# Benthic Macroinvertebrates as Ecological Indicators for Estuarine and Coastal Ecosystems:

## Assessment and Intercalibration

Doctoral dissertation in Biology (Scientific area of Ecology) presented to the University of Coimbra

Dissertação apresentada à Universidade de Coimbra para obtenção do grau de Doutor em Biologia (especialidade Ecologia)

Heliana Lilita Gonçalves Teixeira
University of Coimbra
2010

## This thesis was supported by:

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

PhD grant attributed to Heliana Teixeira (SFRH/BD/24430/2005)

Department of Life Sciences, Faculty of Sciences and Technology,
University of Coimbra, Portugal

IMAR-CMA
Marine and Environmental Research Centre

*Dedicada com todo o meu amor*

*às duas pessoas mais inspiradoras da minha vida,*

*a minha mãe Luísa e a minha avó Linda.*

# Contents

# Abstract

The aim of the research work presented in this thesis is to be a contribution to the field of ecological assessment in coastal and transitional ecosystems. The main goals were: a) to present a method for the assessment of the ecological status of benthic macroinvertebrate communities in Portuguese transitional waters that would meet the requirements of the European Water Framework Directive (WFD); and b) to propose alternatives to harmonize ecological assessments, namely those based on benthic macroinvertebrate, across wide geographic scales.

Chapters I, II and III describe the steps to attain the first major goal. For this purpose, it was used data from the Mondego estuary in Portugal, where biological and environmental data had been routinely collected from 1990 to 2006. Since the 1980's this system has been harassed by anthropogenic pressures commonly observed in coastal environments worldwide, such as eutrophication and physical disturbances that led to hydromorphological changes of the system properties. An evident ecological quality decrease in the estuary triggered the implementation of mitigation measures in the last 12 years, enabling its recovery. Therefore, the Mondego estuary provided an important field lab to test a battery of indices selected for evaluating the ecological status in transitional waters under the scope of the WFD.

Chapter I describes how habitat mapping allowed setting natural expectations for biological communities' distribution along estuarine gradients. Environmental data, such as salinity, sediment grain size composition and organic matter content, from recent years (2002 to 2005, covering all seasons), and relevant for structuring subtidal benthic invertebrate communities, were used to identify six distinct zones within the Mondego estuary (ANOSIM and Principal Component Analysis - PCA, PRIMER Software). The observed environmental trends significantly reflected the patterns of distribution of the invertebrate communities (BIOENV and ANOSIM, PRIMER Software), and allowed accounting for the influence of natural gradients in the performance of ecological assessments tools.

In Chapter II, three ecological indices (Margalef index, Shannon-Wiener index and AZTI's Marine Biotic Index - AMBI), selected to meet the EU WFD requirements, were evaluated for their potential to detect impaired benthic invertebrate communities and their subsequent recovery. To attain this goal, the indices were tested in three periods (Springs of 1990/1992, 2000/2002 and 2005/2006) of distinct pressures intensity in the Mondego estuary. The trends detected by the indices (PERMANOVA) concurred with the history of disturbance of the system, responding both to different types of impacts and to mitigation measures undertaken. This allowed defining approximate reference conditions for the indices proposed that would reflect a quality improvement in the benthic invertebrate compartment, while accounting for the natural gradients acting upon these communities along the system.

Chapter III describes the performance of the BAT – Benthic Assessment Tool, a multimetric tool *sensu* the EU WFD guidelines, which consists of the three indices previously selected for the Mondego estuary merged into a single value – the Ecological Quality Ratio (EQR). This EQR is obtained after a Factor Analysis (PCA extraction method, Statgraphics Software) and Euclidean distance projection, using the reference conditions proposed in Chapter II to limit the scale of EQR from 0 to 1, as described in Bald *et al*. (2005). The method was tested against the effects of anthropogenic disturbance using eight years (Springs between 1990 to 2006) of subtidal benthic invertebrate data from the Mondego estuary. Although the BAT could capture the ecological decline and recovery of the system as reflected by the benthic invertebrates, the indices within the mutlimetric tool were not contributing equally to the final classification. Constraints such as the typical abundance of tolerant species in estuarine ecosystems and the ecological classification of key species in the Mondego estuary have weakened mainly the performance of the AMBI.

Chapters IV and V exemplify and propose distinct ways to intercalibrate ecological assessments across wide geographic scales.

Chapter IV explains the process to adapt the European developed index, AMBI, to a new geography. With macroinvertebrate data from Southern California marine bays, the index was calibrated to the new habitat using local taxonomic expertise to classify local species into the ecological groups used by the AMBI. Then, taking the ranking and classification of samples, AMBI performance was validated against the local Benthic Response Index (BRI) and also using the Professional Judgement of benthic ecologists. The best correlation between the AMBI results and

those of the BRI (*n*= 685: r= 0.70; Kappa: Moderate agreement for samples classification) was obtained applying AMBI based on a mixture of local and previous expertise regarding species ecological classification, and including a weighting factor for abundance data. As for the best agreement of the AMBI with Expert Judgement (*n*= 21: r= 0.93; Kappa: Very Good agreement for samples classification), it was reached using local expertise criteria for the classification of species ecological strategy, for non-transformed abundance data. The AMBI presented however less discriminatory power than Expert judgement for the classification of samples. The study revealed that a significant part of the disagreements between the two indices' assessments resulted from the approaches followed by each to classify species according to their ecological strategies.

Chapter V proposes an approach that uses the consensus among expert best professional judgement (BPJ) to establish a common scaling for benthic ecological assessments. Sixteen benthic ecologists from four regions in Europe and USA were provided macroinvertebrates species-abundance data for twelve sites per region, to rank from best to worst and classify into four categories. Site rankings were highly correlated among experts regardless of whether they were assessing samples from their home region. There was also good agreement on condition category, though agreement was better for samples at extremes of the disturbance gradient. The absence of regional bias and the agreement obtained suggest that expert judgment is a viable means for establishing a uniform scale to calibrate indices consistently across geographic regions.

# Resumo

O intuito do trabalho de investigação apresentado nesta tese é contribuir para a área da avaliação ecológica em ecossistemas costeiros e de transição. Os objectivos principais foram: a) apresentar uma metodologia para a avaliação do estado ecológico das comunidades de macroinvertebrados bentónicas em águas de transição Portuguesas, que cumprisse os requerimentos da Directiva Quadro da Água (DQA); b) propor alternativas para harmonizar avaliações ecológicas, nomeadamente as baseadas em macroinvertebrados, a larga escala geográfica.

Os Capítulos I, II e III descrevem os passos para atingir o primeiro dos objectivos gerais. Com este propósito, foram usados dados do estuário do Mondego em Portugal, onde informação biológica e ambiental foi sendo regularmente recolhida de 1990 a 2006. Desde a década de 80 que este sistema tem sido afectado por pressões antropogénicas vulgarmente observadas em ambientes costeiros mundiais, tais como eutrofização e perturbações físicas que provocam alterações hidromorfológicas nas características do sistema. Uma evidente diminuição da qualidade ecológica do sistema levou à implementação de medidas de mitigação nos últimos 12 anos, permitindo a sua recuperação. Por estas razões, o estuário do Mondego propiciou um importante laboratório de terreno para testar um conjunto de índices, seleccionados para avaliar o estado ecológico em águas de transição no âmbito da DQA.

O Capítulo I descreve como o mapeamento do habitat permitiu criar expectativas naturais para a distribuição das comunidades biológicas ao longo dos gradientes estuarinos. Dados ambientais, tais como salinidade, composição granulométrica do sedimento e conteúdo em matéria orgânica, de anos recentes (2002 a 2005, abrangendo todas as estações do ano), e relevantes para a estruturação das comunidades de macroinvertebrados bentónicos subtidais, foram usados na identificação de seis zonas distintas no estuário do Mondego (ANOSIM e Análise de Componentes Principais – PCA, PRIMER Software). As tendências ambientais observadas reflectiram significativamente os padrões de distribuição das comunidades de invertebrados

(BIOENV e ANOSIM, PRIMER Software), permitindo perceber a influência dos gradientes naturais no desempenho de ferramentas de avaliação ecológica.

No Capítulo II, três índices ecológicos (índice de Margalef, índice de Shannon-Wiener e 'AZTI's Marine Biotic Index' - AMBI), seleccionados para satisfazer os requerimentos da DQA UE, foram avaliados quanto ao seu potencial para detectar comunidades de macroinvertebrados degradadas e também a sua subsequente recuperação. Para atingir este objectivo, os índices foram testados em três períodos (Primaveras de 1990/1992, 2000/2002 e 2005/2006) sob diferentes intensidades de pressão no estuário do Mondego. As tendências detectadas pelos índices (PERMANOVA) foram coincidentes com a história de perturbações do sistema, respondendo não só aos diferentes tipos de impactos, como também, ás medidas de mitigação introduzidas. Isto permitiu definir condições de referência aproximadas para os índices propostos, que reflectiriam uma melhoria de qualidade no compartimento biológico dos macroinvertebrados bentónicos; levando, simultaneamente, em consideração os gradientes naturais que exercem a sua influência sobre estas comunidades ao longo do sistema.

O Capítulo III descreve o desempenho do BAT - Benthic Assessment Tool, uma ferramenta multimétrica segundo as directrizes da DQA UE, que consiste em três índices, previamente seleccionados para o estuário do Mondego, fundidos num único valor – o Ratio de Qualidade Ecológica (EQR). Este EQR é obtido através de uma Análise Factorial (utilizando PCA como método de extracção, Statgraphics Software) e da projecção das distâncias euclidianas, usando as condições de referência propostas no Capítulo II para limitar a escala de EQR entre 0 e 1, como descrito em Bald *et al*. (2005). Este método foi testado contra os efeitos de perturbação antropogénica usando para o efeito oito anos de dados (Primaveras entre 1990 até 2006) de invertebrados bentónicos subtidais do estuário do Mondego. Apesar de o BAT ser capaz de captar o declínio ecológico e a recuperação do sistema através dos invertebrados, os índices que constituem esta ferramenta multimétrica não contribuíram do mesmo modo para a classificação final obtida. Limitações como a abundância típica de espécies tolerantes em sistemas estuarinos e a classificação ecológica de espécies chave no estuário do Mondego enfraqueceram sobretudo a performance do índice AMBI.

Os Capítulos IV e V exemplificam e propõem modos distintos de intercalibrar avaliações ecológicas num contexto de larga escala geográfica.

O Capítulo IV explica o processo utilizado para adaptar o índice AMBI, desenvolvido na Europa, a uma nova geografia. Usando dados de macroinvertebrados das baías marinhas do Sul da Califórnia, o índice foi calibrado para este novo habitat recorrendo a taxonomistas bentónicos locais para classificar as espécies locais de acordo com os 5 grupos ecológicos usados pelo AMBI. Depois, considerando o 'ranking' e a classificação obtidos para as amostras, o desempenho do AMBI foi validado usando um índice ecológico local, o 'Benthic Response Index' (BRI), e ainda através da Avaliação Profissional de ecologistas bentónicos. A melhor correlação entre os resultados do AMBI e os do BRI ($n$= 685: r= 0.70; Kappa: concordância Moderada para a classificação de amostras) foi obtida aplicando o AMBI baseado numa mistura de classificações ecológicas de espécies previamente disponíveis e outras provenientes da reclassificação por especialistas locais, e incluindo ainda um factor de ponderação para os dados de abundância. O melhor acordo do AMBI com a Avaliação Profissional de ecologistas bentónicos ($n$= 21: r= 0.93; Kappa: Muito Boa concordância para a classificação de amostras), foi atingido usando o critério dos taxonomistas locais para a classificação da estratégia ecológica das espécies, para dados de abundância não transformados. O AMBI demonstrou, no entanto, um menor poder discriminatório do que a Avaliação Profissional dos ecologistas bentónicos na classificação das amostras. O estudo revelou ainda que uma parte significativa dos desacordos entre os dois índices resultou das abordagens seguidas por cada um para classificar as espécies de acordo com as suas estratégias ecológicas.

O Capítulo V propõe uma abordagem baseada no consenso entre as Avaliações Profissionais de vários especialistas (best professional judgement - BPJ) como base para estabelecer uma escala comum para avaliações ecológicas bentónicas. Foram fornecidos a dezasseis ecologistas bentónicos de quatro regiões da Europa e dos Estados Unidos dados de abundância de espécies de macroinvertebrados de doze locais por cada uma dessas regiões, para ordenar de melhor a pior estado ecológico e classificar em quatro categorias. A ordenação dos locais esteve fortemente correlacionada entre os especialistas, independentemente das amostras serem da sua própria região ou não. Observou-se ainda existir uma boa concordância nas categorias de condição ecológica, apesar do acordo ser maior para as amostras nos extremos do gradiente de perturbação. A ausência de enviesamento regional e a conformidade obtida sugerem que a Avaliação Profissional é um meio viável para estabelecer uma escala uniforme que permita calibrar índices de forma consistente em todas as regiões geográficas.

# General introduction

*'Basically everything is an indicator of something*
*but nothing is an indicator of everything'*

(Cairns *et al.* 1993)

## The global need to preserve coastal and marine ecosystems

The marine environment presents diverse habitats that support a high level of biodiversity. Of the 29 nonsymbiont animal phyla described until 1999, all but one have representatives in the ocean, 13 of which are endemic, showing clearly that phyletic diversity is highest in the sea (Briggs 1994, Ray & Grassle 1991, Snelgrove 1999). Still, the vast richness of marine biodiversity remains to be discovered, and recent research indicates that marine diversity is much more extensive and vulnerable than previously thought (Hendricks *et al*. 2006). Scientists point to overwhelming numbers of unknown taxa within unexplored habitats, with estimates of total species numbers suggesting that less than 1% of marine benthic species are presently known (May 1992, Briggs 1994, Snelgrove 1999).

Within marine coastal areas there are a wide variety of habitats such as salt marshes, mangrove forests, coral reefs, seagrass meadows, coastal sedimentary habitats, and algal beds that concentrate important biodiversity hotspots, supporting rich species assemblages (Gray 1997, Duarte 2009). In general, coastal ecosystems rank amongst the most productive in the world, with overall rates of primary production comparable to those of the rainforest (Duarte & Cebrián 1996). However, the rates at which some of these key coastal habitats are being lost globally are 2 to 10 times faster than those in tropical forests (Lotze *et al.* 2006).

Coastal habitats are particularly valuable in regulating the cycling of nutrients which control the productivity of marine and terrestrial primary producers (Costanza *et al.* 1999). These highly productive areas sustain a high diversity of macrofauna and enhance secondary production, giving nearshore ecosystems an important role as nurseries for numerous species (Beck *et al.* 2001, Martinho *et al.* 2009). The high production of vegetated coastal areas also renders them important sites for carbon sequestration, while contributing to sediment stabilization, very important to dissipate wave energy and shelter shoreline from severe physical disturbances (Smith 1981, Duarte *et al.* 2005, Duarte 2009).

The described attributes and properties place marine ecosystems among the most ecological and socio-economically vital on the planet, providing essential goods and services to mankind (Costanza *et al.* 1997, Harley *et al.* 2006). It has been estimated that about 63% of the world ecosystems services are provided by marine systems, most of it coming from coastal systems (Costanza *et al.* 1997). Coastal environments, including estuaries, coastal wetlands, beds of seagrasses and algae, coral reefs, and continental shelves are of disproportionately high value. The benefits human populations derive, directly or indirectly, from ecosystems functions go from the most evident, water, food and raw materials supply, to gas and climate regulation; disturbance regulation, flood and erosion control; nutrient cycling and waste detoxification; biological control, habitat and genetic resources; and recreational and cultural uses (Costanza *et al.* 1997, Worm *et al.* 2006).

However, marine ecosystems are increasingly being changed due to human activities and several impacts threat their integrity and sustainable exploitation (Antunes & Santos 1999, Harley *et al.* 2006, Halpern *et al.* 2007, 2008, Diaz & Rosenberg 2008). A recent effort to map the human impact on marine systems indicated that there was no area unaffected and almost half of them (41%) were already affected by multiple anthropogenic drivers, especially intertidal and nearshore habitats (Halpern *et al.* 2008). In coastal ecosystems, the main impact drivers include land reclamation, dredging, pollution (such as sediment discharges, hazardous substances, litter, oil-spills, nutrient enrichment), over exploitation of marine resources (such as fishing, sand extraction, oil and gas), unmanaged tourism, introduction of alien species and climate change (Gray 1997, Costanza *et al.* 1999, Halpern *et al.* 2007).

These anthropogenic pressures represent serious impacts for coastal environments, affecting among other things the global nitrogen cycle and enhancing coastal eutrophication, undoubtedly one of the most pervasive problems in coastal ecosystems (Gray 1997, Carpenter *et al.* 1998, Boesch 1999). Eutrophication is caused by the increase of supply of organic matter to coastal systems (Nixon 1995), and, over the past several decades, nutrient enrichment has increased dramatically due to human-mediated activities that yield elevated concentrations of nutrients (mainly nitrogen). Expressions of eutrophication include an array of cascading environmental responses such as rapid microalgal and macroalgal growth, harmful algal blooms (HABs), reduced biodiversity by massive mortality of benthic organisms and alteration of biotic community structure, imbalanced food webs, diminished harvestable fisheries, loss of essential habitat (*e.g.*, seagrass and shellfish beds), and ultimately decreased ecosystem resilience (Cardoso *et al.* 2004a, Dolbeth *et al.* 2005, 2007, Kennish & Townsend 2007, Vaquer-Sunyer & Duarte 2008).

Pollution is another problem harassing coastal ecosystems. From industrial toxic chemicals, such as Mercury, DDT, PCBs and related toxic degradation products; to radioactive wastes; and oil spills, all sort of chemicals have permeated world's oceans, contaminating water and sediments (McCahon & Pascoe 1990, Chapman & Anderson 2005, Sindermann 2006). This has contributed to habitat degradation and threatened biological communities, ultimately endangering human health. Also important ecosystem services such as fisheries and tourism become immediately affected (Tietenberg 2003). Consequences to biotic communities are recognized at several levels: development and reproductive impairment; reduced fitness of adult individuals by carcinogenicity, imunotoxicity, endocrine disruption; biomagnification through the trophic chain and in many cases death (Sindermann 2006).

Marine habitats are also being degraded in terms of their ability to harbour living resources, and maintain biodiversity. Of particular concern are those habitats that depend on the establishment of long-lived organisms to provide the complex structure of the habitat, such as coral reefs, coastal marshes and mangroves (Boesch 1999, Halpern *et al.* 2007). Their recovery is particularly slow, and in practical terms, from the human perspective such habitat degradation may be considered irreversible (Boesch 1999). Complete loss of habitat is the most serious threat to marine biodiversity, especially if contiguous but different habitats forming landscape diversity are lost (Gray 1997).

The global effects of climate change will also have an impact on ocean and coastal ecosystems (Feely *et al.* 2004, Fabry *et al.* 2008, Guinote & Fabry 2008). There are expected damaging consequences, endangering several coastal habitats, due to increasing temperature, sea level rise, and decrease of oceanic pH (Harley *et al.* 2006, Orth *et al.* 2006, Halpern *et al.*

2007). While the ocean plays an important role in moderating the build up of atmospheric $CO_2$, especially in a climatic change context (Caldeira & Wickett 2003, Sabine *et al.* 2004), the acidification of seawater resulting from oceanic absorption of $CO_2$ will impact negatively on calcifying organisms (Fabry *et al.* 2008, Guinote & Fabry 2008, Hall-Spencer *et al.* 2008). The temperature increase will amplify hypoxia conditions worldwide, while it enhances the respiratory demand of the organisms, reduces oxygen solubility, and reduces the ventilation of coastal waters by affecting stratification patterns (Vaquer-Sunyer & Duarte 2008). Increasing coastal flooding events are linked to sea level rise but were probably accelerated by historical losses of floodplains and erosion control provided by coastal wetlands, reefs, and submerged vegetation (Danielsen *et al.* 2005). Also, changing coastal currents may affect the distribution and recruitment of fish populations, and altered patterns of precipitation and runoff may affect estuarine and coastal management and strategies to control nonpoint sources of nutrients and other pollutants (Boesch 1999).

All these impacts lead, ultimately, to an overall biodiversity decline and, presently, there are already many registers of populations, key species and even entire functional groups being lost in estuaries, coral reefs and many other coastal systems around the world (Hooper *et al.* 2005, Worm *et al.* 2006). By affecting ecosystems properties, biodiversity loss impairs at least three critical ecosystem services: number of viable fisheries; provision of nursery habitats; and filtering and detoxification services provided by suspension feeders, submerged vegetation, and wetlands (Hooper *et al.* 2005, Worm *et al.* 2006). Loss of filtering services contributes eventually to declining water quality and the increasing occurrence of harmful algal blooms, fish kills, shellfish and beach closures, and oxygen depletion (Dame *et al.* 2002). An increased number of species invasions over time coincided also with the loss of native biodiversity (Worm *et al.* 2006). However, invasions usually do not compensate for the loss of native biodiversity and services, because often they comprise other species groups, mostly microbial, plankton, and small invertebrate taxa (Lotze *et al.* 2006). Ultimately, biodiversity plays an important role in enhancing ecosystems ability to withstand recurrent perturbations by either increasing resistance to disturbance or by enhanced recovery afterwards (Tilman *et al.* 1997, Yachi & Loreau 1999, Stachowicz *et al.* 2002, Stachowicz & Byrnes 2006, Worm *et al.* 2006, Rockström *et al*. 2009).

Recognizing the growing threat that anthropogenic activities represent to marine ecosystems, especially in coastal areas, and also human dependence on them, several legislation and ambitious protection strategies are being implemented worldwide to warrant their ecological integrity (Borja *et al.* 2008a). Initiatives such as the ones promoted by the Convention on Biological Diversity concerning marine and coastal ecosystems with signatory countries from all over the world; or the HELCOM, a regional initiative for the protection of the marine environment of the Baltic Sea, gathering countries around the Baltic; or the International Coral Reef Initiative (ICRI), a partnership among governments, international organizations, and non-government organizations to preserve coral reefs and related ecosystems; the OSPAR Convention, an international cooperation on the protection of the marine environment of the Noeth-East Atlantic; or the increasing establishment of Marine Protected Areas (MPAs) worldwide; are good examples of the growing awareness on the need to protect marine and coastal ecosystems. The concept of 'ecological integrity' suggests a meaning beyond human interests and assumes a holistic approach of ecosystem management (Karr 1996, Borja *et al.* 2008a). Indeed, most of the recent strategies to preserve and manage coastal and marine ecosystems already integrate ecosystems structure, function and processes with the anthropogenic impacts and activities (Antunes & Santos 1999, Borja *et al.* 2008a).

## Assessing ecosystems' condition

Ecological assessments have long evolved from the traditional methods where monitoring efforts focused on obvious, discrete sources of stress, and physical-chemical analyses of water column or measures of specific pollutants were used to detect signs of contamination of aquatic systems (Karr 1991). Several parameters such as temperature, pH, dissolved oxygen, or suspension solids, would give a first overview of the system characteristics. Other parameters such as nutrients (phosphorus and nitrogen) and chlorophyll *a* concentration in the water column would proportionate important data on water quality regarding symptoms of eutrophication. Heavy metals; pesticides; or aromatic compounds, would also be monitored when identified discharges or other non-point sources were suspected, such as industrial discharges, agricultural waters runoff or oil spills.

The sediment, where it reflects many of the effects of pollutants or disturbance, was another important component traditionally surveyed (Burton 2002). Sediments tend to be more conservative than the water column and changes in the sediment usually have high implications in the biogeochemical processes. Analysis to sediment composition, namely the presence of contaminants, organic matter content, total organic carbon, redox potentioal and granulometry can provide crucial information on chemical and organic contamination. However, sediment chemistry only provides information on contaminants present, not on their bioavailability or toxicity (Chapman 2007).

In the 1960's, serious environmental and human health disasters such as the Minata Disease in Japan caused by mercury poisoning, or the detrimental effects of pesticides such as DDT on the environment, awaked public concerns with environmental pollution (Carson 1962, Sindermann 2006). The unpredicted consequences of the use of about 100,000 chemicals, associated with a growing anthropogenic development, supported the use of ecotoxicological tests as an alternative way to detect the effects of contaminants, namely chemicals and radiations, and their interactions with living organisms and on the structure of ecosystems (Depledge 1990, Jørgensen 1998).

However, despite that there has been a major effort over many years to measure the sublethal effects of pollutants, toxicity testing provides generally ''worst case'' rather than ''real case'' information on the responses of laboratory test organisms to field-collected sediments (Chapman 2007, Depledge 2009). Also, such tests provide supporting evidence that responses observed in the resident community are associated with the measured component rather than other potential stressors in the environment. So, among the limitations presented by these methods there is this chance of failing to measure all possible factors and variables that interplay in ecological issues (Kurtz *et al.* 2001). This became particularly significant when nonpoint sources were acknowledge being responsible for impairment of most aquatic systems (Karr 1991). In addition, there was also the chance of not being able, with the analytical means available, to detect small quantities of a given substance that could yet affect biota; and there were also problems in controlling ecotoxicological tests *in situ* and with the limitation of unrealistic reproductions *in vitro*. Besides the difficulty of directly measuring the cause of an impact, there was moreover the increasingly prevalence of multiple and often simultaneous uses and impacts in a system, making it harder to isolate and track specific causes (Karr 1991).

It became evident that the best measure of the effect of an action or substance in the system would be the systems' response itself (Karr 1991). Incorporating measures of biotic integrity in index development has the merit of letting communities of organisms 'tell the story' with respect to classifying habitat quality along the continuum from non-degraded to degraded condition (Diaz *et al.* 2004, Chapman 2007). The biological communities, with their capability of integrating every event occurring in the environment and reflect symptoms from any unusual change became an essential part of many modern assessment techniques that integrate multiple lines of evidence (Burton *et al*. 2002). As result, the past 30 years have witnessed a rapid acceleration of scientific interest in the development and application of ecological indicators.

### *Ecological indicators – concepts and types*

Environmental indicators should reflect all the elements of the causal chain that links human activities to their ultimate environmental impacts and the societal responses to these impacts (Smeets & Weterings 1999). Since the use of environmental indicators became routine two decades ago that the discussion over the important properties of an environmental indicator has been widely revisited (O'Connor & Dewling 1986, Cairns *et al.* 1993, Dale & Beyeler 2001, Fisher *et al*. 2001, Salas 2002, ICES 2001, 2005, UNESCO 2003, 2006, EEA 2005, Magni *et al.* 2005, Rees *et al.* 2006, 2008, Salas *et al.* 2006a,b). In general, an ideal indicator would gather the following features:

(i) be linked to a conceptual stressor–response framework (with the ability to communicate potential cause–effect relationships);

(ii) capable of measuring change or its absence with confidence (robust to influences of confounding environmental factors);

(iii) highly sensitive to stress and anticipatory (early warning of potential problems);

(iv) applicable over a variety of spatial scales, conditions and communities or ecological environments (to support global as well as local comparisons);

(v) desirable operationally (easy to measure, reproducible with minimum measurement error, cost-effective);

(vi) integrative (serves multiple indicator purposes);

(vii) non-destructive (measurement does not cause ecosystem damage);

(viii) easy to understand and communicate (non-specialists need to act on findings);

(ix) scientifically and legally defensible (robust to peer review and wider challenge);

(x) and capable of conveying information that is responsive and meaningful to decision-making (directly tied to management questions and linked to thresholds for appropriate action relative to designated ecosystem goals).

Ecological indicators are a subset of environmental indicators that apply to ecological processes, commonly used to provide synoptic information about the state of ecosystems (NRC 2000, Marques *et al.* 2009a). They comprise biological, chemical, or physical measurements, indices or models that attempt to characterize critical components of the ecosystem (Fisher *et al.* 2001, Kurtz *et al.* 2001, Marques *et al.* 2009a). Ecological indicators mostly address the ecosystem's structure (*e.g.*, genetic, population, habitat, and landscape pattern) and/or composition (*e.g.*, genes, species, populations, communities, and landscape types) and/or functioning (*e.g.*, genetic, demographic/life history, ecosystem, and landscape disturbance processes), accounting for a certain aspect or component, for instance nutrient concentration, water flow, macroinvertebrate and/or vertebrate diversity, plant diversity and productivity, erosion symptoms and, on occasion, ecological integrity at a system's level (Young & Sanzone 2002, Niemi & McDonald 2004, Salas *et al.* 2006a).

The concepts of *indicator* and *index* are often used as synonyms but it is important to clarify that in a regulatory context *indicator* may be a proxy for something different from what it actually measures (Rees *et al.* 2008). An indicator is intended to highlight the status of the system and, for e.g., the European Environmental Agency recognizes distinct types of indicators depending on what they address: descriptive indicators, performance indicators, efficiency indicators and total welfare indicators (Smeets & Weterings 1999). The term should therefore be distinguished from *index*, an aggregation of indicators into a single representation (Rees *et al.* 2008). Indices are considered as one possible measure of the systems status, as they relate to a specific qualitative or quantitative feature of the system (Pinto *et al.* 2009). The selection of key indicators, effective at capturing the system condition and announcing changes compared to the specified objectives, leads then to the elaboration of an assemblage of relevant indices used as operational tools.

There are numerous ecological indices for assessing the health status of aquatic ecosystems, but most of the ecological indices used in the coastal and marine ecosystems result, in their whole, from just a few distinct theoretical approaches (Salas 2002, Jørgensen *et al.* 2005, Salas *et al.* 2006a,b, Marques *et al.* 2009a). An overview of the most common approaches *sensu* Salas (2002) and Marques *et al*. (2009a) follows.

***Indices based on indicator species*** Within these indices, information can be taken at community-level or focusing on species composition (Diaz *et al.* 2004). Either way, they centre their attention on the presence/absence of given indicator species: species extremely sensitive to environmental stress, tolerating only a narrow range of environmental conditions; or species whose appearance and dominance is associated to an environmental deterioration, because they are favoured for that feature, or because they are more tolerant to that type of pollution than other less resistant species (*e.g.*, Glémarec & Hily 1981, Borja *et al.* 2000, Smith *et al.* 2001). There are also those indices based on bioaccumulator species, described as capable of resisting and accumulating various pollutant substances in their tissues (*e.g.*, Goldberg *et al.* 1978, Reish 1993, Storelli & Marcotrigiano 2001).

***Indices based on ecological strategies*** These indices aim to assess environmental stress effects taking the ecological strategies followed by different organisms into consideration. That is the case of trophic indices based on organisms' different feeding strategies (*e.g.*, Word 1978, Milovidova & Alyomov 1992). Other indices have as background the concept of taxon as a functional group, where species within each group, not necessarily related, exploit a common resource in a similar way, and it is expected that disturbances would affect all members of the group. Others account for the different behaviour of two taxonomic groups under environmental stress situations (*e.g.*, Raffaelli & Mason 1981, Belsher 1982, Gómez-Gesteira & Dauvin 2000, Dauvin & Ruellet 2007).

***Indices based on diversity*** Diversity is one of the most used concepts in assessing pollution, supported by the fact that in most occasions the relationship between diversity and environmental disturbances can be seen as an inverse one (Connel 1978, Huston 1979, Mackey & Currie 2001, Magurran 2004). There are indices that measure species richness, others using models of species abundance, indices based on the proportional abundance of different species, and dominance indices (*e.g.*, Simpson 1949, Shannon & Weaver 1963, Margalef 1968, Berger & Parker 1970, Gray 1979, Lambshead *et al.* 1983). There are also those taking into account taxonomic, numerical, ecological, genetic and filogenetic aspects of diversity (*e.g.*, Warwick & Clarke 1995, 1998, Clarke & Warwick 1999).

***Indices based on species biomass and abundance*** These approaches centre on the energy variation in the ecosystem accounting for the variation of organism's biomass and abundance as a measure of environmental disturbances (*e.g.*, Pearson & Rosenberg 1978, Warwick 1986, Warwick & Clarke 1994).

***Indices thermodynamically oriented or based on network analysis*** These result from a wider application of theoretical concepts, following the assumption that it is possible to develop a theoretical framework able to explain ecological observations, rules, and correlations based on an accepted pattern of ecosystem theories (Jørgensen & Marques 2001). That is for instance the case of Ascendency (Ulanowicz 1980, 1986, Ulanowicz & Norden 1990), and Exergy (Jørgensen & Mejer 1979, 1981), a concept derived from the field of thermodynamics, which can be seen as energy with a build in measure of quality.

***Multimetric indices*** These indices attempt to integrate information regarding different aspects of the environment, incorporating metrics that span ecological levels from individual through population to community, ecosystem, and landscape (Karr 2002). They may include physicochemical factors, diversity measures, specific richness, taxonomical composition, the system's trophic structure, etc. (*e.g*., Deegan *et al*. 1997, Weisberg *et al.* 1997, Vollenweider *et al*. 1998, Muxika *et al.* 2007).

## *Benthic macroinvertebrates based indicators*

Benthic habitats' overall importance to ecological processes and as providers of key ecosystems services is unquestionable; however, most of them are yet poorly studied (Snelgrove 1999). This explains why many efforts have been dedicated to understand them and evaluate their condition. In a recent review of ecological indices available to assess the quality of aquatic benthic habitats Diaz *et al.* (2004) summarized the most common parameters and methods used in the construction of such indices (Table I). The majority of proposals included mostly invertebrates and/or fish information; others included other groups of organisms; sediment chemistry and contaminants; secondary production or processes such as metabolism. Few integrated representatives of the entire food chain, phyto or zooplankton, or microbial elements (Diaz *et al.* 2004).

**Table I.** Parameters and methods most commonly used in the construction of benthic indices, after Diaz *et al.* (2004).

(i) physical habitat assessments incorporating descriptive parameters such as temperature and sediment type in an attempt to control for covariance;

(ii) indicator thresholds, such as faunal chemical and pollution tolerance; although the derivation and validation of suitable candidate measures or numerical standards that place the habitat within the adopted categorical scale were seldom justified;

(iii) reference sites or conditions, varying from adjacent areas of habitat and multiple regional reference stations to long-term evaluations of historically characteristic conditions (>100 years);

(iv) species counts such as abundance, biomass and species richness;

(v) species traits or functional groups based on feeding strategy or more complex formulations involving multiple characteristics;

(vi) diversity indices, most often the Shannon–Wiener index of diversity, H', but others such as Simpson's D are routinely used;

(vii) other habitat quality indices based on physical, chemical and/or biological criteria;

(viii) univariate data treatment such as regression, t-tests and ANOVA;

(ix) multivariate data treatment such as MDS, PCA and other similarity based statistics.

Particularly benthic macroinvertebrates, due to their direct dependency on the sediment, became a valuable fraction of the ecosystems and are frequently used as bioindicators in ecological assessments (Elliott 1994, Diaz *et al.* 2004, McLusky & Elliott 2004, Dauvin 2007, Pinto *et al.* 2009). Various studies have demonstrated that the macrobenthos responds relatively rapidly to both anthropogenic and natural stress (Pearson & Rosenberg 1978, Dauer *et al.* 2000, Bustos-Baez & Frid 2003). Using benthic communities has certain advantages because they are disturbance indicators with a real effect on the biota at the species-community level, and they are global pollution/disturbance indicators with easily worked elements. Several other characteristics make macrobenthic organisms useful and suitable indicators, such as:

(i) living in bottom sediments, where exposure to contaminants and oxygen stress is most frequent (Kennish 1992, Engle 2000);

(ii) most benthos are relatively sedentary and reflect the quality of their immediate environment (Pearson & Rosenberg 1978, Dauer 1993, Weisberg *et al.* 1997);

(iii) many benthic species have relatively long life spans and their responses integrate water and sediment quality changes over time (Dauer 1993, Reiss & Kröncke 2005);

(iv) they include diverse species with a variety of life features and tolerances to stress, which allow their inclusion into different functional response groups (Pearson & Rosenberg 1978);

(v) they have fundamental role providing links to the higher trophic levels (birds and fishes), and some species are, or are prey of, commercially important species (McLusky & Elliott 2004, Reiss & Kröncke 2005);

(vi) they affect fluxes of chemicals between sediment and water columns through bioturbation and suspension feeding activities, as well as playing a vital role in nutrient cycling (Reiss & Kröncke 2005).

For these reasons indices based on benthic macroinvertebrates have proved to be effective measurements of coastal and estuarine conditions (Elliott 1994, McLusky & Elliott 2004, Pinto *et al.* 2009), and are commonly used to assess the biological quality of the environment worldwide (*e.g.*, Weisberg *et al*. 1997, Borja *et al.* 2000, 2003, Dauer *et al.* 2000, Llansó *et al.* 2002a,b, Simboura & Zenetos 2002, Dauer & Llansó 2003, Salas *et al.* 2004, Borja & Heinrich 2005, Muxika *et al.* 2005, Dauvin *et al.* 2007, Blanchet *et al*. 2008, Hale & Heltshe 2008).

Still, several authors (*e.g*., Olsgard *et al.* 1997, Diaz *et al.* 2003, ICES 2004, Occhipinti-Ambrogi & Forni 2004, Rakocinski & Zapfe 2005) have underlined a number of disadvantages of the existing benthic indices: they represent a static expression of an ecological condition; they are not explicitly linked to changes in ecological function; they may not be specific with respect to different kinds of stressors; they are subject to underlying taxonomic changes, namely across estuarine gradients; their use can be labour intensive; and they are not applied consistently across biogeographic provinces. Acknowledging, furthermore, that, a benthic invertebrate index, as any other biotic index, is unlikely to be universally applicable without some degree of calibration, because not only natural conditions vary, but also organisms' sensitivity to distinct types of anthropogenic disturbance is likely to differ depending on geography and habitat (Jørgensen *et al.* 2005, Dauvin 2007). In addition, another important drawback for the use of most benthic invertebrate indices is the growingly scarcity of taxonomic expertise, since the majority of them requires species level information (Dauvin 2007).

## Ecological indicators and environmental management

Communication demands simplicity and indicators simplify a complex reality (Smeets & Weeterings 1999). The attention on indicators stems from the need to assess ecological condition in supporting regulatory, stewardship, sustainability, or biodiversity decisions (Niemi & McDonald 2004). The growth in the number of these indices has been fuelled by management's desire for a reductionist approach to the assessment of habitat quality (Diaz *et al.* 2004).

Ecological indicators emerged as powerful tools to measure and synthesize information from specific biological or habitat components (Rees *et al.* 2008, Marques *et al.* 2009a). Therefore, monitoring programs began to include indicators that measured characteristics of the most valued ecological components and those that were most responsive to a diversity of stressors (Kurtz *et al.* 2001).

In an assessment framework context, the main attribute of an ecological indicator is to combine numerous environmental factors in a single value (or category), which might be useful in terms of management and for making ecological concepts compliant with the general public understanding (Salas *et al.* 2006a, Rees *et al.* 2008). This final outcome can easily be interpreted by a non-specialist within a 'good' versus 'bad' continuum, often to meet a minimum legislative requirement (Diaz *et al.* 2004). This allows furthermore monitoring the effects of policy responses (Smeets & Weeterings 1999). Therefore, in such a regulatory context, it is important to pair indicators with thresholds, and deriving them is commonly more challenging than developing the indicators themselves (Fisher 2001, Rees *et al.* 2008). Lately, environmental indicators have also been used as a powerful tool to raise public awareness on environmental issues. Providing information on driving forces, impacts and policy responses, is a common strategy to strengthen public support for policy measures (Smeets & Weeterings 1999).

Several policies - the Clean Water Act (1972) and the Oceans Act (2000) in United States of America, the Oceans Act (1996) in Canada, the Water Services Act (1997) and the National Water Act (1998) in South Africa, the Water Law (2002) in China, the Water Act (2007) in Australia, or the Water Framework Directive (2000) and the Marine Strategy (2008) in Europe - started to incorporate the concept of integrated management, recognizing the need of assessing different elements of an ecosystem. The conservation and management strategies have numerous methodologies at their service to evaluate ecosystems. The Sediment Quality Triad

(SQT), the Weight–Of-Evidence (WOE), the Ecological Risk Assessment (ERA), or the Toxicity Identification Evaluation (TIE), are examples of well established integrative techniques to assess targeted parameters and evaluate the quality status of ecosystem components *sensu* Driving force-Pressure-State-Impact-Response - DPSIR approach (Borja *et al.* 2008a, Rees *et al.* 2008, Chapman 2009). As the name suggests, these techniques bring together information from multiple indicators (*e.g.*, including multiple biological endpoints as well as additional data on chemical, biogeochemical, toxicological, physical, and hydrographic conditions) (Magni *et al.* 2004) and ecological indices are commonly used to translate this data.

### *Importance of selecting ideal indicators*

Demonstrating the cause of environmental impact is a much more difficult task than merely observing that impact has occurred (Cairns *et al.* 1993). Thus, the use of ecological indicators is not exempt of criticisms. The first is that the aggregation results in the oversimplification of the ecosystem under observation (Salas *et al.* 2006a). In addition, no single diagnostic method is suitable in all situations (Cairns *et al.* 1993, Jørgensen *et al.* 2005). However, problems also arise from the fact that indicators account not only for numerous specific system characteristics, but also other kind of factors (*e.g.*, physical, biological, ecological, socio-economic, etc). Thus, indicators must be utilised following the right criteria and in situations that are consistent with its intended use and scope; otherwise they may lead to confusing data interpretations (Salas *et al.* 2006a, Marques *et al*. 2009a). Ideally, each indicator in an indicator set should have a particular function in the problem solving logic of the environmental issues that are to be addressed with the use of indicators.

Due to the panoply of indices available, the US EPA prepared *Evaluation Guidelines* (Jackson *et al*. 2000) to assist with the development and selection of indicators for specific environmental programs. Those guidelines are divided into four phases, from conceptual relevance to feasibility of implementation to response variability and finally to interpretation and utility, focusing on fundamental questions (Fisher 2001, Kurtz *et al.* 2001):

> *1. Is the indicator relevant to the assessment question (management concern) and to the ecological resource or function at risk?*

*2. Are the methods for sampling and measuring the environmental variables technically feasible, appropriate, and efficient for use in a monitoring program?*

*3. Are errors of measurement and natural variability over time and space sufficiently understood and documented?*

*4. Will the indicator convey information on ecological condition that is meaningful to environmental decision-making?*

Besides the use of indicators criteria or evaluation guidelines, the selection of adequate indicators may also be aided by indicator frameworks to identify the right combination of specific environmental management needs, such as the *Driving force-Pressure-State-Impact-Response* (DPSIR) model. According to this systems analysis view, social and economic developments exert *Pressure* on the environment and, as a consequence, the *State* of the environment changes, such as the provision of adequate conditions for health, resources availability and biodiversity. Finally, this leads to *Impacts* on human health, ecosystems and materials that may elicit a societal *Response* that feeds back on the *Driving forces*, or on the state or impacts directly, through adaptation or curative action (Smeets & Weeterings 1999, Elliot 2002). This framework articulates the need for indicators in support of each of its five elements and for the flow and feedback of information among them (Elliot 2002, Niemeijer & de Groot 2008, Rees *et al.* 2008). In a recently updated *enhanced DPSIR* (eDPSIR) approach the inter-relation of indicators became also an explicit part of the indicator selection process (Niemeijer & de Groot 2008). Such conceptual indicator frameworks can potentially play an important role in developing consistent indicator sets (Niemeijer & de Groot 2008).

## *The use of ecological indicators in ecosystem management: example from the EU Water Framework Directive (WFD)*

### *WFD - Overview, main objectives and concepts*

The WFD played an important role in bringing ecological indicators to the centre of the discussion between scientists, managers and stakeholders and set off significant research in the field of ecological assessment, not only in Europe but worldwide. Triggered by this European Directive we have witnessed an unprecedented boom of ecological indices to be applied in the context of integrated assessments of the ecosystems, highlighting the importance of ecological indicators in regulatory actions.

The WFD main objective is the achievement of a 'Good Status' for all waters under its legislation by 2015 (EC 2000). This directive addresses pollution through its objective of 'good chemical status' (based on a list of priority substances), but it goes further by recognising that water should also be able to support healthy ecosystems. Thus, the directive set the basic requirements for measuring the health of surface water ecosystems, identifying four common 'quality elements' to be used in determining 'ecological status': phytoplankton; other aquatic flora; benthic (bottom-dwelling) invertebrate fauna; and fish fauna. The definition of the ecological status should also account for the hydromorphological and physicochemical status of the systems under evaluation, as supporting elements of the biological ones (Vincent *et al*. 2003).

For the assessment of the ecological status, natural systems must be organized into typologies, meaning that they should be characterized based on physical and hydromorphogogical features. This is because it is expected that environmental features will determine which biological communities will establish, and thus similar conditions will lead to similar communities. Hence, systems with similar patterns will be classified under the same type, and type specific reference conditions will be defined for the biological elements present. These procedures are crucial steps in the process of assessing the ecological quality status of the water bodies and many guidance documents and scientific literature have been dedicated to them (*e.g*., European Commission Environment site: http://ec.europa.eu/environment/water/water-framework/facts_figures/guidance_docs_en.htm).

The status of the biological elements and supporting elements must then be integrated into a classification of ecological status following the 'one out, all out' principle of the WFD that sets the lowest classification obtained by any biological element as the overall classification for biota in that system (EC 2000). This precautionary approach, chosen to facilitate objective final assessments, has however been largely criticized. According to some authors, it shows the danger of ending up with average environmental situations, where ecological problems may be levelled out (Borja & Heinrich 2005, Sandin 2005). Also, this 'worst-wins' approach entails the risk of imposing recovery costs not proportionate to the achievable ecological improvement and significantly increasing the risk of misclassification (Sandin 2005, Vuori 2007). Moreover, the knowledge on each of the biological elements is often distinct, with some compartments better studied due to operational assessment tools longer implemented (Borja *et al.* 2004a); additionally, there might be the case where information from a specific biological element is better linked to a specific stressor harassing a particular system (Sandin 2005).

### *The role of ecological indicators in the establishment of the ecological status*

The WFD specifies a five-point scale for surface water quality, from 'High' to 'Bad' ecological status (see EC 2000 - WFD Annex V for status classes' definitions), which represent the deviation occurring from a reference condition (measured as an Ecological Quality Ratio - EQR). In the ecosystem-based approach assumed by the WFD, every ecosystem component, from biota, to physicochemical, and hydromorphological is considered a potential ecological indicator of ecosystem health, and their respective information is accounted for when defining the ecological status (EC 2000). For each element to be assessed the directive stipulates which parameters should be addressed (*e.g.*, alterations on abundance and composition of macroinvertebrates, or composition, abundance and age structure of fish fauna) (see EC 2000 - WFD Annex V Section 1.1 for the complete list). The European Member States (MSs) have selected measures or indices able to reflect alterations on those features due to anthropogenic disturbance, which originated several proposals to assess ecological status from fish to hydromorphological aspects for all categories of surface waters under this Directive (Carletti *et al*. 2009, Poikane 2009, van de Bund 2009).

Regarding the benthic macroinvertebrates biological element in coastal and transitional waters, 12 proposals have already been officially accepted across the EU, applying mainly to coastal waters (Borja *et al*. 2009b, Carletti *et al*. 2009). The proposals go from multimetric approaches, *i.e.*, including several metrics into an equation; to multivariate approaches, *i.e.*, integrating several metrics using multivariate analysis; and univariate approaches. These methodologies vary not only in the way they integrate those metrics but also in the metrics they use to express the three main parameters requested by the WFD for this biological element. Hence:

(i) for assessing the *abundance of invertebrate taxa*, distinct proposals came up resultant from different interpretations of this criterion; these go from measures of abundance distribution (taking the degree of deviation from a log-normal distribution as an indicator), density, biomass, or species richness, accounting for the number of species present or using the Margalef index (Margalef 1969);

(ii) for the *level of diversity*, indices such as the Shannon-Wiener (Shannon & Wiener 1963), Simpson (Simpson 1949), or $ES_{100}$ (Hulbert 1971) were chosen; other methods included the newly developed Taxonomic Spread Index (TSI) (MarBIT, Meyer *et al*. 2008) based on the presence/absence of taxa from the reference species lists, considering species composition and their taxonomical spread, which correlates also with the number of taxa present; or another diversity index, SN (NQI, Rygg 2006), accounting for the number of species and abundance of individuals in a sample, but independent of the relative dominance of species in the sample;

(iii) for the *proportion of disturbance-sensitive and indicative of pollution taxa* most methods chose between the AMBI (Borja *et al*. 2000), the BQI (Rosenberg *et al*. 2004), the Bentix (Simboura & Zenetos 2002), or the MEDOCC (Pinedo & Jordana 2008), which are all indices that take the whole community and classify species according to their tolerance to disturbance; other proposals were the ISI index (Rygg 2002), accounting for the relative presence of pollution-sensitive species; or measuring the deviation observed from reference lists of sensitive and tolerant taxa (MarBIT, Meyer *et al*. 2008).

However, not all methods integrate these three parameters required by the WFD. For example, some account only for the proportion of disturbance-sensitive and taxa indicative of pollution, this is the case of the Bentix (in Greece and Cyprus) or the MEDOCC (adopted by Spain for the Mediterranean - Catalonia and the Balearic islands). On the other hand, there are methods that do not classify species in groups of sensitivity, instead they evaluate the overall changes in species composition using measures of similarity (Bray Curtis) with reference communities (*e.g*. BEQI in Belgium and the Netherlands, Van Hoey *et al*. 2007), or include instead feeding guilds structure (*e.g*. ZKI in Estonia, Borja *et al*. 2009b). On the other hand, there are methods that tend to present some redundancy of metrics incorporated, for example the BBI (Perus *et al*. 2007), by using the BQI index in association with species richness, Shannon-Wiener index and density, since species richness is already accounted for in the BQI by a factor which scales with logarithm to the number of species (Josefson *et al*. 2009).

## Developing consistency between ecological assessments

The growing number of tools and methods available for assessing the health of coastal and marine ecosystems poses a major problem to the comparability between ecological assessments based on distinct approaches or indices. However, it is accepted that there are often major differences among various regions and ecosystem types, and any single indicator may not work consistently across all scenarios (Magni *et al.* 2005). Scientists recognized thus that the time has come now to focus on developing consistency between the assessments and on promoting means to compare methodologies and tools (Diaz *et al.* 2004, Borja & Dauer 2008, Borja *et al.* 2008b).

Developing consistency between assessments is a pathway to greater capacity building, information exchange, and quality control and, ultimately, a basis for improving our understanding of global patterns of marine ecosystem health (Magni *et al.* 2005). There are several ways to reach for this intercomparability and intercalibration, and the following three are examples of alternative approaches.

### *The exercise of intercalibration 'sensu' WFD*

Given the wide range of ecosystems found across Europe, using one method to assess all water bodies would not be adequate. Instead, the WFD establishes a common definition of good ecological status, which Member States must use when developing their national assessment methods. This was also an attempt to profit from historical data and well established long-term monitoring programs within some countries. However, although it allowed the use of an important amount of data and already gathered knowledge, it created on the other hand a comparability problem.

For regulatory purposes it is of sum importance that quality status classifications are in the same scale. To ensure this, an intercalibration exercise was undertaken between Member States sharing systems with common typology (CIS 2003). The essence of this intercalibration was to ensure that all assessment methods for biological quality elements corresponded to comparable levels of ecosystem alteration.

The intercalibration exercise will provide a common scale across Europe, by getting all Member States to have a common interpretation of the normative definitions for the WFD five classes of ecological status. The work focuses on defining the upper and lower boundaries of good status. The line between 'good' and 'moderate' status is particularly important, as it defines whether or not a water body will meet the Directive's 2015 goal of good condition, thus playing a crucial role in identifying where action is needed to restore the quality of Europe's waters.

The intercalibration should focus on specific system type/biological quality element/pressure combinations. This exercise is not necessarily about agreeing common ecological quality ratio (EQR) values for the good status class boundaries as measured by different assessment methods. The EQR is how the chosen methods must express the deviation observed from the predefined reference condition. So, agreeing on common values would only make sense in the case that methods are very similar or when results for different assessment methods are normalised with appropriate methods. This is because distinct methods (*e.g.*, using different parameters indicative of a biological element) may show different response curves to pressures and therefore produce different EQR values when measuring the same degree of impact. These different EQR values should nevertheless, after the intercalibration exercise, reflect a comparable level of anthropogenic alteration to the biological quality element.

The WFD presented three options for intercalibration, which Member States should follow depending on their assessment methods' specificities (CIS 2005). As it is stated in the WFD Guidance on Intercalibration Process-Phase II, the main difference between the options is whether the boundary setting procedure is carried out at Geographical Group level using common metrics (option 1 and 2) (see *e.g.*, in Buffagni *et al.* 2006, Carvalho *et al.* 2008), or at member state level using national metrics (option 3) (see *e.g.*, in Birk & Hering 2006, Borja *et al.* 2007).

### *Direct comparison of indices*

The differences in approach and suites of measures included in different benthic indices leads to questions about whether the application of the various indices would yield different results (Ranasinghe *et al.* 2002). With a scientific or regulatory purpose underneath, many studies have tried to deal with the comparability of distinct ecological assessment approaches by directly comparing ecological indices (*e.g*. for marine macroinvertebrate fauna: Ranasinghe *et al.* 2002, 2009, Diaz *et al.* 2003, Reiss & Kröncke 2005, Labrune *et al.* 2006, Quintino *et al.* 2006, Dauvin 2007, Dauvin *et al.* 2007, Blanchet *et al.* 2008, Borja *et al.* 2008b, Bakalem *et al.* 2009, Pinto *et al.* 2009). The level of agreement between them can provide some degree of intercalibration or validation, and, on the other hand, the frequency, magnitude, and nature of the discrepancies can elucidate how distinct premises may influence indices performance. This is useful either to calibrate indices from different geographical origins, for potential use in new scenarios or contexts; to check for the equivalence of the assessments provided by each approach, for example, within a management context; or to adjust the choice of available indices to specific assessment needs, such as the type of disturbance.

### *Expert judgement*

When it comes to ecological indices comparability, some authors have criticized the two approaches previously described, stating the opinion that the different indices can be harmonized but not intercalibrated since often even indices based on the same notion do not interpret the same information in the same way (Ruellet & Dauvin 2007). Added difficulty is found when comparing the results or intercalibrating information from indices that do not communicate the same information (Ruellet & Dauvin 2007).

As Elliot (2002) pointed out, the marine system is so complex that it is unlikely that scientists will ever be able to fully and quantitatively predict all natural and anthropogenic changes and so best (expert) judgement will have to be relied on for decision making. This statement reflects however that the concept of expert judgement carries within it as much recognized value as vagueness and subjectivity. Some studies have however attempted to use this expert knowledge in objective ways, for example, in evaluating sediment quality (Lee & Jones-Lee 2004, Bay *et al.* 2007), in selecting reference locations in stream biological assessments (Whittier *et al.* 2007), in linking water body stressors and biotic integrity (Manolakos *et al.* 2007), and in assessing quality in marine soft-bottom benthic communities (Weisberg *et al.* 2008).

