Universidade de Aveiro Departamento de Biologia
2012

Tânia Marisa
Neves Mendes

Avaliação integrada de rios baseada em diatomáceas e macroinvertebrados

Combined assessment of streams based on diatoms and macroinvertebrates

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Ecologia, Biodiversidade e Gestão de Ecossistemas, realizada sob a orientação científica da Doutora Salomé Fernandes Pinheiro de Almeida, Professora Auxiliar do Departamento de Biologia da Universidade de Aveiro e da Doutora Maria João de Medeiros Brazão Lopes Feio, Investigadora Auxiliar do Instituto do Mar da Universidade de Coimbra.

Dedico este trabalho ao meu pai.

I dedicate this work to my father.

o júri / the jury

Presidente / president	Doutor João António de Almeida Serôdio Professor Auxiliar do Departamento de Biologia da Universidade de Aveiro
Vogais / examiners committee	Doutor Bruce Chessman Principal Research Scientist at the New South Wales Office of Environment and Heritage
	Doutora Salomé Fernandes Pinheiro de Almeida Professora Auxiliar do Departamento de Biologia da Universidade de Aveiro (orientadora)
	Doutora Maria João de Medeiros Brazão Lopes Feio Investigadora Auxiliar do Instituto do Mar – CMA, Universidade de Coimbra (co-orientadora)

agradecimentos / acknowledments

A realização desta dissertação não teria sido possível sem a ajuda de várias pessoas.

Em primeiro lugar agradeço à Professora Doutora Salomé Almeida (orientadora) e à Doutora Maria João Feio (co-orientadora) pela ajuda, orientação e conhecimento transmitido ao longo da minha dissertação e também pela amizade e disponibilidade demonstradas.

À Carmen Elias, Sónia Serra, Ana Calapez e à Ana Luís por toda a ajuda disponibilizada durante as saídas de campo e no laboratório. Obrigada pela amizade, apoio e paciência.

Agradeço a dois amigos especiais, Raquel Silva e Luís Matos, os quais permaneceram sempre do meu lado.

Luís, obrigada por toda a paciência, dedicação, apoio e motivação em todos os momentos.

Por fim, um enorme obrigado à minha família pelo apoio incondicional e porque sem eles jamais teria chegado até aqui.

Palavras chave

Modelos preditivos, diatomáceas, macroinvertebrados, Directiva Quadro da Água, rios, biomonitorização

Resumo

As diatomáceas e os macroinvertebrados fornecem informação complementar na avaliação da qualidade da água. No entanto, os métodos utilizados para esse fim têm sido desenvolvidos separadamente para as duas comunidades. O objetivo deste estudo foi verificar se um modelo preditivo baseado nos dois elementos biológicos produz uma avaliação mais simplista e simultaneamente mais holística e robusta da qualidade dos ecossistemas face aos métodos individuais, os quais necessitam de ser combinados posteriormente, usualmente com base na abordagem “one-out al-out”. Para tal, foram utilizados dois métodos, RIVPACS e BEAST, devido às suas diferentes características, especialmente porque o RIVPACS utiliza dados de presença/ausência enquanto o BEAST utiliza dados de abundância. Foram construídos 6 modelos preditivos para o território português: dois para as diatomáceas, dois para os macroinvertebrados e dois integrando as duas comunidades. Nas primaveras de 2004 e 2005 foram simultaneamente amostradas diatomáceas e macroinvertebrados de 143 locais minimamente perturbados. Foram selecionados 23 locais afetados por contaminação orgânica, efluentes industriais e minas do centro de Portugal para serem utilizados como locais teste. O modelo RIV INV+DIAT atribuiu a mesma classe de qualidade do que o método “one-out all-out” a cerca de 70% dos locais teste, enquanto o BEAST INV+DIAT apenas partilhou cerca de 40% dos locais com a mesma classe. As respostas dos diferentes métodos (incluindo o “one-out all-out”) à degradação ambiental foram avaliadas através de correlações de Spearman. Apesar do RIVPACS ser menos sensível do que o BEAST, demonstrou funcionar melhor quando se combinam as duas comunidades. O tipo de dados influenciou a avaliação dos dois métodos demonstrando ser apenas fiável integrar as diatomáceas e os macroinvertebrados num único método usando dados de presença/ausência.

