium)*

Pressure Indicators	**Type of data**	**Source**
Bank regulation (%)	Percentage of regulated estuarine site bank length	Maps/GE
Dredging	Mean volume and intensity	Port authorities
Interference hydrographic regime	Percentage of area occupied by structures interfering with the hydrographic regime	Maps/GE
River Flow and Dams	Flow (m^3 s^{-1}) and Number of large dams	INAG
Sediment metals concentration	Concentration & ERL and ERM	(Long et al. 1995)
Sediment PAH concentration	Concentration & ERL and ERM	(Long et al. 1995)
Industry	Number of industries in the watershed	INE
Population	Population density of watershed surrounding areas	INE
Shelfish quality	Categories according to national standards	IPIMAR
Agriculture	Used agricultural surface area	INE
Aquaculture	Number and area occupied	IPIMAR/GE
Intensity of port/marina developments	Number of berths in marinas/Port areas	Port authorities
Commercial Fishing	Number of licensed boats/Mean commercial fish landings	DGPA/INE
Recreational fishing	Number of recreational licensed fishermen	DGPA/INE
Pressure index - Pi (Sum)	**Sum of all standardized indicators**	
Aubry & Elliott (A&E) adapted	**Adapted from 15 EII criteria**	

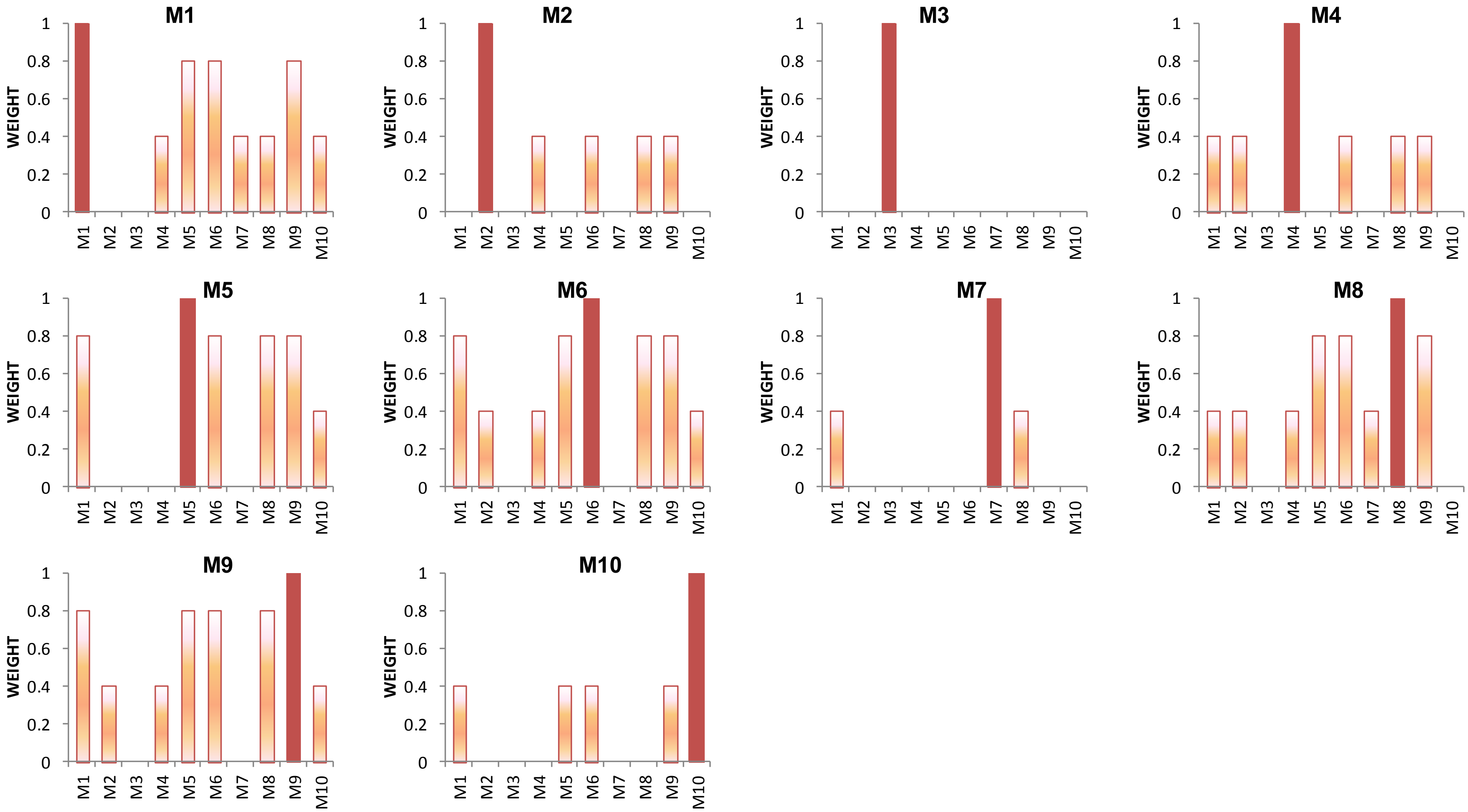
D: Relationsip between metrics defining the different scenarios of change