Standardize expert judgement could help solving the scaling problem of ecological assessments. For this matter, indices intercalibration or validation would be achieved by determining the level of agreement provided by an index with best professional judgment to assess the condition of, for example, biotic communities, since experts will take into account several aspects of community health. Some studies have already showed good evidences that this could be an interesting alternative to calibrate distinct approaches of condition assessment and to weight different lines of evidence when defining the overall ecological condition of natural environments (Bay *et al.* 2007, Ranasinghe *et al.* 2009).

Other scientific issues associated with environmental management can also profit from this approach. For example, it is difficult to determine the reference state of many coastal habitats, because most of these areas have been under the influence of human activities for several decades (Bakalem *et al.* 2009). Within the WFD it is mentioned that Member States could make use of expert judgement to help define quality status objectives representative of natural reference conditions in the absence of undisturbed sites. However, no guidance is proposed on how to use this knowledge in a consistent, practical and measurable way. A scaling agreed by experts, for specific habitats, could help set individual indices thresholds along gradients of disturbance, allowing their application in the absence of reference conditions.

# Objectives and structure of the thesis

Presently, the translation and implementation of scientific research information onto environmental management strategies is one of the most important issues in debate. Scientists are required to make explicit to the public and policymakers the consequences of human activities in an area and so they must be able to derive and make full use of predictive models and/or other decision supporting instruments. The work presented in this thesis was developed to help tackle some of the needs currently identified in this field. In this sense, there were two main objectives to this study:

- To make a contribution for the assessment of the ecological status of Portuguese transitional waters in the scope of the EU WFD, by developing an assessment tool capable of detecting anthropogenic disturbances and effectively measuring the condition of benthic macroinvertebrate communities; and

- To develop consistency between ecological assessments, understanding the ecological principles used to formulate ecological indices, improving therefore our knowledge on benthic communities and our management ability.

To attain these major goals, specific topics were approached, which gave origin to the five chapters in this thesis:

**Chapter I** The WFD recognizes that to warrant an effective ecological assessment it is essential to identify the environmental characteristics, physical and hydromorphological, that characterize natural systems: Typology. This is because these features will determine which biological communities will naturally settle. The first chapter of this thesis deals essentially with this issue when assessing ecological quality in highly variable transitional systems: estuaries. Which natural driving forces structure communities in such environments? How should natural variability be accounted for when assessing estuarine communities? These are important questions to answer to be able to distinguish anthropogenic disturbances from any natural variability associated with the habitat. This chapter is a first step procedure to help attain the first main objective of this thesis of proposing an ecological assessment tool within the scope of the WFD. The Mondego Estuary in Portugal was used as case study.

**Chapter II** Another crucial aspect in the process of ecological assessment is the definition of reference conditions. In the context of the WFD, when it is not possible to derive reference conditions based on a spatial network of reference sites or from modelling, Member States could base reference conditions on expert judgement. In that case, they should use as many sources of information as possible, including monitoring data and relevant information (*e.g.*, historical data) to improve confidence in their understanding of how the biological quality element responds to increased pressure and hence the values for that element under conditions of no or only very minor human disturbance. The second chapter of this thesis deals with the challenge of defining such reference conditions for habitats with a long history of anthropogenic disturbance, as are estuaries. In such cases, reference sites can hardly be found, and to overcome this, expert judgement based on the use of historical data was adopted to set reference conditions expectations. A time series from the Mondego Estuary (Portugal), covering periods of distinct pressures and different types of impacts in the system, was used to help understand ecological indices variation. The positive trends observed allowed identifying approximate reference values for specific indices in such habitats.

**Chapter III** To assess a system's overall ecological status, the WFD defined which parameters should be addressed for each biological element. In the case of macroinvertebrate fauna, alterations to abundance and composition should be reported, including the identification of sensitive and indicator species of pollution. In this chapter, three indices providing complementary information (Margalef index, Shannon-Wiener index and the AZTI's Marine Biotic Index) were used as a multimetric tool – BAT (Benthic Assessment Tool) - taking the reference conditions defined in the previous chapter. Its performance evaluating the status of this biological element *sensu* WFD was tested. The classification results were discussed in the context of estuarine environments constraints, using the Mondego Estuary as case study. With this third chapter the first major goal of this thesis is accomplished.

**Chapter IV** Often there is the need to calibrate indices from different geographical origins, for potential use in new scenarios or contexts. However, the differences in approach and suites of measures included in different benthic indices leads to questions about whether the application of the distinct indices would yield different results. A comparison between indices outputs allows evaluating the level of agreement between them and can provide some degree of intercalibration or validation. On the other hand, the frequency, magnitude, and nature of the discrepancies can clarify on how distinct premises may influence indices performance and hence their adequacy to meet specific objectives. In this chapter we took a widely used European index, the AZTI's Marine Biotic Index (AMBI), and tested its potential application in a new geographical area in the United States: the Southern California marine bays. The AMBI was compared with a local index, the Benthic Response Index (BRI), and also validated against expert judgement. The two approaches, though based on similar ecological principles, classify species regarding their sensitivity/tolerance to disturbance using distinct methods; reintroducing a controversial discussion about the geographical or habitat dependence of species' ecological strategies. This chapter is a contribution to meet the second major goal of this thesis.

**Chapter V** With the implementation of environmental assessment frameworks worldwide and the consequent flourishing of indices proposals, it became important to agree on what defines good ecological condition and how could we develop consistency in the assessments. The use of expert judgement is a valuable instrument that can build up consensus in this field. In this chapter, coastal marine benthic macroinvertebrate communities were used to test whether a wide scale geographical consensus could be achieved in defining ecological condition using Best Professional Judgement. Experts from four geographical areas in Europe and the United States were asked to classify samples from across those regions and elucidate on the rationales for their assessments, on what could be a first step for developing a benthic condition common scaling. This last chapter is an important contribution to the field of ecological quality assessment, where the second major goal of this thesis is fulfilled.

Each of the previous chapters constitutes a scientific paper, either published or in preparation for publication:

**Chapter I**

Teixeira H, Salas F, Borja A, Neto JM, Marques JC, 2008. A benthic perspective in assessing the ecological status of estuaries: the case of the Mondego estuary (Portugal). Ecological Indicators 8: 404-416.
**DOI: 10.1016/j.ecolind.2007.02.008**

**Chapter II**

Teixeira H, Salas F, Neto JM, Patrício J, Pinto R, Veríssimo H, García-Charton JA, Marcos C, Pérez-Ruzafa A, Marques JC, 2008. Ecological indices tracking distinct impacts along disturbance-recovery gradients in a temperate NE Atlantic Estuary - Guidance on reference values. Estuarine, Coastal and Shelf Science 80: 130-140. **DOI: 10.1016/j.ecss.2008.07.017**

**Chapter III**

Teixeira H, Neto JM, Patrício J, Veríssimo H, Pinto R, Salas F, Marques JC, 2009. Quality assessment of benthic macroinvertebrates under the scope of WFD using BAT, the Benthic Assessment Tool. Marine Pollution Bulletin 58: 1477-1486.
**DOI: 10.1016/j.marpolbul.2009.06.006**

**Chapter IV**

Teixeira H, Weisberg SB, Borja A, Ranasinghe JA, Cadien DB, Velarde RG, Lovell LL, Pasko D, Phillips CA, Montagne DE, Ritter K, Salas F, Marques JC. Calibration and validation of the AZTI's Marine Biotic Index (AMBI) for Southern California Marine Bays. ***In preparation.***

**Chapter V**

Teixeira H, Borja A, Weisberg SB, Ranasinghe JA, Cadien DB, Dauer DM, Dauvin J, Degraer S, Diaz RJ, Grémare A, Karakassis I, Llansó RJ, Lovell LL, Marques JC, Montagne DE, Occhipinti-Ambrogi A, Rosenberg R, Sardá R, Schaffner LC, Velarde RG, 2010. Assessing coastal benthic macrofauna community condition using Best Professional Judgement - developing consensus across North America and Europe. Marine Pollution Bulletin 60: 589-600.
**DOI: 10.1016/j.marpolbul.2009.11.005**

# Chapter I

## A benthic perspective in assessing the ecological status of estuaries: the case of the Mondego estuary (Portugal)

**Abstract** In transitional waters the process of defining reference conditions (in the scope of the WFD) must account for the natural great variability of such environments. Therefore stretches reflecting different physical-chemical and biological conditions throughout the system should be defined in order to correctly establish benthic specific reference conditions. Both salinity and sediment structure are major factors controlling physical-chemical conditions and therefore organisms' distribution within an estuary. These environmental variables (salinity, sediment grain size composition and organic matter content) patterns were studied in the Mondego estuary and some clear gradients emerged. Also ecological indices (AMBI, Margalef and Shannon-Wiener) were applied to subtidal benthic communities of the Mondego estuary and, generally, there was evidence of a decrease in diversity in the estuary from the downstream section towards its inner parts, but also differences were found between areas of distinct sediment composition. After comparing environmental patterns with biodiversity trends, the information was used to define homogeneous sectors along a temperate estuary in Portugal. In the Mondego estuary six zones, covering the main physical gradients affecting benthic communities, were defined: four in the northern arm and two in southern arm. Zones established will allow future determination of benthic reference conditions adjusted for each of the sectors, according to their characteristics, and consequently the conditions they provide for benthic assemblages settlement.

## Keywords

Water Framework Directive
Estuaries
Salinity
Sediment composition
Macrozoobenthos
Ecological indices
Zonation

## Introduction

The European Water Framework Directive (WFD) implies the definition of water typologies and the choice of methodologies to evaluate their ecological status. After this first step, the reference conditions for each type need to be determined and further applied to each of the water bodies (EC 2000, Vincent *et al.* 2003, Bettencourt *et al.* 2004). For transitional waters this procedure cannot be effectively implemented just by following WFD's guidelines (Vincent *et al.* 2003). According to this Directive, reference conditions should remain type-specific. But in transitional waters, due to strong salinity gradients, a single water body can show different physical-chemical conditions through its extension (Bald *et al.* 2005). Hence, estuaries are critical examples where conditions at the mouth, areas with higher marine influence, are different from those at the inner parts, essentially polyhaline and mesohaline sections, and both of these from the ones at the head, under oligohaline conditions and fresh tide influence (McLusky 1971, Elliot & McLusky 2002). So, if only a single set of reference values was to be defined for each type, in transitional waters, such as estuaries, one would most probably use conditions at the mouth and lower reaches. This choice would condemn naturally impoverished inner reaches of estuarine systems to very low ecological classifications. On the contrary, it could lead to estuarine downstream parts to be classified always as in an excellent ecological condition, regardless possible existence of impacts, if reference values for that type had been based upon upstream biological communities.

Since transitional waters are more complex than other categories of surface waters, they must be dealt with in a special manner when implementing the scheme of quality assessment of the WFD. Therefore, prior to the use of environmental quality assessment tools, the existence of different sections within a system should be evaluated. If individual partitions of a single water body have particular characteristics that reflect onto biological communities, then specific reference conditions should be considered for each stretch of an estuary (Bald *et al.* 2005, Muxika *et al.* 2007).

This issue was recently addressed in other works, which have proposed a division of systems in water bodies according to their specificities (Prior *et al.* 2004, Bald *et al.* 2005, Ferreira *et al.* 2006). Morphological characteristics; physical-chemical features, such as salinity; human pressures and other impacts, were some of the criteria used to define homogeneous water bodies within some systems (Borja 2005, Borja *et al.* 2006, Ferreira *et al.* 2006).

In some systems of the Iberian Peninsula (Basque Country - Spain), in order to reflect water bodies' specific hydrographical properties, estuaries were split into different water stretches, using salinity gradient as a characterization factor and following the Venice Symposium System for classes definition (Table I) (Anonymous 1959, Bald *et al.* 2005, Muxika *et al.* 2007). In a similar way, Ferreira *et al.* (2006) developed a stepwise methodology which was applied to some Portuguese water bodies. In this methodology water bodies' delimitation was dependent on physical-chemical aspects such as morphology and salinity, and also on pressure and state indicators, such as nutrient loading, dissolved oxygen and chlorophyll *a* levels.

**Table I.** Salinity classes from the Venice Symposium classification system (Anonymous 1959). Classification of the 25 sampling stations of the Mondego estuary, according to Venice Symposium system.

| Salinity ranges | Venice classification system classes | Mondego estuary stations |
|---|---|---|
| < 0.5 | Freshwater | – |
| 0.5 to 5 | Oligohaline | 20, 21, 22, 23, 24, 25 |
| 5 to 18 | Mesohaline | 17, 18, 19 |
| 18 to 30 | Polyhaline | 6, 7, 8, 9, 11, 12, 13, 14, 15, 16 |
| 30 to 34 | Euhaline estuarine * | 1, 2, 10, 3, 4, 5 |
| > 34 | Euhaline sea * | – |

* Adapted by Bald *et al.* (2005) for Atlantic coasts.

It is widely accepted that, one of the main factors influencing species distribution in estuaries is salinity (Bulger *et al.* 1993, McLusky & Elliot 2004). The Venice Symposium salinity system has been widely applied and accepted but, from the biological point of view, its compliance with species distribution does not always gather consensus (Bulger *et al.* 1993, Attrill & Rundle 2002, Chainho *et al.* 2006).

Apart from salinity, other features such as sediments particles size, organic matter content, dissolved oxygen, depth, hydrodynamic conditions and vegetation cover, may also determine species' distribution in estuaries (Marques *et al.* 1993, Schlacher & Wooldridge 1996, Ysebaert *et al.* 2003, Houte-Howes *et al.* 2004). Therefore, in some systems, assigning benthic reference conditions to stretches based solely on this parameter might result difficult. In the Mondego estuary (Portugal) distinct benthic communities occur in zones which have similar salinity but are different regarding habitat (e.g. sediment grain size or presence of vegetation cover) or hydrological regime (Marques *et al.* 1993, Dolbeth *et al.* 2003, Rodrigues 2004).

The descriptive definitions within EUNIS (European Nature Identification System) have been used to suggest qualitative reference conditions for the coastal and transitional waters of United Kingdom (UK) and Republic of Ireland (RoI) (Prior *et al.* 2004). As several habitats can be present within a given water body type, qualitative reference conditions are described for the suggested predominant habitats and associated communities within it (Prior *et al.* 2004, Muxika *et al.* 2007). Nevertheless, to establish quantitative reference conditions for transitional waters, UK and RoI also considered the salinity regime since habitats in lower salinity areas naturally support less diverse faunal assemblages compared to higher salinity areas. Borja *et al.* (2004a) and Muxika *et al.* (2007) used a mixture of historical data, expert judgment and modelling in assessing benthic reference conditions in northern Spain.

Despite the numerous definitions of estuary found in the literature (see review by Elliot & McLusky 2002), it is consensual that the spatial salinity gradient and its temporal variability is an intrinsic characteristic to such systems. Hence, whether or not the Venice classification applies, from the WFD perspective, those natural fluctuations cannot prevent systems from management, whilst they must be accounted for when developing assessment tools. And since the biological quality elements are the most important ones when defining the Ecological Quality Status (EQS) of a water body (EC 2000), these procedures must not be concluded without regarding their compliance with biological communities.

This paper explores extensive data on salinity, sediment grain size and organic matter content and on benthic communities from the Mondego estuary. These features where regarded as major factors driving subtidal benthic communities' distribution within this estuary (Marques *et al.* 1993, Rodrigues 2004), and therefore selected for this study. The objective of this paper is to distinguish zones within the estuary from a benthic perspective, in order to correctly define sections where distinct benthic reference conditions should prevail.

## Methodology

### *Study site*

Within the WFD implementation procedure, the Mondego estuary (Figure 1) was assigned to Portuguese Transitional Waters of the Type A2. This type includes mesotidal (tidal range 1–3 m) polyhaline estuaries of well-mixed waters and under irregular river discharges (Bettencourt *et al.*

2004). The Mondego estuary is located on the Atlantic coast of Portugal (40º 08 N, 8º 50 W) and is influenced by a warm-temperate climate.

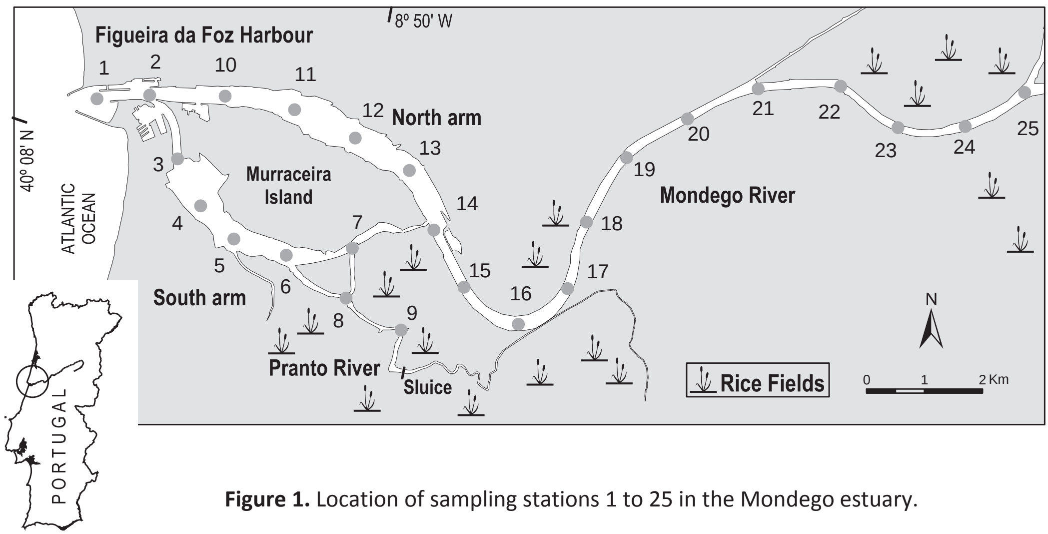



**Figure 1.** Location of sampling stations 1 to 25 in the Mondego estuary.

The Mondego River, which drains a 6670 km$^2$ basin, ends in front of the city of Figueira da Foz. Its estuary is approximately 8.6 Km$^2$ and its upstream limit, defined as a function of the tidal influence, was settled 21 km upstream from the mouth. This estuary has a mean water flow of 79 m$^3$ s$^{-1}$, which in rainy years can reach above 140 m$^3$ s$^{-1}$, dropping to 27 m$^3$ s$^{-1}$ in dry years. The tidal range resembles that of other Portuguese estuaries, varying between 0.35 to 3 m. The last 7 km, near the mouth, consist of two arms separated by a river island, the Murraceira, formed by the deposition of detrital materials transported by the river. The northern arm is the deepest (4–10 m during high tide), and constitutes the main navigation channel and the location of the Figueira da Foz harbour. The southern arm is shallower (2–4 m during high tide) and up to very recently (May 2006) almost silted up in the upper zones; hence, freshwater outflow moves mainly via the northern arm. Hydraulic circulation in the southern arm has been mostly dependent on tides and on the freshwater input from the Pranto River, a small tributary. Discharges from the Pranto are controlled by a sluice and regulated in accordance to the water needs of rice fields placed on the Lower Mondego Valley.

Due to agricultural lands' drainage, apart from being N-enriched ($NO_3$-N, $NO_2$-N and $NH_4$-N), the Pranto water discharge contains several pollutants, including suspended solids and pesticides (Marques *et al.* 1999, Martins *et al.* 2001, Lillebø *et al.* 2005).

In the 1980s, *Zostera noltii* beds extended along the downstream half part of the southern arm (with an estimated surface of 15 ha). By the mid-1990s, *Zostera noltii* had become restricted to a small patch (0.02 ha) located downstream, having been replaced elsewhere by fast-growing green macroalgal blooms (Marques *et al.* 2003, Cardoso *et al.* 2004a). Such shift in the benthic primary producers had effects on the structure and functioning of the biological communities, including the species composition, inducing the emergence of a new selected trophic structure (Marques *et al.* 1997, 2003, Patrício & Marques 2006). More recently (from 2000 onwards), these meadows have shown a recovery and latest data point to an area increase up to 4.2 ha (in 2005), within the downstream zone of southern arm.

### *Data series*

### Biological data

In spring 2002, the subtidal soft-bottom communities were sampled along 14 sampling sites, covering the last 8 km of the estuary, located approximately 1 km apart from each other, along both arms of the Mondego River. From 2003 onwards 11 new sampling sites were added covering the upper parts of this system (Figure 1). These 25 stations were seasonally sampled in the winter, spring, summer and autumn until spring 2005 (except summer 2004).

Samples with five replicates were taken using a Van Veen grab LMG model with an area of 0.078 $m^2$ (except 2002, grab area of 0.0496 $m^2$). Samples were sieved *in situ* through a 1 mm mesh sieve bag and preserved with 4% buffered formalin solution. Specimens collected were later counted and identified at species level, whenever possible.

### Physical-chemical data

*In situ* salinity values, at surface and bottom, were registered at high tide conditions. Monthly measurements took place at 25 stations in the years 2003, 2004 and until May of 2005 (except the period from August to December 2004). In 2002, only bottom waters salinity was measured, and just in the first 14 stations, in spring. Although complete tidal cycle salinity measurements would be the ideal data set to treat for this purpose, there was no such data available for this estuary.

Sediment samples were also taken seasonally at each of the 25 stations, using a Van Veen grab LMG model, according to the procedures described above. As for salinity data in 2002, only spring samples were taken and at the first 14 stations. Sediment organic matter content was quantified by weight difference between sediment after oven drying at 60 °C for 72 h and after combustion at 450 °C for 8 h, and expressed as a percentage of total sample weight. Grain size analysis was carried out by mechanical separation through a column of sieves with different mesh sizes. Sediments were classified as coarse sand (≥0.5mm), fine + medium sand (>0.063 and <0.5mm), silt (>0.038mm and <0.063mm) and clay (≤0.038mm), adapted scale from Brown & McLachland (1990). Grain composition was expressed in percentage of total sample weight.

### *Data analysis*

Salinity data of all seasons were used, thus all seasonal variability could be embraced. These data were used to calculate salinity features of sampling stations such as mean, median, maxima, minima, 25% and 75% quartiles (Q25 and Q75) and standard deviation, which were then used in a cluster analysis (CA) to compare similarity among the 25 sampling stations. The cluster analysis was applied in order to group sites with similar salinity features that consequently will characterise water bodies with similar physical-chemical properties (Bald *et al.* 2005). The CA was performed with PRIMER 5.2.6 © software (Software package from Plymouth Marine Laboratory, UK), similarity distance between groups was calculated by normalised Euclidean distance, using Group Average as clustering method.

These data (mean, median, maxima, minima, Q25 and Q75 and standard deviation) were also used to perform a n-MDS (non-metric multidimensional scaling) to verify how stations were distributed relatively to each other (normalised Euclidean distance was used as similarity measure for similarity matrix calculation, without any transformation or standardization). These salinity descriptors were chosen because it produces a more clear output than if salinity values registered through time are used. Moreover, by using these descriptors more importance is given to stations showing typical conditions than to isolated or extreme values that might have occurred during the study period. Afterwards, sampling sites were labelled according to salinity classes and an ANOSIM (analysis of similarity) was applied to see whether the proposed groups were significantly dissimilar.

PRIMER 5.2.6 © software (Software package from Plymouth Marine Laboratory, UK) was used for n-MDS and ANOSIM.

After estuarine characterization according to salinity features, the other environmental variables (sediment grain size and OM content) were added to stations' description. The previous procedure was followed for these environmental variables (sediment grain size and organic matter content - OM). Sediment parameters were treated as 4 separated variables (% of coarse sand, % of fine + medium sand, % of silt and % of clay). A n-MDS was performed with the descriptors (mean, median, maxima, minima, Q25, Q75 and standard deviation) of each of the 4 sediment parameters and the % of OM.

A principal components analysis (PCA) was applied to the most determinant environmental variables (salinity, grain size, OM content) in benthic communities' distribution according to literature for this estuary. In this analysis full samples data was used, not descriptors of the parameters as in previous statistical analysis above mentioned. To avoid PCA ordination to vary with variables' scale changes, a correlation-based PCA was performed, *i.e.*, normalise all axes, so that they have comparable, dimensionless scales. Therefore, and although grain size categories and organic matter content were expressed in percentage, no data transformation was done before running PCA (Clarke & Warwick 2001). The PCA was performed with PRIMER 5.2.6 © software (Software package from Plymouth Marine Laboratory, UK), data normalised before calculation.

Some ecological indicators, at use in Portuguese WFD benthic assessment tools, were calculated for biological data: Margalef index (d) (Margalef 1968), Shannon-Wiener index (H') (Shannon & Weaver 1963) and the AZTI Marine Biotic Index (AMBI) (Borja *et al.* 2000). The AMBI was calculated by using the software available at www.azti.es and the species-list of July 2006, following the guidelines of Borja and Muxika (2005).

An attempt was made to search for correlation of ecological indices results with the summary provided by each of the first 3 axes of the PCA of environmental variables. A regression analysis (using Statgraphics Plus 5.0) was used to see whether these could explain some of the variability found in the indices. An ANOVA was previously applied to test for the significance of the relationship between each of the indices and three principal components information (scores). Then a linear model approach was used to describe the relationships found.

A BIOENV analysis was applied to link environmental variables selected to biotic data (performed with PRIMER 5.2.6 © software). For the biotic data, the same similarity matrix used for the n-MDS procedure was used. For abiotic parameters, normalised Euclidean distance was used as similarity measure for similarity matrix calculation, without any transformation or standardization. The Spearman rank correlation (ρ) method was chosen, and the maximum number of trials was 6, to allow all parameters to be tested together. The best overall combinations were selected.

Density data of benthic communities (after square root transformation and Bray-Curtis similarity calculation) were used to perform a n-MDS ordination of biotic assemblages during the studied period along the estuary. An ANOSIM was applied in order to see whether the communities in each sector of the estuary were significantly dissimilar. The n-MDS and ANOSIM analysis were performed using PRIMER 5.2.6 © software.

## Results

As it happens for most of the temperate estuaries, the surface and bottom water's salinity patterns were different (Figure 2). Since benthos is more influenced by bottom salinity values, these were adopted in this study.

Some stations could be clearly assigned to a salinity range, while others presented a high variability that made it difficult to assign them to a specific class (Figure 2b). Stations like those in the inner part of the southern arm (7 to 9) or some stations in the upper part of the estuary (16 to 19) registered salinity values across several classes of the Venice Symposium classification. Such large ranges indicate that these stations sometimes presented conditions that do not fit the characteristics described for a single salinity class. The cluster of sites based on bottom salinity values indicated five distinct groups: the group of stations at the head of the estuary (21 to 25), the sites near the mouth of the estuary (1, 2, 10, 3, 4 and 5); sites in the northern and southern arms of the Mondego (11, 12, 13, 14, 15 and 6, 7, 8, 9); a fourth group of stations located upstream in the northern arm (16, 17) and another group comprising stations in the inner parts (18 to 20). An approximate correspondence of the groups formed with the Venice salinity classes is proposed in Table 1.

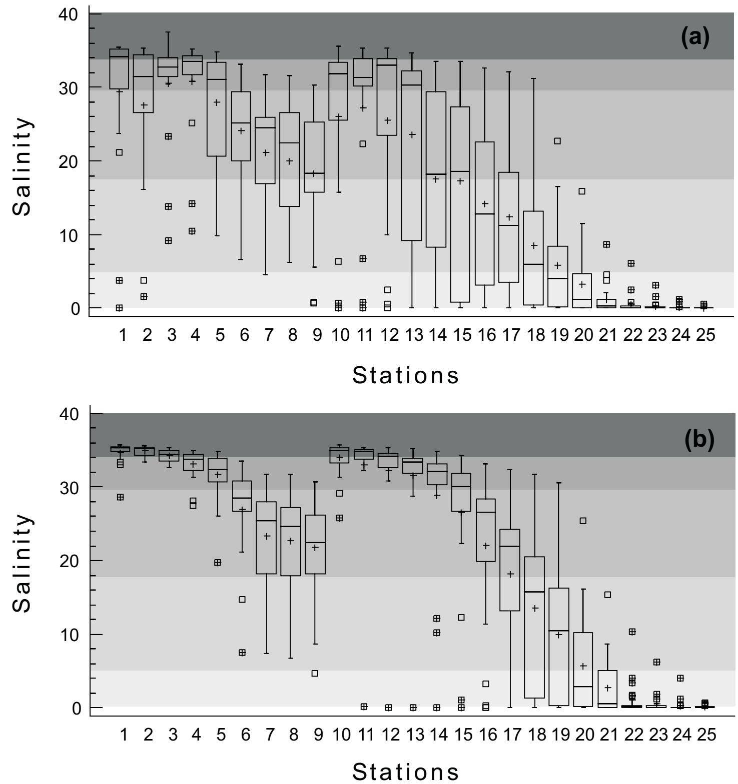



**Figure 2.** Variation of surface (a) and bottom (b) salinity values at each of the sampling stations from 2003 to 2005 (mean: cross; median: horizontal line; Q25 and Q75: box; maximum and minimum: whiskers extending from box; and when present also signed outliers: squares; and far outliers: crossed squares). Venice Symposium salinity classes on grey scale as shown in Table 1 (from light to dark grey: 0.5-5, 5-18, 18-30, 30-34, 34-40).

Stations 16 and 17, showed a wide range of salinity variation, resulting more difficult to be assigned to a class. Thus, since station 16 presented typical average polyhaline conditions, while station 17 presented lower mean salinity values, just in the transition between polyhaline and mesohaline classes, they were split and classified respectively as polyhaline and mesohaline. Also, station 20, despite being grouped close to stations 18 and 19, presented typical oligohaline salinity values, and was therefore included in this class (Table 1).

During the study period, euhaline (E) stations presented a mean bottom salinity of 33.8; polyhaline (P) stations of 27; mesohaline (M) stations of 14 and oligohaline (O) stations of 1.7 (see Table 1 for stations in each category).

The n-MDS plot clearly reflects the spatial distribution of the sampling stations along the estuarine gradient (stress = 0.01) (Figure 3). In fact, the ordination plot of sampling stations based on bottom salinity data, after cluster analysis classification, almost overlaps their real spatial distribution along the estuary from the mouth towards upstream areas (see Figure 1).

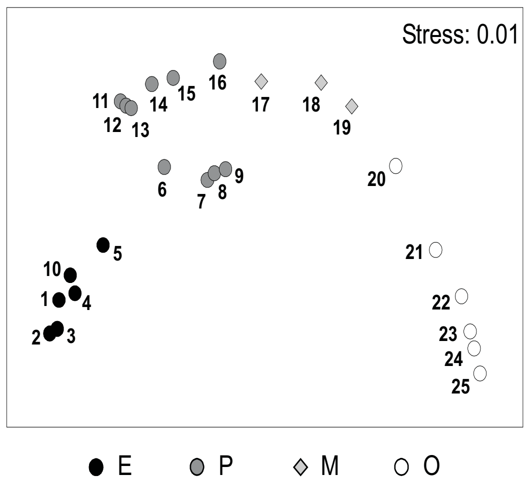



**Figure 3.** n-MDS ordination based on the following variables obtained from bottom salinity data: mean, median, maxima, minima, Q25, Q75 and standard deviation (stress: 0.01). Stations are labelled as a function of salinity characteristics in accordance with the Venice classification classes: E - euhaline, P - polyhaline, M - mesohaline, O - oligohaline.

ANOSIM analysis of bottom salinity features indicated a clear separation of the groups defined according to Venice Symposium salinity classes (R = 0.957; $p$ = 0.001). Regarding pairwise differences, all zones were significantly different ($p < 0.05$) and all presented a high R value, indicating a good segregation of groups (E/P: R = 0.989; $p$ = 0.001; E/M: R = 1; $p$ = 0.012; E/O: R = 1; $p$ = 0.002; P/M: R = 0.829; $p$ = 0.003; P/O: R = 0.998; $p$ = 0.001; M/O: R = 0.870; $p$ = 0.012).

Regarding sediment parameters, evidence of a downstream to upstream stations gradient was also found (Figure 4) (n-MDS stress = 0.07), yet less evident than the one presented by salinity. Four groups can be identified in this n-MDS plot. Station 1, in the estuary's mouth, setting the limits to marine environment, appears isolated in the n-MDS plot. Also southern arm inner stations (6 to 9) are located quite apart from the rest, indicating distinct sediment composition. Then northern arm stations (2, 10, 11, 12, 13, 14 and 15) and southern arm stations located nearer the mouth and with higher marine influence (3, 4, 5) are also aggregated in the n-MDS representation, indicating similar sediment composition. Near this last group, in a continuum, are stations beyond 15, grouped together and apart from both those at the mouth (1) and inner south arm (6 to 9).

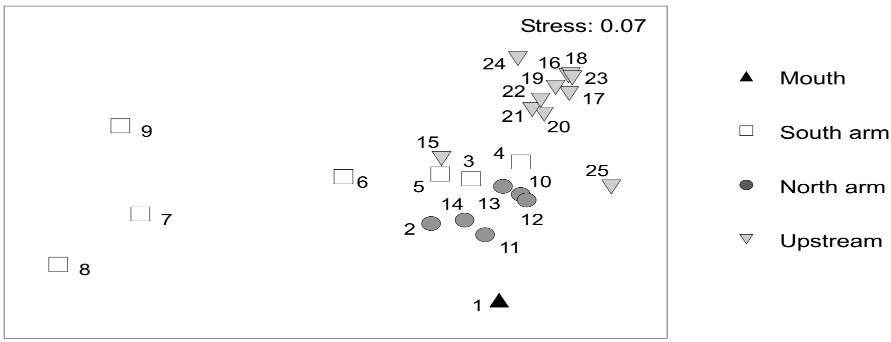



**Figure 4.** n-MDS ordination based on the following variables obtained from sediment grain size and organic matter content data: mean, median, maxima, minima, Q25, Q75 and standard deviation (stress: 0.07). Stations are grouped as a function of their location in the estuary.

Taking information from previous environmental variables into a PC analysis, some patterns regarding these features could be observed among the 25 sampling stations. Figure 5a displays the first two axes (PC1 and PC2) of the PCA ordination on the six environmental variables selected for this estuarine characterization (salinity; % OM, % of coarse sand, % fine + medium sand, % silt and % of clay in the sediments). The first component accounts for 54.1% of the variability in the data set and the first two components account for 82.2%, so the 2-d plot provides a good summary of the sample relationships. In the PCA plot (Figure 5a) two distinct groups of stations can be identified, which represent two gradients, both from the mouth and into the inner parts (stations) of each of the estuarine arms: north and south. In Figure 5 (b to g) some patterns were observed, when superimposed the value of the studied parameters as a bubble plot on the PCA points (stations).

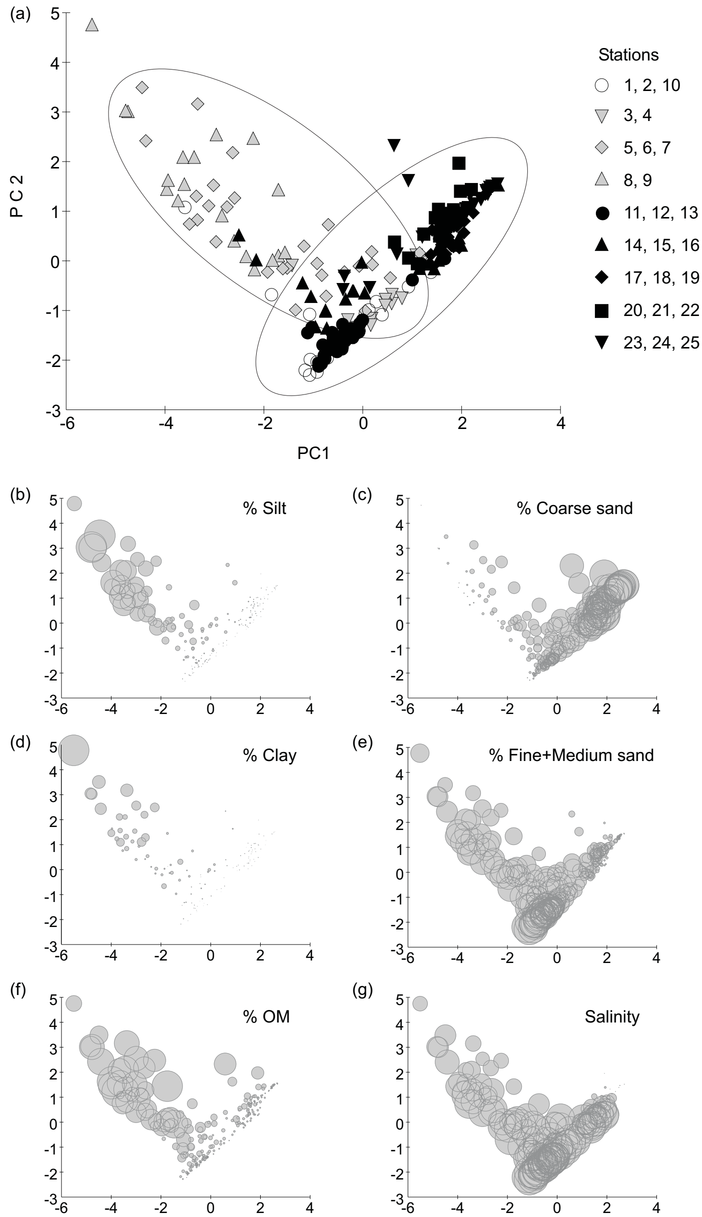



**Figure 5.** a) Two-dimensional PCA ordination of the six environmental variables (salinity, % coarse sand, % fine + medium sand; % silt and % clay in the sediment, and % of organic matter in the sediment) for the stations (1-25) along the estuary (%variance explained= 82.2%). b)-g) the same PCA ordination but with superimposed circles of increasing size with increasing value of each of the six variables mentioned.

PC1 represents an axis of increasing grain size since, except for coarse sand percentage (r = 0.493) (Figure 5c), all coefficients related to sediment grain size are negative (% fine + medium sand: r = -0.421; % silt: r = -0.438 and % clay: r = -0.363) (Figure 5e, b, d). Also % of organic matter in the sediment (r = -0.420) and salinity (r = -0.281) showed an opposite trend (Figure 5f, g).

In the second axis both salinity (r = -0.499) and % of fine + medium sand (r = -0.449) have negative coefficients indicating a decrease of both parameters from stations near the mouth to inner estuarine stations. Roughly equally weighted, but showing an opposite trend, are finer particles (% silt: r = 0.388 and % clay: r = 0.424), increasing along PC2 and towards inner stations of the south arm (Figure 5b, d). As for the % OM in the sediment (r = 0.362) and % coarse sand (r = 0.298), although weaker, they share the trend of parameters previously analysed to increase along this second axis.

Still, almost 20% of the variability in our samples remains unexplained and is only understood when information from PC3 (8.1%) and PC4 6.8%) is regarded. PC3 is dominated mostly by salinity (r = -0.797) and PC4 by % of clay (r = -0.727).

The regression results between the summary (PCA scores) provided by each of the first 3 axes of the PCA of environmental variables and the ecological indices applied to biotic data revealed trends that can be observed in Figure 6. Regression analysis of the three indices with PC1 points to a decrease in terms of ecological quality with increasing grain size of the sediment, towards upstream areas of the estuary. AMBI has a positive slope ($\beta$ = 0.16) due to its logarithmic, where an increase reflects a worse condition. On the contrary, an increase in H' and d values points to higher diversity levels. All three relationships are significant ($p \leq 0.01$) but only Shannon-Wiener presented moderately strong correlation with this axis (AMBI: r = 0.34; H': r = -0.56; d: r = -0.45). This first axis explains approximately 11%, 32% and 23% of the variability ($R^2$) found in the AMBI, H' and d indices, respectively.

Regarding the second axis, driven by a decrease in salinity and % of fine + medium sand and an increase in finer particles in the sediment (% silt and clay), the same trend was observed as in PC1. All three indices show a decreasing ecological quality (AMBI: $\beta$ = 0.36; H': $\beta$ = -0.17; d: r = -0.21) towards inner stations of both arms of the estuary, and an increase as we get closer to the marine limits of the system. Yet, only AMBI presents a moderately strong relationship to PC2 axis (r = 0.53) with 28% of its variability explained. Both H' and d indices are only weakly, though significantly ($p \leq$ 0.01) related (H': r = -0.25; d: r = -0.34) to the second PC. Above it was mentioned that the third component broadly represents an axis of decreasing salinity. Regressions show that AMBI has a

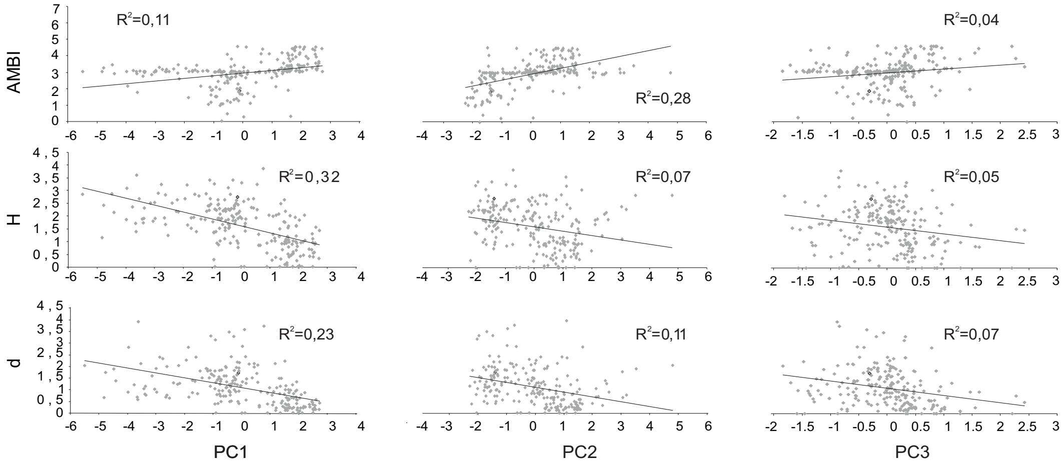



**Figure 6.** Linear regression of the ecological indices AMBI, Margalef (d) and Shannon-Wiener (H’) at the 25 sampling stations, against the first three PC axes scores from the environmental PCA of Figure 5. R-squared values are indicated for each regression analysis ($R^2$).

positive slope while the other two indices have a negative slope. For all indices this means that ecological quality/diversity tends to decrease with salinity, along this axis. Only a small % of the variability in these indices could be explained by PC3 (see Figure 6 for $R^2$ values). These relationships found are relatively weak (AMBI: r = 0.19; H’: r = -0.21; d: r = -0.27) though significant ($p \leq 0.01$).

From the combination of the previous environmental parameters matrix with benthic communities density matrix, through BIOENV procedure, it was found that the best match between the two matrices is achieved with the combination of two variables, explaining 48% of the variability ($\rho$ = 0.483) within macrobenthos. The variables that best group the stations in a manner consistent with faunal patterns are salinity and % of fine + medium sand in the sediment.

Since these environmental factors reasonably captured biotic communities’ patterns, a division of the 25 stations in sectors based on salinity and sediment features is proposed in Figure 7. North and South arms where regarded as distinct subsystems. In each of them, accounting for salinity, grain size and organic matter content of the sediment, zones were established, which reflect gradually different physical environments for benthic communities to settle.

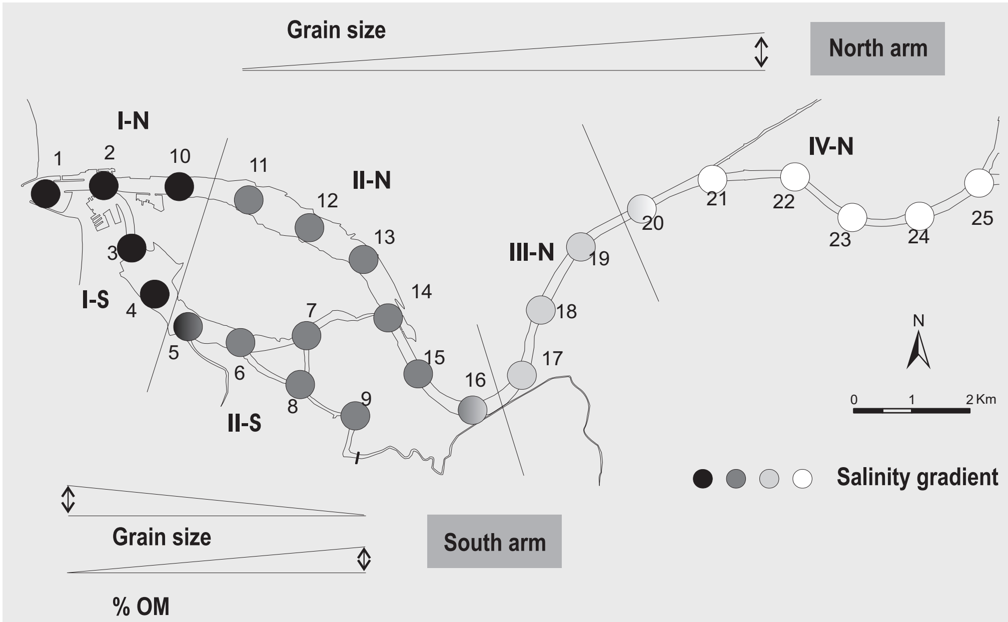



**Figure 7.** Sectors defined in Mondego estuary's North (I - IV) and South (I - II) arms, based upon variation of the six environmental variables along the 25 sampling stations. Decreasing salinity from station 1 to 25 (dark to light circles coloration). Sediment related parameters trends in each arm indicated (grain size and % organic matter - OM).

In the northern arm four sectors were defined: sector I (I-N) with typically euhaline estuarine salinities and sediment mostly composed by fine and medium sand particles; polyhaline sector II (II-N) with sediments composed by fine and medium sand particles but with an increment of the percentage of coarse sand towards inner stations; sector III (III-N) mesohaline with coarse sand dominating sediment composition; and sector IV (IV-N), where sediments are mostly composed by coarse sand. This is a transition area with oligohaline salinities registers and a considerable influence of fresh waters, most upstream stations sometimes register salinity values of 0 and are affected mostly by tidal regime. In the southern arm two sectors where defined: sector I (I-S) with higher marine influence, registering euhaline salinities and with a mixture of fine + medium sand and coarse sand particles in the sediment; and sector II (II-S) with polyhaline salinity and with fine + medium sand particles in the sediment along with an increase of silt and clay contribution towards inner stations. This last sector is the only sector whose organic matter content in the sediments is of importance.

A n-MDS ordination was performed with benthic communities densities', during the whole 2002-2005 period. As expected, it reflected a spatial gradient from the estuary's mouth towards its inner parts, with communities from stations 1 to 25 distributed in a continuum (stress = 0.15). When stations were grouped in sectors (I-N to IV-N and I-S to II-S), the bidimensional representation obtained was that of Figure 8. An ANOSIM analysis, of subtidal benthic communities densities, showed that these 6 groups were significantly different (R = 0.557; $p$ = 0.001).

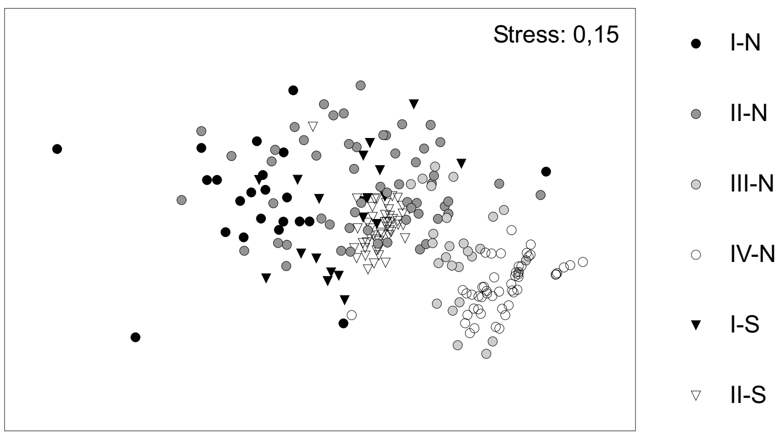



**Figure 8.** n-MDS ordination plot based on benthos densities (ind.m$^{-2}$) observed in the 25 sampling stations in the period of 2002 to 2005 (stress: 0.15). Stations are labelled according to sectors defined in Figure 7.

Pairwise comparisons among those groups indicated that despite significance of their segregation, not all of them were strongly dissimilar from each other (see R values). While some groups were significantly ($p$ = 0.001) more strongly dissimilar (I-N/II-N: R = 0.338; I-N/III-N: R = 0.782; I-N/IV-N: R = 0.918; I-S/II-S: R = 0.618; I-S/III-N: R = 0.607; I-S/IV-N: R = 0.942; II-S/III-N: R = 0.726; II-S/IV-N: R = 0.958; II-N/IV-N: R = 0.693), communities from some of the adjacent sectors, though significantly different, did not show the same degree of segregation (lower R values) (I-N/I-S: R = 0.278; $p$ = 0.001; II-S/II-N: R = 0.225; $p$ = 0.001; II-N/III-N: R = 0.235; $p$ = 0.001; III-N/IV-N: R = 0.248). In Figure 8 it is clear that these communities from adjacent areas follow a gradient and partially overlap. Also, this pairwise analysis of similarity showed no significant differences between benthic communities of sectors I-S and II-N ($p$ = 0.294). In Figure 8 is observed that northern arm II-N community overlaps the range of communities that extend through out the whole southern arm (both I-S and II-S sectors).

Since WFD's reference conditions concept must account for all range of natural variability, for the purpose of this paper all seasons were treated together. Yet, regarding benthic communities' data, group segregations by season was in general more evident. ANOSIM results for the six estuary zones by season were: winter R = 0.613, $p$ = 0.001; spring R = 0.626, $p$ = 0.01; summer R = 0.642, $p$ = 0.01 and autumn R = 0.471, $p$ = 0.01.

## Discussion

Splitting a biological continuum is a difficult task, but to deal with quality assessment and management issues (*e.g.*, in the scope of recent and new coming Directives, such as WFD or European Marine Strategy) defining homogeneous units is essential. These units, referred as water bodies within WFD, must be studied at the whole ecosystem level: water, sediment and biological elements from phytoplankton to fishes (not only benthos).

The WFD methodology for biological elements quality assessment states that one must report to reference conditions. Thus reference values for the parameters to be assessed must be established. An approach to this issue in transitional waters (TW) must account for the strong gradients and constraints that individuals face within estuarine limits and so, in this water category (TW), reference conditions should not be equal through out a system but should accompany natural trends. As already mentioned, this paper objective was to understand and define those trends and propose what might be homogeneous zones for benthos in this estuary (see Borja 2005, for issues relating to research needs within the WFD).

The environmental variables chosen explained satisfactorily the trends within the estuary and also associated biotic assemblage's patterns. Other variables of sum importance for benthos, such as dissolved oxygen, pollutants, nutrients; were not treated though. During the studied period, there were no registers of limiting dissolved oxygen values for the benthic compartment. Also, nutrient and chemical pollutants levels were not considered, because our objective was to account for natural features responsible for benthic distribution and to what extent they were responsible for it. Since, for the definition of reference conditions, natural variability range is what matters, the six parameters treated were selected from the whole data available.

In this way, one of the most important features affecting both physical-chemical variables and biology within an estuary is salinity. Diversity trends in estuaries have long been associated to this factor and several hypotheses suggested.

As a consequence of physiological limits, the benthic macroinvertebrates compartment here studied, due to its low mobility, is directly influenced by salinity (Bulger *et al.* 1993, McLusky & Elliot 2004). But, rather than absolute salinity tolerance, Attrill (2002) found evidence that the major factor influencing the distribution of organisms in estuaries is salinity variation (tidal range). Fluctuating salinity conditions add in fact a level of stress that prevents organisms maximizing their potential distribution. So that species minimum (artenminimum) was registered at points of maximum salinity range. Also, in systems with high seasonal variability in river flow rate and therefore salinity, the upper-middle estuarine fauna may switch each year between oligohaline and mesohaline, resulting in communities that seldom progress beyond early benthic community succession (Santos *et al.* 1996, Ysebaert *et al.* 2005).

So, both tidal and seasonal salinity ranges should be accounted for when distribution patterns of benthic community are evaluated. For this paper only seasonal salinity fluctuations could be studied, yet future research on tidal changes in all extension of the Mondego estuary should be considered.

Salinity trends in the Mondego estuary follow those of most temperate estuaries with significant freshwater contribute during winter. Strong salinity fluctuations between seasons are observed, with significant salinity decreases occurring in rainy periods, which affect mostly the middle-upper estuarine areas (inner southern arm: stations 7-9; and middle-upper northern arm: stations 16-19).

Differences found regarding salinity patterns in the Mondego estuary's north and south arms can be explained namely by the interventions that occurred in this system through the last two decades, altering its hydromorphology. The south arm's hydrological regime is mostly dependent on tidal influence and on small tributary irregular inputs, whose discharges are controlled upon water needs on upstream rice fields and flood related measures. Most of the time it behaves as a coastal lagoon, with a higher residence time than the north arm, and whenever a freshwater discharge occurs, since it is much shallower (4 - 5 m maximum) than northern one, (through the Pranto sluice - Figure 1) its inner polyhaline stations' (7 to 9) bottom salinity drops. Northern arm's depth was increased (until 10 m in downstream areas) and therefore bottom salinity in its downstream zone is less affected by freshwater inputs on the system, even in rainy seasons. This explains the n-MDS ordination (Figure 3), which separates stations 6 to 9, located in the south arm, from the north arm polyhaline stations (11 to 16). The north arm middle stations (17 to 19) are not as deep and freshwater influence increases, so that salinity ranges are higher. Finally, upper estuarine stations range from typically oligohaline to freshwater, in the rainiest seasons, when tidal influence is just perceived on water level variation.

Along with salinity, sediment type is recognized as another very important factor influencing macrobenthic species composition and abundance patterns within salinity zones (Holland *et al.*, 1987). Like salinity, sediment structure also presented a gradient along the estuary, and differences as well were observed between both arms at the lower estuary. Two general patterns can be described for this estuary: one of increasing grain size from the mouth along the northern arm and towards its upstream stations (sediments clearly dominated by coarse sand); and another, inverse, of decreasing grain size from the mouth to southern arm most inner stations. Yet, at the first stations of both arms (2 to 14 in I-N and II-N compared with 3 to 5 in I-S and II-S), closer to the mouth, due to marine influence, differences in sediment composition become less apparent. Sediments are composed by a mixture of fine to coarse sands. Nevertheless, four sectors were defined here (I-S, II-S, I-N and II-N) because different conditions prevail at each one of them. Current velocities are higher along the north arm of the estuary, due to both river discharge and fast tidal penetration, which can explain bottom characteristics. Additionally, the north arm downstream stations' (1 to 11, I-N) bottom is regularly disturbed by dredging activities, leading to direct habitat loss and frequent sediment structure alteration. Therefore evidence of biological communities less structured and characterized by epibenthic fauna were found in these stations (Marques *et al.* 1993, Rodrigues 2004). The south arm, except for sporadic interventions (e.g. construction of a small fishermen dock in 2004, near station 3), is not under such continuous pressure and so the sediment structure is more stable. Moreover, water circulation depends mostly on tides, which favours finer particles and organic matter deposition (McLusky & Elliot 2004), and in fact the south arm stations present considerable organic matter content in the sediment. On previous reports (Marques *et al.* 2004), based on Molvær *et al.* (1997) classification of organic matter content in the sediments, almost all the estuarine area was described as in very good condition, expect for some inner parts of the south arm where organic matter content was considered too high and led to bad classifications. Nevertheless, the existence of finer sandy sediments and mud flats in this arm allows for important infaunal assemblages to settle (Marques *et al.* 1993, Rodrigues 2004). Studies carried out in the Mondego estuary have always emphasized its relative peculiarity with regard to the biological communities (Marques *et al*. 1993, 2003, Cardoso *et al.* 2004a, Chainho *et al.* 2006), pointing namely to clear differences between both arms.