keywords

Predictive models, diatoms, macroinvertebrates, Water Framework Directive, streams, bioassessment

abstract

Diatoms and macroinvertebrates provide complementary information on stream water quality. However, classification methods have been developed separately for the two biological elements. The aim of the present study was to assess if a predictive model based on the evaluation of biodiversity using taxa from both biological elements, produces a simpler and simultaneously more holistic and accurate assessment of stream health than individual methods. These classifications need to be combined later, usually based on "one-out all-out" approach. For that purpose, two different approaches were used, BEAST and RIVPACS, due to their different characteristics, mostly because RIVPACS uses presence/absence data while BEAST uses abundance. Six predictive models were built for the entire Portuguese territory: two for diatoms, two for macroinvertebrates and two combining diatom and macroinvertebrate communities. Data from 143 minimally disturbed sites sampled simultaneously for diatoms and invertebrates in the spring of 2004 and 2005 were used to calibrate and validate the models. For all the six predictive models, 23 impacted streams from central Portugal affected by organic contamination, industrial effluents and mine drainage were used as test sites. The RIV INV+DIAT model shared with "one-out all-out" approach about 70% of the test sites with the same quality class while the BEAST INV+DIAT model only shared about 40%. The responses to the environmental degradation of the different approaches (including the "one-out all-out") were analyzed through a Spearman correlation. In spite of the less sensitive RIVPACS approach results in comparison to BEAST, it showed to work better when the two biological elements were joined. The type of data influenced the assessment of the two approaches and diatoms and macroinvertebrates can be integrated reliably into a single method using only the presence/absence type of data.

Index

Index of Figures

Index of Tables

Acronym list

ANOVA	Analysis of Variance
ASPT	Average Score Per Taxon
BEAST	Benthic Assessment of Sediment
BMWP	Biological Monitoring Working Party
CEE	Índice da Comunidade Económica Europeia
DA	Discriminant Analysis
DFA	Discriminant Function Analysis
DGA	Atlas Digital do Ambiente
E	Expected
EQR	Ecological Quality Ratio
GQA	General Quality Assessment
HMS	Habitat Modification Score
HQA	Habitat Quality Assessment
IBD	Indice Biologique Diatomées
IBI	Index of Biotic Integrity
IBMWP	Iberian Biological Monitoring Working Party
INAG	Instituto Nacional da Água
IPS	Indice de Polluossensibilité Spécifique
IPtI	Invertebrate Portuguese Index
MDS	Non-metric Multidimensional Scaling
Med-GIG	Mediterranean Geographical Intercalibration Group

MN	Mean Value
NRA	National River Authority
O	Observed
O/E	Observed Expected ratio
PC	Probability of Capture
PCA	Principal Component Analysis
RIVPACS	River Invertebrate Prediction And Classification System
RMSE	Root Mean Squared Error
RSSD	Replicate Sampling Standard Deviation
SD	Standard Deviation
SI	Saprobien Index
TDI	Trophic Diatom Index
UK	United Kingdom
UPGMA	Unweight Pair Group Method with Arithmetic mean
USA	United States of America
WFD	Water Framework Directive

Chapter 1 - INTRODUCTION

Water degradation is a result of human activities, as most societies are clustered as close as possible to rivers, facilitating waste disposal (Perry and Vanderklein 1996). The realization that unmanaged ecosystems will soon fail to provide free ecological services, such as drinking water, fish and waste assimilation, has led to a considerably improvement in the water legislation viewing the protection of the aquatic ecosystems (Cairns Jr. and Pratt 1993).

Until the 90s, most water quality monitoring programs were focused only on chemical analysis. This is an accurate approach but presents the disadvantage of providing a fragmented overview of ecosystem health, as well as providing information of the water quality only at the time of sampling (Alba-Tercedor and Sánchez-Ortega 1988, Atazadeh et al. 2007, Bere and Tundisi 2011a). On the other hand, assessment based on biological communities has several advantages over the physical and chemical measurements of water quality: they show the cumulative effects of present and past condition and therefore provide a direct, holistic and integrated assessment of environmental conditions that are highly variable in space and time (Bere and Tundisi 2011a, Stoermer and Smol 1999). Furthermore, the use of biological indicators for assessment of water quality is now mandatory under the European Water Framework Directive of 2000 which should achieve the good ecological status (quantitative and qualitative) until 2015 and ensure the sustainable use of aquatic environments (WFD; EC Parliament and Council 2000).

Biomonitoring is defined as the measurement and evaluation of the ecosystem condition using biological responses to impacts, usually caused by human activities but also implies quality control through corrective and preventive actions when the expected conditions are not achieved (Matthwes et al. 1982). The idea of biological monitoring is not new. In the early days of the industrial revolution, canaries were kept in underground coal mines and if they showed adverse responses to conditions, the miners abandoned the mine (Cairns Jr. and Pratt 1993). The modern history of aquatic biomonitoring began in Europe in the twentieth century. Studies of biological indicators relied on the identification of indicative species of human degradation and biological classification of streams (Cairns Jr. and Pratt 1993). The use of biological communities (fishes, invertebrates, macrophytes and algae) as indicators of water quality is evolving and are becoming more widely used while initially mainly invertebrates were used. Especially in the past two decades, several methods have been developed to assess streams ecological health such as autoecological indices, indices of biotic integrity, predictive models and others.