The figures show the weight applied for each metric in accordance to their correlation to the tested metric (driving the scenario) for the French index ELFI and the British index TFCI. The tested metric driving each scenario is indicated in the title of each graph and by the solid bar. The absence of bar indicates metrics that are uncorrelated with the metric leading the scenario (for these metrics the average score has been considered in the scenario definition). See Appendix 3A for explanation of metrics abbreviations.

French index – ELFI*:*

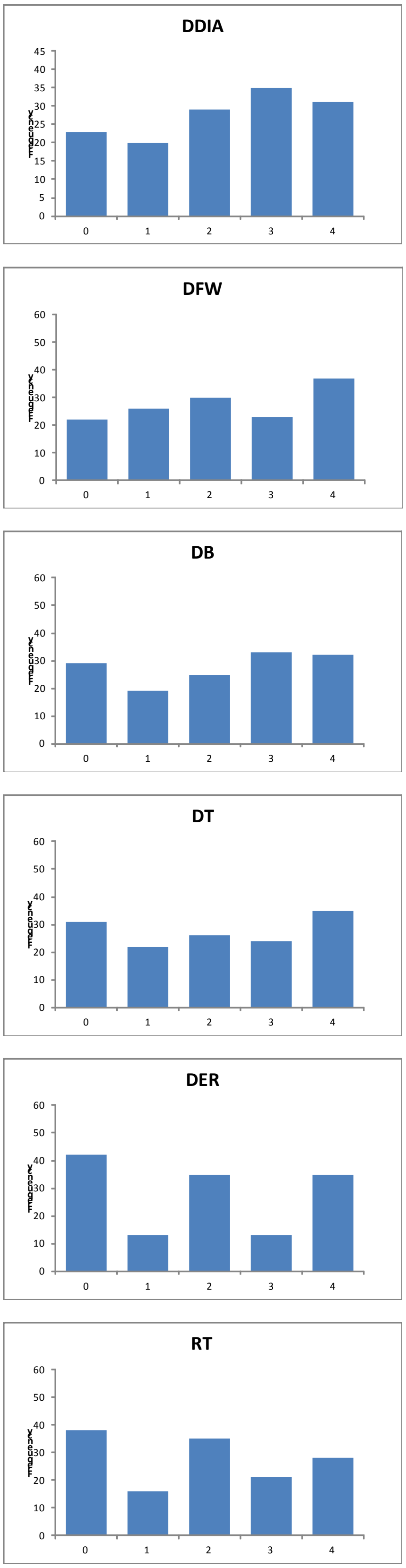


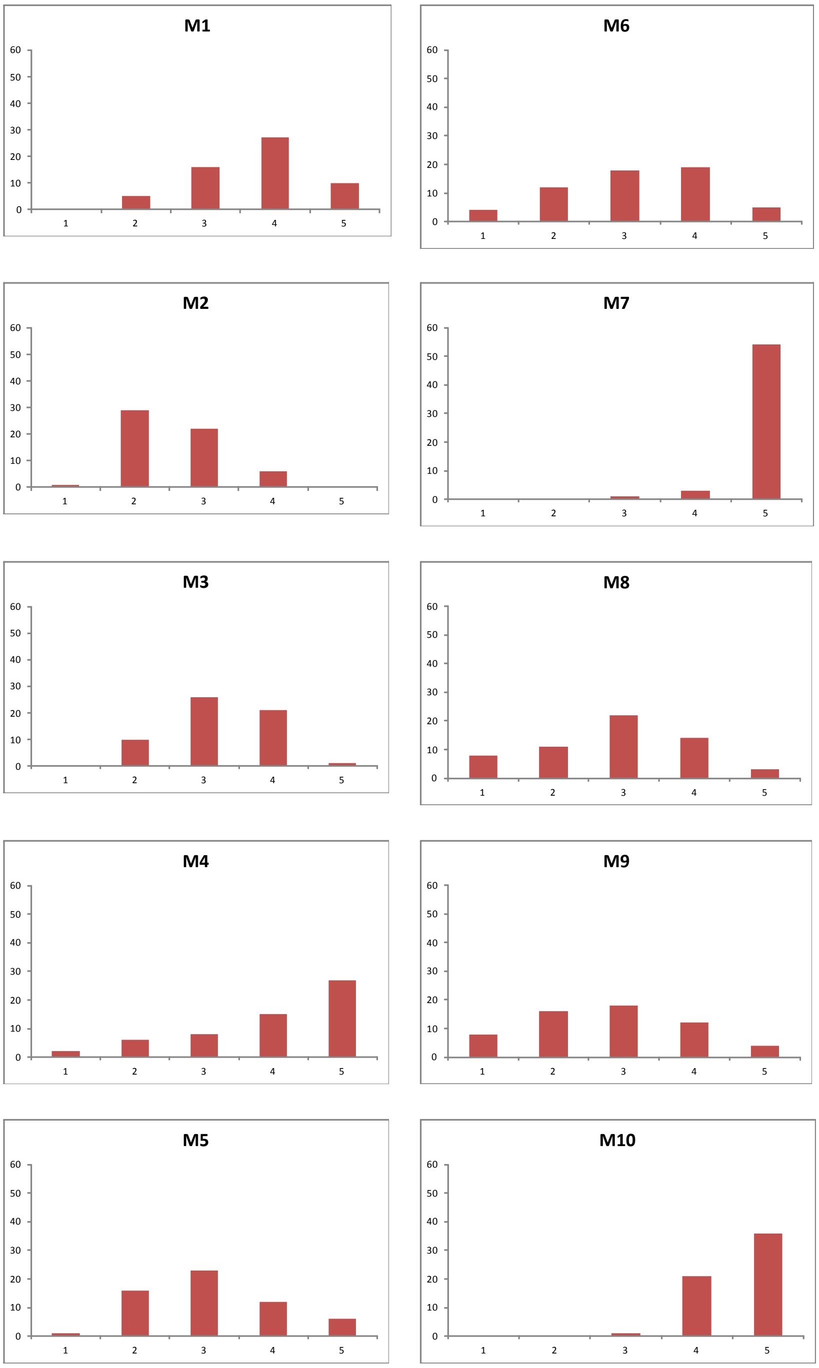
***British index – TFCI*:**



E: Metric scores frequency distribution for the TFCI and ELFI indices across the investigated datasets

ELFI **TFCI**





See Appendix 3A for explanation of metrics abbreviations.