The diversity indices tested, which will be used for quality assessment within WFD monitoring, showed some correlation with the environmental gradients found within the estuary (Figure 6). This shows that benthic communities are changing along the system. Thus the abiotic influence on ecological indices results should be accounted for when defining reference values, for the parameters to be assessed, on the different sectors. Reference conditions must reflect a gradual change, which in estuaries is expected to be a naturally lower species richness and evenness towards their inner parts (McLusky & Elliot 2004). Linear regressions of the indices and PC axes are in agreement with this general trend. All three indices (H', d and AMBI) revealed less diverse communities with increase of coarse sand and also with increase of silt and clay in the sediments. As mentioned in the results, these gradients were identified from downstream (near the mouth) to inner areas in each of the two arms, respectively. Tagged along with these is salinity decrease, most evident in the north arm subsystem where mesohaline and oligohaline characteristics are also encountered.

Salinity and % of fine/medium sand in the sediment, together explain 48% of the variability found within the whole estuarine subtidal benthic community (from the mouth to the upstream limit of the estuary). This result is rather acceptable, taking into account that a 4 year seasonal biotic data set was used and only six variables were selected for attempting to describe its general distribution patterns. Salinity and sediment related parameters are therefore important factors that must be accounted for when sector specific reference conditions are defined within WFD procedures.

In this paper we attempted to account just for natural patterns responsible for general trends regarding diversity variation in the Mondego estuary. Nevertheless, not all differences found among the south arm communities are due to natural habitat diversity. In fact, a number of studies have shown that some of the ecological quality differences observed between downstream and upstream areas in the south arm are mainly due to anthropogenic related pressures that harassed it in the last two decades, for instance nutrient loading (Marques *et al.* 1997, 2003, Pardal *et al.* 2000, Martins *et al.* 2001, Dolbeth *et al.* 2003, Cardoso *et al.* 2004a). From 1998, experimental mitigation measures were applied to test conditions susceptible to reverse eutrophication symptoms in this subsystem (Neto 2004). The main action then were the limited re-establishment of the communication between the two arms, which led to a reduction on the water residence time in the south arm, and a diminution in the freshwater discharge proceeding from the Pranto River sluice, being instead diverted to a channel discharging into the north arm through another sluice located more upstream. The north arm presents in fact a greater water flow and a lower water residence time than the southern one, and therefore the effects of extra nutrient loading are less harmful. Last, as a complementary measure, the existent macrophyte beds were physically protected against stepping on. These mitigation measures implemented in the south arm improved transparency of the water and decreased nutrient loading. In parallel, it allowed some recovery of the *Zostera noltii* meadows, which spread again up to stations 5 and 6 (personal field observation). Moreover, improvements have been observed with regard to the benthic communities along the southern arm of the estuary (Teixeira *et al.* 2007). All this suggests that amelioration may proceed into inner areas. In case the entire south arm undergoes a gradual environmental improvement following the application of extensive mitigation measures (May 2006), we must conclude that diversity might have been constrained not only by salinity/sediment related factors but also by environmental conditions related to human impacts. These two aspects contributed to the distinction found between the south arm communities. On the other hand, benthic communities along the north arm, from the mouth up to mesohaline areas, appear to follow a continuum (Chainho *et al.* 2006), and therefore the strong segregation found between stations regarding salinity factors, also evident in the sediment, may not reflect directly on them.

As expected, and similarly to other estuarine systems, separation of assemblages between salinity zones and sediment was not discrete (Holland *et al.* 1987). These authors also observed that a continuum of slightly different assemblages occurred along salinity and sediment gradients. From our data, despite the weak segregation (low R values), the biotic assemblages grouped are significant (Figure 8), which points towards a change in communities along the north arm gradient. Remind that data from all seasons were used together in our analysis, but if a seasonal displacement occurs within the north arm communities as observed by Chainho *et al.* (2006), this might have contributed to a weaker community distinction between adjacent areas. Our results showed evidence that treating benthic assemblages by season a better segregation between stretches is found.

Thus, the six sectors (Figure 7) that arose from the results are areas where environmental differences encountered reflected in biological communities and so each of them might justify different sets of reference values. Nonetheless, the splitting of these communities by zones will always be, at some extent, artificial. To avoid treating the south arm as two artificial water bodies, Ferreira *et al.* (2006) argued that a division between the estuary's mouth and southern arm should settle mainly on human (pressure and state indicators) criteria instead of on natural ones (morphology and salinity). As showed, according to Bald *et al.* (2005) approach, the downstream section of the southern arm presents euhaline estuarine conditions, with the average salinities at stations 3 and 4, located in this area, reaching the value of 32 during our study period. Also sediment characteristics do not differ much from those at the mouth. Diversity in this highly marine influenced section (I-S) is therefore comparatively stronger than in lower salinity muddy sediment stations located upstream in the south arm (during the study period, average salinity in stations 8 and 9 was 22.5). As a result, for management issues and independently of the water bodies that will be proposed within this estuary, in these two areas of the south arm distinct conditions prevail and, for benthic condition correct assessment, there is a need of splitting this arm in two sections where different reference conditions should rule. By dividing the Mondego estuary into 6 sectors the entire natural benthic diversity within this system should be covered. See that the sectors here proposed with regard to establishing benthic communities' reference conditions are not to be regarded as different water bodies as in the WFD. These sectors are just a tool to guide through benthic condition assessment, meaning, zones where different and adjusted reference conditions should be used when applying European WFD tools (multimetric indices) (Borja *et al.* 2004a, Muxika *et al.* 2007). From the management point of view it would be ineffective to consider these six sectors as requiring distinct management plans, since a water body should be functional in biological and physicochemical terms. And, at this point, from a management perspective, we agree with Ferreira *et al.* (2006) when they argue for the non-splitting of the south arm unit. For this specific estuary, whatever water bodies are considered, a special look is necessary upon the north arm last kilometres (downstream area), where the commercial harbour and a recreational marina operate. Actually, the regular physical disturbance of the bottoms related to these infrastructures' activities are a pressure continuously affecting biotic assemblages.

# Chapter II

## Ecological indices tracking distinct impacts along disturbance-recovery gradients in a temperate NE Atlantic estuary – guidance on reference values

**Abstract** Results gathered from a monitoring programme on the Mondego estuary (Western Coast of Portugal) were compliant with findings drawn from other studies, which pointed out that unstable environments, namely estuarine systems, constitute a great challenge for the use of environmental tools such as ecological indices. The Margalef Index, the Shannon-Wiener Index and *AMBI* were proposed to evaluate the ecological status of benthic communities in the scope of the European Water Framework Directive (WFD) in Portugal and other European member states. In this system these indices were not equally effective in tracking changes in benthic communities that expressed obvious responses to anthropogenic disturbances (eutrophic situations and severe physical disturbance) and to subsequent impacts' cessation. Natural variability played an important role on the indices' response, with estuarine gradient and habitat heterogeneity affecting the ranges of values obtained, and with extreme climate events slowing down the ongoing recovery process. Nonetheless, both natural and anthropogenic variability during the study period could be satisfactorily detected if we accounted for the information provided by all three indices. Based on the whole range of variation observed a) as a response to different kinds of impacts, and b) along recovery gradients, as well as accounting for the natural driving forces acting upon estuarine benthic invertebrate communities, we propose a set of reference values for these three indices. This proposal aims at contributing to the implementation of classification tools within the scope of the European WFD.

## Keywords

Estuary
Benthic invertebrates
Eutrophication
Physical disturbance
Ecological indicators
Reference values

## Introduction

Ecological indicators are meant to illustrate the status of a system based on information from its components. The ecological condition of estuarine systems is rather difficult to assess due to natural stressors that could mask the response of potential ecological indicators. For instance, regarding benthic invertebrate based indicators, natural dominance of species in estuarine communities, as well as strong recruitment events, coupled with longitudinal and vertical gradients along estuaries lead to great variability of indices' behaviour. This variability is not always associated with disturbance, in this way making it difficult to detect such an occurrence, if it was to be present (Mackey & Currie 2001, Dauvin 2007, Elliot & Quintino 2007). The European Water Framework Directive (WFD, EC 2000) establishes an outline for the protection of all waters (including inland surface waters, transitional waters, coastal waters, and groundwater), aiming at achieving a good quality status for all waters by 2015. The concept of ecological status developed in the WFD is defined in terms of the biological community quality, as well as the systems' hydrological and chemical characteristics. Applying the WFD, it requires methods capable of distinguishing between different levels of ecological quality to classify surface waters (Borja *et al.* 2004a, Borja 2005).

Experience demonstrates that none of the available measures of disturbance effects may be considered ideal. But the combination of different measures results as a good toolset for determining the ecological quality status (Bettencourt *et al.* 2004, Borja *et al.* 2004a). In this sense, the results of the TICOR project (Typology and Reference Conditions for Portuguese Transitional and Coastal Waters) (Bettencourt *et al.* 2004) include a method that combines a suite of indices. Recently, several approaches have been considering the possibility of combining different existent metrics into general indices of ecological quality: the M-AMBI in Spain (Borja *et al.* 2004a, Muxika *et al.* 2007); the IQI in the United Kingdom and Republic of Ireland; the DKI in Denmark; the NQI in Norway and the BQI in Sweden (Jowett 2006). Following this concept of multi-metric tools, three indices - Margalef Index, Shannon-Wiener Index, and *AMBI* - were selected to be included in the WFD methodology to assess the benthic compartment Ecological Status in Portuguese Coastal and Transitional systems (Bettencourt *et al.* 2004).

A good example of Portuguese transitional systems is the Mondego Estuary, which has been monitored for almost two decades with relative regularity. The combination of highly variable freshwater discharge and the mesotidal regime that characterises this estuary is common to most of these systems in Portugal, representing approximately 93% of the total area of transitional water systems. Moreover, from the beginning of the 1990s up until now, the Mondego estuary suffered several interventions that caused deterioration of the ecological condition, followed by a gradual recovery of the system as a function of the application of mitigation measures (e.g., Marques *et al.* 1993, 2003, Cardoso *et al.* 2004a, Lillebø *et al.* 2005, 2007, Dolbeth *et al.* 2007). Studies carried out in the system during this period provided a reliable long term data set on the benthic communities, allowing the evaluation of their response to changes in prevailing environmental conditions.

The objectives of this study are: (1) to assess if the indices included in the WFD Portuguese multi-metric methodology are capable of detecting the disturbances and events affecting the subtidal benthic communities; (2) to establish approximate reference conditions for the above mentioned indices, contributing to the implementation of the WFD; and (3) to adjust, taking the Mondego Estuary as an example, those reference values reflecting the natural variability and gradients that can be found in transitional ecosystems such as estuaries.

## Methodology

### *Study site*

The Mondego Estuary (860 ha) is an Atlantic warm-temperate, polyhaline, intertidal system located on the north-western coast of Portugal. In its last extension it divides in two arms (Figure 1): a deeper northern arm and a shallower southern arm, the last characterised by large areas of intertidal mudflats (almost 75% of the area).

The system receives agricultural runoff from 15000 ha of upstream cultivated land (mainly rice and corn fields) and supports a substantial population, industrial activities, salt-works and aquaculture farms, and is also the location of the Figueira da Foz city harbour. All these activities contribute to the nutrient loads entering the Mondego Estuary.

In the first half of the 1990s, the northern channel (location of the stations N1 to N4 in Figure 1) was sunk and the width of the river bed narrowed by about 75%, to increase water flow and consequently the capacity to transport sediment beyond the system, preventing the silting that normally occurred in these parts of the river. From the mid 1990s to 2006, the main pressures harassing Mondego's northern arm resulted mainly from harbour facilities and consequent dredging activities, causing physical disturbance of the bottoms.

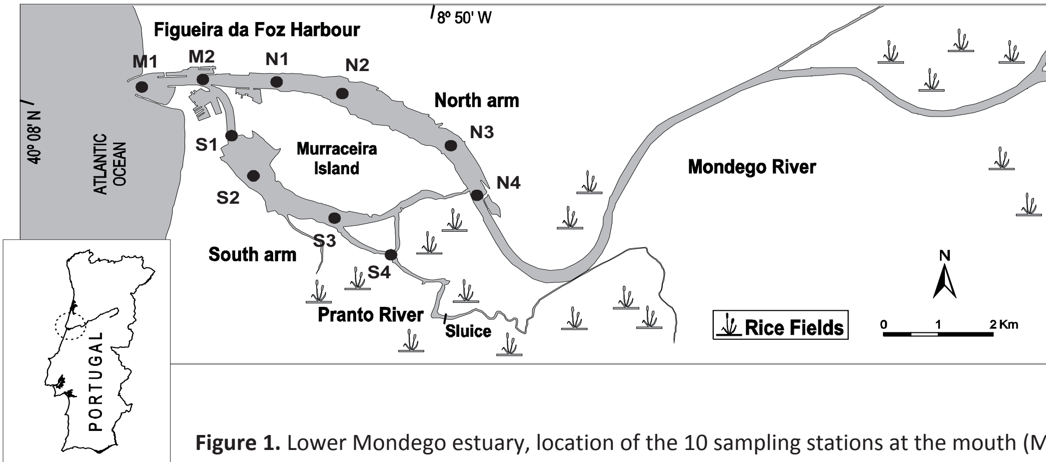



**Figure 1.** Lower Mondego estuary, location of the 10 sampling stations at the mouth (M1 and M2) and on both the northern (N1 to N4) and southern (S1 to S4) arms of the estuary.

In 1990-1992, the communication between the two arms of the estuary was totally interrupted in the upstream area, causing the river discharge to flow essentially through the northern arm. The water circulation in the south arm became dependent on the tides and on the small freshwater input from a tributary, the Pranto River, artificially controlled by a sluice (Martins *et al.* 2001, Marques *et al.* 2003, Lillebø *et al.* 2005). The combined effect of an increased water residence time and of nutrient concentration became a major driving force behind the occurrence of seasonal blooms of *Ulva* spp and a concomitant severe reduction of the area occupied by *Zostera noltii* beds, previously the richest habitat in terms of productivity and biodiversity (Marques *et al.* 1993, 1997, 2003). The shift in benthic primary producers affected the structure and functioning of the biological communities, inducing the emergence of a new selected trophic structure, which has been analysed in abundant literature (e.g., Marques *et al.* 1997, 2003, Cabral *et al.* 1999, Dolbeth *et al.* 2003, Cardoso *et al.* 2004 a,b, Lopes *et al.* 2005, Patrício *et al.* 2004, Patrício & Marques 2006, Martins *et al.* 2007).

From 1998 to 2006 several mitigation measures were carried out to ameliorate the system's condition, namely by improving water circulation through an experimental re-establishment of the upstream connection between the two arms of the estuary and the reduction of the discharges proceeding from the Pranto River sluice. After this intervention, it was observed a partial recovery of the area occupied by *Zostera noltii* and the cessation of green *Ulva* spp. blooms (Lillebo *et al.* 2005, 2007, Verdelhos *et al.* 2005), as well as a recovery of the intertidal macroinvertebrate assemblages (Dolbeth *et al.* 2007).

Regarding the interventions on the estuary above described, we considered three periods, all covered by the available data set: a) 'Period I', the one of greater environmental pressure, between 1990 and 1997, during which several physical interventions took place leading to the total interruption of the upstream communication between the northern and southern arms of the estuary; b) 'Period II', from 1998 to 2002, shortly after the experimental re-establishment of the connection between the two arms; c) 'Period III', from 2003 to 2006, corresponding to a more clear response of the system to mitigation measures, during which pressures and impacts on both arms of the estuary became minimized.

### *Data series*

#### *Biological data*

Subtidal soft-bottom macrobenthic assemblages have been surveyed at ten sampling stations (Figure 1). The stations were situated along the estuarine gradient from the mouth (M1/M2) towards the inner areas, so that in both arms each group of 2 stations represents distinct confinement and salinity conditions. Therefore, N1/N2 and S1/S2 pairs reflect a higher marine influence, with euhaline estuarine/polyhaline salinities; whereas inner estuarine pairs N3/N4 and S3/S4 present lower polyhaline salinities (Teixeira *et al.* 2008a).

To characterize the three periods referred to above, we considered a total of 48 macro-benthic samples collected during the spring at 8 stations from both arms and taking into account six different years, respectively 1990 and 1992 for Period I, 2000 and 2002 for Period II, and 2005 and 2006 for Period III.

The 2 stations close to the mouth (M1 and M2) (Figure 1) were sampled only from 1998 onwards. Nevertheless, in order to account for the entire range of natural variability along the estuarine gradient, which may affect the indices performance, data from these 2 stations were also considered.

Samples were collected using Van Veen grabs, gathering five replicates each time at each station, corresponding to an area of 2340 to 2480 $cm^2$, depending on the grab utilised. Samples were sieved through a 1mm sized mesh and preserved in a 4% buffered formalin solution. Individuals were counted and identified at species level whenever possible. Taxa in the benthic data matrix are all at species or genus level except for the Chironomidae, Nemertea and Oligochaeta individuals. All macrobenthic abundance data were transformed to number of individuals per square meter (indiv. $m^{-2}$).

*Environmental data*

To be associated with the macrobenthos dataset, at each sampling occasion the following physical-chemical parameters of bottom water were measured *in situ*: salinity (Practical Salinity Scale), water temperature (ºC), dissolved oxygen (mg $l^{-1}$) and pH. Additionally, bottom water samples were collected for laboratorial determination of dissolved nutrient concentration (mg $l^{-1}$) (Nitrate-Nitrogen, Nitrite-Nitrogen and Phosphates-Phosphorus), and sediment samples were taken to determine grain size and organic matter content. Nutrient concentration assessment was performed in the laboratory according to APHA (1995) and Strickland & Parsons (1972) standard procedures. Sediment organic matter content was quantified by weight difference between sediment after oven drying at 60 ºC for 72 h and after combustion at 450 ºC for 8 h, and expressed as a percentage of the total sample weight. Grain size analysis was carried out by mechanical separation through a column of sieves with different mesh sizes. Sediments were classified as coarse sand (≥ 0.5mm), fine + medium sand (> 0.063 and < 0.5mm), silt (> 0.038mm and < 0.063mm) and clay (≤ 0.038mm) (scale adapted from Brown & McLachland 1990). Grain composition was expressed in percentage of total sample weight.

*Data analysis*

The values of the three ecological indices, Margalef Index (*d*) (Margalef 1968), Shannon-Wiener Index (*H'*) (Shannon & Wiener 1963) and *AMBI* (Borja *et al.* 2000), were calculated from the benthic data matrix, using the following algorithms:

(1) $d = (S\text{-}1)/\log_e N$;

(2) $H' = -\sum p_i \log_2 p_i$;

(3) $AMBI = [(0)(\%GI)+(1,5)(\%GII)+(3)(\%GIII)+(4,5)(\%GIV)+(6)(\%GV)]/100$;

where $S$ is the number of species, $N$ is the total number of individuals, $p_i$ is the proportion of abundance of species, and %GI to GV is the % of individuals belonging to each of the five ecological groups considered in the construction of the *AMBI* index (Borja *et al.* 2000). The AMBI 4.0 software, freely available at http://www.azti.es, was used to calculate this index, using a species list of July 2006 and following Borja and Muxika (2005) guidelines.

The Margalef Index has no upper limit, which makes it more difficult to predict the thresholds that separate disturbed from undisturbed communities. Regarding the Shannon-Wiener Index, it usually takes values from 0 to 5 bits/individual, values above 5 being very rare (Salas *et al.*, 2006). *AMBI* index values ranges from 0 to 7 (Borja *et al.* 2000). In terms of ecological interpretation, an increase of both the Margalef and the Shannon-Wiener indices is generally accepted as an indication of an ecological quality improvement, while for *AMBI* a better quality would be indicated by a decrease on the index value.

A Permutational Multivariate Analysis of Variance (PERMANOVA) (Anderson 2001, McArdle & Anderson 2001) was applied to the results of each index to determine whether they detected significant ecological differences regarding ecological condition of benthic assemblages through the main gradients studied: estuarine spatial gradient and temporal disturbance-recovery gradient. The experimental design consisted of 3 fixed factors: arm (northern arm *vs.* southern arm), confinement (inner stations *vs.* downstream stations) and period (3 periods of distinct pressures); and the random factor year nested within the period (2 levels), with $n$ = 2 stations per confinement $\times$ arm $\times$ year. All analyses were performed with PERMANOVA software (Anderson 2005). No transformation or data standardization was operated before analysis; Bray-Curtis dissimilarity was the chosen distance measure to perform the PERMANOVA, and distances were maintained (i.e. not replaced by their ranks before proceeding with the analysis). For the tests, at $\alpha$-level of 0.05 or 0.1, 999 permutations were used.

After evaluating ecological indices variability with PERMANOVA, the variation in the distribution of benthic community assemblages was also observed with the non-metric Multidimensional Scaling (nMDS) ordination model of the Bray-Curtis similarity matrix (Clarke & Warwick 2001). An Analysis of Similarity (ANOSIM) was carried through to search for differences in species' assemblages (using the species abundance Bray-Curtis similarity matrix) comparing *a priori* defined groups (as in the previous PERMANOVA experimental design). Initially, data from both arms were treated together (one-way ANOSIM for factors arm, period and confinement/salinity regime); then the arms were separated to make species distribution patterns clearer.

Finally, to assess species data variation and the relationship between species distribution and the environmental conditions during the study period, multivariate ordination techniques were used. Because gradient length in standard deviation units (SD) was higher than 3, as revealed by performing a Detrended Correspondence Analysis (DCA) (detrending by segments) on species data; subsequent analysis was based on the existence of an underlying unimodal species-response model, by using Canonical Correspondence Analysis (CCA). Physical-chemical parameters were included in the analysis as an environmental data matrix.

To evaluate ecological indices' trends against those of environmental parameters within available data, a Principal Components Analysis (PCA) of the values of the three indices' results and the number of species was performed, with the environmental parameters as supplementary variables. Since some environmental information was missing for the year 1992 in the south arm (stations S1, S2, S3 and S4), these four samples were removed from the analyses. Ordinations were carried out using the CANOCO 4.5 for Windows (ter Braak & Smilauer 1998).

The data set used in all previous analyses excluded the stations at the mouth of the estuary.

The response of the indices to the gradients studied (spatial and recovery) guided through the selection of values for each lower estuarine zone, under the best ecological scenario observed. The highest values found for each index were then increased by approximately 15-20% to set approximate reference conditions for these three indices for this type of system.

## Results

Springtime surveys at ten stations during eight sampling years in the Mondego estuary (Figure 1) resulted in the identification of a total of 119 taxa. Of the species identified, 49 of them occurred just once within the 64 samples studied; and only four taxa were present in more than a half of the samples collected (*Crangon crangon, Hydrobia ulvae, Scrobicularia plana* and *Streblospio shrubsolii*). The number of species in a sample ranged from 1 to 26, at stations S2 (1998) and M2 (2006) respectively. The four most abundant taxa alone accounted for more than 60% of the total abundance during the study period (*Alkmaria romijni, Hydrobia ulvae, Scrobicularia plana* and *Streblospio shrubsolii*). The number of species and density observed at the main estuarine areas during the study period is plotted in Figure 2.

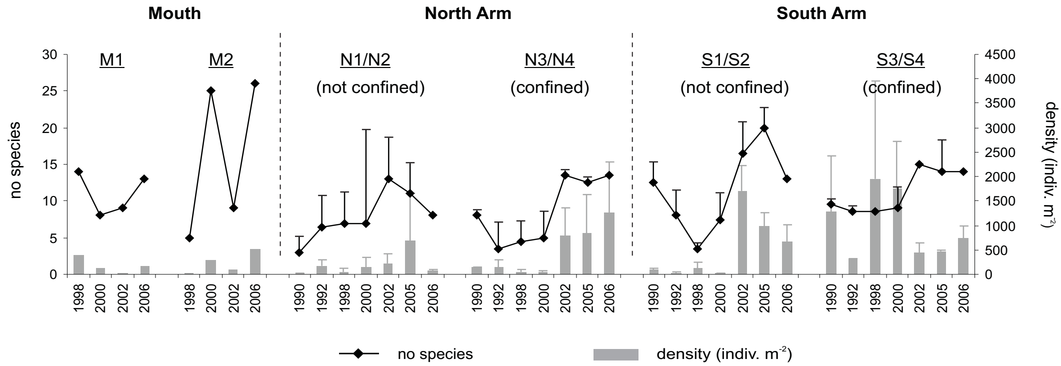



**Figure 2.** Number of species and density (indiv. m$^{-2}$) distribution on stations at the mouth (stations M1 and M2) and at both arms of the estuary during the study period. For stations in each confinement level at both arms, medium values and standard deviation are presented.

Except for the highly variable mouth area, from 2002 onwards, the number of species in all estuarine zones was higher than in previous years. The abundance patterns were different, with no great variations observed at the mouth or downstream stations of the north arm during the study period. At the inner stations of the north arm and outer stations of the south arm, though, an increase was observed from 2002 onwards. In contrast, at the inner south arm stations, which registered the highest abundances of individuals at this estuary during the first years studied, an overall decrease of abundance was observed since 2002.

Generally, environmental registers varied as expected of an estuary (Figure 3). Along the study period, though, occasional changes occurred in some parameters. Bottom water salinity decreased from the mouth towards inner estuarine stations at both arms, with the south arm inner stations (S3 and S4) presenting the lowest values of salinity and also pH. In Period II (2000/2002), both these parameters reached their lowest values during the entire study period. Dissolved oxygen values did not differ much between both arms; yet, a decreasing trend towards the south arm inner stations was also observed. Its lowest values during the studied period were, however, registered at the north arm stations during Period I (1990/1992), but still within the limits above which most biologically requirements are satisfied according to Best *et al.* (2007): 5.7 to 7 mg l$^{-1}$, respectively from marine to freshwater limit. Although through time, no clear patterns were observed regarding nutrients data, the phosphate registers of 1992 were unusually high in the entire estuary, and the nitrite values of 2000 were also considerably higher than in other sampled years. The content of organic matter in the sediment was negligible, except for stations S3 and S4 and N4 (in 2002) and N2 (in 1992). These were the stations with finer particles in their sediment composition.

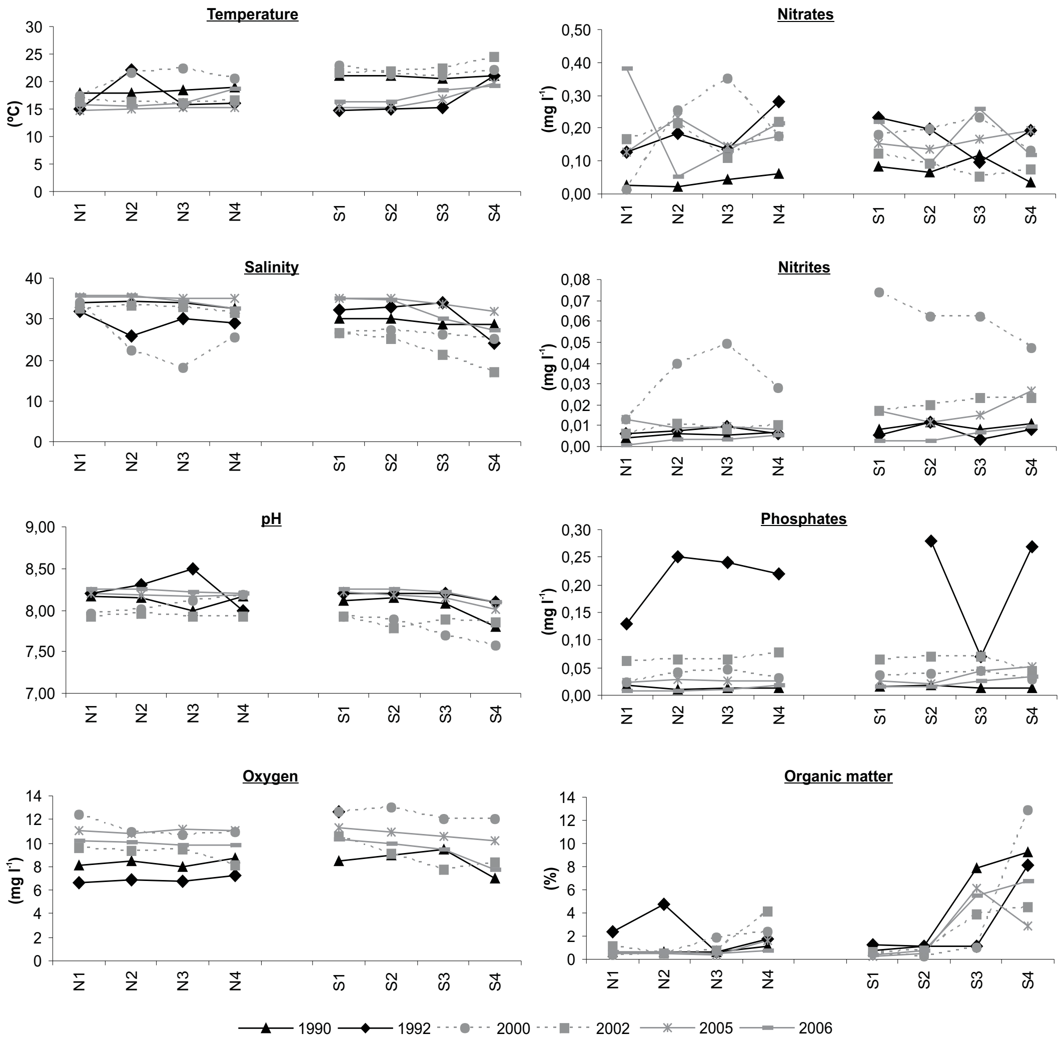



**Figure 3.** Variation of environmental parameters at sampling stations of both arms, during the study period.

*Indices performance*

A wide range of values was found for the ecological indices applied along the study period, which in terms of ecological evaluation covered different levels of quality status (Figure 4). The Margalef Index ranged from 0 (registered at S1 in 1998) to 4.16 (registered at M2 in 2006) and the Shannon-Wiener Index ranged from 0 (registered also at S1 in 1998) to 3.45 bits/individual (at M2 in 2000). The AMBI index's best score was 0.98 and was registered at the mouth of the estuary (M1 in 2002). The worst value for the index was of 3.84 at N1 in 2000.

The analysis of variance regarding ecological indices revealed that some significant differences could be found along the two main gradients proposed for study: a) the estuarine spatial gradient, here accounting furthermore for the singularity of the two arms of the Mondego estuary, and b) the temporal changes, regarding periods of disturbance at both arms followed by a recovery and amelioration of the main pressures acting in each arm.

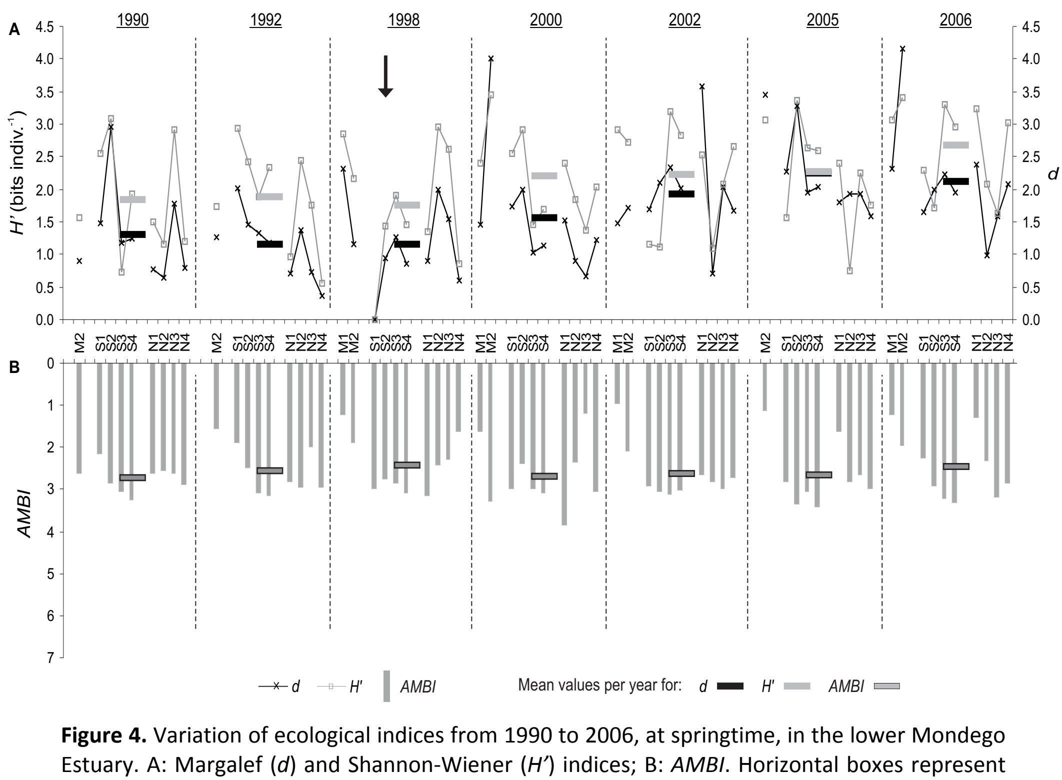



**Figure 4.** Variation of ecological indices from 1990 to 2006, at springtime, in the lower Mondego Estuary. A: Margalef (*d*) and Shannon-Wiener (*H'*) indices; B: *AMBI*. Horizontal boxes represent mean values for each index in each sampling year at all the stations. The arrow in 1998 indicates the moment in time when the southern arm mitigation measures were applied.

For both the Margalef Index and *AMBI*, from 1990 to 2006, significant interactions were observed between factors arm and period (Table I and II). Pair-wise *a posteriori* comparisons showed that significant differences between arms occurred in Period I regarding the Margalef Index, with higher mean values at the south arm ($d_{NA}$ = 0.89, $d_{SA}$ = 1,60); and in Period III regarding *AMBI*, when the south arm presented higher mean values ($AMBI_{NA}$ = 2.47, $AMBI_{SA}$ = 3.04). If we consider the stations' location within the estuary, focusing on the factor level of confinement/salinity regime, significant differences were also found with regard to *AMBI* (Table II), where stations located in inner areas of both arms presented higher mean values than those downstream, closer to the mouth of the estuary ($AMBI_{\text{outer stations}}$ = 2.62, $AMBI_{\text{inner stations}}$ = 2.91).

**Table I.** PERMANOVA on Bray-Curtis dissimilarities for Margalef Index results at 48 subtidal sampling stations distributed along two confinement/salinity conditions (Co) at each of the two arms of the Mondego estuary (Ar) in six years (Year) over three distinct periods (Pe). When the number of possible permutations was small, Monte-Carlo *P*-values were followed.

| **Source** | ***d.f.*** | ***SS*** | ***MS*** | ***F*** |
|---|---|---|---|---|
| Ar | 1 | 2987.77 | 2987.77 | 43.07* |
| Co | 1 | 315.95 | 315.95 | 0.59 |
| Pe | 2 | 4597.17 | 2298.59 | 3.87 |
| Year(Pe) | 3 | 1780.87 | 593.62 | 2.05 |
| ArxCo | 1 | 455.62 | 455.62 | 1.29 |
| ArxPe | 2 | 1005.57 | 502.79 | 7.25* |
| ArxYe(Pe) | 3 | 208.09 | 69.37 | 0.24 |
| CoxPe | 2 | 194.15 | 97.077 | 0.18 |
| CoxYe(Pe) | 3 | 1605.01 | 535.00 | 1.85 |
| ArxCoxPe | 2 | 293.45 | 146.73 | 0.41 |
| ArxCoxYe(Pe) | 3 | 1060.31 | 353.43 | 1.22 |
| Residual | 24 | 6953.87 | 289.74 | |
| Total | 47 | 21457.87 | | |

**Pair-wise *post-hoc* comparisons:**

| | Period I | Period II | Period III |
|---|---|---|---|
| North *vs* South | 5.23* | 0.89 | 1.89 |

| | North Arm | South Arm |
|---|---|---|
| Period I vs Period II | 1.63 | 0.65 |
| Period I vs Period III | 6.27* | 3.29** |
| Period II vs Period III | 1.19 | 1.10 |

* $P \leq 0.05$; ** $P \leq 0.1$. Pair-wise *a posteriori* tests among arms along periods; and among periods within each arm, using the *t*-statistic.

With the Shannon-Wiener Index, some interaction was also found between the confinement and the year (nested in factor period) factors (Table III). Pair-wise *a posteriori* comparisons indicated significant differences between the inner and the outer stations (factor confinement/ salinity regime) of the estuary in the years 2000 and 2002, with the outer stations presenting higher mean values in 2000 ($H'_{\text{outer sations}}$ = 2.43, $H'_{\text{inner stations}}$ = 1.64) and the inner stations in 2002 ($H'_{\text{outer sations}}$ = 1.48, $H'_{\text{inner stations}}$ = 2.69).

**Table II.** PERMANOVA on Bray-Curtis dissimilarities for *AMBI* results at 48 subtidal sampling stations distributed along two confinement/salinity conditions (Co) at each of the two arms of the Mondego estuary (Ar) in six years (Year) over three distinct periods (Pe). When the number of possible permutations was small, Monte-Carlo *P*-values were followed.

| **Source** | ***d.f.*** | ***SS*** | ***MS*** | ***F*** |
|---|---|---|---|---|
| Ar | 1 | 422.91 | 422.91 | 19.77* |
| Co | 1 | 369.84 | 369.84 | 5.19** |
| Pe | 2 | 44.33 | 22.17 | 0.37 |
| Year(Pe) | 3 | 179.93 | 59.98 | 0.55 |
| ArxCo | 1 | 105.65 | 105.65 | 1.23 |
| ArxPe | 2 | 244.53 | 122.26 | 5.71** |
| ArxYe(Pe) | 3 | 64.18 | 21.39 | 0.19 |
| CoxPe | 2 | 511.49 | 255.75 | 3.59 |
| CoxYe(Pe) | 3 | 213.79 | 71.26 | 0.65 |
| ArxCoxPe | 2 | 536.97 | 268.49 | 3.14 |
| ArxCoxYe(Pe) | 3 | 256.15 | 85.38 | 0.78 |
| Residual | 24 | 2620.09 | 109.17 | |
| Total | 47 | 5569.90 | | |

**Pair-wise *post-hoc* comparisons:**

| | Period I | Period II | Period III |
|---|---|---|---|
| North *vs* South | 0.55 | 1.37 | 4.07* |

| | North Arm | South Arm |
|---|---|---|
| Period I vs Period II | 0.65 | 1.93 |
| Period I vs Period III | 3.05** | 2.07 |
| Period II vs Period III | 0.91 | 0.57 |

* $P \leq 0.05$; ** $P \leq 0.1$. Pair-wise *a posteriori* tests among arms along periods; and among periods within each arm, using the *t*-statistic.

Regarding indices behaviour along the three studied periods within each arm, both the Margalef Index and AMBI presented significant differences between Period I and Period III in the northern arm, with mean values indicating a higher quality in Period III ($d_{PeI}$ = 0.89, $d_{PeIII}$ = 1.78; $AMBI_{PeI}$ = 2.68, $AMBI_{PeIII}$ = 2.47). For the south arm only the Margalef Index presented significant differences between these two periods, with higher mean values also during Period III ($d_{PeI}$ = 1.60, $d_{PeIII}$ = 2.17). In addition, significant differences in the Shannon-Wiener Index values were detected along the study period (Table III), namely between Period I (1990 and 1992) and Period II (2000 and 2002), for the whole estuary, with higher mean values during Period II comparatively to Period I ($H'_{PeI}$ = 1.90, $H'_{PeII}$ = 2.06).

**Table III.** PERMANOVA on Bray-Curtis dissimilarities for Shannon-Wiener Index results at 48 subtidal sampling stations distributed along two confinement/salinity conditions (Co) at each of the two arms of the Mondego estuary (Ar) in six years (Year) over three distinct periods (Pe). When the number of possible permutations was small, Monte-Carlo *P*-values were followed.

| **Source** | ***d.f.*** | ***SS*** | ***MS*** | ***F*** |
|---|---|---|---|---|
| Ar | 1 | 1015.73 | 1015.73 | 3.81 |
| Co | 1 | 20.75 | 20.75 | 0.02 |
| Pe | 2 | 1070.47 | 535.24 | 4.39** |
| Year(Pe) | 3 | 365.99 | 121.99 | 0.34 |
| ArxCo | 1 | 30.03 | 30.03 | 0.05 |
| ArxPe | 2 | 658.35 | 329.17 | 1.23 |
| ArxYe(Pe) | 3 | 800.49 | 266.83 | 0.74 |
| CoxPe | 2 | 1023.33 | 511.67 | 0.59 |
| CoxYe(Pe) | 3 | 2581.31 | 860.43 | 2.40** |
| ArxCoxPe | 2 | 883.30 | 441.65 | 0.77 |
| ArxCoxYe(Pe) | 3 | 1708.99 | 569.66 | 1.59 |
| Residual | 24 | 8599.21 | 358.30 | |
| Total | 47 | 18757.95 | | |

**Pair-wise *post-hoc* comparisons:**

| | |
|---|---|
| Period I vs Period II | 2.47** |
| Period I vs Period III | 2.57 |
| Period II vs Period III | 1.40 |

| | 1990 | 1992 | 2000 | 2002 | 2005 | 2006 |
|---|---|---|---|---|---|---|
| Non-confined/higher salinity vs confined/lower salinity | 0.62 | 0.71 | 2.92* | 2.87* | 0.85 | 0.72 |

| | PeI/Conf1 | PeI/Conf2 | PeII/Conf1 | PeII/Conf2 | PeIII/Conf1 | PeIII/Conf2 |
|---|---|---|---|---|---|---|
| Year 1 vs Year 2 | 0.22 | 0.13 | 2.43** | 3.81* | 0.77 | 0.79 |

* $P \leq 0.05$; ** $P \leq 0.1$. Pair-wise *a posteriori* tests among periods; among confinement/salinity conditions within each year; and among years of the same period in each confinement/salinity condition, using the *t*-statistic.

*Spatial and temporal variations in the macrobenthic communities' structure*

Benthic communities showed a weak segregation between assemblages considering both space (the whole estuary), and time (all the different periods analysed together). Nevertheless, significant differences between the benthic assemblages could be found regarding those factors also significantly influencing ecological indices, namely arm, period and confinement. The low strength of the $R$ values confirms the evidence of assemblages' weak segregation regarding those factors (ANOSIM, factor arm: $R = 0.119$, $p = 0.04$; factor confinement: $R = 0.283$, $p = 0.01$; factor period: $R = 0.192$, $p = 0.01$, where pair-wise comparisons indicated both Period I and Period II as significantly different from Period III: $R = 0.298$, $p = 0.01$; $R = 0.279$, $p = 0.01$, respectively).

However, when the northern and the southern arms were observed separately, a clearer distribution of the communities arose, both in space and time (Figure 5A and B). Results gave evidence that the factor confinement exerts a much stronger influence in the south arm benthic communities (ANOSIM, south arm: $R_{\text{conf}} = 0.553$, $p = 0.01$), while those from the north arm are barely distinguishable regarding this same factor (ANOSIM, north arm: $R_{\text{conf}} = 0.182$, $p = 0.03$).

On the other hand, northern arm communities presented clearer distribution patterns along the time factor than southern arm ones (ANOSIM, north arm: $R_{\text{Period}} = 0.29$, $p = 0.01$, where pair-wise comparisons indicated both Period I and Period II as significantly different from Period III: $R = 0.434$, $p = 0.01$; $R = 0.398$, $p = 0.02$, respectively; south arm: $R_{\text{Period}} = 0.195$, $p = 0.04$, where pair-wise comparisons indicated Period I significantly different from Period III: $R = 0.366$, $p = 0.04$). To understand if the confinement factor was masking stronger patterns within the southern arms assemblages, a separate one-way ANOSIM was applied to the inner and the outer stations to see whether each community exhibited different patterns during the studied period. In fact, the southern arm communities from less confined stations and with higher salinity values, presented significant differences along the three studied periods (ANOSIM $R_{\text{Period}} = 0.509$, $p = 0.03$), while communities from the inner stations did not differ significantly along that time (ANOSIM $R_{\text{Period}} = 0.375$, $p = 0.11$) (Figure 5B).

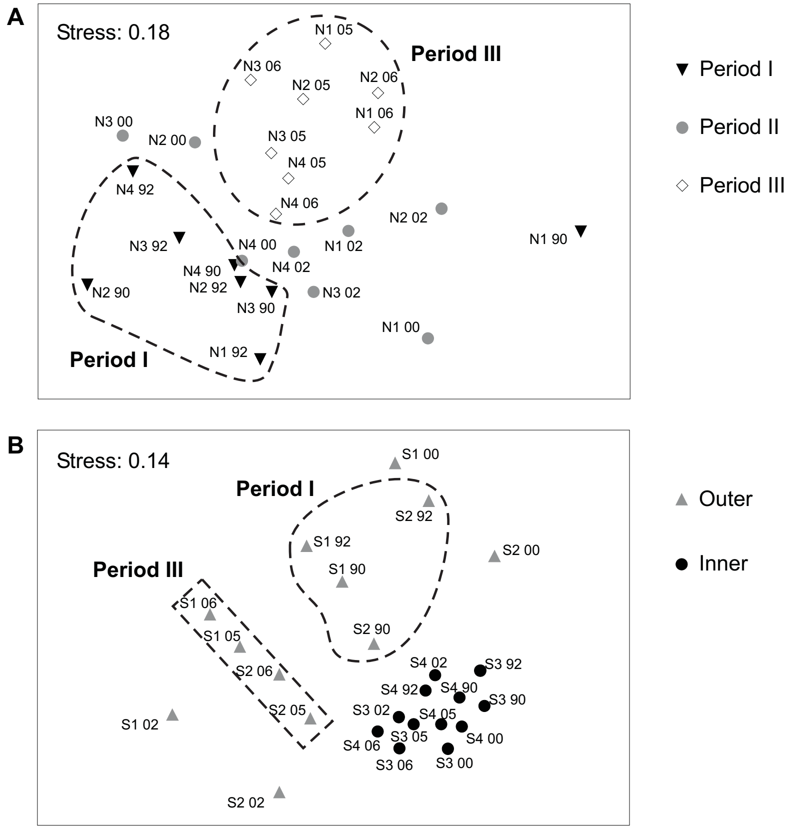



**Figure 5.** nMDS plots of benthic density distribution on the northern (A) and southern (B) arms of the Mondego estuary. The most determinant factors within each assemblage distribution (period in A and confinement in B) were plotted together with a sample label.

### *Physicochemical environmental variables as structuring factors*

The first two axes of the CCA performed on the species density data accounted for 41% of the total variance and 58% of variance due to species-environment relations. The ordination diagram (Figure 6) shows the main pattern of variation in the benthic community representing the sampling stations in relation to environmental variables. The projection along the first axis essentially reflected a gradient of increasing grain size, increasing phosphate and decreasing nitrate concentrations. Samples in the left edge of the diagram appear thus positively correlated to the nutrient enrichment of the system, with almost all benthic assemblages during the three study periods in the estuary consistently associated with increasing nitrate concentrations. On the right bottom of the diagram though, a group of stations from Period II, mainly from the year 2002, from the north arm and the downstream south arm, appear more associated with coarser sediments and phosphate increase.

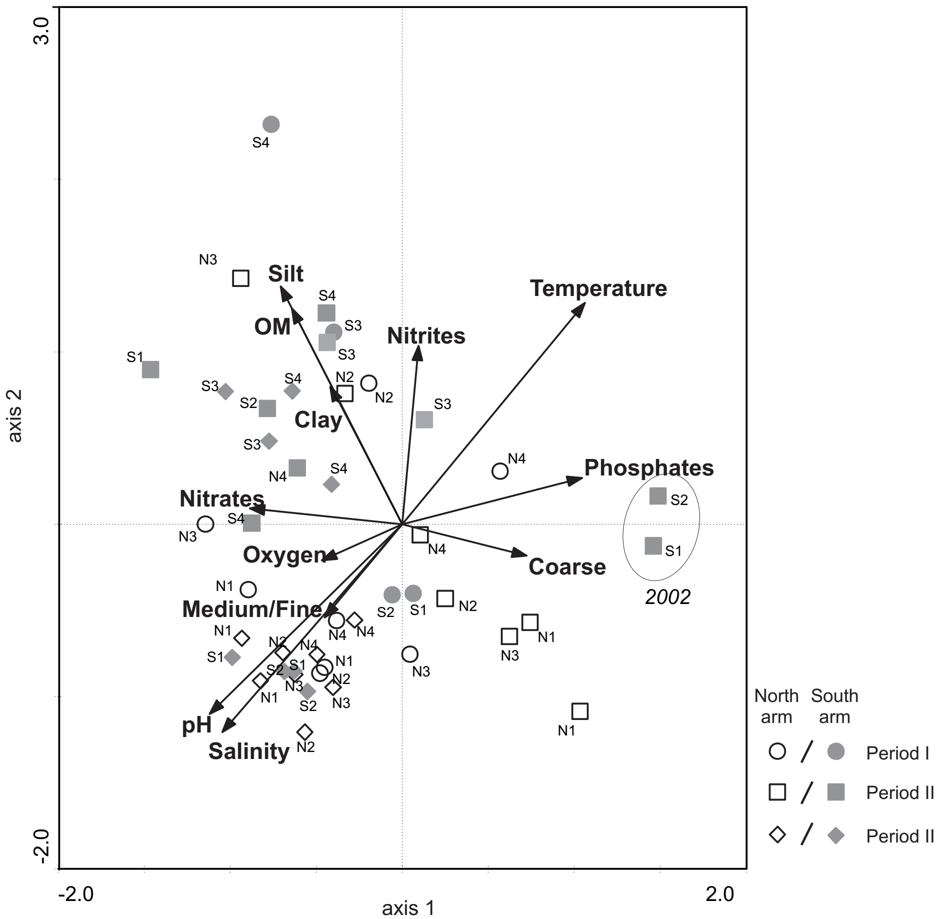



**Figure 6.** Ordination diagram for the first two canonical axes of the Correspondence Analysis on benthic densities and environmental parameters matrices (station labels after Fig. 1; Medium/Fine – medium / fine sand; Coarse – coarse sand; OM – organic matter in sediment).

The second axis of the CCA mainly expresses spatial variation trends related to confinement conditions along the estuary, separating inner south arm areas strongly associated to mud content and organic matter in sediment, from outer areas of both arms plotted along increasing salinity and pH gradients and strongly correlated with medium/fine sediment (Figure 6). Regarding the Period III sampling, communities from all stations, except those at the inner south arm stations, were clearly aggregated pointing to similar species composition.

Regarding the ecological indices and the environmental variables, the first two axes of the PCA (Figure 7) explained 87% of the variability found among the indices. However, the supplementary environmental variables could only explain 27% of the total variability found among the three indices, of which 84% in shown in the first two axes of the diagram. *AMBI* index is apart from the other two and more correlated to organic enrichment, increasing with muddy sediments (silt and clay) and organic matter content in the sediment. Additionally, negative correlation of this

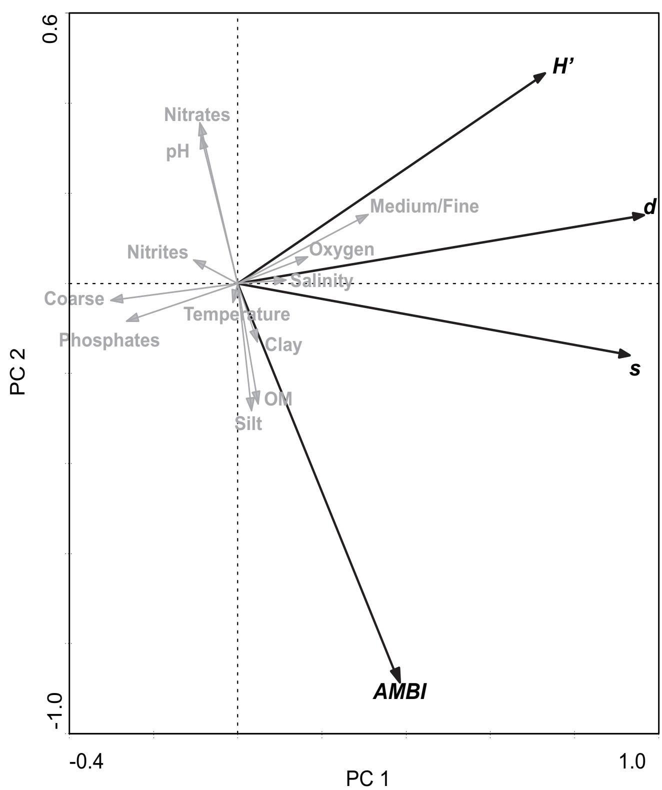



**Figure 7.** Ordination diagram for the first two axes of the Principal Component Analysis of ecological indices results, plotting environmental parameters as supplementary variables.

index with pH indicates decreasing values towards marine estuarine limits. The number of species and both the Margalef and the Shannon-Wiener indices respond to the degree of confinement expressed by the salinity gradient, with decreasing values towards the inner estuary. Likewise, the Margalef and Shannon-Wiener indices presented increasing values along with the increment of medium/fine sand instead of a coarse sand dominance in the sediments. Regarding trophic parameters (water column dissolved nutrients), the Margalef and Shannon-Wiener indices showed an opposite trend to an increase of phosphate or nitrite, while the *AMBI* trend was of increasing quality with a nitrate increase.

### *Reference conditions*

Based on the responses of the indices to the different pressures and impacts along the disturbance-recovery gradient, a set of values for each index studied is suggested, which indicate an improvement in the ecological status of benthic community towards a 'reference condition'. These values were obtained after an increase by approximately 15-20% of the best values found under the best ecological scenario observed (Table IV). Values are proposed for the Lower Mondego estuary and account for both estuarine gradient and habitat characteristics. The mouth, north arm, downstream and upstream south arm will have specific references conditions that reflect the observed differences among benthic communities and respective indices' variation.