1.1 Macroinvertebrates as bio-indicators of water quality

The term macroinvertebrate describes the animals that have no backbone and that can be seen by the naked eye. Normally these organisms exceed 500 µm of body size. They are mostly insects but also decapods, crustaceans, mollusks, leeches, oligochaetes and planarians. The majority of freshwater insects has an amphibiotic life cycle and spends their adult stage on land. Macroinvertebrates commonly inhabit the bottom substrates (sediments, debris, logs, macrophytes, filamentous algae, etc.) and are referred as benthic macroinvertebrates or macrobenthos (Cummins 1992, Rosenberg and Resh 1993).

The macroinvertebrate communities are the most widely used for assessing water quality for several reasons. They are found along the river continuum, are cosmopolitan and respond to changes in water quality resulting from anthropogenic disturbances (Azrina et al. 2006). Because of their limited migration, they are good indicators of localized impacts. These organisms have a complex life cycle of approximately 1 year so they can integrate and reflect the environmental changes that they have gone through (Barbosa et al. 2001). The freshwater macroinvertebrates include representatives of many insect orders that contribute to important ecological functions such as decomposition, nutrient cycling and play an important role in food webs as both consumers and prey (Kenney et al. 2009). They are relatively easy to identify to family level and many taxa can be identified to lower taxonomic levels. There are many species within a community with different ranges of tolerance and sensitiveness to stress that provides information for interpreting the cumulative effects (Abbasi and Abbasi 2011). Sampling of macroinvertebrates in wadeable rivers is relatively easy and inexpensive and has minimal adverse effect on the resident biota (Barbour et al. 1999). In addition, methods for analyzing their data are well established.

Many methods based on macroinvertebrates for evaluating ecosystems health have been developed through time and implemented in Europe in the beginning of the XXth century. Most of them are based on Kolkwitz and Marsson (1908, 1909) and originated new biotic indices (Figueroa et al. 2003).

Biotic indices are numerical expressions combining a quantitative measure of species diversity with the qualitative information on the ecological sensitivity of individual taxa (Bieger et al. 2010). They are based on the assumptions that the number of taxonomic groups decreases and that macroinvertebrates follow this disappearing sequence with the increase of organic pollution: Plecoptera, Ephemeroptera, Trichoptera, *Gammarus*, *Asellus*, red migdes Chironomidae and Tubificidae. The declining order only reflects their tolerance to organic pollution (Czerniawska-Kusza 2005).

Beck was the person who popularized the term “biotic index”. Beck’s Biotic Index (Beck’s BI - 1954) is based on macroinvertebrates tolerances to organic pollution and was developed in Florida. It’s considered the real first biotic index because it included description of field procedures and identification to the species level. Organisms were divided into three classes: “Class I” for the intolerant and “Class II” for the facultative and “Class III” for those tolerant to organic pollution. However, he decided not use the tolerant organisms because they could be found in clean waters, but in lower abundance. The Beck’s indice value can oscillate between 0 and 40, but it not takes into account the organism’s abundance, only attributes the numeric values of 2 and 1 to the “Class I” and “Class II”, respectively. So, final score of the index for a site is calculated by summing the number of species of “Class I” multiplied by two, with the number of species of “Class II” (Davis 1995).

The Trent Biotic Index [TBI – (Woodiwiss 1964)] was developed by the Trent River Authority in England. The sampling included all available habitats during 10 minutes with a hand-net. The index’s value is based on the presence or absence of six “groups” of invertebrates with different degrees of tolerance to organic pollution. The final value can vary from 0 (grossly polluted) to 10 (unpolluted). The Trent Biotic Index served as model to other several biotic indices (Muralidharan et al. 2010).

The Belgian Biotic Index [BBI – (De Paw and Vanhooren 1983)] was developed in Belgium and combined different biotic indices. All available habitats are sampled with a 300-500 µm mesh hand-net during 3 or 5 minutes, depending on the width of the river. The macroinvertebrates are preserved *in situ* and identified to family or genus levels in the laboratory. The final value varies from 0 (very heavily polluted) to 10 (unpolluted) (Abbasi and Abbasi 2011).