Appendix 5: Reference conditions

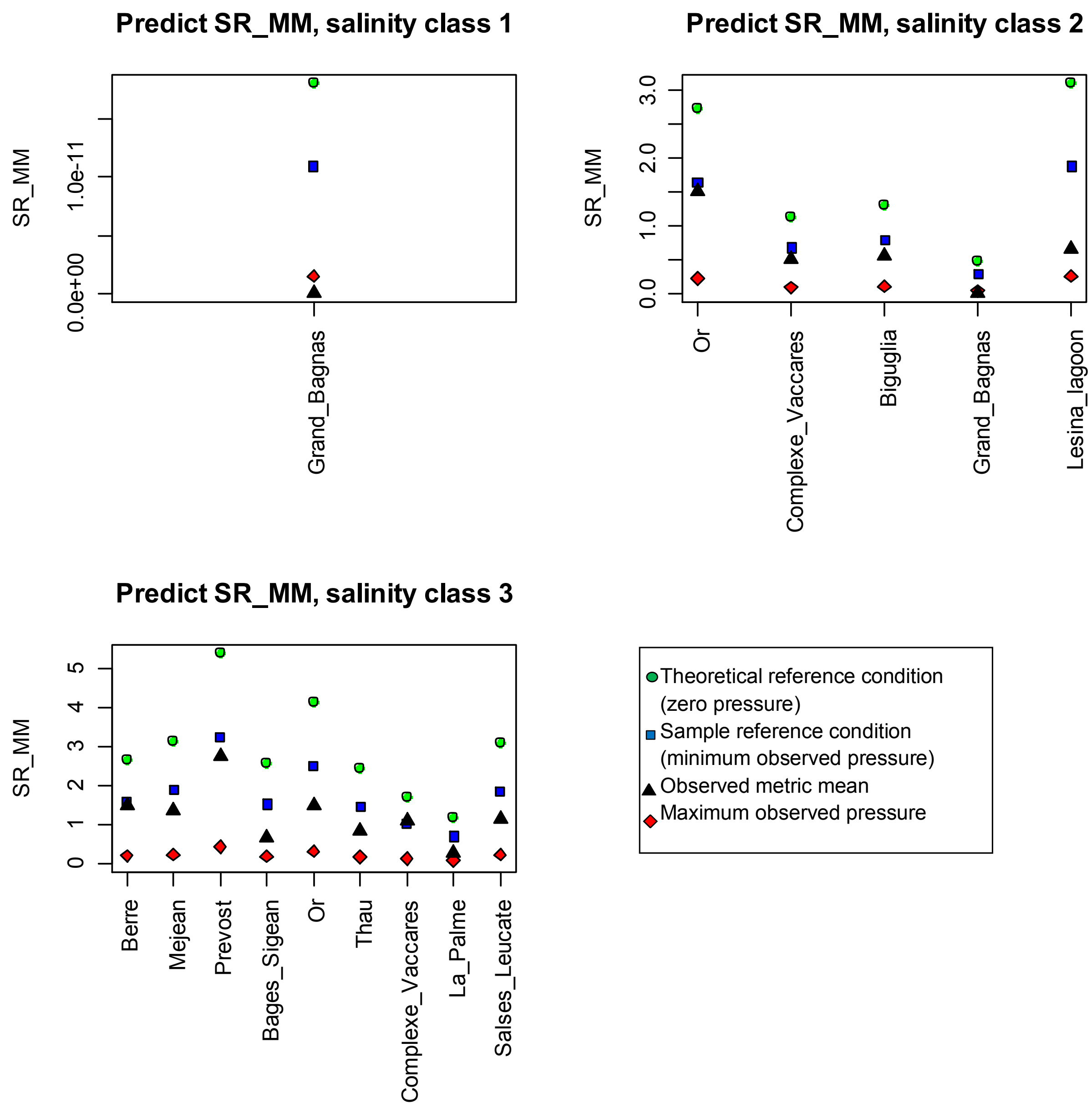
A: List of variables accounting for sources of natural variability included in the models to assess reference conditions for fish metrics

Sources of variability included in the models	Lagoons	Estuaries
Salinity *(Sal class)*	X	X
Depth		X
Bottom temperature *(Temp)*	X	
Season	X	X
Size of lagoon *(Area)*	X	
Size of estuary *(Area, class)*		X
Latitude *(Lat)*		X
Estuary entrance width *(Ent width)*		X
Section of inlets for lagoons (*Sect*)	X	
Continental shelf width *(Shelf width)*		X
Mean annual river discharge *(Discharge)*		X
Intertidal area (% classes)		X

B: Predictions of reference condition according to the models linking fish metrics to human pressures