**Table IV.** Ecological indices' approximate reference values for the Lower Mondego estuary (springtime situation).

<table>
<tr><th colspan="2" rowspan="2">Lower Mondego Estuary</th><th colspan="3">Indices</th></tr>
<tr><th>Margalef</th><th>Shannon-Wiener</th><th>AMBI</th></tr>
<tr><td>Mouth</td><td>euhaline estuarine sandy</td><td>5</td><td>4.1</td><td>0.8</td></tr>
<tr><td>North Arm</td><td>polyhaline sandy</td><td rowspan="2">4</td><td rowspan="2">4</td><td>1</td></tr>
<tr><td>South Arm (downstream)</td><td>polyhaline sandy</td><td>1.5</td></tr>
<tr><td>South Arm (upstream)</td><td>polyhaline muddy sand</td><td>3</td><td>3.8</td><td>2.4</td></tr>
</table>

## Discussion

The data set here presented allowed for indices' comparison at distinct habitats within an estuary and also to observe their variation under different types of disturbance and along recovery gradients. Since our objective was to propose guidance reference values for ecological quality assessment using these indices, an effort was made to integrate these influencing factors, allowing for a better understanding of the system's evolution due to human induced causes.

### *Indices response to disturbance and recovery*

Despite the indices' overall capability of tracking recovery gradients along the three studied periods at both arms, some differences between them were observed. The Margalef and the Shannon-Wiener indices could differentiate between eutrophic and non eutrophic situations, as well as severe physical disturbance from regular dredging activities. The Margalef Index was effective in tracking the ecological improvement that occurred in each arm between the first and last periods studied. The Shannon-Wiener Index also indicated a positive evolution of the system, detecting improvements already in Period II comparatively to Period I, for the entire estuary.

The *AMBI* though could only register significant changes in the north arm, regarding physical disturbances, from Period I to Period III. However, the mean scores classify them both as slightly polluted periods, indicating only a slight improvement in quality during this study. Physical impacts such as sand extraction, dredging activities or re-working of organic – poor sediments at subtidal sandbanks, have already been reported as not causing any major changes in *AMBI* values (e.g., Borja *et al.* 2003, Muxika *et al.* 2005). No increase in opportunistic species occurred on account of these types of interventions. Some authors state that although larvae of pioneer opportunistic species may be adapted to colonise enriched sediments, the adults of this small sized, short life cycles and rapid colonisers of defaunated areas, may not necessarily be tolerant towards oxygen deficiency or physical disturbance (Rosenberg *et al.* 2004) and other species may be more tolerant than the opportunists (Diaz & Rosenberg 1995, Rosenberg *et al.* 2004). Therefore, physical disturbance seems to not favour *r*-strategy macrobenthic species, with the authors of this index recognising that competitive ability of species classified as opportunistic in *AMBI* might not be advantageous in organic-poor or naturally-physically stressed environments such as the inner parts of estuaries (Borja *et al.* 2003, Muxika *et al.* 2005). Several studies found evidence of sediment structure reestablishment and faunal re-colonisation occurring in shorts periods of time (weeks to months) after dredging and, even in situations of severe physical interventions, former macrobenthic community structure almost recovers completely after few months (e.g., Garcia-Guerra *et al.* 2003, Sanchez-Moyano *et al.* 2004, Guerra-García & García-Gómez 2006). This fact might explain the fairly good results obtained with *AMBI* for the Mondego's north arm in the first period of intense physical interventions, and their proximity with the scores found in the second and third periods, where only comparatively small scale dredging impacts occurred.

Regarding the south arm recovery, previous studies have demonstrated its positive evolution at most downstream areas; after severe symptoms of eutrophication. The increase of *Zostera noltii* beds reflected positively within south arm intertidal benthic communities (Cardoso *et al.* 2007, Dolbeth *et al.* 2007, Patrício *et al.* 2007). In this study, the subtidal communities near intertidal *Zostera noltii* beds (south arm downstream stations S1 and S2), showed also significantly different composition between disturbed and undisturbed periods (Figure 5B). Evidence of the positive influence of the proximity of seagrass habitats in communities from unvegetated sediments was found in other studies, showing that macroinvertebrate species were correlated with the dynamics of those macrophyte beds (Turner *et al.* 1999, van Houte-Howles *et al.* 2004). Even though the AMBI was initially designed to be sensitive to trophic induced changes within communities, it did not detect any benthic improvement within south arm less confined areas during the study period. In the absence of low dissolved oxygen events, which is the present case, eutrophication has proven to have a minimal effect on the benthos (Dauer *et al.* 2000). Moreover, one of the factors affecting the distribution of tolerant species is salinity (Rosenberg *et al.* 2004), hence the dominance of typically estuarine tolerant species (ecological group III) seems to be forcing *AMBI* classification. Therefore no significant changes in community composition (ecological groups) along the eutrophication recovery gradient could be identified with the AMBI.

However, the AMBI could identify better the spatial gradient of eutrophication from outer to inner stations than the temporal recovery one within the downstream southern communities. The PCA showed a positive association of *AMBI* with muddy sediments and higher organic matter content. These are parameters strongly associated to the inner stations of Mondego Estuary's south arm, indicating higher values of the index for these sites through out the study period.

On the other hand, the performance of the ecological indices applied was also affected by the occurrence of a massive flood event. In 2000, before the floods of 2001 (Period II), the outer stations presented higher diversity (*H'*) than inner ones, following most estuaries' usual trend of declining diversity and species richness with salinity decrease (Dauer & Alden 1995, Ysebaert *et al.* 2003, Dethier & Schoch 2005, Zettler *et al.* 2007). But in 2002, after the floods, *H'* index registered values at outer stations lower than those presented by confined stations. Secondary production studies on intertidal benthic communities' in the estuary also showed the setback effect this extreme flood has had on the benthic production levels during this recovery process (Dolbeth *et al.* 2007). The CCA showed that in Period II communities from north and downstream south arms' stations were correlated to phosphate increase. The Mondego Estuary is still exporting nutrients as a result of decades of retention due to anthropogenic activities, and if eroded, as by a flood event, the enriched sediments may be a great source of phosphate (Neto *et al.* 2008). Downstream benthic communities typically under higher salinity conditions were more affected by this flood, interfering with species mortality and recruitment events (Teixeira *et al.* 2007). Such events may lead to mistaking classifications of unstressed communities as highly stressed, due to dense recruitment events (Beukema 1988, Dauer *et al.* 1993) or species loss; and might have prevented the Margalef and AMBI indices to detect some improvements in the estuarine communities already in Period II.

### *Indices response to estuarine gradient and habitat heterogeneity*

The results indicate that the indices at some extent also respond to the spatial heterogeneity at the estuary. Previous studies at this estuary already reported differences regarding community structural parameters, such as species composition, total macrofaunal abundance and biomass; secondary production; physicochemical and hydromorphomological factors along the estuarine gradient and for distinct habitats (Marques *et al.* 1993, Pardal *et al.* 1993, Chainho *et al.* 2006, Dolbeth *et al.* 2007, Teixeira *et al.* 2008a).

The estuarine gradient, represented by the confinement/salinity regime factor, seems to take an important role in the results obtained for the *AMBI*, with decreasing scores from outer stations, nearer the estuary mouth, to inner stations of both arms. Similar trend was observed by Borja *et al.* (2008b) at Chesapeake Bay. If in the south arm the eutrophy gradient might act as a confounding factor; in the north arm, outer stations in the two latest periods have been under higher dredging frequency and intensity than inner ones. As for the Margalef and Shannon-Wiener indices, despite no significant differences regarding confinement were clear in the experimental design (except for *H'* in Period II - 2000/2002), PCA showed that both of them are strongly related to increasing salinity and the percentage of medium and fine sand in the sediment. These parameters were strongly associated with the downstream areas of both arms (Teixeira *et al.* 2008a), subsequently indicating an increasing trend of both indices from the inner estuary towards its marine limits. These findings meet common knowledge of estuarine communities' dynamics related to salinity variability, which originates lower specific richness than in adjacent coastal systems (Atrill & Rundell 2002).

The different habitats on the system, defined essentially by hydrodynamics and sediment type, also affected the indices' behaviour. The higher hydrodynamism in the northern arm promotes different sediment structure (Teixeira *et al.* 2008a) and communities characterized by fluctuating marine species which are mostly K-strategy species, from ecological group I and II. In the south arm, where circulation until 2006 depended primarily on a tidal regime, the conditions are favourable for fine particles and organic matter deposition (Flindt *et al.* 1997). Sediment stability and organic matter content are essential to infaunal assemblage's development, promoting enhanced populations, namely of small polychaetes, within a highly productive environment such as the Mondego south arm (Dolbeth *et al.* 2003). It is difficult to evaluate organic enrichment situations in such already naturally rich areas (Elliot & Quintino 2007), nevertheless, the higher values *AMBI* assumes in the south arm, are more a consequence of distinct habitat characteristics rather than a reflection of the worse condition of the downstream areas of this arm, since in Period III the disturbance was minor, eutrophication symptoms had disappeared and even the highly disturbed lower intertidal communities had showed a clear recover (Dolbeth *et al.* 2007).

### *Reference conditions for macrobenthos at the Mondego estuary*

Our results show that even along disturbance gradients, and for different types of impacts, the main spatial estuarine gradients still reflect in the benthic communities' structural parameters and therefore will have an effect on the values obtained when applying ecological indicators. This emphasises findings from other studies that pointed the necessity of establishing different reference conditions for different estuarine habitats before benthic condition assessment (Weisberg *et al.* 1997, Borja *et al.* 2008b, Chainho *et al.* 2008). Moreover, since the data set covered several ecological quality conditions, it allowed further comprehension on: how these indices vary; the range of values that might come up on these distinct scenarios; and their efficiency as quality assessment tools. This was of utmost importance when establishing the reference conditions for this system, accounting furthermore for natural variation according to WFD requests. To reflect the estuarine specificities already mentioned throughout the paper, values for expected undisturbed communities were adjusted for each estuarine zone (Table IV). A general decrease of diversity from marine limits (mouth) towards the inner estuary (upstream south arm) is represented in the reference values. Moreover, AMBI's sensitivity to species composition driven by differences in habitat characteristics is also expressed by the different values expected for each of the four areas.

Due to the highlighted weaknesses and to the different features of the benthic community captured by each index, as Dauer (1993) pointed out, the combined use of multiple variables, methods, or analyses with different assumptions is therefore suggested to ensure more reliable assessments.

In this paper there was also evidence of natural extreme events, such as floods, interfering with the ecological indices performance, but the part that these extreme climate changes take in the quality assessments studies is yet to be further scrutinized.

The data set available to cover the disturbance-recovery events on the Mondego estuary reports to springtime, therefore the reference conditions proposed do not account for the seasonal variation on the benthic communities structural parameters. The AMBI index has been reported as not susceptible to seasonality (Muxika *et al.* 2007); however, evidence of seasonal variation has been described for Margalef and Shannon-Wiener indices (Reiss & Kröncke 2005, Chainho *et al.* 2007). Hence this should be considered before using this set of values out of the mentioned season in any ecological assessment.

# Chapter III

## Quality assessment of benthic macroinvertebrates under the scope of WFD using BAT, the Benthic Assessment Tool

**Abstract** Assessing the health of ecosystems has become a focal point among researchers worldwide. Recently, the European Water Framework Directive intensified the development of approaches to assess ecosystems' ecological quality. The Benthic Assessment Tool (BAT) is a multimetric approach to evaluate condition of subtidal soft bottom macroinvertebrates of coastal and transitional waters. The effects of anthropogenic disturbances on benthic macroinvertebrate communities, from 1990 to 2006, allowed testing BAT performance in Mondego estuary (Portugal). The method was able to detect decrease on ecological quality, induced essentially by eutrophication and physical disturbances, and follow communities' subsequent recovery. It evidenced, nevertheless, some limitations associated with the unstable nature of estuaries. The ecological classification of key species in the community and the balance expected between ecological groups of estuarine communities had great influence in the final ecological assessment. Shortcomings of the method were discussed in the light of its suitability for assessing transitional waters' condition.

**Keywords**
WFD
Ecological Quality
Benthic Macroinvertebrates
Transitional Waters
BAT

## Introduction

Under the scope of the European Water Framework Directive (WFD, EC 2000), the assessment of the ecological quality of a water body must be based upon the status of different biological quality elements and supported by hydromorphological and physicochemical quality elements. The status of these elements is determined by the deviation they exhibit from the expected reference conditions, at undisturbed or nearly undisturbed situations, specific for each type of water body (EC 2000, annex V).

Several tools, for coastal (CW) and transitional waters (TW), have been developed and successfully tested by EU Member States for the ecological assessment of different biological quality elements: phytoplankton (Devlin *et al*. 2007, Revilla *et al*. 2008), angiosperms (Foden & de Jong 2007, Romero *et al*. 2007, Garcia *et al*. 2009) and marine macroalgae (Ballesteros *et al*. 2007, Arévalo *et al*. 2007, Wells *et al*. 2007, Juanes *et al*. 2008, Neto *et al*. submitted), benthic macroinvertebrates (Borja *et al*. 2000, Rosenberg *et al*. 2004, Muxika *et al*. 2007), and fish fauna (this element only for TW) (Breine *et al*. 2007, Coates *et al*. 2007). It is essential that outputs from different assessment methods result in equivalent quality levels for systems under the same ecological condition. This is a necessity underpinning the WFD intercalibration exercise and, therefore, EU Member States are required to compare their assessments and ensure that the metrics are calibrated (e.g., Buffagni *et al*. 2006, Borja *et al*. 2007, Ruellet & Dauvin 2007, Carvalho *et al*. 2008). However, the first need is to ensure that assessment tools can successfully detect environmental oscillations caused by non-natural causes. Within estuarine systems it is more difficult to establish a stressor-response relationship using ecological indices due to their natural variability (Wilson & Welkaim 1992, Elliot & Quintino 2007).

The Benthic Assessment Tool (BAT), presented in this paper, was developed to assess ecological quality based on the macroinvertebrate fauna from subtidal soft bottoms. The tool includes a selection of metrics for abundance and composition of macroinvertebrate fauna as required by the WFD and has been tested in Portuguese CW and intercalibrated with other tools of the NEAGIG for CW (WFD NEAGIG 2007). As for the BAT, most of the methods available were initially developed for marine habitats therefore the assessment of transitional systems requires an adaptation of the existent metrics to their natural dynamics (Dauvin 2007). In a previous paper, Teixeira *et al*. (2008b) concluded that after adjusting reference values of the metrics included in this method (Margalef index, Shannon-Wiener index and AMBI) for estuarine natural gradients they presented significant differences for benthic communities under distinct anthropogenic pressures both spatially and temporally.

The objective of this study was now to evaluate the indices joint performance in the BAT in assessing the ecological condition of benthic macroinvertebrate communities *sensu* WFD in estuaries. As an example, a long term data series from a well documented South-Western European transitional water system was used: the Mondego estuary. This system was under the influence of several pressures (Flindt *et al*. 1997, Marques *et al*. 2003, Neto *et al*. 2008), constituting therefore an adequate test dataset. The response of benthic macroinvertebrate communities to natural oscillations and human induced disturbances, from 1990 to 2006, allowed evaluating the BAT ability to capture changes on their ecological quality status.

## Methodology

### *Benthic Assessment Tool (BAT)*

The BAT is a multimetric methodology using three indices selected from previous works (Bettencourt *et al*. 2004, Teixeira *et al*. 2007) in order to fulfil WFD recommendations of including 'abundance' and 'composition' as measurable attributes for macroinvertebrate benthic fauna. The metrics selected to translate these attributes were the Margalef index (d) (Margalef 1968) and the Shannon-Wiener index (H' $\log_2$) (Shannon & Weaver 1963), since these indices provide complementary diversity measures. To assess the composition of macrobenthos it was chosen the AZTI's Marine Biotic Index (AMBI) (Borja *et al*. 2000), which is based on the presence of sensitive and pollution indicator species.

Within the WFD, the assessment must be presented in a standardized way as an Ecological Quality Status (EQS), translating the results of the Ecological Quality Ratio (EQR) (EC 2000). Therefore a multimetric approach combining the three indices into an EQR was adopted to complete the BAT. The method chosen to combine indices' information was described by Borja *et al*. (2004a) and has been used to assess physicochemical (Bald *et al*. 2005) and macroinvertebrates (Muxika *et al*. 2007) EQS within the scope of WFD. It consists in Factor Analysis (FA), with the Principal Component Analysis (PCA) as extraction method, applied to the three indices' results for all samples. Virtual samples representing the expected values for the indices in 'High' and 'Bad' ecological status need to be included in the analysis. Data are previously standardised by subtracting the mean and dividing by the standard deviation. In the analysis, the 'varimax' rotation method is adopted to facilitate the interpretation of results, and the first 3 factor scores are extracted (using the 'factor loadings' on their calculation). These scores provide samples' relative positions in the new 3-D space created by the FA, and then the projection of the community represented on each sample on the axis connecting both reference stations ('High' and 'Bad' status) can be calculated. The value of 1 is attributed to the distance between both virtual stations ('Bad' and 'High'), accordingly to the definition of EQR (EC 2000). Subsequently, the Euclidean Distance of each projection point to the virtual 'Bad' status situation is measured, so samples in better condition, with higher Ecological Status, achieve EQR values near 1, while those in worse ecological condition will be located near the 'Bad' reference station (values close to 0). Within the EQR scale (0 to 1) five ecological quality status classes were defined to establish the final EQS: 'Bad', 'Poor', 'Moderate', 'Good' and 'High' (EC 2000).

With this method sampling points can change position with the incorporation of new data (Bald *et al*. 2005, Ruellet & Dauvin 2008), leading eventually to different EQS than in previous evaluations. To avoid this effect Borja *et al*. (2008c) recommended using this method with a sampling data set of at least 50 locations to minimize the changes. In addition, each time new data integrate the data matrix, as Bald *et al*. (2005) proposed, Discriminant Analysis (DA) should be used to attribute an EQS instead of re-run the complete FA. For WFD assessment purposes, a complete FA can be conducted every 6 year when a new EQS evaluation is required from Member States.

*Study site*

The Mondego estuary (Figure 1) is a relatively small warm-temperate polyhaline intertidal system (21 km long and 860 ha surface area), located at the western coast of Portugal (40º08'N, 8º50'W).

Since the 1930s, large-scale interventions have occurred in the estuary that modified the riverbed's topography and altered its hydrodynamics. The main modifications were the construction of harbour facilities near the estuary mouth and, in the early 1990s, the erection of stonewalls to create channels and water reservoirs to regulate the river water flow and improve the use of water resources. As a consequence, the southern arm (Figure 1) became strongly silted up in the upstream areas, causing the river discharge (mean annual average 79 $m^3 s^{-1}$) to flow essentially through the North arm. In 1994, the communication between the two arms of the estuary was totally interrupted due to the completion of stonewalls in the northern arm banks. Consequently, water circulation in the southern arm became essentially dependent on tides and small freshwater input from a tributary artificially controlled by a sluice, the Pranto River and eutrophication symptoms in the South arm became evident since (Marques *et al*. 2003).

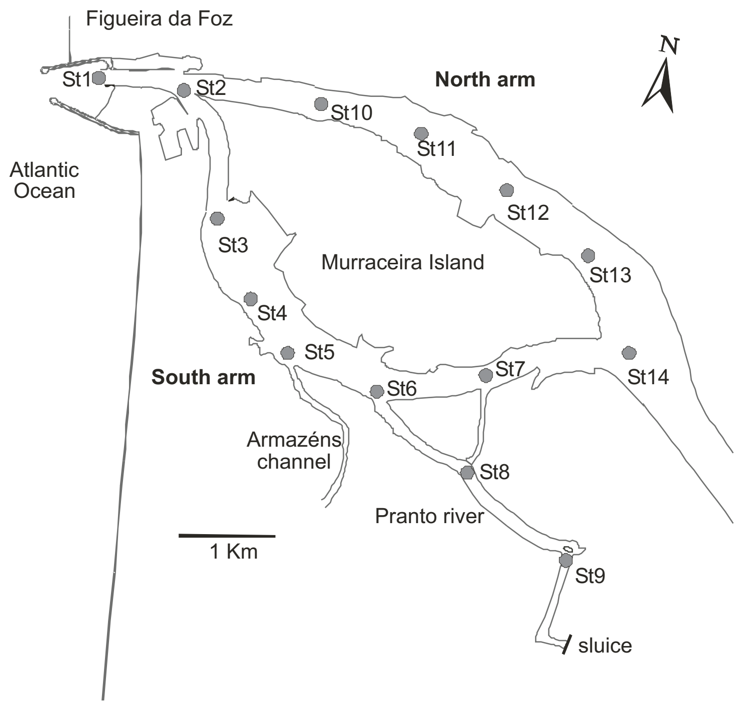



**Figure 1.** The lower Mondego estuary: location of the subtidal sampling stations (St1 to St14).

From the early 1990s to 2006, three distinct periods under distinct anthropogenic influence could be identified at this estuary: (a) a first period before 1994, characterised by the occurrence of strong physical disturbances within the North arm; (b) a second period, from 1994 to 1997/98, during which the interruption of the communication between both arms prevailed and environmental conditions degraded in the southern arm, resultant from higher water residence time and nutrient loading; and (c) a third period after 1997/98, when experimental mitigation measures were implemented in the estuary, namely by re-establishing for a very limited extent the communication between the two arms (only 1.5 to 2 hours before and after each high tide peak and through a section of only 1 $m^2$) and reducing the loading of nutrients into the South arm (Lillebø *et al*. 2005, Neto *et al*. 2008).

The occurrence of natural extreme climatic events was also registered during the study period, namely massive floods during the winters of 1996 and 2000/2001, and two severe droughts in 1992 and 2005.

### *Sampling program*

The dataset used in this study corresponds to samples from subtidal soft-bottom benthic macroinvertebrate communities collected during spring seasons of 1990, 1992, 1998, 2000, 2002, 2004, 2005, and 2006.

Biological samples were collected at 14 stations located at the lower Mondego estuary (Figure 1). At each station three to five replicates were randomly collected from a small boat using a van Veen grab model LMG (5 to 8 L sampling volume, and 0.0496 to 0.0780 $m^2$ sampling surface). At every sampling event, to support biota data, physicochemical parameters were also measured close to the bottom (turbidity; particulated organic matter – POM; total suspended solids – TSS), at high tide situation. Bottom water samples were collected for determination of nutrients (ammonium N-$NH_4$; phosphate P-$PO_4$; nitrite N-$NO_2$) and chlorophyll *a* concentrations. Analyses of concentrations were performed in the laboratory according to Limnologisk Metodik (1992) for N-$NH_4$ and P-$PO_4$, and APHA (1995) and Strickland and Parsons (1972) for N-$NO_2$ and chlorophyll *a*.

In the laboratory, the 1 mm fraction of the biological samples was separated and fixed with 4% buffered formalin solution. Afterwards, animals were sorted and preserved in 70% ethanol for later identification and count to species level (or lowest reliable taxonomic level). Data *per* sample was standardized, by transforming the macrobenthic abundance to number of individuals *per* $m^2$ (indiv. $m^{-2}$) (density), which was then used in the indices' calculations.

The macroinvertebrates matrix was truncated according to rules adopted by NEA-GIG Member States within the intercalibraton exercise (Borja *et al*. 2007), to avoid as much as possible bias in data originated from different procedures. In our case it allowed also to harmonise data collected through a long time period at this estuary.

In order to detect natural disturbances, extreme climatic events were analysed using river flow as a surrogate. Hydrological data was obtained from INAG – Portuguese Water Institute (http://snirh.inag.pt). Annual mean flow registers were obtained from the Coimbra's dam 'Ponte-

Açude' station (12G/01AE), located 38 km inland from the seaside. These data (Figure 2) illustrate the occurrence of two floods, in 1996 and during the winter of 2000/2001; and two severe droughts during 1992 and 2005. The available biological dataset allows only to evaluate the effect of the centenary floods during winter 2000/2001, and severe droughts of 1992 or 2005.

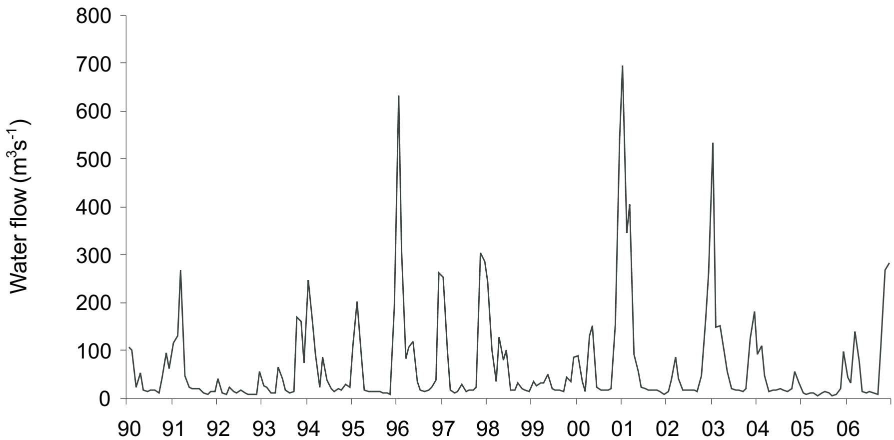



**Figure 2.** Annual mean flow ($m^3\ s^{-1}$), from 1990 to 2006, at Coimbra's dam 'Ponte-Açude' station located upstream in the Low Mondego river valley.

*Data analysis*

Benthic communities were analysed regarding temporal variations of the indices: Margalef (1968), Shannon-Wiener (Shannon & Weaver 1963) and AMBI (Borja *et al*. 2000). The software and guidance to calculate the AMBI are freely available at AZTI's web page (http://www.azti.es) and in Borja and Muxika (2005). For AMBI calculation the species list of December 2007 was used. To evaluate if the ecological indices variation along the study period the Kruskal-Wallis test was applied to check for differences between years in each estuarine arm (St1 was excluded from this analysis since it was not surveyed in all sampling years). To determine which years were different, the Fisher's least significant difference (LSD) method was performed on ranks for multiple pairwise comparisons.

Then the BAT method was applied to assess the EQS of benthic macroinvertebrates at the Mondego estuary during the study period. To calculate the EQR, the reference conditions defined by

Teixeira *et al*. (2008b) for this estuary (Table I) were used. To establish a correspondence between the EQR and a final EQS, the following thresholds were adopted: 0 ≤ Bad ≤ 0.27; 0.27 < Poor ≤ 0.44; 0.44 < Moderate ≤ 0.58; 0.58 < Good ≤ 0.79; and 0.79 < High ≤ 1. These ranges were agreed after the intercalibration exercise for coastal waters in the North East Atlantic region (WFD NEAGIG 2007). Despite these thresholds are valid for the CW type adjacent to this estuary, they are not adjusted to TW types, and thus are used here merely as guidance, since the intercalibration exercise within TW tools is yet to be conducted.

**Table I.** Reference conditions, adapted from Teixeira *et al*. (2008b), for an estuary of the Type A2 within Portuguese transitional waters (Bettencourt *et al*. 2004). 'High' and 'Bad' reference values for Margalef, Shannon-Wiener and AMBI indices within the BAT.

| **Habitats** | **Margalef (*d*)** | **Shannon-Wiener ($H'log_2$)** | **AMBI** |
|---|---|---|---|
| High status | | | |
| Euhaline estuarine* | 5.0 | 4.1 | 0.8 |
| Polyhaline / Sandy** | 4.0 | 4.0 | 1.0 |
| Polyhaline / Sandy-mud *** | 4.0 | 4.0 | 1.5 |
| Polyhaline / Muddy-sand **** | 3.0 | 3.8 | 2.4 |
| Bad status | | | |
| All habitats | 0.0 | 0.0 | 7.0 |

Habitat correspondence within the Mondego estuary: * St 1, 2, 10; ** St 11 to 14; *** St 3, 4; **** St 5 to 9.

The contribution of each index to the final ecological classification was measured by means of Spearman Rank correlations between the three indices of the BAT method and the final EQR.
To assess compliance of the obtained classifications with WFD EQS definitions (EC 2000, annex V), random samples along the disturbance gradient were picked and the ecological groups' distribution in each class were observed.

Finally, to test the effect that classification of key species in the community has in the final assessment, the BAT was recalculated with a change in the AMBI index calculation. EQR assessments produced both ways were then compared. For that, and taking into account the abundance patterns exhibited by the species *Alkmaria romijni*, this species was experimentally reclassified as an opportunistic taxon (of second order – EG IV) in the new AMBI run, changing the EGIII assignment proposed in the AMBI species list.

All statistical analyses were performed using the Statgraphics Plus 5.1 software.

## Results

*Indices' patterns and ecological status assessment*

In the North arm, the Margalef index presented lower values in the beginning of the study period (1990 and 1992) and gradually increased until the last years, reaching the highest mean value by 2006 (Table II; Figure 3). This was confirmed by the significant differences found by the Kruskal-Wallis test along the study period for this index (test statistic = 16.568; $p$-value = 0.02), where *a posteriori* tests showed differences essentially between the initial years (1990 to 2000) and the latest years (2004 to 2006) (see Table III for pairs comparisons). Although less evident, results of the Shannon-Wiener index in the North arm varied temporally in the same way as Margalef's, lower values were found in the first two years of study, followed by a slight increase in the subsequent years (Table II; Figure 3). However, despite the observed trend the differences found between years were not significant for this index (Kruskal-Wallis test statistic = 8.022; $p$-value = 0.331). The AMBI showed the least variation of the three indices throughout the study period in the north arm (Kruskal-Wallis test statistic = 4.657; $p$-value = 0.702). Only at downstream stations (Sts 2 and 10) some slightly higher values (meaning more disturbed communities) were registered during the first sampling years (1990 to 2002) comparatively to subsequent years (2004 to 2006) (Table II). Spatially, except for periods highly disturbed in the North arm, stations closer to the river mouth (Sts 1, 2 and 10) showed, in general, the highest species richness (*d*), while middle stations usually registered the lowest (Table II). This downstream area also presented usually slightly higher diversity (H') and lower AMBI values (meaning less disturbed communities) which tended to increase towards the upstream areas of the North arm (Table II).

In the South arm, the Margalef index decreased from 1990 to 1998 and started to increase again in the following years until 2006 (Figure 3). The Kruskal-Wallis test revealed that the trends observed during the study period for Margalef index in this arm were significant (test statistic = 30.066; $p$-value ≤ 0.0001). In general, the first years of study registered significantly different values

**Table II.** Values of the Margalef (*d*), Shannon-Wiener (*H*’) and AMBI indices registered along the study period in the 14 sampling stations of the lower Mondego estuary.

| | **North arm stations** | **1990** | **1992** | **1998** | **2000** | **2002** | **2004** | **2005** | **2006** |
|---|---|---|---|---|---|---|---|---|---|
| *d* | 1 | ‘ | ‘ | 2.156 | 1.451 | 1.469 | 1.141 | ‘ | 2.323 |
| | 2 | 0.910 | 1.259 | 1.151 | 3.857 | 1.720 | 2.088 | 3.459 | 3.980 |
| | 10 | 0.768 | 0.715 | 0.893 | 1.530 | 3.578 | 1.827 | 1.797 | 2.180 |
| | 11 | 0.949 | 1.377 | 1.987 | 0.898 | 0.711 | 1.324 | 1.924 | 0.996 |
| | 12 | 0.801 | 0.000 | 1.053 | 0.270 | 0.710 | 1.459 | 1.635 | 1.156 |
| | 13 | 1.945 | 0.731 | 1.367 | 0.666 | 2.163 | 1.982 | 1.934 | 1.580 |
| | 14 | 0.790 | 0.362 | 0.310 | 1.231 | 1.678 | 1.576 | 1.566 | 2.072 |
| *H'* (log2) | 1 | ‘ | ‘ | 2.795 | 2.394 | 2.914 | 2.200 | ‘ | 3.068 |
| | 2 | 1.555 | 1.735 | 2.156 | 3.411 | 2.714 | 2.614 | 3.072 | 3.375 |
| | 10 | 1.504 | 0.965 | 1.357 | 2.401 | 2.535 | 2.363 | 2.404 | 3.115 |
| | 11 | 1.845 | 2.447 | 2.962 | 1.842 | 1.088 | 2.250 | 0.738 | 2.076 |
| | 12 | 1.360 | 0.000 | 2.025 | 0.650 | 1.632 | 2.355 | 2.476 | 1.506 |
| | 13 | 2.949 | 1.753 | 2.358 | 1.371 | 2.082 | 2.608 | 2.250 | 1.623 |
| | 14 | 1.205 | 0.548 | 0.469 | 2.038 | 2.653 | 1.418 | 1.672 | 3.013 |
| AMBI | 1 | ‘ | ‘ | 1.223 | 1.622 | 0.978 | 1.154 | ‘ | 1.244 |
| | 2 | 2.622 | 1.543 | 1.875 | 3.270 | 2.077 | 1.083 | 1.141 | 1.908 |
| | 10 | 2.622 | 2.813 | 3.150 | 3.843 | 2.639 | 1.583 | 1.636 | 1.304 |
| | 11 | 2.568 | 2.969 | 2.413 | 2.357 | 2.833 | 2.438 | 2.816 | 2.308 |
| | 12 | 0.387 | 7.000 | 2.250 | 2.500 | 2.985 | 1.375 | 2.250 | 2.880 |
| | 13 | 2.635 | 2.000 | 2.400 | 1.200 | 2.972 | 2.719 | 2.661 | 3.197 |
| | 14 | 2.873 | 2.960 | 1.650 | 3.038 | 2.719 | 4.046 | 2.980 | 2.879 |
| | **South arm stations** | **1990** | **1992** | **1998** | **2000** | **2002** | **2004** | **2005** | **2006** |
| *d* | 3 | 1.468 | 2.013 | 0.000 | 1.736 | 1.822 | 1.318 | 2.264 | 1.674 |
| | 4 | 3.146 | 1.467 | 0.935 | 1.992 | 2.104 | 0.770 | 3.285 | 1.734 |
| | 5 | 2.105 | 1.911 | 1.071 | 1.213 | 1.880 | 1.692 | 2.695 | 3.281 |
| | 6 | 1.177 | 1.321 | 1.254 | 1.026 | 2.334 | 1.567 | 1.952 | 2.233 |
| | 7 | 1.574 | 0.713 | 0.809 | 1.431 | 1.877 | 1.779 | 1.897 | 2.037 |
| | 8 | 1.248 | 1.175 | 0.854 | 1.146 | 2.025 | 0.859 | 2.035 | 1.632 |
| | 9 | 1.373 | 0.381 | 0.545 | 0.799 | 1.290 | 0.000 | 1.973 | 1.826 |
| *H'* (log2) | 3 | 2.556 | 2.940 | 0.000 | 2.550 | 1.174 | 2.219 | 1.566 | 2.316 |
| | 4 | 3.112 | 2.419 | 1.428 | 2.922 | 1.114 | 1.000 | 3.362 | 1.670 |
| | 5 | 1.224 | 2.739 | 2.033 | 2.495 | 2.842 | 1.902 | 3.224 | 3.302 |
| | 6 | 0.719 | 1.881 | 1.910 | 1.461 | 3.194 | 2.137 | 2.642 | 3.310 |
| | 7 | 1.614 | 1.441 | 1.664 | 2.392 | 2.836 | 2.649 | 2.486 | 2.935 |
| | 8 | 1.932 | 2.346 | 1.463 | 1.683 | 2.826 | 1.500 | 2.576 | 2.987 |
| | 9 | 2.305 | 0.346 | 0.587 | 1.380 | 2.483 | 0.000 | 2.817 | 3.183 |
| AMBI | 3 | 2.144 | 1.904 | 3.000 | 3.000 | 2.908 | 2.959 | 2.836 | 2.152 |
| | 4 | 2.865 | 2.500 | 2.769 | 2.400 | 3.045 | 3.000 | 3.347 | 2.922 |
| | 5 | 3.088 | 3.182 | 3.041 | 3.042 | 3.037 | 2.979 | 3.252 | 3.899 |
| | 6 | 3.044 | 3.071 | 2.859 | 2.993 | 3.128 | 3.176 | 3.039 | 3.223 |
| | 7 | 3.307 | 3.075 | 3.000 | 3.056 | 3.421 | 2.914 | 3.061 | 3.153 |
| | 8 | 3.240 | 3.157 | 3.087 | 3.079 | 3.016 | 3.000 | 3.433 | 3.185 |
| | 9 | 3.255 | 2.895 | 3.000 | 3.000 | 3.173 | 7.000 | 3.131 | 2.973 |

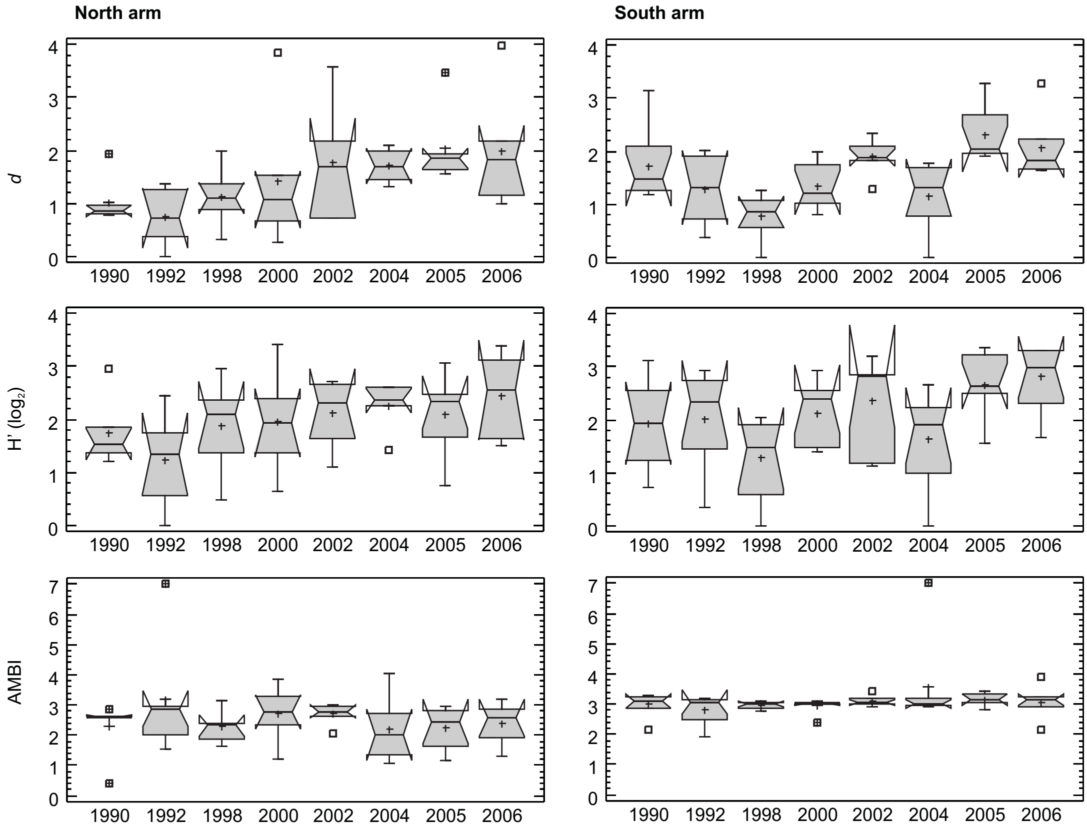



**Figure 3.** Variation of the Margalef, Shannon-Wiener and AMBI indices in the North (left hand) and South (right hand) arms of the estuary, along the study period (mean: cross; median: horizontal line; Q25 and Q75: box; maximum and minimum: whiskers extending from box; and when present also signed outliers: squares; and far outliers: crossed squares).

from those in the year 1998 and from those of the last years. In the last years values were significantly higher than those registered in 1998 and in 2000 (Table III for detailed pairs comparisons). In the year 2004 a punctual decrease in the index values was observed (Figure 3; Table III) with significantly lower values than contiguous years (2002, 2005/06). The Shannon-Wiener index varied similarly, presenting significant differences between years in the south arm (Kruskal-Wallis test statistic = 18.094; *p*-value = 0.012). In 1998 the estuary reached the lowest values for this index, but in 2000 they started to reach values from the initial years and, with exception of 2004, they continued to increase until 2006, where the highest mean value for the index was observed. These changes were gradual and not all consecutive years present significant differences, but in general

**Table III.** *Post hoc* multiple comparisons using the Fisher's least significant difference (LSD) method, for indices where Kruskal-Wallis tests denoted significant differences. Only significantly different pairs of years are presented; *d*: Margalef index; *H'*: Shannon-Wiener index.

| LSD contrast | difference | *p*-value |
|---|---|---|
| **North arm** | | |
| ***d*** | | |
| 1990 vs. 2004 | -15.333 | 0.036 |
| 1990 vs. 2005 | -18.167 | 0.014 |
| 1990 vs. 2006 | -16.333 | 0.026 |
| 1992 vs. 2002 | -16.333 | 0.026 |
| 1992 vs. 2004 | -20.833 | 0.005 |
| 1992 vs. 2005 | -23.667 | 0.002 |
| 1992 vs. 2006 | -21.833 | 0.004 |
| 1998 vs. 2005 | -15.667 | 0.032 |
| 2000 vs. 2005 | -15.833 | 0.030 |
| **South arm** | | |
| ***d*** | | |
| 1990 vs. 1998 | 21.214 | 0.002 |
| 1990 vs. 2005 | -16.286 | 0.013 |
| 1992 vs. 1998 | 12.643 | 0.050 |
| 1992 vs. 2002 | -17.286 | 0.008 |
| 1992 vs. 2005 | -24.857 | 0.0003 |
| 1992 vs. 2006 | -17.857 | 0.007 |
| 1998 vs. 2002 | 29.929 | <0.0001 |
| 1998 vs. 2005 | -37.500 | <0.0001 |
| 1998 vs. 2006 | -30.500 | <0.0001 |
| 2000 vs. 2002 | -17.857 | 0.007 |
| 2000 vs. 2005 | -25.429 | 0.000 |
| 2000 vs. 2006 | -13.429 | 0.005 |
| 2002 vs. 2004 | 20.643 | 0.002 |
| 2004 vs. 2005 | -28.214 | <0.0001 |
| 2004 vs. 2006 | -21.214 | 0.002 |
| ***H'* (log2)** | | |
| 1990 vs. 2005 | -15.429 | 0.049 |
| 1990 vs. 2006 | -19.143 | 0.016 |
| 1992 vs. 2006 | -16.714 | 0.034 |
| 1998 vs. 2002 | -19.357 | 0.015 |
| 1998 vs. 2005 | -25.929 | 0.001 |
| 1998 vs. 2006 | -29.643 | 0.0003 |
| 2000 vs. 2006 | -16.143 | 0.040 |
| 2004 vs. 2005 | -20.071 | 0.012 |
| 2004 vs. 2006 | -23.786 | 0.003 |

initial years (1990 to 1998) presented significant differences from last two years (Table III for detailed pairwise comparisons). As in the North arm, among the indices studied, the AMBI showed also the least variation throughout the study period at the South arm. Hardly any variation was observed in this index (Kruskal-Wallis test statistic = 6.357; *p*-value = 0.499) and only in 1990, 1992 and 2006 slightly lower values were registered at downstream stations (Table II). Spatially, and except for highly disturbed periods in the South arm, the downstream and middle sections presented usually the best Margalef index and AMBI values (Table II), no clear trend could be observed for Shannon-Wiener index. Overall, AMBI values were usually slightly higher in this arm than the northern channel.

The BAT detected an ecological quality decrease during the first years of study at the estuary (Table IV; Figure 4A and B) with maximum expression in 1992 at North arm (Sts 2 to 14), and in 1998 at the southern arm. In fact, subtidal macroinvertebrates communities reached a 'Bad' quality status at these years at sampling stations 12 (North arm) and 3 (South arm). After these years, the general quality of the estuarine macrobenthic

**Table IV.** EQR according to BAT, for the 14 sampling stations, along the study period.

| | Station | 1990 | 1992 | 1998 | 2000 | 2002 | 2004 | 2005 | 2006 |
|---|---|---|---|---|---|---|---|---|---|
| North arm | **1** | | | 0.73 | 0.64 | 0.72 | 0.64 | | 0.76 |
| | **2** | 0.48 | 0.58 | 0.59 | 0.72 | 0.64 | 0.72 | 0.82 | 0.82 |
| | **10** | 0.46 | 0.41 | 0.42 | 0.48 | 0.68 | 0.66 | 0.65 | 0.75 |
| | **11** | 0.46 | 0.53 | 0.66 | 0.46 | 0.36 | 0.53 | 0.44 | 0.49 |
| | **12** | 0.51 | 0.00 | 0.50 | 0.30 | 0.40 | 0.61 | 0.59 | 0.43 |
| | **13** | 0.64 | 0.46 | 0.55 | 0.46 | 0.57 | 0.61 | 0.58 | 0.47 |
| | **14** | 0.37 | 0.28 | 0.33 | 0.48 | 0.59 | 0.41 | 0.48 | 0.65 |
| South arm | **3** | 0.63 | 0.71 | 0.22 | 0.60 | 0.49 | 0.53 | 0.57 | 0.61 |
| | **4** | 0.77 | 0.59 | 0.43 | 0.68 | 0.50 | 0.38 | 0.79 | 0.52 |
| | **5** | 0.65 | 0.75 | 0.63 | 0.68 | 0.77 | 0.68 | 0.85 | 0.84 |
| | **6** | 0.53 | 0.64 | 0.65 | 0.59 | 0.82 | 0.67 | 0.76 | 0.82 |
| | **7** | 0.62 | 0.55 | 0.58 | 0.69 | 0.73 | 0.75 | 0.74 | 0.78 |
| | **8** | 0.62 | 0.65 | 0.56 | 0.61 | 0.78 | 0.57 | 0.72 | 0.74 |
| | **9** | 0.66 | 0.45 | 0.47 | 0.56 | 0.67 | 0.00 | 0.76 | 0.80 |

communities begun to increase exhibiting, nonetheless, punctual signs of disturbance related to specific events. In the surroundings of station 12, during the construction of the Sewage Water Treatment Plant, which entered into functioning during the year 2003, a ‘Moderate’ to ‘Poor’ condition was detected (Figure 4A). At Sts 3 and 4, in 2004, (‘Moderate’ and ‘Poor’ respectively) (Figure 4B), the reduction in EQS coincides with the construction of a small fishing harbour near station 3 around that period (Neto, person. com.). The EQS reduction observed at station 9 (‘Poor’), in 2004, (Figure 4B) matches in time with the lower water transparency and higher concentrations of nutrients, TSS and POM from bottom water samples found during the spring months (Spring mean/Annual mean without Spring months, respectively, for transparency: 0.3/ 0.6 m; N-$NH_4$: 0.44/0.26 mg $L^{-1}$; N-$NO_2$: 0.056/ 0.035 mg $L^{-1}$; $PO_4$: 0.065/ 0.048 mg $L^{-1}$; TSS: 0.075/ 0.037 g $L^{-1}$; POM: 0.009/ 0.005 g $L^{-1}$). In the 2002 Spring, following a rainy 2001 where a massive flood event was registered (Figure 2), it was also observed an EQS decrease in downstream stations of the South arm (Sts 3 and 4) (Figure 4B).

Overall, a gradual increase of the ecological quality in the entire system was registered by the BAT after the main disturbances harassing each estuarine arm were reduced, with higher visibility from 2005 onwards. The upstream stations from the northern arm (Sts 11 to 14) presented however less stable ecological condition along the study period, being more often characterised by deficient ecological conditions (‘Moderate’ or ‘Poor’) (Figure 4A). The Spearman Rank correlations between the three indices that constitute the BAT method and the final EQR revealed that the classification result was mainly driven by Margalef (R = 0.77; $p < 0.001$) and Shannon-Wiener (R = 0.87; $p < 0.001$) indices. The AMBI showed no significant correlation with the final output (R = 0.09; $p > 0.05$).

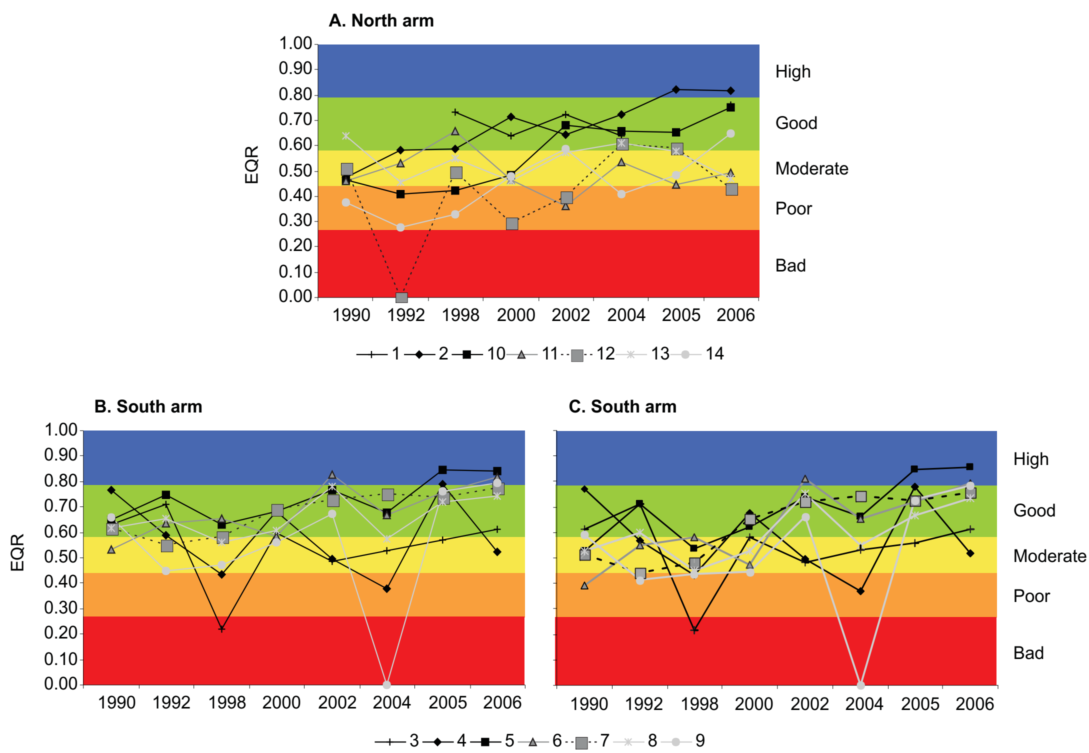



**Figure 4.** EQR and EQS according to BAT method, for the sampling stations (1 to 14) within the two estuarine arms (A and B) during the study period; (C) EQS within South arm after species tolerance reclassification in AMBI.

### *Taxonomic composition component*

AMBI presented a weak variation and hence a low contribution to the final EQR. To analyse this, stations representative of the 5 classes of EQS (with EQRs from 0 to 0.85) were randomly selected in each of the estuarine arms (Figure 5A and B). Analysing the ecological groups' distribution in these samples, according to AMBI classification of the species, in none of the two arms was an increase of opportunistic species observed at either ‘Poor’ or ‘Bad’ conditions (Figure 5A and B). Randomly selected stations with ‘Poor’ and ‘Bad’ status registers (Figure 5A and B) corresponded actually to situations of nearly or total faunal depletion (Table II). On the other hand, tolerant species (EG III) seem to be dominating across all range of ecological status, especially within the South arm (Figure 5B).

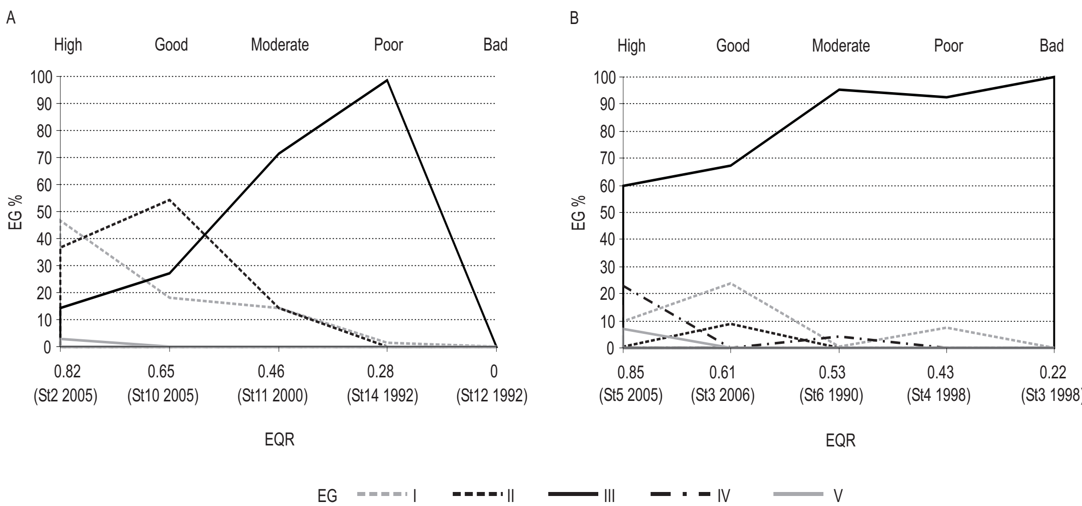



**Figure 5.** Ecological groups (EG) succession along two disturbance gradients at the North (A) and South (B) arms’ communities, from High to Bad EQS according to the BAT assessment. EG: I – very sensitive species, II – indifferent species, III – tolerant species, IV – second-order opportunistic species, and V – first-order opportunistic species, according to their sensitivity to an increasing pollution gradient (Borja *et al*. 2000).

Two relatively common species within this estuary are the polychaetes *Alkmaria romijni* (Horst 1919) and *Streblospio shrubsolii* (Buchanan 1890). The abundance patterns presented by these two species at this estuary differed spatially (Figure 6A and B). *A. romijni* occurred mainly in the South arm subsystem, while *S. shrubsolii* was observed at both arms. Where they co-occurred, it was observed that *A. romijni* reached to a great extent higher density (Figure 6B), exhibiting frequent density peaks during the first study years. *A. romijni* density started decreasing after the implementation of mitigation measures (around 1998) and by 2002 had equal the density registers of *S. shrubsolii*.

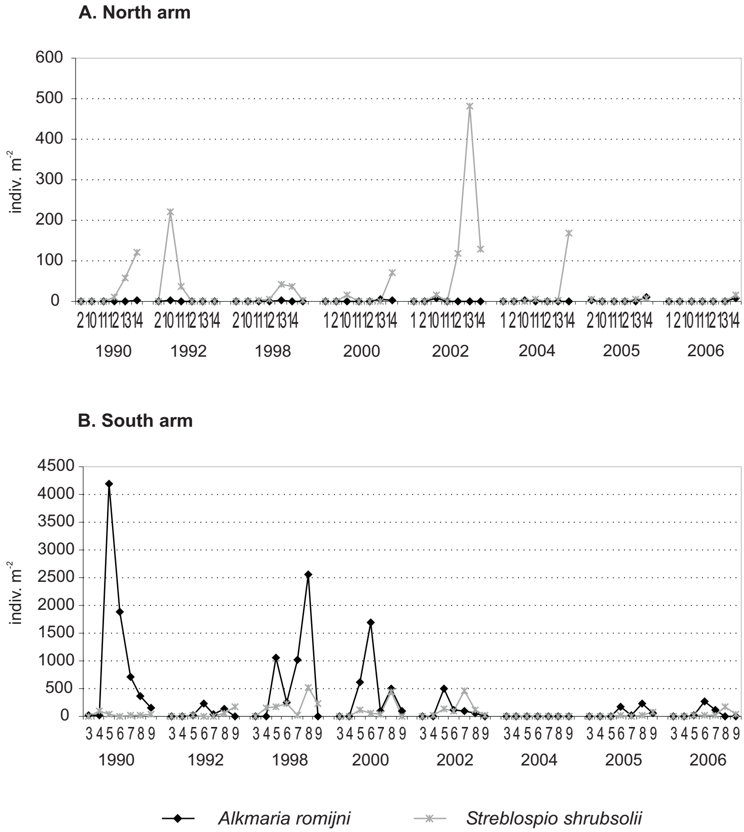



**Figure 6.** Abundance patterns (number of individuals *per* m$^2$) of the AMBI EG-III polychaetes *Alkmaria romijni* and *Streblospio shrubsolii* within the Mondego estuary along the study period at (A) North and (B) South arms' stations.