The Biological Monitoring Working Party (BMWP) Score index (Chesters 1980) was developed in Britain and has been widely applied. The macroinvertebrates are identified to family level and each one receives a score between 1 (most tolerant) and 10 (least tolerant). This index does not take into account the abundance. The BMWP score is the sum of individual scores that can be divided by the number of taxa with score to produce the Average Score Per Taxon (ASPT). The ASPT is less influenced by season and sample size than BMWP score (Muralidharan et al. 2010). The Iberian IBMWP – (Alba-Tercedor et al. 2002) is and adaptation of the BMWP Score System to Iberian rivers. All available habitats are sampled over a 100m stretch with a kick-net with 250 µm mesh size and the invertebrates are identified to family level. The final IBMWP score, number of taxa and IASPT (IBMWP score divided by number of taxa) are calculated for a site based on all the taxa collected and observed (Abbasi and Abbasi 2011). These indices were, until the implementation of the WFD, the most commonly used in Portugal and Spain.

Nowadays, the official index in Portugal is the IPtI (Invertebrate Portuguese Index), established by INAG (2009). This is a multimetric index produced during the Intercalibration Exercise carried out by the Mediterranean Geographical Intercalibration Group (Med-GIG), in which Portugal took part and which aimed the comparability of quality assessments and compliance with the WFD. The index is divided in two indices, the $IPtI_N$, applied to rivers in North of Portugal and the $IPtI_S$, applied to rivers in the South and Littoral. It's calculated as the weighted sum of some metrics, each normalized using the ratio between the obtained values and the corresponding reference values which is dependent on the river type. The final value (Ecological Quality Ratio-EQR) varies between 0 and approximately 1 (for reference sites).

Currently, over the world, the most common methods for assessing water quality are the multimetric indices and the multivariate approaches. The multimetric indices integrate into a single value different metrics (e.g. taxonomic diversity, exposure of the community to stressors) of the biological community that are sensitive to a broad range of human activities. The chosen metrics are calculated from the taxa data matrix at the sites and can be combined (hence “multimetric”) to enhance predictability compared to individual ones (Milner and Oswood 2000). The first multimetric index was developed by Karr (1981) in the USA, the IBI (Index of Biotic Integrity). This index incorporated zoogeographic, ecosystem, community and populations aspects of fish assemblages into a unique ecologically-based index.

The multivariate approaches rely on multivariate statistical methods to uncover patterns in taxonomic composition and will be described in detail later in this chapter.

In spite of the macroinvertebrates being the preferred organisms for assessing water quality (Harding et al. 2005), they also present some disadvantages such as their aggregated distribution which implies many subsamples to collect a representative sample. Moreover, some insects are absent in the water during part of the year and this should be taken into account in the interpretation of the results (Muralidharan et al. 2010). According to Charles (1996), the use of algae for monitoring rivers has increased because of these limitations with benthic invertebrate methods, coupled with significant improvements in technologies for algal assessment. The algal class Bacillariophyceae, the diatoms, is one of the groups of organisms that fulfill the requisites needed for biological monitoring.

1.2 Diatoms as bio-indicators of water quality

The word “diatom” comes from Greek, which means cut in two. The characteristic feature of diatoms is its rigid cell wall composed of silica, called frustule. Each frustule is box-like in structure and made of two parts, the valves. Diatoms are eukaryotic microscopic unicellular organisms, although chains of cells and colonial aggregations may also occur. These algae are pigmented and most are photosynthetic. Diatoms are ubiquitous in their distribution and can be found in all waters except the hottest and most hyper saline. In freshwater, diatoms can live in open water (planktonic) or attached to substrata (periphytic). Periphytic diatoms are found attached to rocks (epilithon), sand grains (epipsammon), plants (epiphyton) and soft sediments (epipelon) (Bold and Wynne 1985, Jones 2007).

Within the algae, diatoms have been the main focus of bioassessment studies (Bold and Wynne 1985, Jones 2007, Lee 1980, Bellinger et al. 2006). These organisms play a crucial role as primary producers in streams and due to their position in food webs it is expected that any disturbance in diatom populations affect the whole aquatic community (Andrén and Jarlman 2008). Diatom assemblages are considered useful tools in water quality monitoring for many other reasons. They form a large part of the benthos (about 90%), are ubiquitous and occur in all types of aquatic systems (Solak and Acs 2011). Due to their short generation time (high reproduction rate), diatoms show quick responses to water quality degradation by changing species composition and diversity (Bere and Tundisi 2011a). For a large number of species ecological information is available and many show narrow ranges of tolerance to several abiotic features. This information in conjunction with the persistence of frustules in sediments has been used for historical reconstruction (Cooper 1995). These organisms can be preserved and stored indefinitely as permanent slides and reinvestigated whenever necessary (Solak and Acs 2011). In addition, diatoms are easy to sample and their identification is possible through taxonomic guides because it is mainly based on frustule morphology (Aboal et al. 2003, Krammer 2000, 2002, 2003, Krammer and Lange-Bertalot 1986, 1988, 1991a, 1991b, Lange-Bertalot 2001, Lange-Bertalot et al. 2003, Levkov 2009, Werum and Lange-Bertalot 2004). Nevertheless, the use of these organisms presents the disadvantage of requiring taxonomic expertise (Solak and Acs 2011).