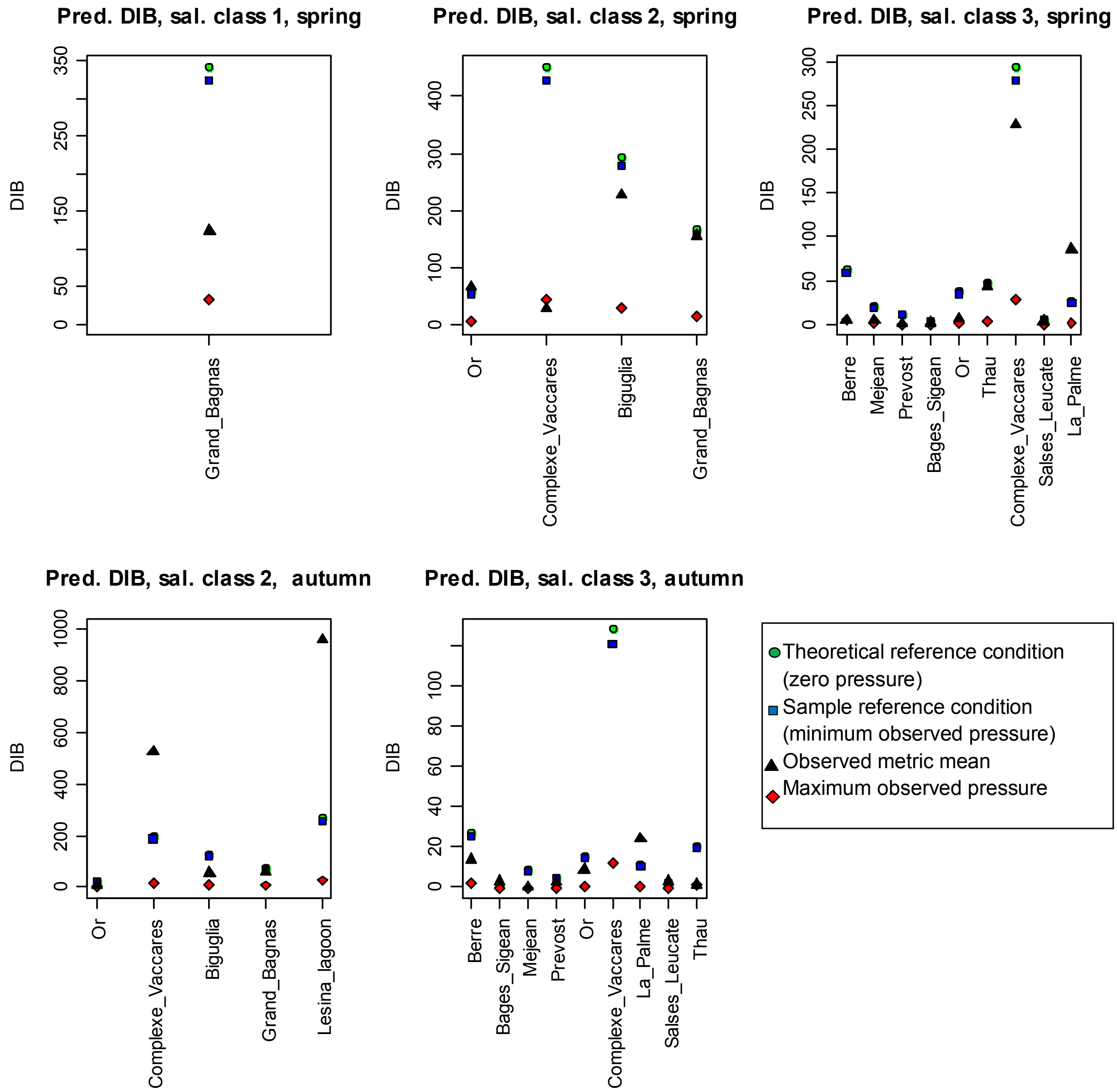
Number of marine migrating species in lagoons