Both species are classified as tolerant species (EG III) within the AMBI list, but considering the opportunistic patterns exhibited by the species *A. romijni*, it was observed that its experimental reclassification as an opportunistic taxon (of second order – EG IV) modified AMBI results and produced important changes in the final EQR. While in the North arm no significant change on the EQR was observed, in the South arm, the new EQR variation was great enough to induce changes in some of the final EQS classifications (Figure 4B and C). These occurred mostly until 1998 with several stations dropping one category and placing all South arm below Good EQS in 1998. The ecological quality decline until mitigation measures and the subsequent recovery was more evident when *A. romijni* was reclassified as a second-order opportunistic (EG IV) species (Figure 4C).

## Discussion

### *Overall BAT performance*

The final assessment provided by the BAT detects the major trends within subtidal benthic communities. In different areas of the system, it followed the ecological condition's decrease, due to anthropogenic influence and the expected improvement for the subsequent recovery, after cessation of the major impacts at both arms. Moreover, the assessment tool proved sensitivity towards decreasing quality associated with different types of disturbances. Not only heavy physical disturbances and enhanced eutrophication related processes could be tracked, but also the influence of smaller impacts could be detected. The BAT classification reflected the shifts that these occurrences have promoted within the benthic communities, oscillating between moderate, poor and bad in the identified situations.

These findings corroborate evidences from monitoring and scientific studies carried out in the system in the last years, namely at the South arm, which reported diversity changes for intertidal macroinvertebrate communities, before and after mitigation measures (Martins *et al*. 2001, Pardal *et al*. 2004, Dolbeth *et al*. 2007, Patrício *et al*. 2009). Changes were triggered by the nutrient enrichment of the system which, together with hydrodynamics, played a major role on macroalgal production, enhancing and controlling eutrophication processes within this arm.

The BAT was also able to detect the disturbance of the subtidal benthic communities due to artificially controlled freshwater runoff ('Moderate'/ 'Poor' status in stations 8 and 9, in 2004). These punctual events rapidly change the physical-chemical environment of these communities, by promoting strong salinity drops, increased turbidity and high availability of nutrients (Martins *et al*. 2001, Neto *et al*. 2008).

Despite trends were satisfactorily detected by the BAT, the discrimination between different degrees of disturbance still needs improvement so that benthic communities' condition is better reflected by the final EQS classification. For example, the heavy physical disturbances, such as large scale dredging and embankment activities in the North arm during the early 90's, had a higher impact in the benthic communities (leading inclusive to faunal depletion) than the smaller scale interventions, such as maintenance dredging activities or the construction of infrastructures as the wastewater treatment plant and the small fishing harbour. These last activities punctually decreased the EQS but communities' recovery was fast. Despite that, the difference in EQR between impacts of such different intensity was lower than the expected. In systems of relatively high natural disturbance, factors as the hydrodynamic regime and the vicinity to recruitment areas (e.g., Kenny & Rees 1996, Boyd *et al*. 2003, 2005), such as the marine limits in an estuary, have been identified as promoting a rapid recolonization. On the other hand, the high frequency of dredging for channel maintenance prevents communities to develop to more mature states. This could partially explain, the low EQS difference found between the communities under severe disturbance in the early 90's and those of subsequent years in the North arm.

### *Adjustment of assessment methods to estuarine environments*

Within the indices that integrate the BAT, those accounting for 'abundance' *sensu* WFD (Margalef and Shannon-Wiener) were, in general, able to capture the roll of events that occurred at the system. During the study period, they revealed a higher capacity to discriminate between undisturbed and disturbed communities, associated with clear identified pressures or interventions in the system. Contrarily, the index selected to measure taxonomic 'composition' (AMBI) was almost invariable within the study period. Except for clear environmental degradation (e.g, St 12 in 1992 and St 9 in 2004) and clearly well balanced communities under higher marine influence, the AMBI showed less discriminatory power to detect more subtle changes overtime within the estuarine subtidal macroinvertebrate communities. Therefore, Margalef and Shannon-Wiener contributed more decisively to the final EQR output of the BAT, as evidenced by the results of the Spearman Rank correlations.

The reclassification of a single key species, such as *A. romijni* in this system, changed the AMBI results improving the ecological classification across the gradient of disturbance (Figure 4C), providing better differentiation between communities most and least affected by eutrophication processes. This suggests that changes can be introduced in the AMBI that make it more adequate to TW environments. Labrune *et al*. (2006) have pointed out that the AMBI sensitivity to dominance of individuals of a given EG will force its final classification. Tolerant species, from EG-III, are expected to occur across a wide range of environmental conditions and in estuaries these species clearly dominate communities (Dauvin *et al*. 2007, Elliot & Quintino 2007) conditioning the performance of any index based on the distribution of ecological groups. In the case of AMBI, assessments tend to fall around medium levels of disturbance, with lower number of cases reaching higher or lower ecological status. This effect was clearly observed in our results, especially within inner areas of the estuary which are additionally naturally organic enriched habitats and where the dominance of such tolerant species is even stronger (Blanchet *et al*. 2008, Lavesque *et al*. 2009). In transitional systems its effectiveness could be improved by weighting ecological groups differently. Lowering the weight of the EG-III or giving an enhanced importance to the presence of EG-I or EG-II, for example, would decrease the index sensitivity to the dominance of tolerant species and increase its sensitivity to other faunal groups in the community.

*Natural vs. anthropogenic disturbances*

Although an assessment tool is required to accurately translate shifts in benthic community due to anthropogenic influence, this is often a hard goal to reach for TW. Ideally, tools should be able to cope with changes caused by natural oscillations or by natural disturbance events. However, due to the unstable nature of TW systems, namely estuaries, too often the changes that occur within macrobenthic communities' parameters are equivalent regardless of the nature of the events that promote them (Dauvin 2007).

The effects of the occurrence of extreme climatic events were not very clear on the results, nevertheless, punctual effects could be recognized, especially those of the centenary floods during winter 2000/2001. The cumulative effects of anthropogenic disturbances occurring simultaneously could have masked the effect of these natural extreme events, especially during the first years of study when the system was under higher pressures.

Given the aforementioned, the first remark goes to the fact that the effects of extreme climatic events on the subtidal benthic community were not equally felt all around the system. Estuarine gradients might, at some extent, have been responsible. For instance, after the 2001 floods, the macroinvertebrate communities of the downstream parts of the estuary South arm showed a decrease in ecological condition not associated with any specific anthropogenic intervention. These communities might have been more disturbed by the unusually low salinities (Teixeira *et al*. 2007) than communities from other habitats of the estuary with higher salinity fluctuations. On the other hand, species' dynamics could act as a compensation mechanism, where marine species could be replaced during floods by species adapted to lower salinity ranges, and vice-versa (Attrill & Power 2000, Attrill & Rundle 2002). In fact, no particular quality variation was noticed comparatively to the previous condition in the remaining areas of the lower estuary after this flood event. Nonetheless, these events interfere with the dynamics of benthic communities and it has been suggested that the occurrence of such extreme climatic events might delay the recovery process of the benthic communities, felt namely on the secondary production levels (Dolbeth *et al*. 2007).

The second remark is that the indices showed different sensitivity to the effects of such a flood event. During 2002, the indices detected unbalanced communities in the downstream stretch of the South arm, resulting mainly from a massive recruitment of bivalves that occurred after the 2001 floods, which altered the physical-chemical conditions in that area (Teixeira *et al*. 2007). Shannon-Wiener index, extremely sensitive to equitability within a sample, was particularly affected by this occurrence.

## Conclusions

The BAT was sensitive to the anthropogenic disturbances described for this estuary. The decline of benthic community's ecological condition, caused by eutrophication and dredging activities, and its recovery, after impacts cessation, could be detected by the method. Some resolution limits were though detected on the tool performance, mostly due to the fact that its final assessment is more influenced by the abundance component than by the taxonomic composition. The AMBI, index that translates community composition within the BAT method, showed little variation along the study period essentially due to the dominance of tolerant species within estuarine communities and to the ecological classification of key species by the index. Suggestions to adjust the AMBI to TW were presented, namely by accounting for the balance of ecological groups expected within estuarine systems and reviewing species classification.

The ecological assessment however cannot be optimized without reviewing also the concept of EQS, which needs careful adaptation to TW specificities prior to the intercalibration exercise. Our results indicate, as in Lavesque *et al*. (2009), that for estuarine systems the complete range of ecological quality as defined within EQS by the expected proportions of sensitive/opportunist taxa (EC 2000, annex V) should be reviewed.

Finally, the extent to which extreme climatic events affect the assessments of the natural systems demands further investigation, namely in understanding how communities under different degrees of disturbance change in response to such shifts of natural conditions. Clarify the power of the tools to discriminate between changes in ecological quality driven by natural or anthropogenic disturbance is critical for an efficient management.

# Chapter IV

## Calibration and validation of the AZTI's Marine Biotic Index (AMBI) for Southern California marine bays

**Abstract** Benthic indices are sensitive and accurate indicators of marine and estuarine sediment condition. However, they are often difficult to calibrate because of the lack of large datasets from a wide range of quality conditions. In a test of an alternative with less onerous habitat-specific data needs, the AZTI's Marine Biotic Index (AMBI) was calibrated and validated using data from Southern California marine bays. The index was applied in three different ways regarding the criteria used to classify local species sensitivity or tolerance to disturbance: 1) the AMBI using the original available list of species' classifications could not classify the majority of the samples, requiring calibration for the new region; 2) alternatively, using in addition classifications of taxonomic related taxa increased considerably its percentage of classified samples; and 3) finally, using local expertise to classify local species allowed for the assessment of almost all dataset. Using the two last criteria, AMBI was evaluated against the local Benthic Response Index (BRI), using the same data set; and also against Expert Judgement, using a 21 samples validation dataset covering a full gradient of disturbance. For the main dataset (685 samples) the best correlation (r = 0.70) with the local BRI was obtained for AMBI based on a mixture of local and previous expertise regarding species classification, and including a weighting factor for abundance data, proving a Moderate agreement in samples categorization. As for the AMBI assessments across the gradient of disturbance, they were strongly correlated with Expert Judgement, showing Very Good agreement for the samples classification. The study revealed however a tendency for AMBI to classify samples in a 'Marginal deviation from reference' condition, being less discriminatory than Expert Judgement. This could be partially overcome by adjusting the quality categories thresholds for the new habitat. Still, a significant part of the disagreements between the indices resulted from the approaches to classify species according to their ecological strategy. Research on critical taxa may contribute to clarify whether or not species ecological behaviour is geographically / habitat dependent and introduce further improvements on indices' performance.

**Keywords**

AZTI's Marine Biotic Index - AMBI
Benthic Response Index - BRI
Macroinvertebrates
Disturbance
Sediment quality assessment

## Introduction

Benthic macrofauna have been used extensively over the past three decades to assess environmental impacts from discharges and outfalls at small spatial scales (e.g., Dauer *et al.* 1979, Dauer & Conner 1980, Stull *et al.* 1986, Tapp *et al.* 1993, Stull 1995, Diaz *et al.* 2004, Hall *et al.* 2005). Over the last decade, assessment science has progressed from local impact assessments to regional assessments of marine and estuarine sediment quality based on benthic macrofaunal community condition (Bergen *et al.* 1998, Paul *et al.* 2001, Hyland *et al.* 2003, Kiddon *et al.* 2003, Ranasinghe *et al.* 2009).

Several benthic index approaches developed to assess benthic community condition in relation to regional reference conditions have proved to be reliable and sensitive indicators of the condition of marine and estuarine sediments (Diaz *et al.* 2004, Borja *et al.* 2008b, 2009b, Marques *et al.* 2009a). Different index approaches have been calibrated and validated in different habitats and, at present, there is no scientific consensus on one or more preferred index approaches (Labrune *et al.* 2006, Quintino *et al.* 2006, Borja & Dauer 2008, Borja *et al.* 2008b, 2009a, Ranasinghe *et al.* 2009). The challenge is now to accomplish sufficient index performance comparisons to reach such a consensus.

Frequently, one of the primary impediments to benthic index development is the need for large numbers of habitat specific data covering the entire disturbance gradient to achieve accurate index calibration. Benthic indices have been successfully developed in many areas where such data exist. However, often there is a need to develop indices and evaluate habitats which have no history of widespread and intensive benthic sampling, and which, therefore, lack the large numbers of data needed for index development.

In this context, a widely applied index has been the AZTI's Marine Biotic Index (AMBI) (Borja *et al.* 2000), which theoretically is a suitable approach for worldwide assessments, and for which local calibration is much less data dependent than most of the available benthic indices. The question is to evaluate to what extent it can be straightforward applied and how less demanding is its calibration in new geographical areas.

In southern California, the Benthic Response Index (BRI, Smith *et al.* 2001, 2003, Ranasinghe *et al.* 2004, 2009) was successfully developed for the mainland shelf and subsequently extended into bays and estuaries. BRI development involves calculating tolerance scores for species regarding pollution or disturbance, using a multivariate approach. The number of species for which tolerance scores can be calculated and the reliability of the pollution tolerance scores increases with the number of samples from a particular habitat available for BRI development, as well as the extent to which the samples reflect the entire disturbance or pollution gradient. Sample condition is assessed as the abundance-weighted average of tolerance scores for species occurring in the samples. To calibrate the index for new habitats and develop accurate species tolerance scores, it is necessary to use large amounts of data from temporal and spatial gradients of disturbance, pollution, and chemical exposure (Smith *et al.* 2003).

On the other hand, the AZTI's Marine Biotic Index (AMBI) (Borja *et al.* 2000) also evaluates benthic community condition based on the sensitivity and tolerance to pollution and disturbance of species found in samples, but the approach is quite different from BRI, and also less data demanding. AMBI assigns benthic species and higher taxa universally to five ecological groups, based on their sensitivity or tolerance to environmental stress and disturbance. Assuming that ecological group assignments are universal, AMBI can be applied in habitats where few benthic data are available, provided that previous ecological group assignments are available for the organisms recorded. AMBI has been widely used in European estuarine and coastal habitats from the Northern Sea to the Mediterranean (Borja *et al.* 2003, 2009b, Salas *et al.* 2004, Muxika *et al.* 2005, Carvalho *et al.* 2006, Grémare *et al.* 2009) and has had successful adaptations to other geographic regions in North and South America, Greenland, North Africa, Southeast Asia or Southwest Indian Ocean (Cai *et al.* 2003, Muniz *et al.* 2005, Afli *et al.* 2008, Bigot *et al.* 2008, Borja *et al.* 2008b, Callier *et al.* 2008, Josefson *et al.* 2008, Bakalem *et al.* 2009, Borja & Tunberg *in press*). The distribution of the percentages of abundances of the five ecological groups in a sample is the basis to estimate AMBI's values, on which communities' quality classifications are based (Borja *et al.* 2000).

The main objective of the present study was to apply the AMBI in Southern California marine bays, evaluating its performance as an alternative to the local BRI. The success of AMBI's application in the new geographical area was assessed by comparison with the BRI assessments and also against the professional judgement of nine benthic ecology experts (Weisberg *et al.*, 2008). The use of expert opinion provides a benchmark to assess index performance and the agreement among experts presents new opportunities for index validation (Weisberg *et al.* 2008, Muxika *et al.* 2007, Teixeira *et al.* 2010).

## Methodology

Since AMBI's adaptation to new geographical areas relies mostly on benthic ecology expertise to classify local species into five ecological groups (Borja & Muxika 2005), different alternatives to apply the index in this dataset allowed testing the effect of such procedure in the index performance. AMBI was tested following a stepwise procedure based on three different versions of the index with respect to ecological group classifications: (a) RUN1 applied the species classifications available for the AMBI in December 2007 (list available by request to the authors), which allowed analysing how much of the local taxa were already addressed by the index; (b) RUN2, supplemented species classifications in RUN1 by applying author's guidelines to local unclassified species, which allowed using knowledge on close related taxa to classify the majority of Southern California bays' taxa; and (c) RUN3, classifications for all the species were provided by local experts, based on their knowledge of species ecology, which allowed testing the effect that expertise worldwide could have on the quality assessment.

### *Index description and calibration procedure*

*AZTI's Marine Biotic Index (AMBI)* The AMBI is based on an *a priori* classification of macroinvertebrate species according to their sensitivity to disturbance. To adapt this index to Southern California infaunal macroinvertebrate communities, six benthic ecologists with taxonomic expertise and familiarized with the local coastal habitats classified taxa from the study dataset according to their tolerance to disturbance. Following AMBI authors' criteria for species classification (Borja *et al.* 2000, 2008b), they were assigned to one of five ecological groups (EG), as summarized by Grall and Glémarec (1997):

- Group I (EGI), species very sensitive to organic enrichment and disturbance, present under unpolluted conditions. They include specialist carnivores, some deposit-feeding tubicolous polychaetes, and species structuring communities. Normally they have long-life cycles;
- Group II (EGII), species indifferent to enrichment or disturbance, always present in low densities with non-significant variations with time. These include suspension feeders, less selective carnivores and scavengers;
- Group III (EGIII), species tolerant to excess organic matter enrichment, that may occur under normal conditions, but their populations are stimulated by organic enrichment. They are surface deposit-feeding species, as tubicolous spionids;

- Group IV (EGIV), second-order opportunistic species. They are mainly small sized polychaetes: subsurface deposit-feeders, such as cirratulids;
- Group V (EGV), first-order opportunistic species, able to resist high disturbance. These are deposit-feeders, which proliferate in reduced sediments.

Species final classification gathered the consensus of all six experts. The list based on local experts consensus used to assign Southern California taxa to an EG in AMBI is available in electronic format as Appendix A.

The index produces a final score in a continuous scale from 1 to 6 (7 in azoic sediments) across which five categories reflect benthic community health status (Borja *et al.* 2000): 'Undisturbed' (<1.2), 'Slightly disturbed' (1.2-3.3), 'Moderately disturbed' (3.3-5), 'Heavily disturbed' (5-6) and 'Extremely disturbed' (>6). In this paper, the last two classes were merged to allow comparison with the four classes used by Expert Judgement (Weisberg *et al.* 2008). AMBI is described in detail in Borja *et al.* (2000). The index was applied using the software version 4.1 (available at http://ambi.azti.es) and following authors' general guidelines (Borja & Muxika 2005), except when mentioned otherwise.

*Benthic Response Index (BRI)* BRI values, which are abundance weighted average tolerance scores for species in a sample, and respective condition categories were calculated, as described in Ranasinghe *et al.* (2009), using tolerance scores for taxa occurring in two or more samples in their calibration data. This included 373 of the 928 taxa encountered in this study. BRI values range from 0 to 100, from less to more disturbed infaunal communities. Across that range, four condition categories ('Reference/Undisturbed', 'Low disturbance', 'Moderate disturbance' and 'High disturbance') were established based on loss of 25%, 75%, and 95% of the reference species pool (Ranasinghe *et al.* 2009). BRI condition categories correspond to the Expert Judgement categories of Weisberg *et al.* (2008).

*AMBI calibration for Southern California marine bays*

To evaluate AMBI suitability to the new habitats, the index was applied on the dataset in three distinct ways:

(i) An initial AMBI run (RUN1) was conducted using AMBI's list of species classification available at the time of study: species list of December 2007 (available at www.azti.es), which includes assignments of taxa from several regions in the world, namely Europe, North and South America and Southeast Asia. In this first run, any taxa in the dataset was only assigned an EG if that taxa was included in the species list of the AMBI software, otherwise it was 'not assigned';

(ii) A second run (RUN2) was performed using the same AMBI's species list, but this time when a taxa was not included in the list then authors' guidelines were followed (as in Borja *et al.* 2008b): when the same genus was present in the list, and all species were assigned to the same group, the species was assigned to that group; when not the species was assigned to a higher taxonomic level present on AMBI's list;

(iii) Finally, a third run of AMBI (RUN3) was conducted using local expertise on the classification of species ecological behaviour (Appendix A species list).

Before calculating AMBI, the abundance data matrix was truncated and 70 taxa above genus level were ignored, with the exception of the following taxonomic groups: Actiniaria, Cerianthiaria, Chironomidae, Diptera, Dolichopodidae, Halcampidae, Hirudinea, Lineidae, Nemertea, Oligochaeta, Pennatulacea, Phoronida, Runcinidae, Sipuncula Sipunculidae, Tubulanidae, Tubulariidae and Turbellaria. Local experts, in their classification (used in RUN3), removed additionally 64 taxa which they considered mainly out of habitat. Overall, dropped taxa represented less than 1.5 % of the total abundance in the dataset.

The three different runs of AMBI were compared with each other for the percentage of samples classified and for the number of taxa assigned to an EG. Additionally, scores and classifications of samples for the distinct AMBI RUNs were also compared, using Spearman rank correlations and Kappa analysis, to test the influence of different criteria in the index output. Samples with more than 20% of the individuals not assigned to an EG were not included in any of the subsequent analyses, following AMBI guidelines (Borja & Muxika 2005).

*Index validation*

The AMBI performance on Southern California marine bays was evaluated by comparison with the BRI and with Expert Judgement assessments in two ways: first, taking the indices scores and experts' ranking of samples and verifying how well they correlated; and second, evaluating the level of agreement between the approaches regarding the categorization of samples.

The AMBI and BRI were applied to 685 samples collected in Southern California marine bays by eight sampling programs from 1998 to 2005 (Table I). At each sampling site, sediments were collected with a $0.1m^2$ Van Veen grab and sieved through 1 mm mesh screen. Materials retained on the screen were placed in a relaxant solution of 1 kg $MgSO_4$ or 30 ml propylene phenoxytol *per* 20 L of seawater for at least 30 minutes, and then fixed in buffered 10% formalin. In the laboratory, organisms retained on the screens were sorted from debris, identified to the lowest practical taxon (most often species), and counted.

**Table I.** Data sources from Southern California marine bays, used in the present study.

| Program | Location | Period | Samples |
|---|---|---|---|
| Bight'98 | Southern California marine bays | 1998 | 113 |
| Marina Del Rey | Marina Del Rey | 1998-2003 | 60 |
| WEMAP'99 | Southern California marine bays | 1999 | 24 |
| SPAWAR | Chollas and Paleta Creeks, San Diego Bay | 2001-2002 | 61 |
| NAASCO | San Diego Bay | 2001 | 175 |
| Huntington Harbour | Los Alamitos Bay, Huntington Harbour | 2001-2003 | 118 |
| Bight'03 | Southern California marine bays | 2003 | 119 |
| WEMAP'05 | Southern California marine bays | 2005 | 15 |
| **Total** | | | **685** |

A subset of 21 of those samples, covering the entire gradient of disturbance, had been previously evaluated by nine benthic ecologists using professional judgment Weisberg *et al.* (2008). The 21 samples subset was comprised in the 24 southern California marine bay samples of Weisberg *et al.* (2008) except for samples No.12, No.23 and No.30; two that the experts split approximately equally on 'good' vs. 'bad' sample status, and one with only one species present. The experts ranked the 21 samples from least to most disturbed and classified them in four categories based on narrative descriptions of their ecological condition: (1) ‘Unaffected’, a community that would occur at a reference site for that habitat; (2) ‘Marginal deviation from reference’, a community that exhibits some indication of stress, but is within measurement variability of reference condition; (3) ‘Affected’, a community that exhibits clear evidence of physical, chemical, natural, or anthropogenic stress; (4) ‘Severely affected’, a community exhibiting a high magnitude of stress. Affected and severely affected communities show clear evidence of anthropogenic or natural disturbance, while unaffected and marginal communities do not. The same scale was used as a benchmark to validate AMBI in the new geography and compare AMBI and BRI performances against Expert Judgment.

*Correlation analyses between scores assessments*

Spearman rank correlation analysis was used to compare the assessments between the distinct AMBI runs and the BRI, using indices’ scores. To further validate AMBI for the new geography and intercalibrate the two indices, each index was then compared with Expert Judgement. Spearman rank correlation analysis was used to compare indices’ scoring to the experts’ mean ranks of the 21 samples subset.

*Kappa analyses between category assessments*

As both indices provide also categorical classifications, the level of agreement between such classifications was measured with Kappa analysis (Cohen 1960, Landis & Koch 1977). For the Kappa analysis, ‘Moderate’, ‘Good’, ‘Very good’, and ‘Almost perfect’ levels of agreement were established using the equivalence table of Monserud and Leemans (1992). Fleiss-Cohen weights were used in computing the kappa statistic (Fleiss & Cohen 1973), where greater weight is given to those disagreements that are close (*e.g.*, between ‘Unaffected’ and ‘Marginal deviation from reference’, or ‘Affected’ and ‘Severely affected’) and less weight is given to disagreements between further categories (*e.g*. between ‘Unaffected’ and ‘Affected’, or ‘Unaffected’ and ‘Severely affected’).

Using the entire dataset, the AMBI category assessments from each RUN were compared with the BRI to evaluate the level of agreement between the two indices when using different criteria to classify species. Category classifications of both indices were additionally compared with Expert Judgment classifications, using the 21 samples subset.

The agreement between the approaches when identifying samples as undisturbed ('Unaffected' and 'Marginal deviation from reference') or disturbed ('Affected' and 'Severely affected') was also evaluated by contrasting the percentage of samples in each group.

*Abundance data transformation*

The BRI takes weighted individuals' abundance while the AMBI uses raw abundance data. For the Southern California marine bays, a fourth root transformation of the abundance data was found to provide the best correlation of the BRI with a disturbance gradient (Ranasinghe *et al.* 2009), and hence was adopted for applying the index in such habitats. To allow evaluating AMBI's response in the same conditions as the BRI, the AMBI was also tested using the abundance data similarly transformed. This weighting effect was tested using AMBI RUN2 and RUN3, which were then compared to the BRI following the same statistical analyses described above.

Due to this data transformation, AMBI categories' boundaries were adjusted according to predicted values from the fitted model (cubic regression) between abundance data not transformed and fourth root transformed. Adjusted boundaries became: 'Undisturbed' (<1.75), 'Slightly disturbed' (1.75-2.58), 'Moderately disturbed' (2.58-3.7), 'Heavily disturbed' (3.7-4.94) and 'Extremely disturbed' (>4.94).

*Comparison between the two methods for species ecological classification*

Since both indices rely on the classification of species regarding their tolerance to disturbance, the two types of information, AMBI EG assignments and BRI tolerance scores, were compared to verify if there were differences between the classifications of species resultant from the two methods.

The agreement between the approaches regarding the ecological classification of species was evaluated by taking 342 species of the 373 possessing tolerant scores and their correspondent EG classification by local experts. Cumulative distribution frequency curves (CDFs) were used to show the range of distribution of the tolerance scores along the five EG of AMBI. The Kolmogorov-Smirnov test was used to compare the CDFs.

Additionally, to allow an approximate correspondence between the two methods of classifying species, equivalence between the AMBI's Ecological Groups (EG) and the BRI tolerance scores classifications was attempted by selecting a set of thresholds to create BRI tolerance categories that would result in the maximum possible weighted kappa statistic as measured between AMBI and BRI categorizations. Because the number of taxa in EG categories IV and V were low, they were merged into a single category. Standard linear weights for the kappa statistic were used (Cicchetti & Allison 1971) to allow greater flexibility for categorizations when optimizing partial agreement. The 'optimal' set of BRI tolerance thresholds was selected by computing weighted kappa statistics for a large set of possible candidates. These candidates were selected by choosing all permutations of three thresholds, taken at 5% increments of the tolerance range. In addition, distances between individual thresholds within each set were constrained to be no less than 10% of the range. These conditions ensured that optimization converged and resulting thresholds within a set were not too close to one another. The set of thresholds that yielded the largest weighted kappa value when compared with the EG categories (across all 294 taxa) were selected as optimal.

## Results

### *Influence of distinct criteria for applying the AMBI*

The potential of AMBI for classifying Southern California marine bays' samples differed depending on the criteria used to classify species into EG. Since only approximately 24% of the taxa in this data set were included in the original AMBI's list of species (Table II), using RUN 1 approach allowed classifying only 11% of the total number of samples with good confidence. This number increased substantially when AMBI authors' criteria were followed to assign species not in the AMBI list to an EG, resulting in the classification of 630 taxa. This allowed classifying 75% of the samples (Table II). When local expertise criteria were used to assign species to an EG almost all Southern California samples could be classified using the AMBI (Table II).

The distribution of taxa from the Southern California marine bays along the five EG was also different according to the criteria used to classify species (Table III). While according to the three criteria most of the species belong to EG I and II, in the local experts' assignments there is a much higher concentration in EG II, representing more than half of the taxa assigned.

Despite this, among the disagreements observed between AMBI authors and local experts', accounting for 56% of the taxa classified by both criteria (n = 552), no evidence was found for bias in the assignments of one criterion in relation to the other, i.e., one criteria systematically classifying

**Table II.** Summary of AMBI runs in the Southern California bays dataset: percentage of samples with more than 80% of its individuals assigned an ecological group; number (*n*) of taxa assigned an ecological group and ignored.

| AMBI | Samples classified | *n* taxa assigned | *n* taxa ignored |
|---|---|---|---|
| RUN1 (exact taxa) | 10.9 % | 219 | 70 |
| RUN2 (AMBI authors) | 75.0 % | 630 | 70 |
| RUN3 (local experts) | 98.0 % | 748 | 134 |
| | (Total *n* samples: 685) | (Total *n* taxa: 928) | |

species in worse or better EG than the other (species assignments for each criterion in Appendix A). Moreover, the disagreements found were not associated to any particular taxonomic group. For the main Phyla, the disagreements are in proportion with their frequency of occurrence: Annelida, accounting for 37% of the total taxa in the dataset, registered 40% of the disagreements found; Arthropoda, with 32% of the taxa, registered 25% of the disagreements; Mollusca, representing 23% of the total taxa, registered 26% of the disagreements; and the remaining Phyla, with 8% of the total taxa, registered 9% of the disagreements) (for complete species assignments see Appendix A).

**Table III.** Number (*n*) and percentage (%) of the total taxa in each ecological group (EG) according to the three criteria.

| | AMBI RUN1 (exact taxa) | | AMBI RUN2 (AMBI authors) | | AMBI RUN3 (local experts) | |
|---|---|---|---|---|---|---|
| **Ecological Groups** | *n* | % | *n* | % | *n* | % |
| EG I | 70 | 32.0 | 235 | 37.3 | 162 | 21.7 |
| EG II | 79 | 36.1 | 235 | 37.3 | 485 | 64.8 |
| EG III | 39 | 17.8 | 98 | 15.6 | 77 | 10.3 |
| EG IV | 28 | 12.8 | 56 | 8.9 | 19 | 2.5 |
| EGV | 3 | 1.4 | 6 | 1.0 | 5 | 0.7 |
| **Total *n* taxa assigned** | **219** | | **630** | | **748** | |

Despite the differences in the assignments, when the samples in RUN2 and RUN3 were compared (n = 501, samples with less than 20% of individuals not assigned an EG), it was observed that the two criteria produced quite correlated assessments (r = 0.63, $p \leq 0.001$), agreeing on the condition of approximately 84% of the samples. Nonetheless, for 14.5% of the samples the two AMBI runs disagreed on the undisturbed or disturbed condition of the samples, indicating therefore a 'Moderate' agreement (Kappa = 0.42, $p < 0.0001$) for the overall classification of samples.

### *Correlation between assessments*

AMBI's values for the three runs performed were all significantly ($p < 0.01$) correlated with the BRI scores (RUN1: r = 0.30, n = 75; RUN2: r = 0.27, n = 514; RUN3: r = 0.16, n = 670), however, the low coefficients observed reveal the weakness of this positive correlation. The correlation between the AMBI and BRI was weaker when using local experts' assignments (RUN3) than when using AMBI authors' assignments (RUN2).

Since RUN2 and RUN3 are the only criteria that allow for the classification of a meaningful proportion of the Southern Californian samples, these criteria were thoroughly analysed.

The use of fourth root transformed abundance data on the AMBI calculation significantly ($p < 0.0001$) improved its correlation with the BRI assessments (using RUN2: r = 0.53, n = 447; using RUN3: r = 0.64, n = 682).

When considering the subset of samples representing a well defined disturbance gradient, the AMBI and BRI assessments were very similar to Expert Judgement. Significant ($p \leq 0.0003$) strong positive correlations (AMBI RUN2: r = 0.81; AMBI RUN3: r = 0.93; BRI: r = 0.89) were found between the experts' mean ranking of the 21 samples and the scores of the two indices.

### *Level of agreement between category assessments*

The classification of samples into four categories ('Unaffected', 'Marginal deviation from reference', 'Affected', 'Severely affected') varied according to the index used (Table IV). A concentration of samples in the 'Marginal deviation from reference' category was evident for the AMBI. The two indices' categorizations for a sample matched for only 41.9% or 37.6% of the cases depending on which criteria for applying AMBI was considered (RUN2 or RUN3, respectively) (Table IV). The AMBI categorization presented a 'Low' agreement with that of the BRI as showed by the

Kappa analysis (considering RUN2: Kappa = 0.25, $p < 0.0001$, n = 514; or RUN3: Kappa = 0.25, $p < 0.0001$, n = 670).

The agreement between the indices classification of samples increased when undisturbed *vs.* disturbed samples was analysed, instead of considering the four categories. In this case the indices agreed in the condition of a sample in 67.2% (if RUN2) and 74.1% (if RUN3) of the cases.

Similarly to the observed for the indices' scores, the use of transformed abundance data on AMBI calculations improved significantly the agreement with the BRI category assessments (using RUN2: Kappa = 0.39 'Low', $p < 0.0001$, n = 447; using RUN3: Kappa = 0.43 'Moderate', $p < 0.0001$, n = 682).

**Table IV.** Frequency table for samples distribution by the four categories according to AMBI (RUN2: n = 514; and RUN3: n = 670) *vs*. the BRI: each cell shows the percentage of the total table it represents. The totals by category for each method are also presented.

| **Categories** | **BRI** | | | | |
|---|---|---|---|---|---|
| **AMBI RUN2** | Unaffected | Marginal deviation from reference | Affected | Severely affected | Row Totals |
| Unaffected | 0.6 % | 0 % | 0.8 % | 0.2 % | 1.6% |
| Marginal deviation from reference | 22.8 % | 32.3 % | 23.7 % | 2.1 % | 80.9% |
| Affected | 3.1 % | 2.7 % | 8.2 % | 2.1 % | 16.1% |
| Severely affected | 0 % | 0.2 % | 0.4 % | 0.8 % | 1.3% |
| Column totals | 26.5% | 35.2% | 33.1% | 5.3% | **100 %** |
| **AMBI RUN3** | | | | | |
| Unaffected | 0.3 % | 0 % | 0 % | 0.1 % | 0.4% |
| Marginal deviation from reference | 34.8 % | 34.0 % | 23.3 % | 1.8 % | 93.9% |
| Affected | 0.1 % | 0.5 % | 2.5 % | 1.2 % | 4.3% |
| Severely affected | 0 % | 0.1 % | 0.5 % | 0.8 % | 1.3% |
| Column totals | 35.2% | 34.6% | 26.3% | 3.9% | **100 %** |

Comparing the indices' classification of samples with Expert Judgment classification, it was observed that the BRI assessment produce the most similar output (Table V), with an 'Almost perfect' agreement between both classifications (Kappa = 0.89, $p$ < 0.0001). The AMBI runs presented higher discrepancies regarding the distribution of the 21 samples by the four classes (Table V), with kappa indicating a slightly lower agreement between AMBI approaches and Expert Judgement (Kappa = 0.71, 'Very good' agreement for RUN2; Kappa = 0.84, 'Very good' agreement for

**Table V.** Frequency table for the 21 samples distribution by the four categories according to Expert Judgement and the indices BRI and AMBI (RUN2 and RUN3): each cell shows the number of samples and the percentage of the total it represents. The totals by category for each method are also presented.

| **Categories** | **Expert Judgement** | | | | |
|---|---|---|---|---|---|
| **AMBI RUN2** | Unaffected | Marginal deviation from reference | Affected | Severely affected | Row Totals |
| Unaffected | 0 % | 0 % | 0 % | 0 % | 0 % |
| Marginal deviation from reference | 28.6 % | 14.3 % | 9.5 % | 0 % | 52.4 % |
| Affected | 0 % | 9.5 % | 14.3 % | 14.3 % | 38.1 % |
| Severely affected | 0 % | 0 % | 0 % | 9.5 % | 9.5 % |
| | | | | | **100 %** |
| **AMBI RUN3** | | | | | |
| Unaffected | 0 % | 0 % | 0 % | 0 % | 0 % |
| Marginal deviation from reference | 28.6 % | 23.8 % | 9.5 % | 0 % | 61.9 % |
| Affected | 0 % | 0 % | 14.3 % | 9.5 % | 23.8 % |
| Severely affected | 0 % | 0 % | 0 % | 14.3 % | 14.3 % |
| | | | | | **100 %** |
| **BRI** | | | | | |
| Unaffected | 23.8 % | 9.5 % | 0 % | 0 % | 33.3 % |
| Marginal deviation from reference | 4.8 % | 9.5 % | 0 % | 0 % | 14.3 % |
| Affected | 0 % | 4.8 % | 19.1 % | 14.3 % | 38.1 % |
| Severely affected | 0 % | 0 % | 4.8 % | 9.5 % | 14.3 % |
| Expert Judgement totals | 28.6 % | 23.8 % | 23.8 % | 23.8 % | **100 %** |

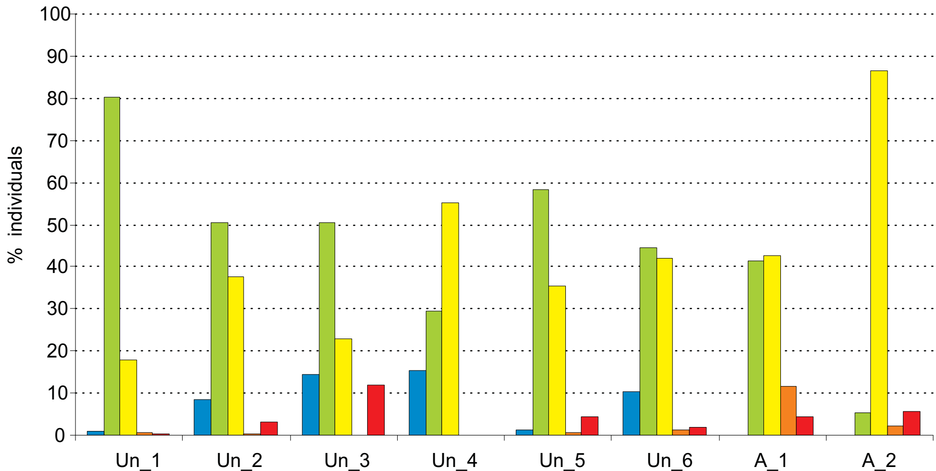
RUN3). Moreover, the AMBI, using either authors or local experts' assignments, could not identify any 'Unaffected' sample and classified more than half of the samples as 'Marginal deviation from reference' (Table V).

If the undisturbed or disturbed status of a sample was considered, instead of four categories, the agreement between both indices and Expert Judgment was achieved for more than 80% of the samples. The BRI agreed with Expert Judgment in 20 out of 21 samples, the AMBI RUN2 in 17 out of 21, and the AMBI RUN3 in 19 out of 21 (Table V).

### *AMBI mismatched assessments with Expert Judgement*

Given the concentration of samples in the 'Marginal deviation from reference' category resulting from the application of AMBI, the index was compared with Expert Judgment for the mismatched samples in the good side of the undisturbed *vs*. disturbed threshold. The AMBI classification based on local expertise assignments (RUN3) presented 13 samples classified as 'Marginal deviation from reference', eight of which Expert Judgment evaluated as either 'Unaffected' or 'Affected' (Table V). These eight mismatched samples were all essentially dominated or co-dominated by individuals of species classified in EG II and III of AMBI (Figure 1). However, the six 'Unaffected' samples according to Expert Judgement had all registered individuals from sensitive taxa *sensu* AMBI (EG I: between 1 and 15.4%), which were not present in the two 'Affected' samples. On the other hand, these two 'Affected' samples registered higher percentages of individuals classified as opportunistic species (EG IV and V: between 8 to 16%), while the 'Unaffected' ones either had no register or had lower percentages of opportunistic taxa (EG IV and V: between 0 and 12%).

The range of AMBI values for these Expert Judgement 'Unaffected' samples (from 1.8 to 2.2; Figure 1) was lower than the range of values observed for the Expert Judgement 'Marginal deviation from reference' samples (ranging from 2.204 to 3.153). To the other side of the gradient however, AMBI values for the Expert Judgement 'Affected' samples were within the range of values observed in the Expert Judgement 'Marginal deviation from reference' samples (2.7 and 3.1; Figure 1).




| | Un_1 | Un_2 | Un_3 | Un_4 | Un_5 | Un_6 | A_1 | A_2 |
|---|---|---|---|---|---|---|---|---|
| Expert rank | 1.9 | 2.7 | 3.4 | 3.9 | 5.4 | 5.5 | 12.9 | 15.1 |
| AMBI score | 1.8 | 2.1 | 2.2 | 2.1 | 2.2 | 2.1 | 2.7 | 3.1 |
| BRI score | 26.1 | 25.7 | 20.8 | 31.9 | 46.8 | 29.5 | 54.2 | 73.1 |

**Figure 1.** Distribution of individuals (%) along the five EG (EG I to V) of AMBI for the 'Marginal deviation from reference' samples of AMBI RUN3, classified as 'Unaffected' (Un) or 'Affected' (A) by Expert Judgement. For each sample the mean Expert rank, the AMBI RUN3 score and the BRI score are shown.

### *Comparison of the methods for species classification*

Taking the 342 species with both tolerance scores and EG assignment (local experts' EG assignments), it was observed that species in EG II presented the lowest mean tolerance scores, followed by those in EG I, and increasing consecutively until EG V (Figure 2).

It was also observed that the tolerance scores' ranges of the five groups overlapped greatly. The Kolmogorov-Smirnov test applied to the cumulative distribution curves (Figure 2) evidenced that the only distinguishable EG were I and IV; II and III; and II and IV (K-S statistic = 1.42, $p$ = 0.035; K-S statistic = 1.82, $p$ = 0.003; K-S statistic = 1.76, $p$ = 0.004, respectively). Hence, considering the EG classification of the local experts, the distribution of the tolerance scores along the five EG did not constitute distinct sets of values. The species classified as EG IV and V were however very few (12 and 3, respectively) comparatively to those in the remaining EG (EG I: 48, EG II: 222, EG III: 55), for which some caution is needed on the interpretation of the tendencies expressed for these two EG. The same previous pattern was observed when only the 193 species present in the 21 samples validation subset were used for comparison.

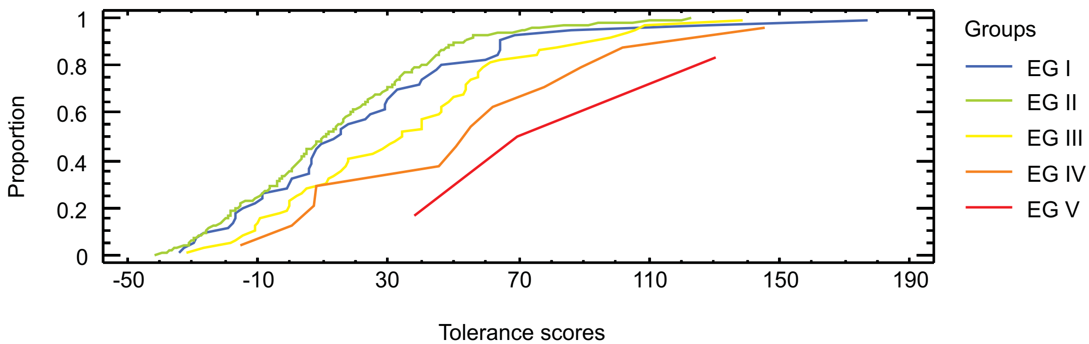



**Figure 2.** Relative cumulative distributions (CDFs) of BRI tolerance scores along the five EG (EG I to V) (local experts assignments) for 342 species in the dataset. The plot shows the proportions that fall below specified tolerance scores for each ecological group.

There were a great number of observations for which AMBI and BRI approaches were contradictory with regard to a species tolerance/sensitivity to disturbance (Table VI, Appendix A). The curves of tolerance scores distribution (Figure 2) show, for example, that the highest tolerance score in BRI (176.74) was actually registered for a species classified in AMBI EG I, the polychaete *Paraonella platybranchia*.

Taking the taxa representing approximately 80% of the total abundance in the dataset (n = 34), important mismatches were detected between species classifications in the two methods, involving approximately 40% of them (Table VI). Such correspondence between the two methods of classifying species was obtained by splitting the range of tolerance scores obtained for local species according to their distribution for the five EG. The following BRI tolerance thresholds were obtained, using Kappa optimization procedure: -41.7 < EG I < 8.8 > EG II < 48.1 > EG III < 66.8 > EG IV/V > 176.7, weighted kappa = 0.15 (poor agreement). The complete species list for California marine bays, as well as tolerance scores and EG assignments by each criterion are available in electronic format in Appendix A.

**Table VI.** Comparison of the AMBI Ecological Group (EG) classifications (by local experts as used in AMBI RUN3) and the BRI tolerance scores of the 34 taxa representing 80% of the total abundance in the data set. Species for which there is a clear mismatch between the classifications in bold. Total abundance of the taxa and percentage of samples on which it occurs are also presented. (n.a. – not assigned).

| Taxa | Total abund. (ind. m$^{-2}$) | Presence (% of samples) | AMBI EG (local experts) | BRI tolerance scores |
|---|---|---|---|---|
| ***Pseudopolydora paucibranchiata*** | 77992 | 74.6 | **III** | **81.7** |
| *Mediomastus* sp. | 53115 | 85.8 | III | 57.8 |
| *Kalliapseudes crassus* | 32578 | 6.4 | III | -- |
| *Exogone lourei* | 31208 | 63.8 | II | 41.9 |
| ***Euchone limnicola*** | 20569 | 58.1 | **III** | **76.3** |
| ***Musculista senhousia*** | 20186 | 57.2 | **II** | **68.0** |
| *Scoletoma* sp. A | 15337 | 44.5 | II | 19.7 |
| *Leitoscoloplos pugettensis* | 13511 | 84.2 | III | 64.4 |
| ***Grandidierella japonica*** | 13340 | 38.7 | **III** | **106.0** |
| ***Streblospio benedicti*** | 13167 | 13.3 | **IV** | **61.8** |
| *Capitella capitata* Cmplx | 12801 | 19.1 | V | 130.8 |
| Oligochaeta | 12066 | 41.9 | V | 70.0 |
| *Scoletoma* sp. | 11014 | 48.5 | II | -- |
| ***Theora lubrica*** | 10496 | 63.6 | **III** | **46.6** |
| *Prionospio (Prionospio) heterobranchia* | 10078 | 72.6 | II | 37.5 |
| *Pista percyi* | 9142 | 58.0 | II | 42.0 |
| ***Synaptotanais notabilis*** | 8209 | 30.7 | **III** | **75.3** |
| *Fabricinuda limnicola* | 7935 | 17.7 | II | 50.0 |
| *Cossura* sp. | 7785 | 23.8 | III | -- |
| *Monocorophium acherusicum* | 7678 | 9.5 | III | 61.7 |
| ***Amphideutopus oculatus*** | 7150 | 49.8 | **III** | **-0.4** |
| Phoronida | 5464 | 35.9 | n.a. | 99.3 |
| *Tagelus subteres* | 4738 | 35.5 | II | 37.3 |
| *Dorvillea (Schistomeringos)* sp. | 4663 | 40.9 | **V** | **38.1** |
| *Scyphoproctus oculatus* | 4562 | 15.9 | **III** | **25.4** |
| *Spiophanes duplex* | 4249 | 41.9 | **III** | **31.8** |
| *Neanthes acuminata* Cmplx | 3942 | 30.9 | III | 58.9 |
| *Dipolydora* sp. | 3779 | 11.5 | n.a. | 56.6 |
| *Euphilomedes carcharodonta* | 3228 | 47.2 | III | 57.2 |
| *Polyophthalmus pictus* | 3007 | 6.1 | I | 0.9 |
| *Diplocirrus* sp. SD1 | 2921 | 33.7 | **III** | **46.6** |
| ***Armandia brevis*** | 2822 | 26.3 | **IV** | **7.1** |
| *Podocerus fulanus* | 2563 | 23.4 | II | 21.6 |
| ***Acteocina inculta*** | 2516 | 27.0 | **II** | **110.1** |

### *AMBI final run calibrated for Southern California Marine bays*

A final revision of all species assignments resulted on a final EG acceptance that reflects a compromise between local experts' opinion and AMBI authors' previous species assignments (Appendix A). Such contribution is available through the AMBI most recently revised species list

(AMBI software web page (http://ambi.azti.es), in the global species list of the index). This final AMBI run (Table VII) outperformed previous runs when compared to the local index, especially if adding a weighting factor to the abundance data. It did not contribute to an improvement of the agreement already observed with Expert Judgement though; being slightly lower than RUN3 for the ranking and classification samples.

**Table VII.** Comparison of the AMBI Final RUN for the Southern California marine bays dataset with the BRI and Expert Judgement. The table shows Spearman rank correlations (r) between the indices scores and with Expert Judgement ranks, and Kappa agreement (Kappa) for the categories; *p*-values < 0.05 indicate statistically significant non-zero correlations at the 95% confidence level and 'n' the number of pairs of data values used to compute each coefficient.

| | **AMBI Final RUN** | |
|---|---|---|
| | Abundance data not transformed | Abundance data $4^{th}$ root transformed |
| ***vs*. BRI** | **Indices scores** | |
| | r = 0.33<br>*p* < 0.0001; n=670 | r = 0.70<br>*p* < 0.0001; n=684 |
| | **Quality categories** | |
| | Kappa = 0.32<br>*p* < 0.0001; n=670<br>'Low' agreement | Kappa = 0.55<br>*p* < 0.0001; n=684<br>'Moderate' agreement |
| ***vs*. Expert Judgement** | **Indices scores** | |
| | r = 0.88<br>*p* < 0.0001; n=21 | r = 0.77<br>*p* < 0.0005; n=21 |
| | **Quality categories** | |
| | Kappa = 0.71<br>*p* < 0.001; n=21<br>'Very Good' agreement | Kappa = 0.66<br>*p* < 0.01; n=21<br>'Good' agreement |

## Discussion

Calibration and validation are fundamental steps for the successful adaptation of ecological indices to new habitats and geographies. Regarding AMBI, there are already several examples of its adaptation to many regions worldwide; however, in none of them a complete review of the local species ecological assignments was done, as the one carried out by local benthic ecologists for the Southern California marine bays dataset. Since this index is based on fundamental ecological characteristics of benthic species, it is thus important to understand how the species classification procedure interferes with the index performance in a new region.

Overall, AMBI performed well when applied to Southern California marine bays datasets. Nevertheless, the comparison with the local BRI and its validation against Expert Judgement showed that its performance could be improved namely by: a) adjusting the index thresholds for the quality categories; b) changing data format in the index algorithm; and c) checking critical species ecological classifications. Each point is thoroughly discussed.

### *AMBI performance in Southern California Marine Bays*

This study showed that the AMBI could not be applied in the new region without previous local calibration. After calibration, though, the AMBI could be successfully validated against Expert Judgement regarding the ranking of samples covering a full gradient of disturbance. The index exhibited however less discriminatory power than Expert Judgement regarding the categorization of samples, with a weaker performance on the good end of the disturbance gradient. This aspect could have contributed to the higher divergences observed with the local BRI assessments for the bulk of samples in the new geography. Unlike the balanced distribution of samples across the gradient of disturbance in the Expert Judgement validation subset, the entire dataset was an uneven and unpredictable representation of the disturbance gradient. And according to the local index, over 2/3 of the samples were classified as 'Unaffected' or 'Marginal deviation from reference', which given such a distribution of samples could have contributed to the lower agreement observed.

Results evidenced that the default AMBI quality thresholds (Borja *et al.* 2000) used in this study, need also to be habitat adjusted to match real ecological expectations. Ecological assignments constraints apart, both the taxonomic composition of the mismatched samples and the range of observed AMBI values (Figure 1) suggest that the classification could be substantially improved by adjusting those boundaries at least in the good side of the gradient. Other studies in transitional habitats have also evidenced that the natural dominance of tolerant (EG III) individuals in such samples prevented AMBI to reach an 'Unaffected' condition, with values < 1.2 (Borja *et al.* 2008b, de Paz *et al.* 2008, Teixeira *et al.* 2008b). Instead of redefining the boundaries of the categories for the index, AMBI's authors suggested that an alternative way to overcome this would be the complementary use of several indices, as in M-AMBI (Muxika *et al.* 2007), and setting the reference conditions for each index according to the expectations for specific habitats, as already done in other parts of USA, such as in the Chesapeake Bay (Borja *et al.* 2008b) or in a Florida estuary and lagoon (Borja & Tunberg *in press*). The quality thresholds would then be defined on the final score for the samples, after a multivariate procedure, reflecting a departure from those reference conditions.

Another important difference between the indices studied is that, while both depend only on the relative abundance of species for which EG assignments or pollution tolerance values are available, BRI uses a factor to transform the abundance weights and avoids overemphasis on one or a few relatively high abundance values (Smith *et al.* 2001, 2003). AMBI is therefore much more sensitive to the dominance of a few species, making it less effective in habitats with influence from adjacent transitional habitats where such patterns are more typical. Ultimately, the transformation applied to the abundance data altered the relative contribution of the EG within AMBI and greatly improved its agreement with BRI, as well as with Expert Judgement. Such improvement, observed for both scores correlation and the classification of samples, was too significant to neglect this transformation step in future AMBI applications for this region. Warwick *et al.* (2010) have also recently tested several transformations for AMBI, with evidences that the impact 'signal' is better captured with modestly-transformed data. Such findings apply to fully marine habitats, for which Smith *et al.* (2001; Appendix A) also observed that less severe transformations were necessary for the BRI to reproduce more accurately the identified pollution gradients. Nevertheless, the fourth root transformation was adopted to compare the two indices in this study because it gave the highest correlation between the BRI values and the disturbance vector for this specific habitat during a previous calibration of this index for Southern California marine bays (Ranasinghe *et al.* 2009).

### *Species ecological classification*

The results showed clearly that the disparity in the ecological assessments between the two indices' can largely be attributed to the differences in species ecological classification (Appendix A). This is important since species classification constituted the basis of the calibration of AMBI to the new ecoregion. The discrepancy within AMBI results, when criteria to classify species within AMBI itself are changed, reveals more accurately the extent to which this procedure interferes with the assessments. Differences on 615 taxa, either on each EG assignment, or not assigning or ignoring the taxa by one of the criteria (details in Appendix A) led to a lower correlation between the runs than the expected when applying exactly the same index to the data. And despite most of the samples maintained their ecological quality classification, 15% of them had their status changed from a good to bad condition or vice-versa due to the classification procedure. AMBI is based on fundamental ecological characteristics of benthic species and despite the theoretical considerations on species ecological behaviour are well understood, the procedure to define sensitivity/tolerance level of a species (Borja *et al.* 2000) is mostly expert knowledge based and not entirely objective. Moreover, one of the AMBI assumptions is that species ecological behaviour is intrinsic and is maintained irrespective of habitat or geography. Therefore each species is bonded to a single ecological group (EG) classification worldwide. However, for 308 species the local experts do not seem to agree with previous AMBI classifications. Such mismatch is either consequence of distinct interpretation of the EG concept or results from distinct knowledge on species behaviour in the new geography. For the 42 taxa (Appendix A) where the disparity goes beyond one level of EG, this might just be the case.

On the other hand, it would be expected that the use of local expertise to classify species would result in a higher agreement between AMBI and BRI. However, it is interesting to notice that there are still important local disagreements on species ecological classification between the two methods, EG assignments and tolerance scores (Figure 2). Although the approach undertaken to establish the correspondence between AMBI EG assignments (local experts' criteria) and BRI tolerance scores is merely a guiding tool, it was observed that: a) overall, the two methods result in very different sets of species; observing the range of species tolerance scores across all EG there is not a significant differentiation of all five groups (Figure 2); b) second, there were evidences that some EG definitions might be ambiguous; species classified in EG II by local experts corresponded to species that presented the lowest mean tolerance scores, which according to the BRI are therefore more associated with less polluted stations than those that experts actually classified in EG I (Figure 2); c) and finally, results further suggest that the present knowledge on species ecological strategies might be insufficient, since for many species the local experts' perception of their ecological behaviour is actually very distinct from what the local BRI tolerance scores point out.

Unlike AMBI, BRI assumes that species behaviour is habitat dependent, with the numbers and kinds of benthic animals that occur in reference areas varying naturally by habitat (Smith *et al.*, 2003). Therefore different pollution tolerance values are empirically developed for individual species for each habitat according to the distribution of species along known gradients of disturbance (Ranasinghe *et al.* 2009). One of the aims of a pan-European scale study (Grémare *et al.* 2009) was to test the validity of the use of a single list of sensitivity/tolerance levels by comparing BQI $E(S_{50})_{0.05}$ between subareas, covering both marine and estuarine habitats. Corroborating the Smith *et al.*

(2001) studies, they too gathered strong evidence that the species sensitivity/tolerance levels presented marked changes according to geographical location. This questions the universal principle of species ecological behaviour beneath AMBI and helps elucidate findings of the study, namely of species for which discrepant classifications were found across methods and criteria to classify them (Appendix A).

## Conclusions

After calibration, the adaptation of AMBI to Southern California marine bays resulted on assessments comparable to those of the test approaches used: a local validated index and an Expert Judgement based validation scale. The agreement achieved for this dataset allows the use of AMBI (final RUN) in the new geography with a classification associated error of approximately 19% of the samples, according to the local index and BPJ validation.

However, findings from this study may improve AMBI's performance, not only for the new geography but also worldwide. The two indices approaches for classifying species according to their sensitivity/tolerance to disturbance presented obvious inconsistencies that go beyond methodological subtleties. This reinforces that species ecological strategies might be geographically/habitat dependent, feeding the controversy around the plasticity or not of species ecological behaviour. The critical species should obviously be object of study in the future, not only to clarify their sensitivity/tolerance level and allow for more robust ecological assessments, but to better understand which factors interfere with species ecology.

# Chapter V

## Assessing coastal benthic macrofauna community condition using best professional judgement - developing consensus across North America and Europe

**Abstract** Benthic indices are typically developed independently by habitat, making their incorporation into large geographic scale assessments potentially problematic because of scaling inequities. A potential solution is to establish common scaling using expert best professional judgment (BPJ). To test if experts from different geographies agree on condition assessment, sixteen experts from four regions in USA and Europe were provided species-abundance data for twelve sites per region. They ranked samples from best to worst condition and classified samples into four categories. Site rankings were highly correlated among experts, regardless of whether they were assessing samples from their home region. There was also good agreement on condition category, though agreement was better for samples at extremes of the disturbance gradient. The absence of regional bias suggests that expert judgment is a viable means for establishing a uniform scale to calibrate indices consistently across geographic regions.

**Keywords**

Best professional judgment
Coastal benthic infauna
Anthropogenic disturbance
Quality assessment
North America
Europe

## Introduction

Benthic invertebrate community condition has been used worldwide to assess the effects of many impacts, including physical disturbance, organic loading and chemical contamination (Pearson & Rosenberg 1978, Dauer *et al.* 2000, Borja *et al.* 2000, 2003, Muxika *et al.* 2005). These assessments often use benthic indices, which translate community composition into a quality classification (Weisberg *et al.* 1997, van Dolah *et al.* 1999, Borja *et al.* 2000, 2004b, Rosenberg *et al.* 2004, Dauvin & Ruellet 2007, Dauvin *et al.* 2007, Muxika *et al.* 2007, Weisberg *et al.* 2008, Ranasinghe *et al.* 2009). Benthic indices have proved to be accurate and sensitive indicators of the condition of the sediments in which benthos live (Diaz *et al.* 2004, Pinto *et al.* 2009).

Using benthic indices for assessment over large geographic areas is problematic, though, because benthic indices are usually developed within specific habitats and ecoregions (Borja & Dauer 2008). Benthic species composition varies naturally across habitat gradients and expectations for reference conditions vary accordingly (de Paz *et al.* 2008, Borja *et al.* 2009a). Consequently, there is no certainty that indices developed in different regions or habitats are assessing biological condition on the same scale. Interpreting different benthic indices developed for different habitats to yield a common assessment for management purposes is further complicated when the indices are based on different combinations of metrics (Diaz *et al.* 2004, Borja *et al.* 2009a,c).

One potential solution is to apply expert best professional judgment (BPJ) to establish a set of samples across regions that provide a uniform scale for calibrating any index, but this assumes there is consensus about benthic community condition classifications among experts across regions. Weisberg *et al.* (2008) found a high level of agreement in expert BPJ in a benthic quality assessment for two United States west coast habitats, but that assessment was limited to experts from within the region making an assessment of biota with which they had great familiarity. Agreement in benthic condition assessments of experts with varying familiarity with resident benthic fauna would be necessary for establishment of a credible scale applicable across broader geographic regions.

Here, we evaluate the level of agreement among experts using BPJ to assess the condition of benthic communities from four widely separated geographic regions. Our objectives were to evaluate whether (1) BPJ assessments were independent of the home regions of the experts, and (2) whether the level of agreement among expert BPJ was sufficient to establish a universal benthic assessment scale for the four regions that could be used to intercalibrate benthic indices and assessment methodologies across habitat boundaries.

## Methodology

Sixteen benthic experts from four geographic regions were provided species-abundance data for twelve sites from each region and asked to determine the condition of the benthos at each site. The four regions included the West (W) and East (E) coasts of the United States (US), and the Atlantic (A) and Mediterranean (M) coasts of Europe. Of the 16 benthic ecologists, nine were from academic institutions, four from municipalities that implement benthic monitoring programs to assess the effect of discharge outfalls, two from non-profit research organizations, and one from a private consulting firm. Their experience in benthic monitoring ranged from 16 to 38 years. Each benthic ecologist was provided species abundance data for each sample and limited habitat data (region, salinity, depth, and percent fines as a measure of sediment grain size) sufficient to establish an expectation for what kinds of organisms should occur there under undisturbed conditions.

The experts were asked to rank the relative condition of the sites from 'best' to 'worst' within each region as well as across all four regions. 'Best' means least likely to have been disturbed while 'worst' means most likely to have been subjected to disturbance, with ties designated as liberally as each expert desired. The experts were also asked to assign each site to one of four condition categories based on narrative descriptions: (1) 'unaffected': a community at a least affected or unaffected site; (2) 'marginal deviation from unaffected': a community that shows some indication of stress, but within the measurement error of unaffected condition; (3) 'affected': where there is confidence that the community shows evidence of physical, chemical, natural, or anthropogenic stress; and (4) 'severely affected': where the magnitude of stress is high. The experts were also asked to identify the criteria they used to evaluate the benthos and rate their importance as follows: (1) very important; (2) important, but secondary; (3) marginally important; (4) useful, but only to interpret other factors. Criteria that were not used by an expert were assigned a rank of 5 for the purpose of calculating an average importance of that attribute among the experts. Since many of the experts identified tolerant and sensitive indicator species as evaluation criteria, they were also asked to list their indicator species and rank their importance on the same scale.

In each of the four regions, the twelve samples were selected to encompass a range of conditions from unimpacted to highly disturbed, from continental shelf and near shore areas with salinity >30 psu. The US West Coast, European Atlantic coast, and Mediterranean coast samples were collected with 0.1 $m^2$ Van Veen grabs and sieved through 1-mm screens, while the US East Coast samples were collected with 0.04 $m^2$ Young grabs and sieved through 0.5-mm screens. For consistency, abundances for the US East Coast samples were standardized to 0.1 $m^2$. The data sets from which the samples were selected, and the assessment measures used to order them, are described below.

*United States West Coast*

Twelve samples were selected from 493 in the data set used by Smith *et al.* (2001) to develop the Benthic Response Index (BRI). These samples were collected between 1973 and 1994, from 25-130 m depths along the southern California mainland shelf. Samples were ordered by their BRI values and selected at even BRI intervals.

*United States East Coast*

Samples were selected from a 338 sample data set collected between Cape Cod, Massachusetts and the mouth of Chesapeake Bay, Virginia, by the US Environment Protection Agency (EPA) for the Virginian Province Coastal Environmental Monitoring and Assessment Program (Strobel *et al.* 1995), the New York-New Jersey Harbor Regional Environmental Monitoring and Assessment Program (Adams *et al.* 1998), and the Mid Atlantic Integrated Assessment (US EPA 1998). Samples were selected by arranging the data set according to their Effects-Range Median (ERM) quotients (Long *et al.* 2000, 2006) and picking twelve samples at even ERM quotient intervals.

*European Atlantic Coast*

Twelve samples from Spain (2), the United Kingdom (5), Ireland (1), Belgium (2), Denmark (1) and Norway (1) were selected from the European dataset of 589 samples used to intercalibrate four different methodologies for assessing benthic quality within the Water Framework Directive (WFD) (Borja *et al.* 2007, 2009c). Samples were ordered from best to worst using the Ecological Quality Ratio (EQR; EC 2000) and selected at even intervals. Only samples classified in the same WFD ecological status for all four methodologies and with EQR standard error <0.1 among the four methodologies were included.

*European Mediterranean Coast*

Twelve samples were selected from published (Muxika *et al.*, 2005) and unpublished data compiled by AZTI-Tecnalia from 3 areas in Spain and 3 areas in Greece. Samples were ordered from best to worst and selected at even intervals using several measures, with generally coincident assessments using biotic indices such as the AZTI's Marine Biotic Index (AMBI) (Borja *et al.* 2000); trophic indices, such as the Infaunal Trophic Index (ITI) (Word 1978, 1980a, 1980b, 1990); and multivariate analyses.

### *Data analysis*

Patterns attributable to familiarity of experts with 'home region' fauna were evaluated in three ways, using regional assessments. First, the sample categorization of each expert was compared to the median categorization of experts from that region, and was quantified as the sum of the deviations (including the positive or negative sign) from the median category for each set of regional samples. Second, Permutational Multivariate Analysis of Variance (PERMANOVA) was used to determine whether there were significant differences in category assignments among groups of experts. The experimental design for this PERMANOVA (Anderson 2001, McArdle & Anderson 2001) included 'Sample Region' (four ecoregions) and 'Expert Region' (four ecoregions) as fixed factors, and a third 'Experts' (4 levels) fixed factor nested within the 'Expert Region' factor, with $n$ = 12 samples for each 'Sample Region' × 'Expert Region'× 'Experts' block. Bray-Curtis dissimilarities were used as distance measures in the PERMANOVA and distances were maintained (i.e. not replaced by their ranks) in the analysis. 4999 permutations were used to achieve an α-level of 0.05 (Anderson 2005). Third, Spearman rank correlation coefficients (ρ) were used to assess whether levels of agreement in categorizing and ranking sites differed between experts' home regions and other regions. Categories and rankings of experts for each region were compared with the respective regional medians.

The level of agreement on condition categories among all the experts was evaluated using Kappa analysis (Cohen 1960, Landis & Koch 1977) by establishing moderate, good, very good, and almost perfect levels of agreement using the equivalence table of Monserud & Leemans (1992). Fleiss–Cohen weights were applied (Fleiss & Cohen 1973) because misclassifications between distant categories (e.g. between 'unaffected' and 'affected', or 'unaffected' and 'severely affected') are more important than misclassifications between closer categories (e.g., between 'unaffected' and 'marginal deviation from unaffected', or 'affected' and 'severely affected').

The level of agreement in ranking sites among all the experts was evaluated using Spearman rank correlation analysis to measure associations between sample rankings by each expert and the median of the expert rankings. The variability of the expert rankings for each sample was measured by the median absolute deviation (MAD). Samples were ordered by median rank across all experts and MADs determined as the median of the absolute values of differences between expert ranks and this rank order.

## Results

There was substantial agreement in condition categories assigned by the experts (Table I). At least half of the experts agreed on sample condition category for 42 out of the 48 samples. Although there was complete agreement among the experts for only two samples and agreement among 15 of the 16 experts for only one other, at least seven experts agreed on the condition category for every sample. In contrast, there were seven samples that were assessed in all four condition categories, but for five of them at least 11 of the 16 experts agreed on their good (′unaffected′ or ′marginal deviation from unaffected′) or bad (′affected′ or ′severely affected′) condition. For 32 of the 48 samples, more than 87% of the experts agreed on whether the sample was in good or bad condition.

There also was a great deal of consensus in ranking of samples (Table II) among the experts. There were a few samples that different experts ranked at opposite extremes of the range (e.g., EU_A1, EU_A12, US_E11, US_W8, US_W11), but most of the discrepant ranks were attributable to only a few experts.

**Table I**. Condition categories assigned by the benthic experts to each of the 48 samples. EU: Europe, US: United States, A: Atlantic, M: Mediterranean, E: East Coast; W: West Coast. Key to condition categories: 1 – 'unaffected'; 2 – 'marginal deviation from unaffected'; 3 – 'affected'; 4 – 'severely affected'.

| | EU Atlantic experts | | | | Mediterranean experts | | | | US East Coast experts | | | | US West Coast experts | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Samples** | **A1** | **A2** | **A3** | **A4** | **M1** | **M2** | **M3** | **M4** | **E1** | **E2** | **E3** | **E4** | **W1** | **W2** | **W3** | **W4** |
| EU_A1 | 3 | 1 | 3 | 3 | 1 | 1 | 3 | 3 | 3 | 3 | 3 | 1 | 1 | 1 | 2 | 2 |
| EU_A2 | 2 | 1 | 3 | 2 | 2 | 1 | 3 | 3 | 1 | 3 | 3 | 2 | 3 | 1 | 2 | 3 |
| EU_A3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 |
| EU_A4 | 4 | 3 | 4 | 3 | 4 | 3 | 4 | 4 | 3 | 4 | 4 | 3 | 4 | 4 | 3 | 4 |
| EU_A5 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| EU_A6 | 3 | 2 | 4 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 2 | 3 |
| EU_A7 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 3 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 3 |
| EU_A8 | 4 | 3 | 4 | 2 | 4 | 4 | 4 | 4 | 3 | 4 | 4 | 3 | 4 | 4 | 3 | 4 |
| EU_A9 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 3 | 2 | 2 | 1 | 2 | 2 |
| EU_A10 | 3 | 3 | 4 | 2 | 4 | 3 | 4 | 3 | 3 | 4 | 3 | 2 | 4 | 4 | 3 | 4 |
| EU_A11 | 1 | 1 | 1 | 1 | 2 | 3 | 1 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 |
| EU_A12 | 2 | 2 | 4 | 4 | 3 | 3 | 3 | 3 | 4 | 4 | 3 | 3 | 3 | 1 | 3 | 3 |
| EU_M1 | 1 | 2 | 4 | 1 | 3 | 1 | 1 | 3 | 1 | 3 | 1 | 1 | 2 | 2 | 2 | 3 |
| EU_M2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 3 | 2 |
| EU_M3 | 2 | 1 | 3 | 2 | 2 | 1 | 2 | 3 | 2 | 3 | 1 | 1 | 2 | 2 | 2 | 3 |
| EU_M4 | 4 | 3 | 4 | 3 | 4 | 4 | 4 | 4 | 3 | 4 | 3 | 3 | 4 | 4 | 3 | 4 |
| EU_M5 | 4 | 2 | 4 | 3 | 3 | 3 | 4 | 4 | 3 | 4 | 3 | 1 | 3 | 4 | 3 | 4 |
| EU_M6 | 3 | 3 | 4 | 3 | 3 | 3 | 4 | 4 | 3 | 4 | 3 | 2 | 3 | 2 | 3 | 4 |
| EU_M7 | 4 | 3 | 4 | 3 | 4 | 3 | 4 | 4 | 3 | 4 | 3 | 2 | 4 | 4 | 3 | 4 |
| EU_M8 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| EU_M9 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 1 | 3 |
| EU_M10 | 2 | 2 | 1 | 3 | 1 | 1 | 2 | 2 | 2 | 3 | 2 | 1 | 2 | 1 | 3 | 1 |
| EU_M11 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| EU_M12 | 1 | 1 | 3 | 1 | 2 | 1 | 1 | 2 | 1 | 3 | 1 | 2 | 2 | 1 | 2 | 3 |
| US_E1 | 2 | 1 | 3 | 1 | 3 | 2 | 2 | 3 | 1 | 3 | 2 | 4 | 2 | 2 | 2 | 2 |
| US_E2 | 1 | 1 | 1 | 2 | 2 | 3 | 1 | 2 | 1 | 2 | 1 | 3 | 1 | 2 | 1 | 1 |
| US_E3 | 2 | 1 | 2 | 1 | 2 | 3 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 1 |
| US_E4 | 2 | 1 | 3 | 2 | 3 | 2 | 3 | 3 | 1 | 3 | 3 | 3 | 3 | 3 | 3 | 2 |
| US_E5 | 2 | 1 | 2 | 1 | 2 | 1 | 3 | 3 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 |
| US_E6 | 4 | 3 | 3 | 3 | 1 | 1 | 3 | 3 | 3 | 3 | 3 | 1 | 1 | 1 | 2 | 2 |
| US_E7 | 2 | 2 | 3 | 2 | 3 | 2 | 3 | 3 | 2 | 4 | 3 | 4 | 3 | 3 | 3 | 2 |
| US_E8 | 2 | 2 | 3 | 2 | 3 | 2 | 3 | 3 | 3 | 3 | 3 | 4 | 3 | 3 | 3 | 4 |
| US_E9 | 2 | 1 | 3 | 1 | 2 | 1 | 2 | 3 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 2 |
| US_E10 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 |
| US_E11 | 2 | 2 | 4 | 2 | 4 | 4 | 2 | 2 | 2 | 2 | 1 | 2 | 3 | 3 | 3 | 3 |
| US_E12 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 |
| US_W1 | 2 | 2 | 2 | 2 | 3 | 3 | 1 | 2 | 2 | 3 | 3 | 2 | 3 | 3 | 2 | 3 |
| US_W2 | 3 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 |
| US_W3 | 3 | 1 | 3 | 3 | 3 | 4 | 3 | 3 | 2 | 3 | 3 | 3 | 3 | 3 | 2 | 3 |
| US_W4 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 1 | 2 | 1 | 2 |
| US_W5 | 4 | 3 | 4 | 4 | 4 | 3 | 4 | 4 | 3 | 4 | 4 | 4 | 4 | 4 | 3 | 4 |
| US_W6 | 2 | 1 | 2 | 2 | 3 | 3 | 1 | 2 | 2 | 3 | 3 | 2 | 2 | 3 | 2 | 3 |
| US_W7 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 2 | 1 | 2 | 3 | 1 | 1 | 1 | 1 | 1 |
| US_W8 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 3 | 1 | 1 | 1 | 1 | 1 |
| US_W9 | 4 | 3 | 4 | 3 | 4 | 3 | 4 | 4 | 3 | 4 | 4 | 4 | 4 | 4 | 3 | 4 |
| US_W10 | 3 | 2 | 3 | 3 | 3 | 4 | 3 | 3 | 2 | 4 | 3 | 4 | 3 | 3 | 3 | 3 |
| US_W11 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 3 | 1 | 2 | 1 | 1 | 1 |
| US_W12 | 4 | 3 | 4 | 4 | 4 | 3 | 4 | 4 | 3 | 4 | 3 | 3 | 4 | 4 | 3 | 4 |

**Table II.** Expert condition rankings for all 48 samples. EU: Europe, US: United States, A: Atlantic, M: Mediterranean, E: East Coast, W: West Coast, S.D.: standard deviation.

| | **EU Atlantic experts** | | | | **Mediterranean experts** | | | | **US East Coast experts** | | | | **US West Coast experts** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Samples** | **A1** | **A2** | **A3** | **A4** | **M1** | **M2** | **M3** | **M4** | **E1** | **E2** | **E3** | **E4** | **W1** | **W2** | **W3** | **W4** | **S.D.** |
| EU_A1 | 37 | 21 | 27 | 42 | 1 | 10 | 35 | 31 | 36.5 | 20 | 26.5 | 11 | 11 | 2 | 27 | 14.5 | 12.8 |
| EU_A2 | 29 | 20 | 31 | 29 | 23 | 10 | 31 | 36 | 18 | 33 | 28 | 23 | 27 | 8 | 28 | 29 | 7.8 |
| EU_A3 | 4 | 2 | 3 | 2 | 11 | 10 | 7 | 2 | 4.5 | 2 | 17.5 | 2 | 3 | 19 | 8 | 2 | 5.6 |
| EU_A4 | 41 | 40 | 44 | 39 | 46 | 41 | 43 | 43 | 32.5 | 43 | 46.5 | 39.5 | 41 | 38 | 33 | 38.5 | 3.9 |
| EU_A5 | 48 | 48 | 48 | 48 | 48 | 48 | 47.5 | 46 | 48 | 48 | 48 | 45 | 48 | 48 | 48 | 48 | 0.9 |
| EU_A6 | 35 | 32 | 38 | 27 | 31 | 30 | 35 | 34 | 44 | 22 | 41.5 | 23 | 36 | 35 | 26 | 35 | 6.1 |
| EU_A7 | 7 | 13 | 13 | 14 | 24 | 10 | 18 | 27 | 18 | 11 | 11 | 11 | 17 | 7 | 14 | 29 | 6.6 |
| EU_A8 | 46 | 44 | 47 | 31 | 45 | 46 | 43 | 47 | 32.5 | 46 | 46.5 | 39.5 | 46 | 42 | 44 | 38.5 | 5.0 |
| EU_A9 | 12 | 5 | 6 | 12 | 7 | 10 | 18 | 10 | 18 | 3 | 26.5 | 11 | 15 | 9 | 16 | 14.5 | 5.9 |
| EU_A10 | 34 | 45 | 43 | 24 | 39 | 38 | 39 | 44 | 32.5 | 39 | 41.5 | 32.5 | 44 | 39 | 34 | 40 | 5.5 |
| EU_A11 | 8 | 12 | 10 | 7 | 22 | 32 | 7 | 15 | 18 | 12 | 11 | 11 | 8 | 5 | 9 | 6 | 7.0 |
| EU_A12 | 30 | 36 | 36 | 46 | 34 | 31 | 35 | 35 | 46 | 47 | 25 | 32.5 | 32 | 6 | 40 | 29 | 9.7 |
| EU_M1 | 10 | 27 | 34 | 13 | 26 | 10 | 18 | 29 | 11.5 | 29 | 7 | 16 | 24 | 24 | 21 | 24.5 | 8.2 |
| EU_M2 | 11 | 7 | 7 | 9 | 8 | 10 | 7 | 16 | 28 | 5 | 19.5 | 11 | 5 | 15 | 39 | 16.5 | 9.2 |
| EU_M3 | 14 | 11 | 26 | 30 | 20 | 10 | 20.5 | 28 | 32.5 | 28 | 7 | 23 | 21 | 23 | 17 | 24.5 | 7.5 |
| EU_M4 | 42 | 39 | 46 | 38 | 44 | 45 | 43 | 48 | 43 | 45 | 31.5 | 39.5 | 45 | 45 | 43 | 43.5 | 4.0 |
| EU_M5 | 44 | 31 | 39 | 36 | 35 | 40 | 37.5 | 37 | 40.5 | 37 | 29.5 | 23 | 38 | 40 | 35 | 43.5 | 5.3 |
| EU_M6 | 36 | 46 | 33 | 41 | 36 | 35 | 37.5 | 38 | 45 | 42 | 31.5 | 39.5 | 37 | 37 | 42 | 36.5 | 4.0 |
| EU_M7 | 39 | 41 | 40 | 37 | 40 | 39 | 43 | 40 | 42 | 38 | 29.5 | 23 | 39 | 41 | 36 | 43.5 | 5.2 |
| EU_M8 | 3 | 6 | 1 | 4 | 3 | 10 | 7 | 3 | 4.5 | 6 | 7 | 2 | 2 | 16 | 3 | 2 | 3.8 |
| EU_M9 | 6 | 18 | 9 | 6 | 18 | 10 | 7 | 17 | 11.5 | 13 | 7 | 16 | 7 | 17 | 7 | 24.5 | 5.7 |
| EU_M10 | 13 | 30 | 8 | 40 | 2 | 10 | 20.5 | 11 | 29.5 | 27 | 19.5 | 32.5 | 20 | 11 | 41 | 12 | 11.8 |
| EU_M11 | 47 | 47 | 45 | 47 | 47 | 47 | 47.5 | 45 | 47 | 44 | 43 | 45 | 47 | 47 | 47 | 47 | 1.4 |
| EU_M12 | 9 | 19 | 29 | 10 | 21 | 10 | 7 | 12 | 11.5 | 24 | 7 | 16 | 22 | 18 | 18 | 24.5 | 6.8 |
| US_E1 | 28 | 24 | 30 | 11 | 25 | 23 | 25.5 | 22 | 11.5 | 21 | 15.5 | 32.5 | 25 | 27 | 25 | 19 | 6.1 |
| US_E2 | 5 | 4 | 11 | 19 | 15 | 29 | 7 | 4 | 4.5 | 14 | 2.5 | 23 | 9 | 25 | 10 | 6 | 8.3 |
| US_E3 | 22.5 | 10 | 15 | 8 | 19 | 28 | 7 | 19 | 11.5 | 15 | 2.5 | 23 | 19 | 26 | 11 | 9.5 | 7.3 |
| US_E4 | 22.5 | 14 | 23 | 23 | 30 | 26 | 24 | 23 | 18 | 31 | 23.5 | 32.5 | 29 | 30 | 32 | 21.5 | 5.2 |
| US_E5 | 18 | 23 | 20 | 16 | 13 | 10 | 29 | 24 | 29.5 | 17 | 13.5 | 23 | 10 | 12 | 23 | 19 | 6.2 |
| US_E6 | 45 | 43 | 22 | 43 | 9 | 10 | 33 | 30 | 38.5 | 23 | 21.5 | 6 | 12 | 10 | 22 | 19 | 13.2 |
| US_E7 | 24 | 28 | 24 | 28 | 33 | 25 | 28 | 20 | 26.5 | 34 | 23.5 | 45 | 30 | 31 | 45 | 21.5 | 7.3 |
| US_E8 | 19.5 | 35 | 25 | 32 | 37 | 24 | 31 | 25 | 40.5 | 32 | 21.5 | 45 | 31 | 34 | 46 | 36.5 | 7.9 |
| US_E9 | 17 | 22 | 28 | 15 | 14 | 10 | 22 | 26 | 18 | 16 | 13.5 | 6 | 18 | 22 | 12 | 16.5 | 5.8 |
| US_E10 | 21 | 9 | 14 | 5 | 5 | 10 | 15 | 7 | 18 | 10 | 15.5 | 6 | 13 | 20 | 13 | 9.5 | 5.1 |
| US_E11 | 27 | 33 | 35 | 25 | 38 | 42 | 23 | 33 | 26.5 | 18 | 2.5 | 32.5 | 28 | 36 | 29 | 33 | 9.2 |
| US_E12 | 19.5 | 15 | 17 | 26 | 6 | 10 | 25.5 | 14 | 11.5 | 8 | 2.5 | 32.5 | 23 | 13 | 20 | 11 | 8.1 |
| US_W1 | 25 | 29 | 18 | 22 | 29 | 27 | 7 | 18 | 23.5 | 25 | 36 | 32.5 | 33 | 33 | 19 | 33 | 7.6 |
| US_W2 | 33 | 26 | 19 | 18 | 12 | 22 | 15 | 13 | 4.5 | 7 | 17.5 | 16 | 14 | 1 | 1 | 2 | 9.1 |
| US_W3 | 31.5 | 25 | 21 | 33 | 27 | 43 | 27 | 21 | 23.5 | 30 | 38.5 | 32.5 | 34 | 29 | 24 | 29 | 6.1 |
| US_W4 | 2 | 1 | 4 | 3 | 16 | 21 | 7 | 8 | 4.5 | 19 | 11 | 2 | 4 | 21 | 4 | 13 | 7.1 |
| US_W5 | 43 | 42 | 42 | 45 | 43 | 36 | 43 | 42 | 38.5 | 41 | 44.5 | 32.5 | 40 | 43 | 37 | 43.5 | 3.4 |
| US_W6 | 15 | 16 | 16 | 17 | 28 | 34 | 15 | 5 | 23.5 | 26 | 38.5 | 23 | 26 | 28 | 15 | 29 | 8.6 |
| US_W7 | 26 | 8 | 12 | 20 | 10 | 10 | 7 | 9 | 4.5 | 9 | 34 | 6 | 6 | 14 | 6 | 6 | 8.2 |
| US_W8 | 1 | 3 | 2 | 1 | 17 | 20 | 7 | 1 | 4.5 | 1 | 34 | 16 | 1 | 3 | 2 | 6 | 9.5 |
| US_W9 | 38 | 37 | 37 | 34 | 41 | 37 | 43 | 39 | 35 | 36 | 44.5 | 45 | 43 | 46 | 30 | 43.5 | 4.6 |
| US_W10 | 31.5 | 34 | 32 | 35 | 32 | 44 | 31 | 32 | 23.5 | 35 | 38.5 | 45 | 35 | 32 | 38 | 33 | 5.2 |
| US_W11 | 16 | 17 | 5 | 21 | 4 | 10 | 7 | 6 | 4.5 | 4 | 34 | 6 | 16 | 4 | 5 | 6 | 8.4 |
| US_W12 | 40 | 38 | 41 | 44 | 42 | 33 | 43 | 41 | 36.5 | 40 | 38.5 | 45 | 42 | 44 | 31 | 43.5 | 4.0 |

### *Regional consistency of ecological assessments*

No regional bias in expert category assignments was observed (Table III). The distribution of deviations from regional median categories was similar for experts' home regions and other regions. More importantly, regional deviations were less than individual deviations (Table III). A slight negative deviation was detected in Atlantic expert assessments, with samples from other regions evaluated in better ecological condition categories than the regional medians (Table III).

**Table III**. Deviation of expert categories from the median for local experts at each set of regional samples. Sums of expert category deviations for regional groups of experts at each regional data set are also presented. Home region results are highlighted. EU: Europe, US: United States.

| | **Experts** | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Samples sets** | **A1** | **A2** | **A3** | **A4** | **EU Atlantic** | **M1** | **M2** | **M3** | **M4** | **Mediterranean** | **E1** | **E2** | **E3** | **E4** | **US East Coast** | **W1** | **W2** | **W3** | **W4** | **US West Coast** |
| EU Atl. | 1.5 | -5.5 | 5.5 | -2.5 | **-1** | 2.5 | -0.5 | 3.5 | 7.5 | **13** | -0.5 | 6.5 | 5.5 | -3.5 | **8** | 3.5 | -1.5 | -1.5 | 5.5 | **6** |
| EU Med. | -1.5 | -5.5 | 4.5 | -3.5 | **-6** | 0.5 | -5.5 | -0.5 | 6.5 | **1** | -3.5 | 6.5 | -4.5 | -8.5 | **-10** | -0.5 | -2.5 | 0.5 | 6.5 | **4** |
| US East | -1.5 | -9.5 | 4.5 | -6.5 | **-13** | 0.5 | -3.5 | -0.5 | 4.5 | **1** | -7.5 | 3.5 | -2.5 | 2.5 | **-4** | -2.5 | -1.5 | -2.5 | -3.5 | **-10** |
| US West | 2.0 | -9.0 | -1.0 | 0.0 | **-8** | 2.0 | 2.0 | -4.0 | 3.0 | **3** | -7.0 | 4.0 | 6.0 | -1.0 | **2** | 1.0 | 1.0 | -6.0 | 1.0 | **-3** |
| **Total** | 0.5 | -29.5 | 13.5 | -12.5 | **-28** | 5.5 | -7.5 | -1.5 | 21.5 | **18** | -18.5 | 20.5 | 4.5 | -10.5 | **-4** | 1.5 | -4.5 | -9.5 | 9.5 | **-3** |

Variability in the category assignments was unrelated to whether the assessments were for home regions. There was no statistical significance for any factor related to 'Expert Region' in the PERMANOVA (Table IV), indicating that expert category assignments were independent of the regions in which the experts worked. These results also indicated that patterns of US East Coast category assessments were significantly different from patterns for other sets of regional samples.

High correlations were observed among individual expert category assignments and the regional median category for the European Atlantic, Mediterranean, and US West Coast samples, with few Spearman correlation coefficients less than 0.80 (Table V). In contrast, for the US East Coast samples, 12 of 16 experts' Spearman correlation coefficients were less than 0.80 and 6 were not statistically significant. However, PERMANOVA (Table IV) showed that category assignments were

**Table IV.** Results of PERMANOVA on Bray-Curtis distances between category assessments of 48 samples from four regions (factor Sample Region; 4 levels, n=12 samples each), by groups of experts from those regions (factor Expert Region; 4 levels, each group with 4 experts).

| **Source** | ***d.f.*** | ***SS*** | ***MS*** | ***F*** |
|---|---|---|---|---|
| Sample Region | 3 | 5867.64 | 1955.88 | 3.56* |
| Expert Region | 3 | 2633.24 | 877.75 | 1.60 |
| ExpReg (Experts) | 12 | 26083.13 | 2173.59 | 3.96** |
| Sample Region x Expert Region | 9 | 1613.36 | 179.26 | 0.33 |
| Sample Region x ExpReg (Experts) | 36 | 14934.57 | 414.85 | 0.76 |
| Residual | 704 | 386401.10 | 548.87 | |
| Total | 767 | 437533.03 | | |
| **Pair-wise *post-hoc* comparisons:** | | | | |
| Sample Region | | | | |
| Atlantic *vs.* Mediterranean | 0.96 | | | |
| Atlantic *vs.* East Coast | 3.14** | | | |
| Atlantic *vs.* West Coast | 0.84 | | | |
| Mediterranean *vs.* East Coast | 2.26* | | | |
| Mediterranean *vs.* West Coast | 0.52 | | | |
| East Coast *vs.* West Coast | 2.30* | | | |
| | Expert Region | | | |
| Experts | Atlantic | Mediterranean | East Coast | West Coast |
| Expert 1 *vs.* Expert 2 | 3.00* | 1.23 | 3.71** | 0.79 |
| Expert 1 *vs.* Expert 3 | 0.96 | 0.98 | 2.25* | 0.76 |
| Expert 1 *vs.* Expert 4 | 1.14 | 2.44* | 0.67 | 0.61 |
| Expert 2 *vs.* Expert 3 | 3.59* | 0.35 | 1.40 | 0.61 |
| Expert 2 *vs.* Expert 4 | 1.87 | 3.79** | 2.99* | 1.30 |
| Expert 3 *vs.* Expert 4 | 1.93 | 3.39** | 1.57 | 1.38 |

* $P \leq 0.05$; ** $P \leq 0.001$. Pair-wise *a posteriori* tests, between Sample Region, and between Experts within each region, using the *t*-statistic.

similar regardless of whether the experts were assessing their home regions, although mean correlations among experts were slightly higher within home region samples, except for US East Coast experts (Table V).

The patterns observed for regional rank evaluations (Table VI) were similar to those for condition category assignments (Table V). Correlation coefficients for rankings were higher, on average, than for category assignments indicating that consensus between experts was higher when ranking samples than assigning condition categories. For both categorization and ranking, US West Coast experts presented a higher level of within group concordance compared to the remaining regional groups of experts (Table V, Table VI), being the only regional group of experts with no significant differences observed between any of its experts' categorizations (Table IV).

**Table V**. Spearman correlation coefficients between expert category assignments and the regional median category (n= 12). A: Atlantic, M: Mediterranean, E: East Coast; W: West Coast.

| Experts | Atlantic | Mediterranean | East Coast | West Coast |
|---|---|---|---|---|
| A1 | 0.92 | 0.89 | 0.43* | 0.75 |
| A2 | 0.87 | 0.87 | 0.67 | 0.88 |
| A3 | 0.88 | 0.90 | 0.63 | 0.91 |
| A4 | 0.86 | 0.80 | 0.50* | 0.82 |
| mean | 0.88 | 0.86 | 0.56 | 0.84 |
| M1 | 0.78 | 0.96 | 0.52 | 0.97 |
| M2 | 0.72 | 0.88 | 0.08* | 0.76 |
| M3 | 0.94 | 0.88 | 0.85 | 0.88 |
| M4 | 0.89 | 0.94 | 0.72 | 0.88 |
| mean | 0.83 | 0.92 | 0.54 | 0.87 |
| E1 | 0.87 | 0.80 | 0.66 | 0.97 |
| E2 | 0.92 | 0.94 | 0.92 | 0.93 |
| E3 | 0.89 | 0.75 | 0.82 | 0.61 |
| E4 | 0.79 | 0.62 | 0.53* | 0.86 |
| mean | 0.87 | 0.78 | 0.73 | 0.84 |
| W1 | 0.78 | 0.96 | 0.50* | 0.91 |
| W2 | 0.69 | 0.94 | 0.44* | 0.99 |
| W3 | 0.91 | 0.61 | 0.82 | 0.94 |
| W4 | 0.80 | 0.92 | 0.68 | 0.99 |
| mean | 0.80 | 0.86 | 0.61 | 0.96 |

* non-significant correlations: $p \geq 0.05$

**Table VI.** Spearman correlation coefficients between expert regional sample ranks and the median regional rank (n= 12). A: Atlantic, M: Mediterranean, E: East Coast; W: West Coast.

| Experts | Atlantic | Mediterranean | East Coast | West Coast |
|---|---|---|---|---|
| A1 | 0.94 | 0.83 | 0.55* | 0.62 |
| A2 | 0.99 | 0.84 | 0.74 | 0.76 |
| A3 | 0.98 | 0.93 | 0.64 | 0.79 |
| A4 | 0.80 | 0.68 | 0.58* | 0.82 |
| mean | 0.92 | 0.82 | 0.63 | 0.75 |
| M1 | 0.83 | 0.96 | 0.55* | 0.90 |
| M2 | 0.84 | 0.87 | 0.02* | 0.70 |
| M3 | 0.98 | 0.87 | 0.73 | 0.77 |
| M4 | 0.94 | 0.99 | 0.54* | 0.81 |
| mean | 0.90 | 0.93 | 0.46 | 0.79 |
| E1 | 0.84 | 0.81 | 0.73 | 0.92 |
| E2 | 0.92 | 0.89 | 0.98 | 0.92 |
| E3 | 0.85 | 0.75 | 0.86 | 0.85 |
| E4 | 0.94 | 0.74 | 0.59 | 0.82 |
| mean | 0.89 | 0.80 | 0.79 | 0.88 |
| W1 | 0.91 | 0.93 | 0.69 | 0.92 |
| W2 | 0.62 | 0.96 | 0.44* | 0.99 |
| W3 | 0.92 | 0.63 | 0.92 | 0.90 |
| W4 | 0.85 | 0.92 | 0.88 | 0.92 |
| mean | 0.82 | 0.86 | 0.73 | 0.93 |

* non-significant correlations: $p \geq 0.05$

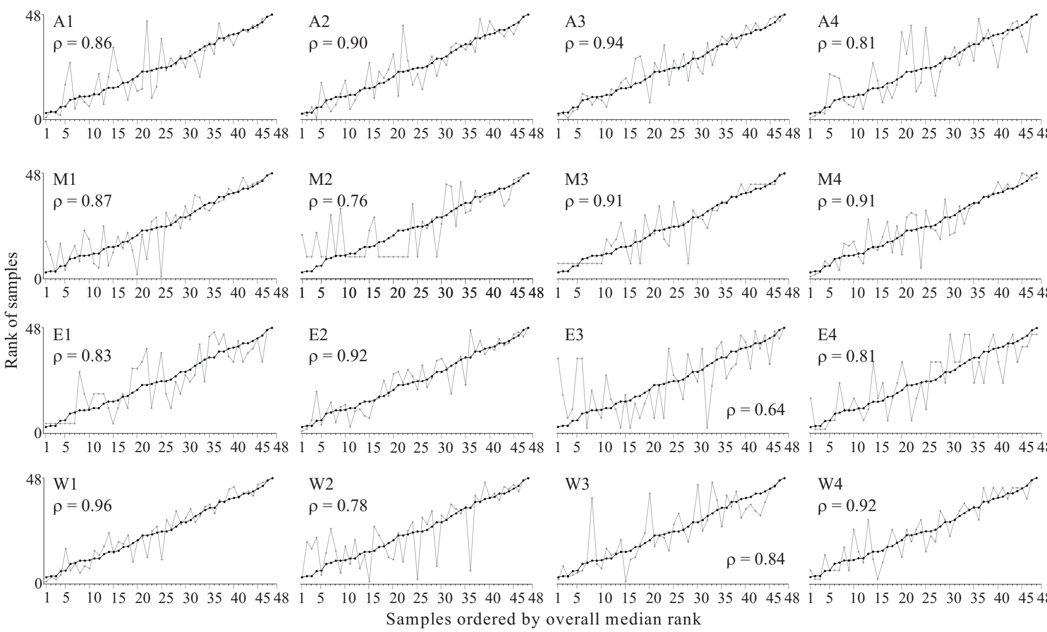
### *Level of agreement on the ecological assessments*

Kappa analysis indicated a high degree of agreement among experts in their condition category assignments (average kappa value of 0.65), with levels of agreement varying from moderate to almost perfect and 78.5% of the comparisons agreeing at 'good' or better (Table VII). Mismatches >30% occurred in less than 10% of the comparisons. At the level of good ('unaffected' / 'marginal deviation from unaffected') or bad condition ('affected' / 'severely affected'), the experts agreed on approximately 80% of the comparisons.

**Table VII.** Kappa values with level of agreement in parentheses (lower left) for condition category assignments, and percentage of mismatch between expert classifications (upper right). A: Atlantic, M: Mediterranean, E: East Coast; W: West Coast. Level of agreement: AP – 'Almost Perfect'; VG – 'Very Good'; G – 'Good' and M – 'Moderate'. The percentage of mismatch is related to the relative number of cases in which one of the experts classified a station as 'unaffected' or 'marginal deviation from unaffected' and the other as 'affected' or 'severely affected'.

| | **Percentage of mismatch** | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **A1** | **A2** | **A3** | **A4** | **M1** | **M2** | **M3** | **M4** | **E1** | **E2** | **E3** | **E4** | **W1** | **W2** | **W3** | **W4** |
| **A1** | | 10.6 | 25.0 | 10.6 | 26.5 | 22.4 | 14.6 | 25.0 | 10.4 | 27.1 | 25.0 | 30.6 | 22.4 | 22.4 | 23.4 | 27.7 |
| **A2** | 0.77 (VG) | | 33.3 | 16.7 | 29.2 | 25.0 | 22.9 | 33.3 | 10.4 | 35.4 | 33.3 | 25.0 | 25.0 | 25.0 | 20.8 | 33.3 |
| **A3** | 0.65 (G) | 0.47 (M) | | 24.4 | 16.7 | 34.7 | 14.6 | 6.4 | 22.9 | 8.5 | 28.0 | 24.4 | 14.9 | 23.4 | 21.7 | 20.8 |
| **A4** | 0.80 (VG) | 0.68 (G) | 0.61 (G) | | 29.2 | 25.0 | 18.8 | 29.2 | 14.6 | 27.1 | 29.2 | 25.0 | 25.0 | 30.6 | 20.8 | 33.3 |
| **M1** | 0.58 (G) | 0.56 (G) | 0.79 (VG) | 0.51 (M) | | 16.7 | 17.0 | 19.1 | 21.3 | 12.8 | 22.4 | 15.6 | 8.3 | 8.3 | 16.7 | 16.7 |
| **M2** | 0.58 (G) | 0.52 (M) | 0.48 (M) | 0.55 (M) | 0.77 (VG) | | 25.5 | 36.2 | 17.8 | 29.8 | 30.6 | 27.7 | 16.7 | 16.7 | 23.4 | 20.8 |
| **M3** | 0.83 (VG) | 0.64 (G) | 0.79 (VG) | 0.70 (VG) | 0.76 (VG) | 0.59 (G) | | 10.4 | 12.5 | 16.7 | 14.6 | 19.6 | 10.4 | 17.0 | 18.8 | 27.1 |
| **M4** | 0.69 (G) | 0.48 (M) | 0.88 (AP) | 0.54 (M) | 0.73 (VG) | 0.44 (M) | 0.84 (VG) | | 22.9 | 14.6 | 25.0 | 27.7 | 20.8 | 27.7 | 29.2 | 25.0 |
| **E1** | 0.77 (VG) | 0.79 (VG) | 0.62 (G) | 0.78 (VG) | 0.64 (G) | 0.68 (G) | 0.79 (VG) | 0.61 (G) | | 25.0 | 22.9 | 27.1 | 18.8 | 24.5 | 18.8 | 27.1 |
| **E2** | 0.64 (G) | 0.46 (M) | 0.88 (AP) | 0.59 (G) | 0.80 (VG) | 0.55 (M) | 0.78 (VG) | 0.80 (VG) | 0.60 (G) | | 18.8 | 28.3 | 18.8 | 22.9 | 27.1 | 17.0 |
| **E3** | 0.68 (G) | 0.49 (M) | 0.47 (M) | 0.58 (G) | 0.57 (G) | 0.44 (M) | 0.74 (VG) | 0.60 (G) | 0.63 (G) | 0.65 (G) | | 33.3 | 16.7 | 20.8 | 29.2 | 29.2 |
| **E4** | 0.40 (M) | 0.45 (M) | 0.54 (M) | 0.50 (M) | 0.75 (VG) | 0.55 (G) | 0.63 (G) | 0.51 (M) | 0.46 (M) | 0.56 (G) | 0.49 (M) | | 19.1 | 19.1 | 19.1 | 34.8 |
| **W1** | 0.67 (G) | 0.62 (G) | 0.81 (VG) | 0.61 (G) | 0.90 (AP) | 0.73 (VG) | 0.84 (VG) | 0.70 (VG) | 0.68 (G) | 0.76 (VG) | 0.72 (VG) | 0.67 (G) | | 8.3 | 12.5 | 16.7 |
| **W2** | 0.67 (G) | 0.60 (G) | 0.67 (G) | 0.47 (M) | 0.89 (AP) | 0.75 (VG) | 0.76 (VG) | 0.62 (G) | 0.57 (G) | 0.65 (G) | 0.64 (G) | 0.64 (G) | 0.88 (AP) | | 16.7 | 20.8 |
| **W3** | 0.63 (G) | 0.66 (G) | 0.67 (G) | 0.67 (G) | 0.72 (VG) | 0.60 (G) | 0.72 (VG) | 0.55 (G) | 0.72 (VG) | 0.62 (G) | 0.55 (M) | 0.66 (G) | 0.78 (VG) | 0.70 (VG) | | 29.2 |
| **W4** | 0.63 (G) | 0.51 (M) | 0.75 (VG) | 0.48 (M) | 0.81 (VG) | 0.60 (G) | 0.68 (G) | 0.70 (G) | 0.61 (G) | 0.79 (VG) | 0.53 (M) | 0.48 (M) | 0.80 (VG) | 0.71 (VG) | 0.58 (G) | |

Sample rankings (Table II) were highly correlated among experts, with an average Spearman correlation coefficient of 0.85 between expert rank and the median rank (Figure 1). Seven experts (A2, A3, M3, M4, E2, W1, W4) deviated little from the median ranks (Figure 1). Of the nine that deviated more, five deviated throughout the range (A4, E1, E3, E4, W3) and four differed primarily for samples in the lower and intermediate ranks (A1, M1, M2, W2).




| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| US_W8 | EU_A3 | EU_M8 | US_W4 | US_W11 | US_W7 | US_E2 | EU_M2 | EU_A11 | EU_M9 | EU_A9 | US_E10 | EU_A7 | US_E12 | US_W2 | US_E3 | US_E9 | EU_M12 | US_E5 | EU_M10 | EU_M3 | US_E6 | EU_M1 | US_W6 |

| 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| EU_A1 | US_E4 | US_E1 | US_W1 | EU_A2 | US_E7 | US_W3 | US_E11 | US_E8 | US_W10 | EU_A6 | EU_A12 | EU_M5 | EU_M6 | US_W9 | EU_A10 | EU_M7 | EU_A4 | US_W12 | US_W5 | EU_M4 | EU_A8 | EU_M11 | EU_A5 |

**Figure 1.** Deviation of each expert's samples rank from median rank along the disturbance gradient (samples ordered by median ranks). Key: A, EU Atlantic; M, Mediterranean; E, US East Coast; W, US West Coast; ρ, Spearman rank correlation; Dev, total absolute deviation from median ranks; grey dots, expert ranks; black dots, median ranks.

Overall, the level of agreement between experts was higher at the extremes of the gradient of disturbance than at the centre (Figure 2). Disagreements with respect to good or bad condition occurred mostly in the intermediate third of samples, where the MAD also was higher, showing that rankings had also higher dispersion near the centre of the gradient (Figure 2). The three samples with ranking standard deviations > 10 (Table II) were in the middle third of the gradient (EU_A1, EU_M10 and US_E6). Samples with higher median absolute deviations from the median rank (Figure 2) were often assigned to three or four categories (Table I).

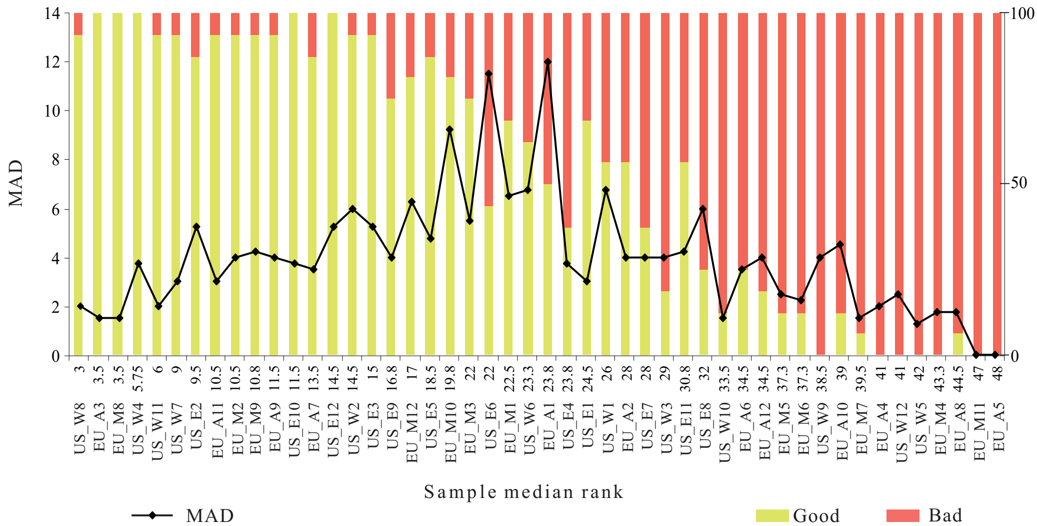



**Figure 2.** Samples' median absolute deviations (MAD) along the disturbance gradient (samples ordered by median ranks), and percentage of experts classifying each sample as 'Good' ('unaffected' or 'marginal deviation from unaffected') and 'Bad' ('affected' or 'severely affected') condition. Key: EU, Europe; US, United States; A, Atlantic; M, Mediterranean; E, East Coast; W, West Coast.

The results indicated tendencies in individual experts unrelated to home regions. Assessments by four experts (A2, E1, E2 and M4) deviated from regional medians, with A2 and E1 consistently negative (classifying in better condition than the median), and E2 and M4 consistently positive, classifying in worse condition than the median (Table III). Within regional groups of experts, *a posteriori* tests showed statistically significant differences between these four experts' category assessments and some of the remaining (Table IV).

## *Criteria used by experts*

The experts used eight criteria for assessing benthic assemblage condition. Six were used by more than half of the experts, with the other two used by only two experts (Table VIII). The three most widely used criteria were 'Dominance by tolerant taxa', 'Presence of sensitive taxa', and

‘Biodiversity number of taxa measures’. However, they were not equally important to experts from different regions. Mediterranean and US East Coast experts, respectively, considered ‘Biodiversity number of taxa measures’ and ‘Presence of sensitive taxa’ only marginally important. In turn, two other attributes, ‘Biodiversity community measures’ and ‘Abundance dominance patterns’ also were considered important by Mediterranean and US West Coast experts, respectively.