Model: SR_MM ~ Sal class + Agr



Density of benthic invertebrate feeding fishes in lagoons

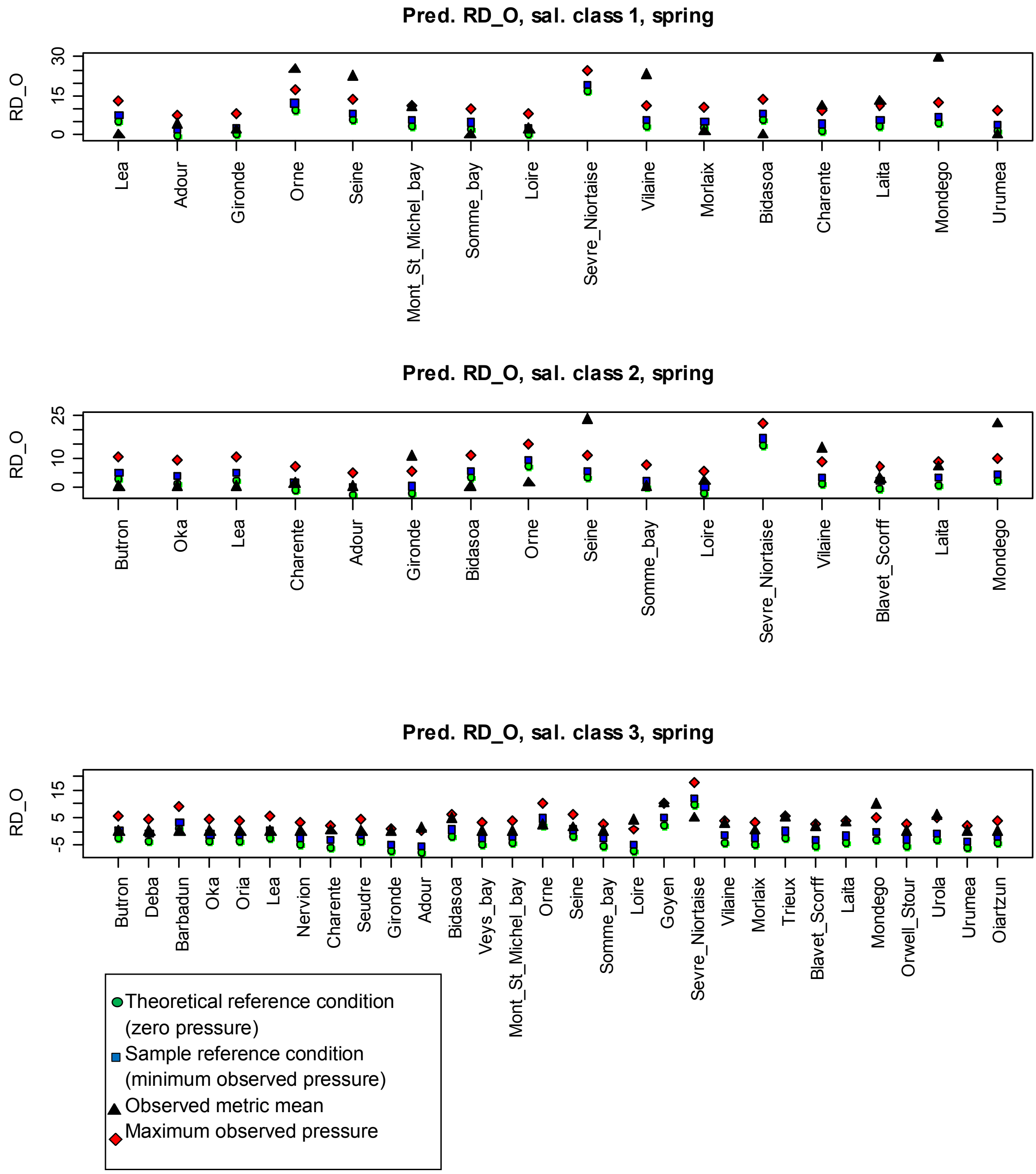
Model: DIB ~ Sal class + Season + Sect+ Area + Urb