**Table VIII.** Criteria used by benthic experts to rank and categorize samples. EU: Europe, US: United States, S.D.: standard deviation. ‘Importance’ is the average importance for all 16 experts, where: 1, very important; 2, important, but secondary; 3, marginally important; 4, useful but only to interpret the other factors; 5, not used. *N* is the number of experts that used the criterion.

| | | | | Regional Average Importance | | | |
|---|---|---|---|---|---|---|---|
| **Criteria** | **Importance** | **S.D.** | ***N*** | **EU Atlantic** | **Mediterranean** | **US East Coast** | **US West Coast** |
| Dominance by tolerant taxa | 1.8 | 0.4 | 14 | 1.4 | 1.5 | 2.1 | 2.0 |
| Presence of sensitive taxa | 2.6 | 0.5 | 11 | 2.8 | 2.0 | 3.3 | 2.3 |
| Biodiversity number of taxa measures | 2.7 | 0.5 | 13 | 2.8 | 3.3 | 2.0 | 2.7 |
| Abundance dominance patterns | 3.0 | 0.9 | 10 | 3.3 | 3.8 | 3.3 | 1.6 |
| Biodiversity community measures | 3.4 | 1.1 | 9 | 3.0 | 2.0 | 4.0 | 4.5 |
| Abundance | 3.7 | 0.9 | 10 | 4.0 | 4.8 | 3.0 | 3.0 |
| Complex analyses | 4.6 | 0.9 | 2 | 5.0 | 3.3 | 5.0 | 5.0 |
| Invasive & Introduced species | 4.8 | 0.2 | 2 | 5.0 | 5.0 | 4.5 | 4.8 |

On average, experts deviating from their peers used less than the average of 5.3 criteria used by the others. Experts who consistently assessed samples in a worse condition than the median used an average of 5.0 criteria compared to an average of 2.8 criteria used by experts assessing samples in better condition than the median. The experts indicated that it was not more difficult to evaluate data from non-home regions because genus and family associations across regions permitted extrapolation from knowledge of local fauna. Most experts identified tolerant taxa at the species or genus level, but mostly relied on the presence of higher-level taxonomic groups for sensitive taxa (Table IX).

**Table IX.** Indicator taxa identified by the experts. S.D.: standard deviation. Importance: average importance for all 16 experts (1-very important; 2-important, but secondary; 3-marginally important; 4-useful but only to interpret the other factors; 5-taxa mentioned but not its importance). *N:* number of experts referring the taxa.

| Tolerant taxa | Importance | S.D. | *N* | Sensitive taxa | Importance | S.D. | *N* |
|---|---|---|---|---|---|---|---|
| *Solemya reidi* | 1.0 | 0.0 | 3 | *Lanice conchilega* | 1.0 | | 1 |
| *Solemya togata* | 1.0 | 0.0 | 4 | Sabellidae | 1.0 | | 1 |
| *Schistomeringos longicornis* | 1.0 | 0.0 | 3 | Terebellidae | 1.0 | | 1 |
| *Ophryotrocha* spp. | 1.2 | 0.4 | 5 | Trichobranchidae | 1.0 | | 1 |
| *Armandia brevis* | 1.3 | 0.6 | 3 | *Amphiura* spp. | 1.3 | 0.5 | 4 |
| *Mulinia lateralis* | 1.5 | 0.7 | 2 | *Amphiodia* spp. complex | 1.4 | 0.5 | 5 |
| *Raricirrus beryli* | 1.5 | 0.7 | 2 | *Ampelisca* spp. | 1.5 | 0.6 | 4 |
| *Capitella capitata* complex | 1.6 | 1.2 | 14 | *Proclea* spp. | 1.5 | 0.7 | 2 |
| *Macoma carlottensis* | 1.8 | 0.5 | 4 | Gammaridea (Haustoriidae/Phoxocephalidae) | 1.8 | 1.8 | 5 |
| *Parvilucina tenuisculpta* | 1.8 | 0.5 | 4 | *Tellina agilis* | 1.8 | 0.8 | 5 |
| *Mediomastus* spp. | 1.8 | 0.8 | 5 | *Cyathura burbancki* | 2.0 | | 1 |
| *Streblospio benedicti* | 1.8 | 0.8 | 6 | Echinoidea | 2.0 | 1.5 | 6 |
| Mollusca | 2.0 | | 1 | *Anadara transversa* | 2.0 | | 1 |
| *Corbula gibba* | 2.0 | 0.8 | 4 | *Mercenaria mercenaria* | 2.0 | | 1 |
| Thyasiridae | 2.0 | 0.0 | 2 | *Mya arenaria* | 2.0 | 0.0 | 3 |
| *Cossura longocirrata* | 2.0 | | 1 | *Nemocardium centifilosum* | 2.0 | | 1 |
| *Armandia* spp. | 2.0 | | 1 | *Plagiocardium papillosum* | 2.0 | 1.4 | 2 |
| *Eteone heteropoda* | 2.0 | | 1 | *Tellina* spp. | 2.0 | 1.0 | 3 |
| *Euchone incolor* | 2.0 | | 1 | *Timoclea ovata* | 2.0 | 1.4 | 2 |
| *Levinsenia gracilis* | 2.0 | | 1 | Ophiuroidea (other than Ophiuridae) | 2.0 | 1.5 | 6 |
| *Nephtys hombergii* | 2.0 | | 1 | Ampharetidae | 2.0 | | 1 |
| *Nucula annulata* | 2.2 | 0.4 | 5 | Maldanidae | 2.0 | | 1 |
| *Malacoceros fuliginosus* | 2.2 | 1.1 | 5 | *Pectinaria* spp. | 2.0 | 1.4 | 2 |
| *Polydora* spp. | 2.3 | 0.5 | 4 | *Ensis directus* | 2.3 | 0.6 | 3 |
| *Prionospio steenstrupi* | 2.3 | 0.5 | 4 | *Macoma balthica* | 2.3 | 0.6 | 3 |
| *Axinopsida serricata* | 2.3 | 0.6 | 3 | *Listriella goleta* | 2.5 | 2.1 | 2 |
| Oligochaeta | 2.4 | 1.4 | 7 | *Spisula* spp. | 2.5 | 2.1 | 2 |
| *Dipolydora* spp. | 2.5 | 0.7 | 2 | Arthropoda | 2.7 | 2.1 | 3 |
| *Thyasira flexuosa* | 2.7 | 1.2 | 3 | *Spisula solidissima* | 2.7 | 1.2 | 3 |
| *Ampelisca* spp. | 3.0 | 1.7 | 3 | Mollusca | 3.5 | 1.7 | 4 |
| *Euphilomedes* spp. | 3.0 | 1.4 | 2 | *Chaetopterus variopedatus* | 3.7 | 0.6 | 3 |
| *Mysella* spp. | 3.0 | 0.0 | 2 | Brachiopoda | 3.7 | 1.2 | 3 |
| *Mytilus edulis* | 3.0 | 1.0 | 3 | *Edwardsia* spp. | 4.0 | 1.4 | 2 |
| *Nassarius mendicus* | 3.0 | 1.4 | 2 | Crustacea | 5.0 | | 1 |
| Amphiuridae | 3.0 | 2.8 | 2 | Bivalvia | 5.0 | | 1 |
| *Chaetopterus variopedatus* | 3.0 | | 1 | Polychaeta | 5.0 | | 1 |
| *Aphelochaeta* spp. | 3.0 | 1.2 | 4 | *Lumbrineris* spp. | 5.0 | | 1 |
| *Chaetozone* spp. | 3.0 | 1.0 | 3 | *Pista* spp. | 5.0 | | 1 |
| *Cirratulus* spp. | 3.0 | 1.0 | 3 | | | | |
| *Tharyx* spp. | 3.0 | 1.0 | 5 | | | | |
| Cossuridae | 3.0 | | 1 | | | | |
| *Streblospio* spp. | 3.0 | | 1 | | | | |
| Lucinidae | 3.3 | 1.5 | 3 | | | | |
| Paraonidae | 3.3 | 0.6 | 3 | | | | |
| *Pseudopolydora* spp. | 3.5 | 1.3 | 4 | | | | |
| *Prionospio* spp. | 3.7 | 1.2 | 3 | | | | |
| Spionidae | 3.8 | 1.5 | 4 | | | | |
| *Ericthonius brasiliensis* | 4.0 | | 1 | | | | |
| Polychaeta | 4.0 | 1.4 | 2 | | | | |
| Cirratulidae | 4.0 | 1.0 | 3 | | | | |
| *Polygordius* spp. | 4.0 | 1.4 | 2 | | | | |
| *Malacoceros* spp. | 4.0 | | 1 | | | | |
| *Nucula* spp. | 4.5 | 0.7 | 2 | | | | |

Most frequently recognized as tolerant taxa were Polychaetes from the *Capitella capitata* complex, *Streblospio benedicti*, *Ophryotrocha* sp., and *Malacoceros fuliginosus*, oligochaetes, and the bivalve *Nucula annulata*. Most commonly identified as sensitive taxa were the Echinoidea, Ophiuroidea (other than Ophiuridae) and Gammaridea higher taxonomic groups, *Amphiodia* spp., and *Tellina agilis*. Different indicator taxa were considered for samples from different regions and, therefore, this list of indicator taxa is not universally applicable throughout all four regions.

## Discussion

No systematic difference in assessments based on experts' regions of origin was observed, though the level of agreement here was slightly lower than that achieved by Weisberg *et al.* (2008) in a single region. The slightly higher correlations within the West Coast group of experts were possibly driven by the particularly close professional ties, since three of them are from the same agency.

There was greater agreement on sample ranks than on sample condition categories. While the experts largely agreed on the relative positions of samples along the disturbance gradient, they had more difficulty establishing assessment thresholds to assign categories. For example, experts A2 and E2 did not differ from the median expert in sample rankings, but there was consistent directional deviation in their condition categories. Other examples of threshold setting being less consensual than ordering of samples were observed when experts gave the same rank to a sample, but disagreed on sample condition (e.g., experts E3 and E4 on samples EU_M5 or EU_M7; Table I, Table II). For both types of evaluation, the consensus was less clear near the middle of the disturbance gradient. From a management perspective, having good agreement at the ends of the gradient is of much less utility than having good agreement near its centre. This agreement is of particular importance in categorizations, since the classification of a site has practical implications whose consequences are most apparent at the good / bad threshold (Borja *et al.* 2009c).

The experts differed in the number of criteria they used for their assessments and those using more criteria generally showed less directional deviation in their category assignments. This is consistent with recommendations to use multiple metrics when assessing ecological status (Weisberg *et al.* 1997, Borja *et al.* 2004a, 2007, 2009a, Dauvin *et al.* 2007, Muxika *et al.* 2007, Borja & Dauer 2008, Lavesque *et al.* 2009). The most widely used criteria, used by the experts, are very similar to those determined by Alden *et al.* (2002), studying the relative discriminatory power of individual metrics within the Benthic Index of Biotic Integrity (B-IBI). However, the number of criteria used was not the only factor affecting individual expert tendencies. Experts who placed higher importance on dominance of tolerant, or presence of sensitive taxa often rated sites more negatively than the average expert. Contrarily, those who tended to classify samples in a better condition than the median, besides using considerably fewer attributes, often disregarded tolerant species presence, or sensitive species presence or both, or did not give any of these criteria the top importance. This manifested itself most in samples that had low species richness but with high quality species present; or those with high species richness but with a high percentage of *Capitella capitata* or other indicators of poor condition (e.g., US_E1; US_W3; EU_M5). The use of complementary criteria that allow measuring different attributes of the benthic communities is therefore advised, and the presence/dominance of indicator species should be among the attributes evaluated over the risk of misclassifying disturbed communities as undisturbed.

Some of the differences in how much emphasis experts placed on use of sensitive and tolerant taxa may have had to do with their comfort in identifying relevant taxa for these criteria outside of their home region. The experts suggested that this was less of a problem for sensitive taxa, which they tended to identify at higher taxonomic levels. In contrast, tolerant taxa tended to be identified at the species level and required local knowledge. For instance, in sample EU_A1 most European Atlantic and Mediterranean and US East Coast experts associated the dominance of *Amphiura filiformis* with organic enrichment, while US West Coast experts considered it a sensitive species. This raises the possibility that species occurring over wide geographical areas may indicate different ecological conditions in different regions. Benthic indices based on indicator species (e.g., AMBI; Borja *et al.* 2000) may need to adjust accordingly when expanding geographic application (Borja *et al.* 2008b).

Individual expert approaches to assigning condition categories and dealing with uncertainty explain many condition category differences. Some experts, assuming a balanced gradient of disturbance from good to bad, simply split the ranked samples in four classes; others divided the range of values for different metrics by four; and others assigned categories depending on how well benthic community characteristics fit their conceptual view of the categories. Additionally, in doubt, some experts attributed ties to samples. For samples in between categories or on category boundaries, some experts always chose the lower condition category.

However, the full gradient of disturbance might not have been truly achieved for all regional datasets, weakening the assumption of balanced samples across the four categories and contributing to the lower overall level of agreement on categorizations. The number of 'unaffected' and 'marginal deviation from unaffected' categories was higher for US East Coast samples (Table I) than the other regions, which had samples more evenly distributed across categories (Table IV). While the other regional samples were selected based on characteristics of the biological communities, US East Coast samples were selected based on abiotic factors, and the ERM values used as proxy for disturbance may not have accurately reflected the condition of the local benthic communities.

Another factor that contributed to discrepancies among experts was the challenge of distinguishing anthropogenic disturbance from natural stresses, which has been largely debated in estuaries (Dauvin 2007, Elliott & Quintino 2007). This was particularly notable for the US East Coast data, which were largely samples from coarse sediments subject to high wave energy or strong currents about which the experts had more disagreements than for the other regions. The high energy led to lower species richness (Hall 1994) than might otherwise be expected for euhaline areas in that geography. Some of the experts ranked the samples as stressed because of the lesser species richness, independent of the origin of that stress, whereas others recognized the communities as dominated by high energy species, such as bivalves, and modified their species richness expectations for the samples accordingly. Thus, the differences in evaluations for these samples can be attributable largely to differences in interpretation of guidelines on how to deal with natural versus anthropogenic stress.

This challenge associated with natural stress illustrates that the level of agreement among experts we observed was probably a minimum estimate, as we withheld information that they might usually have had when making an assessment. In typical assessments, the experts would know the specific sample locations, which we did not share to avoid interference due to local expert knowledge about particular sites. For example, the experts may have used location specific information to lower their species richness expectations based on susceptibility to wave energy stress. We probably also underestimated the true level of consensus because we asked the experts to conduct their assessments in isolation, where normally they would probably confer with their peers. Following submittal of their site assessments, we held a conference call among the experts to investigate factors that led to differences among them. In many cases, experts deviating from the median indicated that hearing the perspectives (such as the potential for wave energy influence) of the other experts would have caused them to change their assessment toward the median, if they had been provided that opportunity.

While there were some sites where the experts disagreed, the generally high level of agreement in our study seems to confirm the European WFD suggestion that BPJ is a viable means for calibrating indices of ecosystem condition (Borja 2005). More importantly, the agreement we observed across large geographies suggests a means for creating a common calibration scale that allows national and international comparisons of benthic condition. While the data set from this study has value in that context, it also needs to be expanded. There are many other geographies and habitats that were not included here, including estuarine habitats. More importantly, we used the four assessment categories used by Weisberg *et al.* (2008) and there is a need to map those to the five ecological quality classes on which the WFD is based or to any other new assessment scheme. Category classifications are important because they usually are the basis for different environmental regulatory and management requirements, which may be associated with substantially different cost. Based on the consistency in sample ranking among the experts in the present study, we expect this mapping will easily be accomplished.

# General discussion

This last section of the thesis synthesizes and discusses its contribution to increase the knowledge on the use of macroinvertebrates as indicators of ecosystems' ecological condition, taking profit from the study of both marine and estuarine communities from a wide geographical range. Major findings will be discussed in the scope of current approaches to environmental assessment worldwide, aiming at understanding the most important patterns involved in species response to anthropogenic disturbance.

## Natural driving forces structuring benthic macroinvertebrate communities and reference conditions

When benthic macroinvertebrates are used to identify possible impacts on marine environments, previous knowledge on what we should expect to find under natural undisturbed conditions is required. However, the wide geographic scale study presented in Chapter V has shown that, at least for temperate regions, the ignorance of pristine conditions is balanced by the application of sound theoretical concepts on balanced marine benthic communities. This was reflected on the consistency found among scientists at the time to select attributes to measure the health of coastal marine macroinvertebrate communities from across the Pacific, Atlantic and Mediterranean. It was almost consensual that the dominance by tolerant taxa and/or the presence of sensitive taxa are crucial at the time to evaluate biological communities' response to disturbance. Also, some of the main attributes selected (biodiversity number of taxa measures; abundance dominance patterns; biodiversity community measures; and abundance) concur with the different facets of diversity as pointed by several authors (e.g., Gaston 1996, Gray 2000, Magurran 2004). Diversity is widely perceived to correlate with environmental well being, and therefore diversity measures of various kinds are playing an increasing role in environmental assessment (Magurran 2004). The achieved consensus allowed that benthic experts from different geographies, provided with limited information about habitat (region, salinity, depth, and percent fines in sediment), and setting their expectations for balanced macroinvertebrate communities, could reach high agreement when evaluating their ecological condition along a disturbance gradient.

Nevertheless, although several important ecological principles that explain communities' response to disturbance are universal (e.g., the Pearson-Rosenberg (1978) conceptual model of response of benthic communities to organic enrichment), species composition, number of species and abundance are shaped by regional processes (Gaston 2000, Gaston & Blackburn 2000). In fact, from estuarine, to marine bays to costal marine environments, the different case studies approached in this thesis demonstrate that geographic, ecosystem and habitat specificities contribute to determine local macroinvertebrate communities, and therefore should be taken into account at the time of assessing their ecological condition. For example, the US East Coast data, used in the Best Professional Judgement study, consisted mostly of samples from coarse sediments subject to high wave energy or strong currents. The high energy led to lower species richness (Hall 1994) than might otherwise be expected for euhaline areas in that geography. More accurate assessments would be achieved if information on such local processes that determine communities' structure and interfere with species distribution had been taken into account.

However, from the studied environments, the estuarine one was undoubtedly the most challenging from an ecological assessment perspective. Any approach to the reference conditions issue in such transitional waters ecosystems must account for strong gradients, coupled furthermore with the great daily and seasonal natural oscillations that organisms must cope with. Hence, reference conditions cannot be the same throughout the entire system but should accompany natural trends instead (Weisberg *et al.* 1997, de Paz *et al.* 2008). Understand and define those trends and propose homogeneous habitats for benthos in the Mondego estuary was a fundamental step on the process of defining an adequate assessment tool in the scope of the WFD. As for most temperate estuaries (Holland *et al.* 1987), salinity and sediment type associated to the hidrological regime, were major factors influencing macrobenthic species composition and abundance patterns in the Mondego. As generally observed in estuaries (McLusky & Elliot 2004, Attrill & Rundle 2002), a natural lower species richness and evenness is observed towards their inner parts. Moreover, tolerant species, expected to occur across a wide range of environmental conditions, assume particular importance in estuaries, where they clearly dominate communities (Dauvin *et al.* 2007, Elliot & Quintino 2007, Blanchet *et al.* 2008, Lavesque *et al.* 2009). This effect was also observed in the Mondego, especially in its inner areas, which are additionally naturally organic enriched habitats, where the dominance of such tolerant species is even stronger. The ecological indices tested (Shannon-Wiener, H'; Margalef, d; and AMBI), which will be potentially used for quality assessment in the scope of the WFD, were partially correlated with the environmental gradients found within the estuary. Such abiotic influence on the communities, and consequently on the ecological indices, was accounted for when defining reference values for the parameters to be assessed as they interfere with species richness (measured by the Margalef index), with the community equitability (expressed by the Shannon-Wiener diversity), and with the balance between sensitive/tolerant/opportunistic species (measured by the AMBI).

This thesis work evidenced also that for temperate estuaries, the complete range of ecological quality (EQS) as defined by the expected proportions of sensitive/opportunist taxa within the Water Framework Directive (EC 2000, annex V) should be reviewed to meet real expectations for transitional waters communities. In such unstable environments, as also demonstrated by other studies (Borja *et al*. 2008a, Lavesque *et al*. 2009), biological communities show more difficulty in reaching the climax succession stage. Similar observations were registered for the Southern California, where differences between the mainland shelf and embayment fauna prevented the use of identical thresholds in the two habitats (Smith *et al*. 2003). For example, hardly any of the echinoderm species used to establish the threshold for loss of community function in the mainland shelf BRI occur in bays (Smith *et al*. 2003). In this case, marine bays assessment thresholds were based on increasing losses in biodiversity from pre-defined reference condition. Under higher oscillations due to stronger land influence than more marine areas, those undisturbed biological communities hardly attain taxonomic compositions matching those of undisturbed communities as defined by the AMBI (Borja *et al*. 2000).

Despite the fact that splitting biological continua is a difficult task and will always be at some extent artificial (Gray 2000, Magurran 2004), such procedure will allow measuring the departure of metrics addressing comparable ecological aspects of the community from the expected reference condition. Still, reference conditions that support robust ecological assessments are difficult to define, not only due to the difficulty in understanding all processes that determine species occurrence but also to define when and where does a situation actually correspond to a truly unimpacted scenario. In the present work, two approaches were tested which are particularly useful for long disturbed ecosystems, for which pristine conditions will hardly be found. The first approach used the history of disturbance of the system to track how selected ecological indices captured its degradation/recovery through time. Relating the indices with specific events allows understanding the approximate range of values one might expect for that system with a given set of indicators and how they relate to different levels of stress in a benchmarking kind of approach (Birk & Hering 2009). This provided us however only with least disturbed scenarios and will oblige an extended monitoring to gradually adjust it to what would be the real reference state. The Mondego estuary case study showed furthermore that this approach increased the knowledge on the ecological evolution of its benthic macroinvertebrate communities and such results along with other information supported some of the management options already undertaken in the system. A second and completely distinct approach proposed the use of expert judgement to define a scale of disturbance that relates to a particular habitat. Temperate coastal marine habitats were used as an example and the promising level of agreement achieved among experts enables such scale to provide, among other uses, a comparable reference state description that can be used in the absence of undisturbed sites.

We must always keep in mind that the global changes and prolonged multiple pressures might lead to thresholds trespassing and probably new equilibria establishment (Scheffer *et al*. 2001, Marques *et al*. 2003, Andersen *et al*. 2009). So the evolution observed within biological communities might be happening in a new direction from that expected.

## Natural variability *vs.* anthropogenic disturbance

Although an assessment tool is required to accurately translate shifts in benthic community due to anthropogenic influence, this is often a hard goal to reach. Ideally, tools should be able to cope with changes caused by natural oscillations or by natural disturbance events. However, especially in transitional waters ecosystems, such as estuaries, due to their unstable nature, too often the changes observed in macrobenthic communities' parameters are equivalent regardless of the nature of the events that promote them (Dauvin 2007). In the Mondego estuary, natural variability leads to the establishment of naturally stressed benthic macroinvertebrate communities which can mask the effects of light or initial stages of anthropogenic disturbance. Such misevaluations can also occur in marine environments when natural stress, such as the high wave energy effect in the US East Coast samples, leads to lower species richness than one might otherwise expect for euhaline areas in that geography (Hall 1994).

Natural variability is even more difficult to discern from anthropogenic effects when it involves the occurrence of extreme climatic events, which are moreover increasing in the present global climatic change scenario. Also, depending on the kind of event and system considered, it will interfere differently in the biological communities. For example, a flood event in the Mondego estuary produced more significantly impacts within downstream communities than on the upstream ones, which are more adapted to freshwater inputs. These phenomena have been reported to delay the recovery process of benthic communities as measured by secondary production levels (Dolbeth *et al.* 2007). Here, it was observed that the different ecological indices were also sensitive to them, although not equally affected; for instance, the Shannon-Wiener index, which is very sensitive to equitability, was the most responsive to alterations in the community due to flood related events, such as following recruitments.

Clarify the power of the tools to discriminate between changes in ecological quality driven by natural or anthropogenic disturbance is critical for an efficient management. However, since the effects of natural changes on benthic communities usually mime those of anthropogenic disturbances, climatic, physical, and hydromorphological supporting descriptors cannot be forgotten at the time to unravel the origin of the disturbance.

## Ecological assessment tools and indices performance

A 'metric' is defined as a measurable part or process of a biological system empirically shown to change in value along a gradient of human influence (Karr & Chu 1997). Chosen parameters must reflect specific and predictable responses of the community to human activities, either to a single impact factor or to the cumulative effects of multiple human impairments within a given system (Schmedtje *et al*. 2009).

The Margalef, the Shannon-Wiener and the AMBI indices applied in different case studies in this thesis have shown to be capable of detecting shifts within benthic macroinvertebrate communities due to anthropogenic pressures. In the Mondego estuary study case species richness and heterogeneity diversity measures were the most effective in detecting two very common pressures in estuaries worldwide: eutrophication and physical disturbances due to dredging activities. Both the Margalef and the Shannon-Wiener indices were able to track temporal differences in the benthic communities' ecologic condition, responding to periods of higher impact and subsequent system recovery; and also spatial differences within the system between areas under the influence of distinct impacts. The AMBI, which accounts for taxonomic composition, showed less effective in this type of systems and the major causes were found to be a) the dominance of tolerant species within such naturally stressed environments which seems to be forcing its classification and b) the classification of key species ecological behaviour (e.g. *Alkmaria romijni*). The testing of the AMBI against the BRI index and expert judgement in a different environment, demonstrated also that both these aspects can interfere greatly with this index performance. In the two case studies, the Mondego estuary and the Southern California marine bays, tolerant species from the AMBI EG III were the most abundant individuals and greatly influenced the classification into slight to mid levels of disturbance. Hence, it is suggested that, regarding its application to estuarine and marine bays environments, the index would perform better if highly abundant species that dominate communities would be down-weighted, for e.g. using abundance data previously fourth root transformed. This is in agreement with what Warwick *et al*. (2010) observed when testing other transformations for the AMBI, when applied to fully marine habitats. They found evidences that the impact 'signal' is better captured with modestly-transformed data. Smith *et al*. (2001, Appendix A) also observed that less severe data transformations were necessary for the BRI when applied to such environments to reproduce more accurately the identified pollution gradients. On the contrary, for marine bays and estuaries, severe abundance data transformations gave BRI a better performance. Despite these evidences, other studies argue that in marine ecosystems such as the Mediterranean (Simboura & Argyrou *in press*), an increase in the weight of the EGIII group (critical for this ecoregion) in the MEDOCC and the BENTIX (two indices which share AMBI principles) would enhance their effectiveness. Indices must therefore be calibrated to the geographies/habitats they are to assess to ensure more accurate results.

Due to the highlighted weaknesses and to the different features of the benthic community captured by each index, the combined use of multiple variables, methods, or analyses with different assumptions is therefore suggested to ensure more reliable assessments (Dauer 1993, Gray 2000). The Best Professional Judgement exercise showed precisely that the experts using more criteria generally showed less directional deviation in their category assignments, which is consistent with recommendations to use multiple metrics when assessing ecological status (Weisberg *et al*. 1997, Alden *et al*. 2002, Borja *et al.* 2004a, 2007, 2009a, Dauvin *et al*. 2007, Muxika *et al*. 2007, Borja & Dauer 2008, Lavesque *et al*. 2009). Also experts that placed higher importance on the dominance of tolerant species or presence of sensitive taxa often had more demanding assessments than the average expert. This explains partially why the three indices applied in the Mondego did better when pooled together in the BAT multimetric tool than when used independently to track anthropogenic disturbance in the estuary. The BAT multimetric tool captures more efficiently the different aspects of the biotic communities that may be affected by disturbance, since each index respond in distinct ways to distinct intensities and/or types of disturbance. The BAT tool still needs some refinement for its use in estuarine systems, not only with respect to the AMBI as already referred before, but also regarding the boundaries between disturbance classes for the final quality classification.

The classification of the biological communities' ecological condition is indeed a crucial step in assessment procedures. One of the difficulties in applying indices is to define the thresholds between status classes along a gradient of disturbance. The Best Professional Judgement exercise demonstrated that, despite the high agreement achieved between experts ranking samples according to their level of disturbance, the definition of their condition categories was a challenge on its own. Moreover it was observed that for middle levels of disturbance, which is were commonly ecological indices tend to perform worse (see for instance Chapter III), even experts have experienced the greater difficulty in setting boundaries. Nevertheless, the generally high level of agreement in our study seems to confirm the European WFD suggestion that Best Professional Judgement is a viable approach for calibrating indices of ecosystem condition (Borja 2005). More importantly, the agreement we observed across large geographies suggests a manner of creating a common calibration scale that allows national and international comparisons of benthic condition. Based on the consistency in sample ranking among the experts, and following the rational for the assessments, this mapping will easily be accomplished. Moreover it can be extended to habitats and biotic elements other than the ones referred in this thesis. In the end, in any type of conservation strategy, determining limits or a safe distance involves normative judgements of how societies choose to deal with risk and uncertainty (Rockström *et al*. 2009).

## Species ecological strategies

The taxonomic composition of the communities is valuable information at the time to assess their ecological condition. Many experts and worldwide conservation strategies advise for its inclusion in environmental assessments. However, in the present thesis it was demonstrated that such procedure is as determinant in the success of benthic assessments as it is difficult to achieve.

There are several possible approaches to determine species ecological strategies and hence their sensitive or opportunistic character in face of anthropogenic impacts. Experimental field evidence of species ecology would be the most reliable way to achieve this information, but for rarer species, harsher habitats, or even wide scale assessments requiring the classification of the majority of taxa, this would not a feasible option. The present work allowed seeing that across geographies and methodologies there are many discrepancies exist regarding the classification of species, which reveals that this is a topic demanding further research.

The BPJ exercise revealed that differences in the emphasis put by experts on using sensitive and tolerant taxa may have been related to their level of uncertainty regarding the identification of relevant taxa for these criteria outside of their home region. In fact, experts suggested that this was less of a problem for sensitive taxa, more often identified at higher taxonomic levels, than for the tolerant ones, which usually require to be identified at the species level, implying sound local knowledge. Indeed, the example of the species *Amphiura filiformis* showed that this distinction between sensitive/opportunist taxa is not always clear. The fact that US West Coast experts considered it a sensitive species, while most European Atlantic and Mediterranean and US East Coast experts associated it to organic enrichment, raises the possibility that populations of the same species, occurring over a wide range of geographical areas, may indicate different ecological conditions in different regions.

Still, even uncertainties regarding species behaviour within the same geographical area can be observed. For instance, studies on the Mondego estuary revealed that key species in this system classified as tolerant according to the AMBI, exhibited actually an opportunistic strategy (Pardal *et al*. 2004). Consequently, its reclassification would improve the tracking of organic enrichment disturbance events within the system. In fact, the AMBI was primarily developed for marine habitats and its use in estuarine environments, even of the same geographical area, needs adjustments. The classification of species is one of the procedures that require revision.

Does this mean that a given species is able to exhibit a distinct ecological strategy according to geography, habitat or even type of disturbance? In a much wider scale, the study comparing the European born AMBI with the American BRI showed that for approximately one third of the species from the Southern California marine bays local experts do not seem to agree with AMBI prior classifications. Such mismatch may be either consequence of distinct interpretation of the Ecological Group (EG) concept or resultant from distinct knowledge on species behaviour in the new geography. However, using either AMBI previous classifications or the local experts' ones, there are still important disagreements on species ecological classification between the AMBI and the BRI methods, even if we would expect that local experts would converge more with local BRI tolerance scores for the species. The BRI has however been validated for the region and its performance tested against several indices and expert judgement. Thus, these findings keep feeding the discussion on whether or not a species ecological strategy is volatile or if there is a lack of knowledge of species behaviour across wide ranges of environmental conditions to correctly classify them.

The AMBI assumes clearly that species ecological behaviour is intrinsic and maintains irrespective of habitat or geography, and therefore each species is bonded to a single ecological group (EG) classification worldwide. On the other hand, the BRI assumes that species behaviour is habitat dependent, with the numbers and kinds of benthic animals that occur in reference areas varying naturally by habitat (Smith *et al*. 2003). One of the aims of a pan-European scale study (Grémare *et al*. 2009) was to test the validity of the use of a single list of sensitivity/tolerance levels by comparing another index using species classification, the BQI $E(S_{50})_{0.05}$, between subareas, covering both marine and estuarine habitats. Corroborating the Smith *et al*. (2001) studies, they too gathered strong evidence that the species sensitivity/tolerance levels presented marked changes according to the geographical location. This questions the universal principle of species ecological behaviour beneath the AMBI and helps in elucidating findings of the present study, namely regarding species for which important discrepant classifications were found across methods and criteria to classify them (Annex A).

# Suggestions for future research

Through the course of this work several questions often arose, which remain unanswered. Due to time and data constraints it was not possible, in the scope of this thesis, to explore them further. Nonetheless, such issues constitute important topics to the field of ecological assessment, and are referred here as suggestions for future research:

a) The BAT multimetric tool was validated in a system where the major pressures were eutrophication and physical disturbance. But how would it respond in face of other types of disturbance (e.g. chemical pollution, hypoxia events, *etc*)?

b) One key issue in ecological assessments is determining cause/effect relationships between disturbances and ecological indices responses. Often the impacts are difficult to quantify due to problems in interpreting the proxies studied or to the lack of adequate ones. Clarify these aspects is important to unravel differences between indices performance, which could contribute to a more robust choice of indicators, especially in multiple impacted sites.

c) How will continued pressures and global climatic changes affect biological communities, namely macroinvertebrates? Will those conditions affect their resilience? And, if such shifts lead to new equilibria, how will this affect the reference conditions against which we measure disturbance related departures?

d) Finally, further research will be necessary on critical species to clarify their sensitivity/tolerance level, allowing more robust ecological assessments, and understanding, from a macroecological perspective, which factors interfere with species ecological strategies.

# References


Adams DA, O'Connor JS, Weisberg SB, 1998. Sediment quality of the NY/NJ Harbor system. Final report. An investigation under the Regional Environmental Monitoring and Assessment Program (REMAP). U.S. Environmental Protection Agency, EPA/902/R-98/001, Edison, NJ.

Afli A, Ayari R, Zaabi S, 2008. Ecological quality of some Tunisian coast and lagoon locations, by using benthic community parameters and biotic indices. Estuarine, Coastal and Shelf Science 80: 269-280.

Alden RW, Dauer DM, Ranasinghe JA, Scott LC, Llansó R, 2002. Statistical verification of the Chesapeake Bay benthic index of biotic integrity. Environmetrics 13: 473-498.

Andersen T, Carstensen J, Hernández-García E, Duarte CM, 2009. Ecological thresholds and regime shifts: approaches to identification. Trends in Ecology and Evolution 24: 49-57.

Anderson MJ, 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecology 6: 32-46.

Anderson MJ, 2005. PERMANOVA: a FORTRAN computer program for permutational multivariate analysis of variance. Department of Statistics, University of Auckland, New Zealand.

Anonymous, 1959. Symposium on the classification of brackish waters, Venice 8-14th April 1958. Archivio di Oceanografia e Limnologia 11 (supplemento).

Antunes P, Santos R, 1999. Integrated environmental management of the oceans. Ecological Economics 31: 215-226.

APHA–American Public Health Association. 1995. Standard Methods for the Examination of Water and Wastewater. American Public Health Association, American Water Works Association & Water Environment Federation (eds). United Book Press, Baltimore, Maryland.

Arévalo R, Pinedo S, Ballesteros E, 2007. Changes in the composition and structure of Mediterranean rocky-shore communities following a gradient of nutrient enrichment: Descriptive study and test of proposed methods to assess water quality regarding macroalgae. Marine Pollution Bulletin 55: 104-113.

Attrill MJ, 2002. A testable linear model for diversity trends in estuaries. Journal of Animal Ecology 71: 262-269.

Attrill MJ, Power M, 2000. Effects on invertebrate populations of drought-induced changes in estuarine water quality. Marine Ecology Progress Series 203: 133-143.

Attrill MJ, Rundle SD, 2002. Ecotone or ecocline: ecological boundaries in estuaries. Estuarine, Coastal and Shelf Science 55: 929-936.

Bakalem A, Ruellet T, Dauvin JC, 2009. Benthic indices and ecological quality of shallow Algeria fine sand community. Ecological Indicators 9: 395-408.

Bald J, Borja A, Muxika I, Franco J, Valencia V, 2005. Assessing reference conditions and physico-chemical status according to the European Water Framework Directive: a case-study from the Basque Country (Northen Spain). Marine Pollution Bulletin 50: 1508-1522.

Ballesteros E, Torras X, Pinedo S, García M, Mangialajo L, de Torres M, 2007. A new methodology based on littoral community cartography dominated by macroalgae for the implementation of the European Water Framework Directive. Marine Pollution Bulletin 55: 172-180.

Bay S, Berry W, Chapman PM, Fairey R, Gries T, Long E, McDonald D, Weisberg SB, 2007. Evaluating consistency of best professional judgment in the application of a multiple lines of evidence sediment quality triad. Integrated Environmental Assessment and Management 3: 491-497.

Beck MW, Heck Jr KL, Able KW, Childers DL, Eggleston DB, Gillanders BM, Halpern B, Hays CG, Hoshino K, Minello TJ, Orth RJ, Sheridan PF, Weinstein MP, 2001. The Identification, Conservation, and Management of Estuarine and Marine Nurseries for Fish and Invertebrates. Bioscience 51: 633-641.

Belsher T, 1982. Measuring the standing crop of intertidal seaweeds by remote sensing. In: Last FT, Hotz MC, Bell B (Eds.), Land and its Uses Actual and Potential: An Environmental Appraisal. NATO Seminar on Land and its Uses, Edinburgh 19 Sep.-1 Oct. 1982, 10: 453-456.

Bergen M, Weisberg SB, Cadien D, Dalkey A, Montagne D, Smith RW, Stull JK, Velarde RG, 1998. Southern California Bight Pilot Project: IV. Benthic infauna. Southern California Coastal Water Research Project. Westminster, CA.

Berger WH, Parker FL, 1970. Diversity of planktonic Foraminifera in deep-sea sediments. Science 168: 1345-1347.

Best MA, Wither AW, Coates S, 2007. Dissolved oxygen as a physico-chemical supporting element in the Water Framework Directive. Marine Pollution Bulletin 55: 53-64.

Bettencourt A, Bricker SB, Ferreira JG, Franco A, Marques JC, Melo JJ, Nobre A, Ramos L, Reis CS, Salas F, Silva MC, Simas T, Wolff W, 2004. Typology and reference conditions for Portuguese transitional and coastal waters. INAG - Instituto da Água, IMAR - Institute of Marine Research, Portugal, 99 pp.

Beukema JJ, 1988. An evaluation of the ABC-method (abundance/biomass comparison) as applied to macrozoobenthic communities living on tidal flats in the Dutch Wadden Sea. Marine Biology 99: 425-433.

Bigot L, Grémare A, Amouroux J, Frouin P, Maire O, Gaertner JC, 2008. Assessment of the ecological quality status of soft-bottoms in Reunion Island (tropical Southwest Indian Ocean) using AZTI marine biotic indices. Marine Pollution Bulletin 56: 704-722.

Birk S, Hering D, 2006. Direct comparison of assessment methods using benthic macroinvertebrates: a contribution to the EU Water Framework Directive intercalibration exercise. Hydrobiologia 566: 401-415.

Birk S, Hering D, 2009. A new procedure for comparing class boundaries of biological assessment methods: A case study from the Danube Basin. Ecological Indicators 9: 528-539.

Blanchet H, Lavesque N, Ruellet T, Dauvin JC, Sauriau PG, Desroy N, Desclaux C, Leconte M, Bachelet G, Janson AL, Bessineton C, Duhamel S, Jourde J, Mayot S, Simon S, De Montaudouin X, 2008. Use of Biotic Indices in semi-enclosed coastal ecosystems and transitional waters habitats—implications for the implementation of the European Water Framework Directive. Ecological Indicators 8: 360-372.

Boesch DF, 1999. The role of science in ocean governance. Ecological Economics 31: 189-198.

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

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

Borja A, Heinrich H, 2005. Implementing the European Water Framework Directive: the debate continues. Marine Pollution Bulletin 50: 486-488.

Borja A, Muxika I, 2005. Guidelines for the use of AMBI (AZTI's Marine Biotic Index) in the assessment of the benthic ecological quality. Marine Pollution Bulletin 50: 787-789.

Borja A., Tunberg BG, *in press*. Assessing benthic health in stressed subtropical estuaries, eastern Florida, USA using AMBI and M-AMBI. Ecological Indicators.

Borja A, Franco J, Perez V, 2000. A marine Biotic Index to establish the ecological quality of soft-bottom benthos within European estuarine and coastal environments. Marine Pollution Bulletin 40: 1100-1114.

Borja A, Muxika I, Franco J, 2003. The application of a Marine Biotic Index to different impact sources affecting soft-bottom benthic communities along European coasts. Marine Pollution Bulletin 46: 835-845.

Borja A, Franco J, Valencia V, Bald J, Muxika I, Belzunce MJ, Solaun O, 2004a. Implementation of the European Water Framework Directive from the Basque Country (northern Spain): a methodological approach. Marine Pollution Bulletin 48: 209-218.

Borja A, Franco J, Muxika I, 2004b. The biotic indices and the Water Framework Directive: the required consensus in the new benthic monitoring tools. Marine Pollution Bulletin 48: 405-408.

Borja A, Galparsoro I, Solaun O, Muxika I, Tello EM, Uriarte A, Valencia V, 2006. The European Water Framework Directive and the DPSIR, a methodological approach to assess the risk of failing to achieve good ecological status. Estuarine, Coastal and Shelf Science 66: 84-96.

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

Borja A, Bricker SB, Dauer DM, Demetriades NT, Ferreira JG, Forbes AT, Huchins P, Jia X, Kenchington R, Marques JC, Zhu C, 2008a. Overview of integrative tools and methods in assessing ecological integrity in estuarine and coastal systems worldwide. Marine Pollution Bulletin 56: 1519-1537.

Borja A, Dauer D, Diaz R, Llanso RJ, Muxika I, Rodríguez JG, Schaffner L, 2008b. Assessing estuarine benthic quality conditions in Chesapeake Bay: a comparison of three indices. Ecological Indicators 8: 395-403.

Borja A, Mader J, Muxika I, Rodriguez JG, Bald J, 2008c. Using M-AMBI in assessing benthic quality within the water framework directive: some remarks and recommendations. Marine Pollution Bulletin 56: 1377-1379.

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

Borja A, Rodríguez JG, Black K, Bodoy A, Emblow C, Fernandes TF, Forte J, Karakassis I, Muxika I, Nickell TD, Papageorgiou N, Pranovi F, Sevastou K, Tomassetti P, Angel D, 2009b. Assessing the suitability of a range of benthic indices in the evaluation of environmental impact of fin and shellfish aquaculture located in sites across Europe. Aquaculture 293: 231-240.

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

Boyd SE, Limpenny DS, Rees HL, Cooper KM, Campbell S, 2003. Preliminary observations of the effects of dredging intensity on the re-colonisation of dredged sediments off the southeast coast of England (Area 222). Estuarine, Coastal and Shelf Science 57: 209-223.

Boyd SE, Limpenny DS, Rees HL, Cooper KM, 2005. The effects of marine sand and gravel extraction on the macrobenthos at a commercial dredging site (results 6 years post-dredging). ICES Journal of Marine Science 62: 145-162.

Breine JJ, Maes J, Quataert P, van den Berg 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.

Briggs JC, 1994. Species diversity: land and sea compared. Systematic Biology 43: 130-135.

Brown AC, McLachland A, 1990. Ecology of Sandy Shores. Elsevier. Amsterdam, 328 pp.

Buffagni A, Erba S, Cazzola M, Murray-Bligh J, Soszka H, Genoni P, 2006. The STAR common metrics approach to the WFD intercalibration process: Full application for small, lowland rivers in three European countries. Hydrobiologia 566: 379-399.

Bulger AJ, Hayden BP, Monaco ME, Nelson DM, McCormick-Ray MG, 1993. Biologically-based estuarine salinity zones derived from a multivariate analysis. Estuaries 16: 311-322.

Burton Jr GA, 2002. Sediment quality criteria in use around the world. Limnology 3: 65-76.

Burton Jr GA, Chapman PM, Smith EP, 2002. Weight of evidence approaches for assessing ecosystem impairment. Human and Ecological Risk Assessment 8: 1657-1673.

Bustos-Baez S, Frid C, 2003. Using indicator species to assess the state of macrobenthic communities. Hydrobiologia 496: 299-309.

Cabral JA, Pardal MA, Lopes RJ, Múrias T, Marques JC, 1999. The impact of macroalgal blooms on the use of the intertidal area and feeding behaviour of waders (Charadrii) in the Mondego estuary (West Portugal). Acta Oecologica 20: 417-428.

Cai L, Tam NFY, Wong TWY, Ma L, Gao Y, Wong YS, 2003. Using Benthic Macrofauna to Assess Environmental Quality of Four Intertidal Mudflats in Hong Kong and Shenzhen Coast. Acta Oceanologica Sinica 22: 309-319.

Cairns J, McCormick PV, Niederlehner BR, 1993. A proposed framework for developing indicators of ecosystem health. Hydrobiologia 263: 1-44.

Caldeira K, Wickett ME, 2003. Oceanography: Anthropogenic carbon and ocean pH. Nature 425: 365.

Callier MD, McKindsey CW, Desrosiers G, 2008. Evaluation of indicators used to detect mussel farm influence on the benthos: Two case studies in the Magdalen Islands, Eastern Canada. Aquaculture 278: 77-88.

Cardoso PG, Pardal MA, Lillebø AI, Ferreira SM, Raffaelli D, Marques JC, 2004a. Dynamic changes in seagrass assemblages under eutrophication and implications for recovery. Journal of Experimental Marine Biology and Ecology 302: 233-248.

Cardoso PG, Pardal MA, Raffaelli D, Baeta A, Marques JC, 2004b. Macroinvertebrate response to different species of macroalgal mats and the role of disturbance history. Journal of Experimental Marine Biology and Ecology 308: 207-220.

Cardoso PG, Raffaelli D, Pardal MA, 2007. Seagrass beds and intertidal invertebrates: an experimental test to the role of habitat structure. Hydrobiologia 575: 221-230.

Carletti A, Heiskanen AS (Eds), 2009. Water Framework Directive Intercalibration Technical Report. Part 3: Coastal and Transitional Waters. JRC Scientific and Technical Reports, EUR 23838 EN/3: 240 pp.

Carpenter SR, Caraco NF, Correll DL, Howarth RW, Sharpley AN, Smith VH, 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. Ecological Applications 8: 559-568.

Carson R, 1962. Silent Spring. Houghton Mifflin, United States.

Carvalho S, Gaspar MB, Moura A, Vale C, Antunes P, Gil O, Fonseca LC, Falcão M, 2006. The use of the marine biotic index AMBI in the assessment of the ecological status of the Obidos lagoon (Portugal). Marine Pollution Bulletin 52: 1414-1424.

Carvalho L, Solimini A, Phillips G, van den Berg M, Pietiläinen O-P, Solheim AL, Poikane S, Mischke U, 2008. Chlorophyll reference conditions for European lake types used for intercalibration of ecological status. Aquatic Ecology 42: 203-211.

Chainho P, Costa JL, Chaves ML, Lane MF, Dauer DM, Costa MJ, 2006. Seasonal and spatial patterns of distribution of subtidal benthic invertebrate communities in the Mondego River, Portugal – a poikilohaline estuary. Hydrobiologia 555: 59-74.

Chainho P, Costa JL, Chaves ML, Dauer DM, Costa MJ, 2007. Influence of seasonal variability in benthic invertebrate community structure on the use of biotic indices to assess the ecological status of a Portuguese estuary. Marine Pollution Bulletin 54: 1586-1597.

Chainho P, Chaves ML, Costa JL, Costa MJ, Dauer DM, 2008. Use of multimetric indices to classify estuaries with different hydromorphological characteristics and different levels of human pressure. Marine Pollution Bulletin 56: 1128-1137.

Chapman PM, 2007. Do not disregard the benthos in sediment quality assessments! Marine Pollution Bulletin 54: 633-635.

Chapman PM, Anderson J, 2005. A Decision-Making Framework for Sediment Contamination. Integrated Environmental Assessment and Management 1: 163-173.

Chapman PM, 2009. Borja *et al.*'s (2008) ''Overview of integrative tools and methods… worldwide' is curiously deficient. Marine Pollution Bulletin 59: 456.

Cicchetti DV, Allison T, 1971. A new procedure for assessing reliability of scoring EEG sleep recordings. American Journal of EEG Technology 11: 101-109.

CIS, 2003. Towards a Guidance on Establishment of the Intercalibration Network and the Process on the Intercalibration Exercise. Guidance Document No. 6. Working Group 2.5 – Intercalibration. Office for Official Publications of the European Communities. Luxembourg, 54 pp.

CIS, 2005. Guidance on the Intercalibration Process 2004-2006. Guidance Document No. 14. 2005. Office for Official Publications of the European Communities. Luxembourg, 31 pp.

Clarke KR, Warwick RM, 1999. The taxonomic distinctness measures of biodiversity: weighting of steps lengths between hierarchical levels. Marine Ecology Progress Series 184: 21-29.

Clarke KR, Warwick RM, 2001. Changes in Marine Communities: an Approach to Statistical Analysis and Interpretation. 2nd edition, Primer-E Ltd, Plymouth, UK.

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

Cohen J, 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20: 37-46.

Connell JH, 1978. Diversity in tropical rain forests and coral reefs. Science 199: 1302-1309.

Costanza R, 1992. Toward an Operational Definition of Ecosystem Health. *in* Costanza R, Norton B, Haskell B, (Eds), Ecosystem Health: New Goals for Environmental Management, Island Press, Washington, US. 269 pp.

Costanza R, 1999. The ecological, economic, and social importance of the oceans. Ecological Economics 31: 199-213.

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

Costanza R, Andrade F, Antunes P, van den Belt M, Boesch D, Boersma D, Catarino F, Hanna S, Limburg K, Low B, Molitor M, Pereira JG, Rayner S, Santos R, Wilson J, Young M, 1999. Ecological economics and sustainable governance of the oceans. Ecological Economics 31: 171-187.

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

Dame R, Bushek D, Allen D, Lewitus A, Edwards D, Koepfler E, Gregory L, 2002. Ecosystem response to bivalve density reduction: management Implications. Aquatic Ecology 36: 51-65.

Danielsen F, Sørensen MK, Olwig MF, Selvam V, Parish F, Burgess ND, Hiraishi T, Karunagaran VM, Rasmussen MS, Hansen LB, Quarto A, Suryadiputra N, 2005. The Asian Tsunami: A Protective Role for Coastal Vegetation. Science 310: 643.

Dauer DM, 1993. Biological criteria, environmental health and estuarine macrobenthic community structure. Marine Pollution Bulletin 26: 249-257.

Dauer DM, Alden RW, 1995. Long-Term Trends in the Macrobenthos and Water Quality of the Lower Chesapeake Bay (1985-1991). Marine Pollution Bulletin 30: 840-850.

Dauer DM, Conner WG, 1980. Effects of moderate sewage input on benthic polychaete populations. Estuarine and Coastal Marine Science 10: 335-346.

Dauer DM, Llansó RJ, 2003. Spatial scales and probability based sampling in determining levels of benthic community degradation in the Chesapeake Bay. Environmental Monitoring and Assessment 81: 175-186.

Dauer DM, Robinson WW, Seymour CP, Leggett Jr AT, 1979. The environmental impact of nonpoint pollution on benthic invertebrates in the Lynnhaven River System. Virginia Water Resources Research Center Bulletin 117, 112 pp.

Dauer DM, Luckenbach MW, Rodi Jr AJ, 1993. Abundance biomass comparison (ABC method): effects of an estuarine gradient, anoxic/hypoxic events and contaminated sediments. Marine Biology 116: 507-518.

Dauer DM, Ranasinghe JA, Weisberg SB, 2000. Relationships between benthic community condition, water quality, sediment quality, nutrient loads, and land use patterns in Chesapeake Bay. Estuaries 23: 80-96.

Dauvin JC, 2007. Paradox of estuarine quality: benthic indicators and indices in estuarine environments, consensus or debate for the future. Marine Pollution Bulletin 55: 271-281.

Dauvin JC, Ruellet T, 2007. Polychaete/amphipod ratio revisited. Marine Pollution Bulletin 55: 215-224.

Dauvin JC, Ruellet T, Desroy N, Janson AL, 2007. The ecological quality status of the Bay of Seine and the Seine estuary: use of biotic indices. Marine Pollution Bulletin 55: 241-257.

Dawson TP, Henderson SJ, Banks S, 2009. Galapagos Coral Conservation: impact, mitigation, mapping and monitoring. Galapagos Research 66: 3-5.

de Paz L, Patricio J, Marques JC, Borja A, Laborda AJ, 2008. Ecological status assessment in the lower Eo estuary (Spain). The challenge of habitat heterogeneity integration: A benthic perspective. Marine Pollution Bulletin 56: 1275-1283.

Deegan L, Finn JT, Ayvazlan SG, Ryder-Kieffer CA, Buonaccoesi J, 1997. Development and validation of an Estuarine Biotic Integrity Index. Estuaries 30: 601-617.

Depledge MH, 1990. New approaches in ecotoxicology: can inter-individual physiological variability be used as a tool to investigate pollution effects? Ambio 19: 251-252.

Depledge MH, 2009. Novel approaches and technologies in pollution assessment and monitoring: A UK perspective. Ocean & Coastal Management 52: 336-341.

Dethier MN, Schoch GC, 2005. The consequences of scale: assessing the distribution of benthic populations in a complex estuarine fjord. Estuarine, Coastal and Shelf Science 62: 253-270.

Devlin M, Best M, Coates D, Bresnan E, O'Boyle S, Park R, Silke J, 2007. Establishing boundary classes for the classification of UK marine waters using phytoplankton communities. Marine Pollution Bulletin 55: 91-103.

Diaz RJ, Rosenberg R, 1995. Marine benthic hypoxia: a review of its ecological effects and the behavioral responses of benthic macrofauna. Oceanography and Marine Biology: An Annual Review 33: 245-303.

Diaz RJ, Rosenberg R, 2008. Spreading dead zones and consequences for marine ecosystems. Science 321: 926-929.

Diaz RJ, Cutter GR, Dauer DM, 2003. A comparison of two methods for estimating the status of benthic habitat quality in the Virginia Chesapeake Bay. Journal of Experimental Marine Biology and Ecology 285/286: 371-381.

Diaz RJ, Solan M, Valente RM, 2004. A review of approaches for classifying benthic habitats and evaluating habitat quality. Journal of Environmental Management 73: 165-181.

Dolbeth M, Pardal MA, Lillebø AI, Azeiteiro U, Marques JC, 2003. Short term and long term effects of eutrophication on the secondary production of an intertidal macrobenthic community. Marine Biology 143: 1229-1238.

Dolbeth M, Lillebø AI, Cardoso PG, Ferreira SM, Pardal MA, 2005. Annual production of estuarine fauna in different environmental conditions: an evaluation of the estimation methods. Journal of Experimental Marine Biology and Ecology 326: 115–127.

Dolbeth M, Cardoso PG, Ferreira SM, Verdelhos T, Raffaelli D, Pardal MA, 2007. Anthropogenic and natural disturbance effects on a macrobenthic estuarine community over a 10-year period. Marine Pollution Bulletin 54: 576-585.

Duarte CM (Ed.) 2009. Global loss of coastal habitats. Rates, causes and consequences. Fundación BBVA.

Duarte CM, Cebrián J, 1996. The fate of marine autotrophic production. Limnology and Oceanography 41: 1758-1766.

Duarte CM, Middleburg J, Caraco N, 2005. Major role of marine vegetation on the oceanic carbon cycle. Biogeosciences 2: 1-8.

Duffy JE, Stachowicz JJ, 2006. Why biodiversity is important to oceanography: potential roles of genetic, species, and trophic diversity in pelagic ecosystem processes. Marine Ecology Progress Series 311: 179-189.

EC, 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for community action in the field of water policy. Official Journal of the European Communities, L327 (volume 43): 1-73.

EC, 2008. Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive). Official Journal of the European Union, L164 (volume 51): 19-40.

EEA, 2005. EEA core set of indicators—guide. European Environment Agency Technical Report, 1/005. Office for Official Publications of the European Communities, Luxembourg, 38 pp.

Elliot M, 1994. The analysis of macrobenthic community data. Marine Pollution Bulletin 28: 62-64.

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

Elliot M, McLusky DS, 2002. The need for definitions in understanding estuaries. Estuarine, Coastal and Shelf Science 55: 815-827.

Elliott M, Quintino V, 2007. The Estuarine Quality Paradox, Environmental Homeostasis and the difficulty of detecting anthropogenic stress in naturally stressed areas. Marine Pollution Bulletin 54: 640-645.

Engle VD, 2000. Application of the indicator evaluation guidelines to an index of benthic condition for Gulf of Mexico estuaries. In: Jackson LE, Kurtz JC, Fisher WS (Eds.). Evaluation Guidelines for Ecological Indicators. EPA/620/R-99/005. U.S. Environmental Protection Agency, Office of Research and Development, Research Triangle Park, NC. 107 pp.

Fabry VJ, Seibel BA, Feely RA, Orr JC, 2008. Impacts of ocean acidification on marine fauna and ecosystem processes. ICES journal of Marine Science 65: 414-432.

Feely RA, Sabine CL, Lee K, Berelson W, Kleypas J, Fabry VJ, Miller FJ, 2004. Impact of Anthropogenic $CO_2$ on the $CaCO_3$ System in the Oceans. Science 305: 362-366.

Ferreira JG, Nobre AM, Simas TC, Silva MC, Newton A, Bricker SB, Wolf WJ, Stacey PE, Sequeira A, 2006. A methodology for defining homogeneous water bodies in estuaries – Application to the transitional systems of the EU Water Framework Directive. Estuarine, Coastal and Shelf Science 66: 468-482.

Fisher WS, Jackson LE, Suter II GW, Bertram P, 2001. Indicators for Human and Ecological Risk Assessment: A.U.S. Environmental Protection Agency Perspective. Human and Ecological Risk Assessment 7: 961-970.

Fleiss JL, Cohen J, 1973. The equivalence of weighted Kappa and the intraclass correlation coefficient as measures of reliability. Educational and Psychological Measurement 33: 613-619.

Flindt MR, Kamp-Nielsen L, Marques JC, Pardal MA, Bocci M, Bendoricchio G, Salomonsen J, Nielsen SN, Jorgensen SE, 1997. Description of the three shallow estuaries: Mondego River (Portugal), Roskilde Fjord (Denmark), and the Lagoon of Venice (Italy). Ecological Modelling 102: 17-31.

Foden J, de Jong DJ, 2007. Assessment metrics for littoral seagrass under the European Water Framework Directive; outcomes of UK intercalibration with the Netherlands. Hydrobiologia 579: 187-197.

Garcia P, Zapico E, Colubi A, 2009. An angiosperm quality index (AQI) for Cantabrian estuaries. Ecological Indicators 9: 856-865.

Gaston KJ, 1996. Measuring Biodiversity: Complex Problems and a Simple Solution. Blackwell Science, 396 pp.

Gaston KJ, 2000. Global patterns in biodiversity. Nature 405: 220-227.

Gaston KJ, Blackburn TM, 2000. Pattern and process in macroecology. Blackwell Science, 377 pp.

Glémarec M, Hily C, 1981. Perturbations apportées à la macrofaune benthique de la baie de Concarneau par les effluents urbains et portuaires. Acta Oecologica Oecologia Applicata 2: 139-150.

Goldberg E, Bowen VT, Farrington JW, Harvey G, Martin PL, Parker PL, Risebrough RW, Robertson W, Schnider E, Gamble E, 1978. The Mussel Watch. Environmental Conservation 5: 101-125.

Gómez-Gesteira JL, Dauvin JC, 2000. Amphipods are good bioindicators of the impact of oil spills on soft bottom macrobenthic communities. Marine Pollution Bulletin 40: 1017-1027.

Grall J, Glémarec M, 1997. Using biotic indices to estimate macrobenthic community perturbations in the Bay of Brest. Estuarine, Coastal and Shelf Science 44: 43-53.

Grapentine L, Anderson J, Boyd D, Burton GA, DeBarros C, Johnson G, Marvin C, Milani D, Painter S, Pascoe T, Reynoldson T, Richman L, Solomon K, Chapman PM, 2002. A decision-making framework for sediment assessment developed for the Great Lakes. Human and Ecological Risk Assessment 8: 1641-1655.

Gray JS, 1979. Pollution-induced changes in populations. Philosophical Transactions of the Royal Society London Biological Sciences 286: 545-561.

Gray JS, 1997. Marine biodiversity: patterns, threats and conservation needs. Biodiversity and Conservation 6: 153-175.

Gray JS, 2000. The measurement of marine species diversity, with an application to the benthic fauna of the Norwegian continental shelf. Journal of Experimental Marine Biology and Ecology 250: 23-49.

Grémare A, Labrune C, Berghe EV, Amouroux JM, Bachelet G, Zettler ML, Vanaverbeke J, Fleischer D, Bigot L, Maire O, Deflandre B, Craeymeersch J, Degraer S, Dounas C, Duineveld G, Heip C, Herrmann M, Hummel H, Karakassis I, Kedra M, Kendall M, Kingston P, Laudien J, Occhipinti-Ambrogi A, Rachor E, Sardá R, Speybroeck J, Van Hoey G, Vincx M, Whomersley P, Willems W, Wlodarska-Kowalczuk M, Zenetos A, 2009. Comparison of the performances of two biotic indices based on the MacroBen database. Marine Ecology Progress Series 382: 297-311.

Guerra-García JM, García-Gómez JC, 2006. Recolonization of defaunated sediments: Fine versus gross sand and dredging versus experimental trays. Estuarine, Coastal and Shelf Science 68: 328-342.

Guerra-García JM, Corzo J, García-Gómez JC, 2003. Short-term benthic recolonization after dredging in the harbour of Ceuta, North Africa. Marine Ecology 24: 217-229.

Guinote JM, Fabry VJ, 2008. Ocean Acidification and Its Potential Effects on Marine Ecosystems. Annals of the New York Academy of Sciences 1134: 320-342.

Hale SS, Heltshe JF, 2008. Signals from the benthos: Development and evaluation of a benthic index for the nearshore Gulf of Maine. Ecological Indicators 8: 338-350.

Hall SJ, 1994. Physical disturbance and marine benthic communities: life in unconsolidated sediments. Oceanography and Marine Biology: an Annual Review 32: 179-239.

Hall Jr LW, Dauer DM, Alden III RW, Uhler A, Burton DT, 2005. An Integrated Case Study for Evaluating the Impacts of an Oil Refinery Effluent on Aquatic Biota in the Delaware River: Sediment Quality Triad Studies. Human and Ecological Risk Assessment 11: 657-770.

Hall-Spencer JM, Rodolfo-Metalpa R, Martin S, Ransome E, Fine M, Turner SM, Rowley SJ, Tedesco D, Buia MC, 2008. Volcanic carbon dioxide vents show ecosystem effects of ocean acidification. Nature 454: 96-99.

Halpern BS, Selkoe KA, Micheli F, Kappel CV, 2007. Evaluating and Ranking the Vulnerability of Global Marine Ecosystems to Anthropogenic Threats. Conservation Biology 21: 1301-1315.

Halpern BS, Walbridge S, Selkoe KA, Kappel CV, Micheli F, D'Agrosa C, Bruno JF, Casey KS, Ebert C, Fox HE, Fujita R, Heinemann D, Lenihan HS, Madin EMP, Perry MT, Selig ER, Spalding M, Steneck R, Watson R, 2008. A Global Map of Human Impact on Marine Ecosystems. Science 319: 948-952.

Harley CDG, Hughes AR, Hultgren KM, Miner BG, Sorte CJB, Thornber CS, Rodriguez LF, Tomanek L, Williams SL, 2006. The impacts of climate change in coastal marine Systems. Ecology Letters 9: 228-241.

Heath MR, 2008. Comment on 'A global map of human impact on marine ecosystems'. Science 321: 1446b.

Hendricks IE, Duarte CM, Heip CHR, 2006. Biodiversity Research Still Grounded. Science 312: 1715.

Holland AF, Shaughnessy AT, Hiegel MH, 1987. Long-term variation in mesohaline Chesapeake Bay macrobenthos: spatial and temporal patterns. Estuaries 10: 227-245.

Hooper DU, Chapin FS, Ewel JJ, Hector A, Inchausti P, Lavorel S, Lawton JH, Lodge DM, Loreau M, Naeem S, Schmid B, Setälä H, Symstad AJ, Vandermeer J, Wardle DA, 2005. Effects of biodiversity on ecosystem functioning: A consensus of current knowledge. Ecological Monographs 75: 3-35.

Houte-Howes KSS, Turner SJ, Pilditch CA, 2004. Spatial differences in macroinvertebrate communities in intertidal seagrass habitats and unvegetated sediment in three New Zealand estuaries. Estuaries 27: 945-957.

Huber H, Hohn MJ, Rachel R, Fuchs T, Wimmer VC, Stetter KO, 2002. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. Nature 417: 63-67.

Hulbert SH, 1971. The non-concept of species diversity: a critique and alternative parameters. Ecology 52: 577-586.

Huston MA, 1979. A general hypothesis of species diversity. American Naturalist 113: 81-101.

Hyland JL, Balthis WL, Engle VD, Long ER, Paul JF, Summers JK, Van Dolah RF, 2003. Incidence of stress in benthic communities along the U.S. Atlantic and Gulf of Mexico coasts within different ranges of sediment contamination from chemical mixtures. Environmental Monitoring and Assessment 81: 149-161.

ICES, 2001. Report of the ICES Advisory Committee on Ecosystems. ICES Cooperative Research Report 249, 75 pp.

ICES, 2004. Report of the Study Group on ecological quality objectives for Sensitive and for Opportunistic Benthos Species (SGSOBS). Copenhagen, ICES. 37 pp.

ICES, 2005. Ecosystem effects of fishing: impacts, metrics, and management strategies. ICES Cooperative Research Report 272, 177 pp.

Jackson LE, Kurtz JC, Fisher WS (Eds), 2000. Evaluation Guidelines for Ecological Indicators. EPA/620/R-99/005. U.S. Environmental Protection Agency, Office of Research and Development, Research Triangle Park, NC. 107 pp.

Jørgensen SE, 1998. Ecotoxicological Research - Historical development and perspectives. *in* Ecotoxicology: Ecological Fundamentals, Chemical Exposure, and Biological Effects. Shüürmann G, Markert B (Eds.), Environmental Science and Technology. Wiley, New York, 900 pp.

Jørgensen SE, Marques JC, 2001. Thermodynamics and ecosystem theory. Case studies from Hydrobiology. Hydrobiologia 445: 1-10.

Jørgensen SE, Mejer H, 1979. A holistic approach to ecological modelling. Ecological Modelling 7:169-189.

Jørgensen SE, Mejer H, 1981 Exergy as a key function in ecological models. In: Mitsch W, Bosserman RW, Klopatek JM (Eds.), Energy and ecological modelling. Developments in environmental modelling, Elsevier. Amsterdam, 1: 587-590.

Jørgensen SE, Xu FL, Salas F, Marques JC, 2005. Application of Indicators for the Assessment of Ecossystem Health. In: Handbook ecological indicators for assessment of ecosystem health. Jørgensen SE, Costanza R, Xu FL. (Eds). CRC Press, Boca Raton, FL, pp. 6-66.

Josefson AB, Hansen JLS, Asmund G, Johansen P, 2008. Threshold response of benthic macrofauna integrity to metal contamination in West Greenland. Marine Pollution Bulletin 56: 1265-1274.

Josefson AB, Blomqvist M, Hansen JLS, Rosenberg R, Rygg B, 2009. Assessment of marine benthic quality change in gradients of disturbance: comparison of different Scandinavian multi-metric indices. Marine Pollution Bulletin 58: 1263-1277.

Jowett D, 2006. NEA-GIG Milestone 6 report, available online at: http://forum.europa.eu.int/Public/irc/jrc/jrc_eewai/library (28/05/2008).

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

Karr JR, 1991. Biological integrity: a long-neglected aspect of water resource management. Ecological Applications 1: 66-84.

Karr JR, 1996. Ecological integrity and health are not the same. Pages 97-109 *in* Schulze PC (Ed.), Engineering Within Ecological Constraints. National Academy of Engineering, National Academy Press, Washington, DC, 213 pp.

Karr JR, 2002. Understanding the consequences of human actions: indicators from GNP to IBI. In: Miller P, Westra L (Eds.). Just Ecological Integrity: The Ethics of Maintaining Planetary Life. Rowman & Littlefield, Lanham, Maryland, pp. 98–110.

Karr JR, Chu EW, 1997. Biological monitoring: Essential foundation for ecological risk assessment. Human and Ecological Risk Assessment: An International Journal 3: 993-1004.

Kennish MJ, 1992. Ecology of Estuaries: Anthropogenic Effects. Marine Science Series. CRC Press, Boca Raton, FL.

Kennish MJ, Townsend AR, 2007. Nutrient Enrichment and Estuarine Eutrophication. Ecological Applications 17, Supplement Eutrophication: S1-S2.

Kenny AJ, Rees HL, 1996. The effects of marine gravel extraction on the macrobenthos: results 2 years post-dredging. Marine Pollution Bulletin 32: 615-622.

Kiddon JA, Paul JF, Buffum HW, Strobel CS, Hale SS, Cobb D, Brown BS, 2003. Ecological condition of US mid-Atlantic estuaries, 1997-1998. Marine Pollution Bulletin 46: 1224-1244.

Kurtz JC, Jackson LE, Fisher WS, 2001. Strategies for evaluating indicators based on guidelines from the Environmental Protection Agency's Office of Research and Development. Ecological Indicators 1: 49-60.

Labrune C, Amouroux JM, Sarda R, Dutrieux E, Thorin S, Rosenberg R, Grémare A, 2006. Characterization of the ecological quality of the coastal Gulf of Lions (NW Mediterranean). A comparative approach based on three biotic indices. Marine Pollution Bulletin 52: 34-47.

Lambshead PJ, Platt HM, Shaw KM, 1983. The detection of differences among assemblages of marine benthic species based on an assessment of dominance and diversity. Journal of Natural History 17: 859-847.

Landis JR, Koch GG, 1977. The measurement of observer agreement for categorical data. Biometrics 33: 159-174.

Lavesque N, Blanchet H, de Montaudouin X, 2009. Development of a multimetric approach to assess perturbation of benthic macrofauna in Zostera noltii beds. Journal of Experimental Marine Biology and Ecology 368: 101–112.

Lee GF, Jones-Lee A, 2004. Appropriate use of chemical information in a best professional judgment triad weight-of-evidence evaluation of sediment quality. Aquatic Ecosystem Health and Management 7: 351-356.

Lillebø AI, Neto JM, Martins I, Verdelhos T, Leston S, Cardoso PG, Ferreira SM, Marques JC, Pardal MA, 2005. Management of a shallow temperate estuary to control eutrophication: the effect of hydrodynamics on the system's nutrient loading. Estuarine, Coastal and Shelf Science 65: 697-707.

Lillebø AI, Teixeira H, Pardal MA, Marques JC, 2007. Applying quality status criteria to a temperate estuary before and after the mitigation measures to reduce eutrophication symptoms. Estuarine, Coastal and Shelf Science 72: 177-187.

Limnologisk Metodik, 1992. Københavns University, Ferksvandsbiologisk Laboratorium, Akademisk Forlag, Denmark.

Llansó RJ, Scott LC, Dauer DM, Hyland JL, Russell DE, 2002a. An estuarine benthic index of biotic integrity for the mid-Atlantic region of the United States. I. Classification of assemblages and habitat definition. Estuaries 25: 1219-1230.

Llansó RJ, Scott LC, Hyland JL, Dauer DM, Russell DE, Kutz FW, 2002b. An estuarine benthic index of biotic integrity for the mid-Atlantic region of the United States. II. Index Development. Estuaries 25: 1231-1242.

Long ER, MacDonald DD, Severn CG, Hong CB, 2000. Classifying probabilities of acute toxicity in marine sediments with empirically derived sediment quality guidelines. Environmental Toxicology and Chemistry 19: 2598-2601.

Long ER, Ingersoll CG, MacDonald DD, 2006. Calculation and uses of mean sediment quality guideline quotients: A critical review. Environmental Science & Technology 40: 1726-1736.

Lopes RJ, Múrias T, Cabral JA, Marques JC, 2005. A ten year study of variation, trends and seasonality of shorebird community in the Mondego estuary, Portugal. Waterbirds 28: 8-18.

Loreau M, Naeem S, Inchausti P, Bengtsson J, Grime JP, Hector A, Hooper DU, Huston MA, Raffaelli D, Schmid B, Tilman D, Wardle DA, 2001. Biodiversity and ecosystem functioning: current knowledge and future challenges. Science 294: 804–808.

Lotze HK, Lenihan HS, Bourque BJ, Bradbury R, 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.

Mackey RL, Currie DJ, 2001. The diversity-disturbance relationship: is it generally strong and peaked? Ecology 82: 3479-3492.

Magni P, Hyland J, Manzella G, Rumohr H, Viaroli P, Zenetos A (Eds.), 2005. Proceedings of the Workshop 'Indicators of Stress in the Marine Benthos', Torregrande-Oristano (Italy), 8–9 October 2004. UNESCO/IOC, Workshop Reports, 195. IMC Special Publication, UNESCO, Paris, 46 pp.

Magurran AE, 2004. Measuring biological diversity. Blackwell Science, Oxford, 256 pp.

Manolakos E, Virani H, Novotny V, 2007. Extracting knowledge on the links between the water body stressors and biotic integrity. Water Research 41: 4041-4050.

Margalef R, 1968. Perspectives in Ecological Theory. University of Chicago Press, Chicago, Illinois, USA.

Marques JC, Maranhão P, Pardal MA, 1993. Human impact on the subtidal macrobenthic community structure in the Mondego estuary (western Portugal). Estuarine, Coastal and Shelf Science 37: 403-419.

Marques JC, Pardal MA, Nielsen SN, Jørgensen SE, 1997. Analysis of the properties of exergy and biodiversity along an estuarine gradient of eutrophication. Ecological Modelling 62: 155-167.

Marques JC, Pardal MA, Lillebø AI, 1999. Comunidades biológicas do estuário do Mondego. Cadernos de Ecologia 2: 46-76.

Marques JC, Nielsen SN, Pardal MA, Jørgensen SE, 2003. Impact of eutrophication and river management within a framework of ecosystem theories. Ecological Modelling 166: 147-168.

Marques JC, Baeta A, Trancoso AR, Marcotegui A, Rodrigues ET, Martinho F, Salas F, Patrício J, Alves J, Neto JM, Sebastião L, Santos M, Fontes MG, Martins MJ, Pardal MA, Neves R, Deus R, Leitão R, Verdelhos T, 2004. Caracterização do estuário do Mondego e estabelecimento de cenários de gestão em termos da qualidade da água e da integridade do ecossistema. IMAR/INAG, 239 pp.

Marques JC, Salas F, Patrício J, Neto J, Teixeira H, 2009a. Ecological Indicators for Coastal and Estuarine Environmental Assessment - A User Guide. WIT PRESS, Southampton, Boston, 183 pp.

Marques JC, Basset A, Brey T, Elliot M, 2009b. The ecological sustainability trigon – A proposed conceptual framework for creating and testing management scenarios. Marine Pollution Bulletin 58: 1773-1779.

Martinho F, Dolbeth M, Viegas I, Teixeira CM, Cabral HN, Pardal MA, 2009. Environmental effects on the recruitment variability of nursery species. Estuarine, Coastal and Shelf Science 83: 460-468.

Martins I, Pardal MA, Lillebø AI, Flindt MR, Marques JC, 2001. Hydrodynamics as a major factor controlling the occurrence of green macroalgal blooms in a eutrophic estuary: a case study on the influence of precipitation and river management. Estuarine, Coastal and Shelf Science 52: 165-177.

Martins I, Lopes RJ, Lillebø AI, Neto JM, Pardal MA, Marques JC, 2007. Significant variations on the productivity of green macroalgae in a mesotidal estuary: implications to the nutrient loading of the system and the adjacent coastal area. Marine Pollution Bulletin 54: 678-690.

May RM, 1992. Bottoms up for the oceans. Nature 357: 278-279.

McArdle BH, Anderson MJ, 2001. Fitting multivariate models to community data: a comment on distance based redundancy analysis. Ecology 82: 290-297.

McCahon CP, Pascoe D, 1990. Episodic Pollution: Causes, Toxicological Effects and Ecological Significance. Functional Ecology 4: 375-383.

McLusky DS, 1971. Ecology of estuaries. Heineman, London, 144 pp.

McLusky DS, Elliot M, 2004. The estuarine ecosystem – ecology, threats and management. Oxford University Press, 214 pp.

Meyer TH, Berg T, Fürhaupter K, 2008. Ostsee Makrozoobenthos Klassifizierungssystem für die Wasserrahmenrichtlinie. Referenz-Artenlisten, Bewertungsmodell und Monitoring. MARILIM Gewässeruntersuchung: 131 pp. http://www.marilim.de/marbit/download/files/MarBIT.pdf

Milovidova NY, Alyomov SV, 1992. Physical/chemical characteristics of bottom sediments and soft bottom zoobenthos of Sevastopol region. In: Polykatpov G (Ed.), Molysmology of the Black Sea. Naukova Dumka, Kiev, 216-282 pp.

Molvær J, Knutzen J, Magnusson J, Rygg B, Skei J, Sørensen J, 1997. Classification of environmental quality in fjords and coastal waters. SFT guidelines 97: 03, 36 pp.

Monserud R, Leemans R, 1992. Comparing global vegetation maps with the Kappa statistic. Ecological Modelling 62: 275-293.

Muniz P, Venturini N, Pires-Vanin AMS, Tommasi LR, Borja A, 2005. Testing the applicability of a Marine Biotic Index (AMBI) for assessing the ecological quality of soft-bottom benthic communities in the South America Atlantic region. Marine Pollution Bulletin 50: 624-637.

Muxika I, Borja A, Bonne W, 2005. The suitability of the marine biotic index (AMBI) to new impact sources along European coasts. Ecological Indicators 5: 19-31.

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

Neto JM, 2004. Nutrient enrichment in a temperate macro-tidal system. Scenario analysis and contribution to coastal management. Ph.D. Thesis. University of Coimbra, Portugal, 139 pp.

Neto JM, Flindt MR, Marques JC, Pardal MA, 2008. Modelling nutrient mass balance in a temperate meso-tidal estuary: Implications for management. Estuarine, Coastal and Shelf Science 76: 175-185.

Neto JM, Gaspar R, Pereira L, Marques JC. Ecological quality assessment of intertidal rocky shores. The use of marine macroalgae under the scope of European Water Framework Directive. Ecological Indicators (*accepted*).

Neto JM, Teixeira H, Patrício J, Baeta A, Veríssimo H, Pinto R, Marques JC. The response of estuarine macrobenthic communities to natural and human-induced changes: dynamics and ecological quality. *Submitted* for publication to Estuaries and Coasts.

Niemeijer D, de Groot RS, 2008. A conceptual framework for selecting environmental indicators sets. Ecological Indicators 8: 14-25.

Niemi GJ, McDonald ME, 2004. Application of Ecological Indicators. Annual Review of Ecology, Evolution and Systematics 35: 89-111.

Nixon SW, 1995. Coastal marine eutrophication: A definition, social causes, and future concerns. Ophelia 41: 199-219.

NRC (National Research Council), 2000. Ecological Indicators for the Nation. National Academy Press, Washington, DC.

Occhipinti-Ambrogi A, Forni G, 2004. Biotic indices. In: Gambi M.C., Dappiano M. (Eds.), Mediterranean marine benthos: a manual of methods for its sampling and study. Biologia Marina Mediterranea 11: 545-572.

O'Connor JS, Dewling RT, 1986. Indices of marine degradation: their utility. Environmental Management 10: 335-343.

Olsgard F, Somerfield PJ, Carr MR, 1997. Relationships between taxonomic resolution and data transformation in analyses of a macrobenthic community along an established pollution gradient. Marine Ecology Progress Series 149: 173-181.

Orth RJ, Carruthers TJB, Dennison WC, Duarte CM, Fourqurean JW, Heck KL. Jr., Hughes AR, Kendrick GA, Kenworthy WJ, Olyarnik S, Short FT, Waycott M, Williams SL., 2006.A Global Crisis for Seagrass Ecosystems. Bioscience 56: 987-996.

Pardal MA, Marques JC, Bellan G, 1993. Spatial distribution and seasonal variation of the subtidal polychaete populations in the Mondego estuary (western Portugal). Cahiers de Biologie Marine 34 : 497-512.

Pardal MA, Marques JC, Metelo I, Lillebø AI, Flindt MR, 2000. Impact of eutrophication on the life cycle, population dynamics and production of *Ampithoe valida* (Amphipoda) along an estuarine spatial gradient (Mondego estuary, Portugal). Marine Ecology Progress Series 196: 207-219.

Pardal MA, Cardoso PG, Sousa JP, Marques JC, Raffaelli D, 2004. Assessing environmental quality: a novel approach. Marine Ecology Progress Series 267: 1-8.

Patrício J, Ulanowicz R, Pardal MA, Marques JC, 2004. Ascendency as ecological indicator: A case study on estuarine pulse eutrophication. Estuarine Coastal and Shelf Science 60: 23-35.

Patrício J, Marques JC, 2006. Mass balanced models of the food web in three areas along a gradient of eutrophication symptoms in the south arm of the Mondego estuary (Portugal). Ecological Modelling 197: 21-34.

Patrício J, Teixeira H, Neto JM, Marques JC, 2007. Opportunistic macroalgae metrics for transitional waters. Testing tools to assess ecological quality status in Portugal. Marine Pollution Bulletin 54: 1887-1896.

Patrício J, Neto JM, Teixeira H, Salas F, Marques JC, 2009. The robustness of ecological indicators to detect long-term changes in the macrobenthos of estuarine systems. Marine Environmental Research 68: 25-36.

Paul JF, Scott KJ, Campbell DE, Gentile JH, Strobel CS, Valente RM, Weisberg SB, Holland AF, Ranasinghe JA, 2001. Developing and applying a benthic index of estuarine condition for the Virginian biogeographic province. Ecological Indicators 1: 83-99.

Pearson TH, Rosenberg R, 1978. Macrobenthic succession in relation to organic enrichment and pollution of the marine environment. Oceanography and Marine Biology: An Annual Review 16: 229-311.

Pergent-Martini C, Leoni V, Pasqualini V, Ardizzone GD, Balestri E, Bedini R, Belluscio A, Belsher T, Borg J, Boudouresque CF, Boumaza S, Bouquegneau JM, Buia MC, Calvo S, Cebrian J, Charbonnel E, Cinelli F, Cossu A, Di Maida G, Dural B, Francour P, Gobert S, Lepoint G, Meinesz A, Molenaar H, Mansour HM, Panayotidis P, Peirano A, Pergent G, Piazzi L, Pirrotta M, Relini G, Romero J, Sanchez-Lizaso JL, Semroud R, Shembri P, Shili A, Tomasello A, Velimirov B, 2005. Descriptors of Posidonia oceanica meadows: Use and application. Ecological Indicators 5: 213-230.

Perus J, Bonsdorff E, Bäck S, Lax H-Gr, Villnäs A, Westberg V, 2007. Zoobenthos as Indicators of Ecological Status in Coastal Brackish Waters: A Comparative Study from the Baltic Sea. AMBIO: A Journal of the Human Environment 36: 250-256.

Pinto R, Patricio J, Baeta A, Fath BD, Neto JM, Marques JC, 2009. Review and evaluation of estuarine biotic indices to assess benthic condition. Ecological Indicators 9: 1-25.

Pinedo S, Jordana E, 2008. Spain (Catalonia and Balearic Islands). In: Carletti A, Heiskanen AS (Eds), 2009. Water Framework Directive Intercalibration Technical Report. Part 3: Coastal and Transitional Waters. JRC Scientific and Technical Reports, EUR 23838 EN/3: 240 pp.

Pires A, Paxton H, Quintino V, Rodrigues NA, 2010. *Diopatra* (Annelida: Onuphidae) diversity in European waters with the description of *Diopatra micrura*, new species. Zootaxa 2395: 17-33.

Poikane S (Ed.), 2009. Water Framework Directive intercalibration technical report. Part 2: Lakes. JRC Scientific and Technical Reports, EUR 23838 EN/2, 176 pp.

Primer 5 for Windows, Version 5.2.6. Copyright © 2001, Primer-E Ltd.

Prior A, Miles AC, Sparrow AJ, Price N, 2004. Development of a classification scheme for the marine benthic invertebrate component, water framework directive. Phase I & II - Transitional and Coastal waters. R & D Interim Technical Report E1-116, E1-132.

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

Raffaelli DG, Mason CF, 1981. Pollution monitoring with meiofauna using the ratio of nematodes to copepods. Marine Pollution Bulletin 12: 158-163.

Rakocinski CF, Zapfe GA, 2005. Macrobenthic process indicators of estuarine condition. In: Estuarine Indicators. S.A. Bortone (Ed). CRC Press, Boca Raton, FL, pp. 315-331.

Ranasinghe JA, Frithsen JB, Kutz FW, Paul JF, Russell DE, Batiuk RA, Hyland JL, Scott J, Dauer DM, 2002. Application of two indices of benthic community condition in Chesapeake Bay. Environmetrics 13: 499-511.

Ranasinghe JA, Thompson B, Smith RW, Lowe S, Schiff K, 2004. Evaluation of Benthic Assessment Methodology in Southern California Bays and San Francisco Bay. Technical Report No. 432. Southern California Coastal Water Research Project. Westminster, CA. 70 pp.

Ranasinghe JA, Weisberg SB, Smith RW, Montagne DE, Thompson B, Oakden JM, Huff DD, Cadien DB, Velarde RG, Ritter KJ, 2009. Calibration and evaluation of five indicators of benthic community condition in two California bay and estuary habitats. Marine Pollution Bulletin 59: 5-13.

Rapport DJ, Costanza R, McMichael AJ, 1998. Assessing ecosystem health. Tree 13: 397-402.

Ray GC, Grassle, JF, 1991. Marine biological diversity. Bioscience 41: 453-461.

Rees HL, Boyd SE, Schratzberger M, Murray LA, 2006. Role of benthic indicators in regulating human activities at sea. Environmental Science and Policy 9: 496-508.

Rees HL, Hyland JL, Hylland K, Mercer Clarke CSL, Roff JC, Ware S, 2008. Environmental Indicators: utility in meeting regulatory needs. An overview. ICES Journal of Marine Science 65: 1381-1386.

Reish DJ, 1993. Effects of metals and organic compounds on survival and bioaccumulation in two species of marine gammaridean amphipod, together with a summary of toxicological research on this group. Journal of Natural History 27: 781-794.

Reiss H, Kröncke I, 2005. Seasonal variability of benthic indices: an approach to test the applicability of different indices for ecosystem quality assessment. Marine Pollution Bulletin 50: 1490-1499.

Revilla M, Franco J, Bald J, Borja A, Laza A, Seoane S, Valencia V, 2008. The assessment of the phytoplankton ecological status in the Basque coast (northern Spain) according to the European Water Framework Directive. Journal of Sea Research 61: 60-67.

Rockström J, Steffen W, Noone K, Persson Å, Chapin III FS, Lambin E, Lenton TM, Scheffer M, Folke C, Schellnhuber H, Nykvist B, De Wit CA, Hughes T, van der Leeuw S, Rodhe H, Sörlin S, Snyder PK, Costanza R, Svedin U, Falkenmark M, Karlberg L, Corell RW, Fabry VJ, Hansen J, Walker B, Liverman D, Richardson K, Crutzen P, Foley J, 2009. A safe operating space for humanity. Nature 461: 472-475.

Rodrigues E, 2004. Variação das comunidades macrobentónicas subtidais do estuário do Mondego ao longo de mais de uma década. MSc Thesis, University of Coimbra, Portugal, 74 pp.

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

Rosenberg R, Blomqvist M, Nilsson HC, Cederwall H, Dimming A, 2004. Marine quality assessment by use of benthic species-abundance distributions: a proposed new protocol within the European Union Water Framework Directive. Marine Pollution Bulletin 49: 728-739.

Ruellet T, Dauvin JC, 2007. Benthic indicators: Analysis for thresholds values of ecological quality classifications for transitional waters. Marine Pollution Bulletin 54: 1707-1714.

Ruellet T, Dauvin JC, 2008. Comments on Muxika et al. ‘‘Using historical data, expert judgement and multivariate analysis in assessing reference conditions and benthic ecological status, according to the European Water Framework Directive’ [Marine Pollution Bulletin 55 (2007) 16–29] Marine Pollution Bulletin 56: 1234-1235.

Rygg B, 2002. Indicator species index for assessing benthic ecological quality in marine waters of Norway. Norwegian Institute for Water Research, Report no. 40114, pp. 1–32.

Rygg B, 2006. Developing indices for quality-status classification of marine softbottom fauna in Norway. NIVA report 5208-2006, 33 pp.

Sabine CL, Feely RA, Gruber N, Key RM, Lee K, Bullister JL., Wanninkhof R, Wong CS, Wallace DWR, Tilbrook B, Millero FJ, Peng TH, Kozyr A, Ono T, Rios AF, 2004. The Oceanic Sink for Anthropogenic $CO_2$. Science 305: 367-371.

Salas F, 2002. Valoración y aplicabilidad de los índices ecológicos y bioindicadores de contaminación orgánica en la gestión del medio marino.Tesis Doctoral. Universidad de Murcia.

Salas F, Neto JM, Borja A, Marques JC, 2004. Evaluation of the applicability of a Marine Biotic Index to characterise the status of estuarine ecosystems: The case of Mondego estuary (Portugal). Ecological Indicators 4: 215-225.

Salas F, Marcos C, Neto JM, Patrício J, Pérez-Ruzafa A, Marques JC, 2006a. User-friendly guide for using benthic ecological indicators in coastal and marine quality assessment. Ocean & Coastal Management 49: 308-331.

Salas F, Patrício J, Marques JC, 2006b. Ecological indicators in coastal and estuarine environmental quality assessment. A user friendly guide for practitioners. University of Coimbra Press, 131 pp.

Sanchez-Moyano JE, Estacio FJ, García-Adiego EM, García-Gómez JC, 2004. Dredging impact on the benthic community of an unaltered inlet in southern Spain. Helgoland Marine Research 58: 32-39.

Sandin L, 2005. Testing the EC Water Framework Directive 'one-out, all-out' rule - simulating different levels of assessment errors along a pollution gradient in Swedish streams. Verh. Internat. Verein. Limnol. 29: 334-336.

Santos PJP, Castel J, Souza-Santos LP, 1996. Seasonal variability of meiofaunal abundance in the oligo-mesohaline area of the Gironde Estuary, France. Estuarine, Coastal and Shelf Science 43: 549-563.

Scheffer M, Carpenter S, Foley JA, Folke C, Walker B, 2001. Catastrophic shifts in ecosystems. Nature: 413: 591-596.

Schlacher TA, Wooldridge TH, 1996. Axial zonation patterns of subtidal macrozoobenthos in the Gamtoos Estuary, South Africa. Estuaries 19: 680-696.

Schmedtje U, Birk S, Poikane S, van De Bund W, Bonne W, 2009. Guidance document on the intercalibration process 2008-2011. Guidance Document No. 14. Implementation Strategy for the Water Framework Directive (2000/60/EC): 55 pp.

Schüürmann G, Ebert RU, Nendza M, Dearden JC, Paschke A, Kühne R, 2007. Predicting Fate-Related Physicochemical Properties. In: C.J. van Leeuwen and T.G. Vermeire (Eds.), Risk Assessment of Chemicals: An Introduction. Springer, Netherlands, pp. 375-426.

Shannon CE, Weaver W, 1963. The Mathematical theory of communication. University of Illinois Press.

Simboura N, Argyrou M, *in press*. An insight into the performance of benthic classification indices tested in Eastern Mediterranean coastal waters. Marine Pollution Bulletin. doi:10.1016/j.marpolbul.2009.12.005.

Simboura N, Zenetos A, 2002. Benthic indicators to use ecological quality classification of Mediterranean soft bottom marine ecosystems, including a new Biotic Index. Mediterranean Marine Science 3/2: 77-111.

Simpson EH, 1949. Measurement of diversity. Nature 163: 688.

Sindermann CJ, 2006. Coastal Pollution: Effects on Living Resources and Humans. CRC Press, Boca Raton, FL.

Smeets E, Weterings R, 1999. Environmental Indicators: Typology and Overview. European Environment Agency, Copenhagen. Report No. 25, 19 pp.

Smith SV, 1981. Marine Macrophytes as a Global Carbon Sink. Science 211: 838-840.

Smith RW, Bergen M, Weisberg SB, Cadien DB, Dalkey A, Montagne DE, Stull JK, Velarde RG, 2001. Benthic response index for assessing infaunal communities on the southern California mainland shelf. Ecological Applications 11: 1073-1087.

Smith RW, Ranasinghe JA, Weisberg SB, Montagne DE, Cadien DB, Mikel TK, Velarde RG, Dalkey A, 2003. Extending the southern California Benthic Response Index to assess benthic condition in bays. Technical Report No 410. Southern California Coastal Water Research Project. Westminster, CA. 36 pp + 4 Appendices.

Snelgrove PVR, 1999. Getting to the Bottom of Marine Biodiversity: Sedimentary Habitats: Ocean bottoms are the most widespread habitat on Earth and support high biodiversity and key ecosystem services. BioScience 49: 129 -138.

Stachowicz JJ, Byrnes JE, 2006. Species diversity, invasion success and ecosystem functioning: disentangling the influence of resource competition, facilitation, and extrinsic factors. Marine Ecology Progress Series 311: 251–262.

Stachowicz JJ, Fried H, Osman RW, Whitlatch RB, 2002. Biodiversity, invasion resistance and marine ecosystem function: reconciling pattern and process. Ecology 83: 2575–2590.

Statgraphics Plus 5.0. Copyright © 1994-2000, Statistical Graphics Corp.

Storelli MM, Marcotrigiano GO, 2001. Persistent organochlorine residues and toxic evaluation of polychlorinated biphenyls in sharks from the Mediterranean Sea (Italy). Marine Pollution Bulletin 42: 1323-1329.

Strickland JDH, Parsons TR, 1972. A Practical Handbook of Seawater Analysis (2nd ed). Bulletin of the Fisheries Research Board of Canada 167, 311 pp.

Strobel CJ, Buffum HW, Benyi SJ, Petrocelli EA, Reifsteck DR, Keith DJ, 1995. Virginian Province Statistical Summary 1990-1993. U.S. Environmental Protection Agency, National Health and Environmental Effects Research Laboratory, Atlantic Ecology Division., EPA/620/R-94/026, Narragansett, RI.

Stull J, 1995. Two decades of marine biological monitoring, Palos Verdes, California, 1972yo 1992. Bulletin of Southern California Academy of Sciences 94: 21-45.

Stull J, Haydock CI, Smith RW, Montagne DE, 1986. Long-term changes in the benthic community on the coastal shelf of Palos Verdes, Southern California. Marine Biology 91: 539-551.

Tapp JF, Shillabeer N, Ashman CM, 1993. Continued observations of the benthic fauna of the industrialised Tees estuary, 1979-1990. Journal of Experimental Marine Biology and Ecology 172: 67-80.

Teixeira H, Salas F, Pardal MA, Marques JC, 2007. Applicability of ecological evaluation tools in estuarine ecosystems: the case of the lower Mondego estuary (Portugal). Hydrobiologia 587: 101–112.

Teixeira H, Salas F, Borja A, Neto JM, Marques JC, 2008a. A benthic perspective in assessing the ecological status of estuaries: The case of the Mondego estuary (Portugal). Ecological Indicators 8: 404-416.

Teixeira H, Salas F, Borja A, Neto JM, Patrício J, Pinto R, Veríssimo H, García-Charton JA, Marcos C, Pérez-Ruzafa Á, Marques JC, 2008b. Ecological indices tracking distinct impacts along disturbance-recovery gradients in a temperate NE Atlantic Estuary – Guidance on reference values. Estuarine, Coastal and Shelf Science 80: 130-140.

Teixeira H, Borja A, Weisberg SB, Ranasinghe JA, Cadien DB, Dauer DM, Dauvin J-C, Degraer S, Diaz RJ, Grémare A, Karakassis I, Llansó RJ, Lovell LL, Marques JC, Montagne DE, Occhipinti-Ambroggi A, Rosenberg R, Sardá R, Schaffner LC, Velarde RG, 2010. Assessing Coastal Benthic Macrofauna Community Condition Using Best Professional Judgement - developing consensus across North America and Europe. Marine Pollution Bulletin 60: 589-600.

ter Braak CJF, Smilauer P, 1998. CANOCO Reference Manual and User's Guide to Canoco for Windows: Software for Canonical Community Ordination (version 4). Microcomputer Power (Ithaca, NY USA) 352 pp.

Tietenberg T, 2003. Environmental and Natural Resource Economics 6$^{th}$ Ed. Pearson Education, Inc. 646 pp.

Tilman D, Lehman C, Thompson K, 1997. Plant diversity and ecosystem productivity: theoretical considerations. Proceedings of the National Academy of Sciences 94: 1857–1861.

Turner SJ, Hewitt JE, Wilkinson MR, Morrisey DJ, Cummings VJ, Funnel G, 1999. Seagrass Patches and Landscapes: The Influence of Wind-wave Dynamics and Hierarchical Arrangements of Spatial Structure on Macrofaunal Seagrass Communities. Estuaries 22: 1016-1032.

Ulanowicz RE, 1980. An hypothesis on the development of natural communities. Journal of Theoretical Biology 85: 223-245.

Ulanowicz RE, 1986. Growth and development ecosystems phenomenology. Springer-Verlag, New York.

Ulanowicz RE, Norden JS, 1990. Symmetrical overhead in flow and networks. International Journal of Systems Science 21: 429-437.

UNESCO, 2003. A reference guide on the use of indicators for an integrated coastal management. ICAM Dossier 1-IOC Manuals and Guides 45.

UNESCO, 2006. A Handbook for Measuring the Progress and Outcomes of Integrated Coastal and Ocean Management. IOC Manuals and Guides 46. ICAM Dossier, 2, Paris.

US EPA (U.S. Environmental Protection Agency), 1998. Condition of the Mid-Atlantic Estuaries. U.S. Environmental Protection Agency, Office of Research and Development, EPA/600/R-98/147, Washington, DC.

van de Bund W (Ed.), 2009. Water Framework Directive intercalibration tecnical report. Part 1: Rivers. JRC Scientific and Technical Reports, EUR 23838 EN/1, 136 pp.

van Dolah RF, Hyland JL, Holland AF, Rosen JS, Snoots TR, 1999. A benthic index of biological integrity for assessing habitat quality in estuaries of the southeastern USA. Marine Environmental Research 48: 269-283.

van Hoey G, Drent J, Ysebaert T, Herman P, 2007. The Benthic Ecosystem Quality index (BEQI), intercalibration and assessment of Dutch Coastal and Transitional Waters for the Water Framework Directive. NIOO rapport 2007-02.

van Houte-Howes KSS, Turner SJ, Pilditch CA, 2004. Spatial differences in macroinvertebrate communities in intertidal seagrass habitats and unvegetated sediment in three New Zealand estuaries. Estuaries 27, 945-957.

Vaquer-Sunyer R, Duarte CM, 2008. Thresholds of hypoxia for marine biodiversity. Proceedings of the National Academy of Science 105: 15452-15457.

Verdelhos T, Neto JM, Marques JC, Pardal MA, 2005. The effect of eutrophication abatement on the bivalve *Scrobicularia plana*. Estuarine Coastal and Shelf Science 63: 261-268.

Vincent C, Heinrich H, Edwards A, Nygaard K, Haythornthwaite J, 2003. Guidance on typology, classification and reference conditions for transitional and coastal waters. European Commission, report of CIS WG2.4 (COAST), 119 pp.

Vollenweider RA, Giovanardi F, Montanari G, Rinaldi A, 1998. Characterisation of the trophic conditions of marine coastal waters with special reference to the NW Adriatic Sea: Proposal for a trophic scale, turbidity and generalised water quality index. Environmetrics 9: 329-357.

Vuori K-M, 2007. Applying Ecological Risk Assessment Approaches in River Basin Management Planning. Finnish Environment Institute and University of Oulu, Oulu, Finland. *in* Ulvi T, Visuri M, Hellsten S, (Eds.), 2007. Proceedings of the European Symposium of Spatial Planning Approaches towards Sustainable River Basin Management. May 14-15, Rovaniemi, Finland. Reports of Finnish Environment Institute 12.

Warwick RM, 1986. A new method for detecting pollution effects on marine macrobenthic communities. Marine Biology 92: 557-562.

Warwick RM, Clarke KR, 1994. Relearning the ABC: taxonomic changes and abundance/biomass relationships in disturbed benthic communities. Marine Biology 118: 739-744.

Warwick RM, Clarke KR, 1995. New ‘biodiversity’ measures reveal a decrease in taxonomic distinctness with increasing stress. Marine Ecology Progress Series 129: 301-305.

Warwick RM, Clarke KR, 1998. Taxonomic distinctness and environmental assessment. Journal of Applied Ecology 35: 532-543.

Warwick RM, Clarke KR, Somerfield PJ, 2010. Exploring the marine biotic index (AMBI): variations on a theme by Ángel Borja. Marine Pollution Bulletin 60: 554-559.

Weisberg SB, Ranasinghe JA, Dauer DM, Schaffner LC, Díaz RJ, Frithsen JB, 1997. An estuarine benthic index of biotic integrity (B-IBI) for Chesapeake Bay. Estuaries 20: 149-158.

Weisberg SB, Thompson B, Ranasinghe JA, Montagne DE, Cadien DB, Dauer DM, Diener DR, Oliver JS, Reish DJ, Velarde RG, Word JQ, 2008. The level of agreement among experts applying best professional judgment to assess the condition of benthic infaunal communities. Ecological Indicators 8: 389-394.

Wells E, Wilkinson M, Wood P, Scanlan C, 2007. The use of macroalgal species richness and composition n intertidal rocky seashores in the assessment of ecological quality under the European Water Framework Directive. Marine Pollution Bulletin 55: 151-161.

WFD NEAGIG, 2007. Report section 2 (benthic inverts) Final draft June 07. http://circa.europa.eu/Public/irc/jrc/jrc_eewai/library?l=/milestone_reports/milestone_reports_2007/coastaltransitional/coast_nea_gig&vm=detailed&sb=Title: 9/03/2009.

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

Whittier TR, Stoddard JL, Larsen DP, Herlihy AT, 2007. Selecting reference sites for stream biological assessments: Best professional judgment or objective criteria. Journal of the North American Benthological Society 26: 349-360.

Wilson JG, Welkaim B, 1992. Estuarine bioindicators – a case for caution. Acta Oecologica 13: 345-358.

Woodward RT, Wui YS, 2001. The economic value of wetland services: a meta-analysis. Ecological Economics 37: 257-270.

Word JQ, 1978. The infaunal trophic index. California Coastal Water Research Project. Annual Report, 19-39 pp.

Word JQ, 1980a Extension of the infaunal trophic index to a depth of 800 meters. pp. 95-101 in: Southern California Coastal Water Research Project Biennial Report 1979-1980. Long Beach, CA.

Word JQ, 1980b. Classification of benthic invertebrates into Infaunal Trophic Index feeding groups. pp. 103-121 in: Southern California Coastal Water Research Project Biennial Report 1979-1980. Long Beach, CA.

Word JQ, 1990. The infaunal trophic index, a functional approach to benthic community analyses. Ph.D. dissertation. University of Washington. Seattle, WA.

Worm B, Barbier EB, Beaumont N, Duffy JE, Folke C, Halpern BS, Jackson JBC, Lotze HK, Micheli F, Palumbi SR, Sala E, Selkoe KA, Stachowicz JJ, Watson R, 2006. Impacts of Biodiversity Loss on Ocean Ecosystem Services. Science 314: 787-790.

Yachi S, Loreau M, 1999. Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. Proceedings of the National Academy of Sciences 96: 1463-1468.

Young TF, Sanzone S (Eds.), 2002. A framework for assessing and reporting on ecological condition: A SAB report. EPA-SAB-EPEC-02–009. U.S. Environmental Protection Agency, Washington, DC, 142 pp.

Ysebaert T, Herman PMJ, Meire P, Craeymeersch J, Verbeek H, Heip CHR, 2003. Large-scale spatial patterns in estuaries: estuarine macrobenthic communities in the Schelde estuary, NW Europe. Estuarine, Coastal and Shelf Science 57: 335-355.

Ysebaert T, Fettweis M, Meire P, Sas M, 2005. Benthic variability in soft-sediments in the mesohaline part of the Schelde estuary. Hydrobiologia 540: 197-216.

Zettler ML, Schiedek D, Bobertz B, 2007. Benthic biodiversity indices versus salinity gradient in the southern Baltic Sea. Marine Pollution Bulletin 55: 258-270.

# Acknowledgements / Agradecimentos

Dá-me especial prazer escrever esta secção pois é um bom indicador de que os objectivos a que me propus há quase 5 anos atrás chegaram a bom porto! Este trajecto foi muito e bem acompanhado e por isso quero relembrar todos aqueles que de vários modos contribuíram para o trabalho aqui apresentado.

Professor João Carlos e Fuen, muito obrigada por tudo! E o 'tudo' é muita coisa na verdade! Por me darem desde logo a oportunidade de trabalhar nesta área de investigação e de acreditarem em mim para os mais variados projectos! Pude crescer com todos os desafios e espero que considerem a aposta em mim ganha! Conto com os dois para as próximas etapas!

Fuen, és um exemplo de capacidade de trabalho incansável… Obrigada pela exigência e pela atenção comigo! De tanto acreditares em mim, no final convenceste-me de que sou capaz de tudo a que me propuser! Foste sempre muito mais do que uma orientadora!

Professor João Carlos, obrigado pela confiança depositada, pelo empenho em nos fazer sentir parte de toda a engrenagem e pelas oportunidades de crescer como investigadora! Não há palavras para o carinho com que sempre embarca as nas nossas ideias mais malucas!

This PhD would not have been possible without all the foreign research groups that took me and supported me during my visits.

Agradezco a Ángel Perez-Ruzafa y a los compañeros en el Departamento de Ecología de la Universidad de Murcia por recibirme en el grupo: Cármen, Mercedes, Vicky, Jonhy, Guiomar, Carlos, Marta, Jorge, y un especial 'obrigada' a José Chartón por la paciencia con la estadística (que al final me enganché y se me abrió el Tercer Ojo con el R)!

También a los padres de Fuen, Mari Paz y Andrés, gracias por el cariño y por los detalles. Una segunda familia en Murcia!

Ángel Borja, te agradezco la oportunidad de trabajar a tu lado desde el primer año de Doctorado hasta el último! Incluso con todo tu trabajo, siempre te mostraste muy calmo y disponible. Sólo espero que con los años haya crecido para afrontar mejor tus desafíos! Te agradezco también a ti y a toda la gente que conocí en AZTI vuestra cálida acogida: Iñigo, Juan, German, a todos los compañeros de la sala de becarios del año 2006, a los colegas de muestreo en los estuarios vascos, al equipo de fútbol de los Martes… muchas amistades que me hacen sentir como en casa cuando os visito!

Steve Weisberg, I am truly grateful to you for the confidence and opportunity you gave me to work as a peer with so many researchers from whom I could learn so much! I could never imagine that a 3 month staying at SCCWRP could result so profitable! Thanks, also to all the people at SCCWRP, especially Ananda, Chris Beegan, Keith, Jim, Steve Bay and Kerry for some after work beers, that made being away from home easier!

Ananda, you do deserve a special acknowledgement! Thank you so much for always being there through all the time I was in Cali, you were true friend! All the help since day one, the daily support at work, all the advices, all the talks, the trips, the Lakers, the fish tacos… Just hope I can return all of this here in Europe one of these days!

… E após cada viagem sempre se regressa a casa e, definitivamente, alguém que sempre cá está a zelar por nós é a nossa querida Gabi! Não é só o que seria do IMAR sem ti… é também o que seria de nós sem ti! Obrigada pelo ânimo, pelos conselhos, pelas gargalhadas, por tudo!

A minha gratidão vai também para todos os co-autores dos trabalhos aqui apresentados e ainda não mencionados, sobretudo à minha gente no IMAR: João Neto, Joana, Rute e Lena.

Agradeço também ao Professor Zé Paulo a sempre amável disponibilidade para consultas estatísticas e a oportunidade de me infiltrar nos cursos que foi organizando!

A todos os amigos do IMAR, aos que ainda estão ou por lá passaram, um sincero obrigada pela entreajuda no trabalho, nas saídas de campo, nos laboratórios, os 'brainstormings' na sala dos alunos, e os regeneradores cafés ou fins de tarde no Garcia, … no fundo, pela cumplicidade de quem está no mesmo caminho! Nomeio apenas alguns dos quais abuso(ei) especialmente da 'cumplicidade': Maria João, Filipinho, Patrick, Zazu, Rui Margalho, João Franco, Alexandra, Emerson, Sara… e Matildinha, a quem agradeço ainda a prontidão e o carinho na revisão final desta tese!

…. E sobretudo um obrigada sem tamanho a todos os amigos!!

À Marininha, Adriana e Fuen, por serem tão especiais e por, sem cerimónia nenhuma, terem saltado do trabalho e invadido todos os recantos da minha vida!

Aos amigos que orbitam em torno do eixo 'Coimbra/Vila Soeiro Team!', e aos quais esta tese me roubou durante largos fins-de-semana, obrigada pelo ânimo e pela paciência! Vocês foram um bálsamo estes 4 anos! Estou de volta!

À Sónia, Catarina, Bojeja e Pastel por anos e anos de amizade e por me manterem actualizada em Guimarães! Fica aqui a promessa de voltar a encurtar a distância! Um agradecimento especial ainda ao Pastel pelas dicas 'estético-gráficas' e ao Bojeja pela edição da capa desta Tese!

… E à minha família!
Aos meus pais, Luísa e Zé, e à minha Fanizinha! Imagino a alegria com que me vêem terminar este Doutoramento! É bom sentir o orgulho que têm em mim! Obrigada todos os dias!

E por fim, chegado à minha vida para trazer a mesma cor que deu à capa da Tese com esta foto maravilhosa… agradeço especialmente ao Òscar! Sobreviveste comigo a este último ano de Doutoramento… o resto da vida será muito mais fácil! Amo-t'estimo!