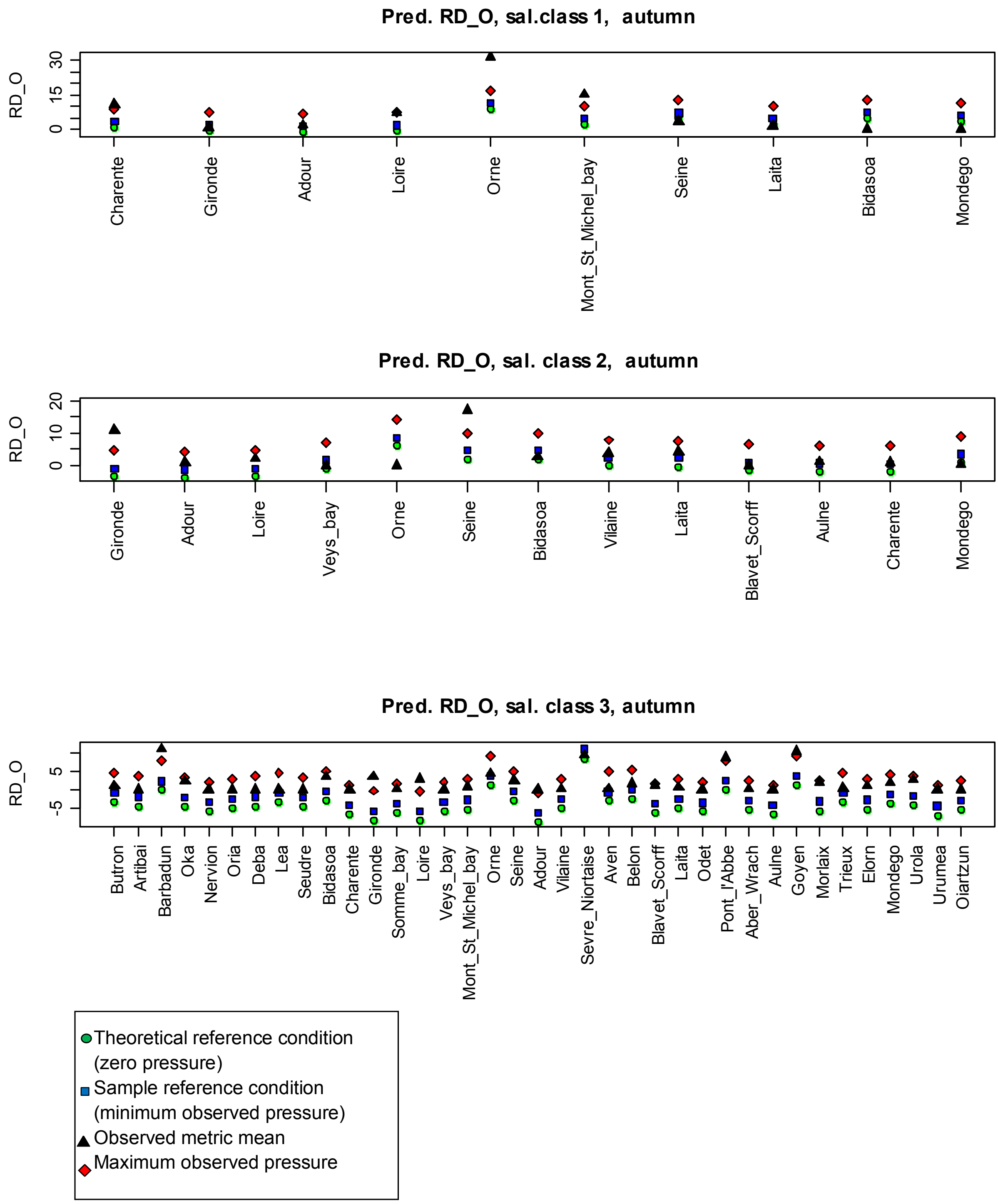
Note: For the Vaccares lagoon complex, salinity class 3, autumn, the true mean value of DIB observed in fish data is 1260.5 (far beyond the maximum value of the y axis).

Percentage density of omnivorous fishes in estuaries

Model: RD_O ~ Sal class + Season + Discharge + Intertidal area + Nat



Note: For the estuary Sevre Niortaise, salinity class 1 and 2 - spring, the true mean value of RD_O observed in fish data is respectively 62.1 % and 88.5% (far beyond the maximum values of the y axis).


Pred. RD_O, sal.class 1, autumn
RD_O
0
15
30
Charente
Gironde
Adour
Loire
Orne
Mont_St_Michel_bay
Seine
Laita
Bidasoa
Mondego
Pred. RD_O, sal. class 2, autumn
RD_O
0
10
20
Gironde
Adour
Loire
Veys_bay
Orne
Seine
Bidasoa
Vilaine
Laita
Blavet_Scorff
Aulne
Charente
Mondego
Pred. RD_O, sal. class 3, autumn
RD_O
-5
5
Butron
Artibai
Barbadun
Oka
Nervion
Oria
Deba
Lea
Seudre
Bidasoa
Charente
Gironde
Somme_bay
Loire
Veys_bay
Mont_St_Michel_bay
Orne
Seine
Adour
Vilaine
Sevre_Niortaise
Aven
Belon
Blavet_Scorff
Laita
Odet
Pont_l'Abbe
Aber_Wrach
Aulne
Goyen
Morlaix
Trieux
Elorn
Mondego
Urola
Urumea
Oiartzun
Theoretical reference condition (zero pressure)
Sample reference condition (minimum observed pressure)
Observed metric mean
Maximum observed pressure