The assessment of water quality in freshwater habitats with benthic diatoms has a long history and the first studies date back a century ago (Kireta et al. 2012). These methods have been reviewed by many authors (Lowe and Pan 1996, Patrick 1973, Rosen 1995, Stevenson and Lowe 1986, Whitton and Kelly 1995, Whitton et al. 1991). Within the last decades diatom-based indices became popular worldwide, especially in Europe (Bere and Tundisi 2011a).

According to Stevenson and Pan (1999), two different approaches using diatoms have been developed: the autoecological indices based on Kolkwitz and Marsson (1908, 1909) works (Butcher 1947, Descy 1979, Slàdecek 1973, Zelinka and Marvan 1961) and the studies centred on the diversity of diatoms as an indicator of river health based on Patrick's monitoring studies (Patrick 1949, Patrick and Strawbridge 1963, Patrick et al. 1954). The autoecological indices use the relative abundance of species and are based on the assumption that species have specific optima and tolerances, sensitivities or preferences for environmental conditions (Stevenson 1998). Diatoms are known to respond to eutrophication, organic pollution, heavy metals, salinity, pH, pesticides, and their sensitivity/tolerance to those environmental characteristics differ among species (Stevenson and Pan 1999). Most of those indices are based on the weight average equation of Zelinka and Marvan (1961) and, according to Rimet et al. (2005), there are as many indices as the number of researchers working in the field. The most significant development during the 80's was the Indice de Polluossensibilité Spécifique (IPS – Cemagref 1982) that provides integrated assessment of a range of water quality variables such as organic pollution, eutrophication, salinity and toxic substances (Solak and Acs 2011). More indices were developed in other countries, like the Trophic Diatom Index (TDI) in UK (Kelly and Whitton 1995), the Saprobienindex (SI) in Austria (Rott et al. 1997) and the Indice Biologique Diatomées (IBD) in France (Lenoir and Coste 1996).

Some characteristics of diatom communities have also been used to assess the ecological integrity of streams such as biomasss, morphology, chemical ratios (chl *a*, N, P, heavy metals, etc), growth, dispersal and metabolic rates. Usually these features are used together with the characteristics of the entire periphyton or plankton assemblages (Stevenson and Pan 1999).

In Portugal, under WFD legislation, two different diatom-based indices were adopted, the IPS for the North of Portugal and the European Index (CEE – Descy and Coste 1990) for the South of Portugal. The IPS index is based on Descy's method and differs only on the indicator and sensitivity values of taxa. The species were grouped in 5 classes from 1 (tolerant species) to 5 (sensitive species). The final values of IPS are then converted into EQRs by dividing them by the reference value for their river type (as established in the WFD implementation). Finally intervals of these values correspond to quality classes (high = 1, good = 2, moderate = 3, poor = 4 and bad = 5) (INAG, I. P. 2009). The CEE index is based on a two-way entry table, which includes 208 taxa. In this table, taxa are grouped into 8 groups arranged in descending order of sensitivity to pollution (group 1 more sensitive and group 8 more tolerant). Vertically, there are 4 subgroups of taxa (9 to 12) with restricted geographic distribution based on alkalinity and mineralization. The index value is obtained by crossing the median values of the group and subgroup (those containing 50% or more of abundance of the taxa involved in the calculation), which is then normalized and can vary from 1 (strongly polluted) to 20 (unpolluted) (INAG, I. P. 2009).

Biotic indices are useful tools for rapid bioassessments but they should be wisely interpreted because of their limitations. The most important is the restricted applicability due the geographic area that they are built for (Abbasi and Abbasi 2011). There are evidences that indices developed for one area are less successful when applied in others because of the floristic differences among regions (Bere and Tundisi 2011a). According to Besse-Lototskaya et al. (2011), European indices use different ecological profiles for the same species because most of them are rare and difficult to define, and therefore, not robust.

1.3 Predictive Models overview

More recently, predictive models appeared as an alternative to the traditional indices in some regions of the world. The predictive models are based on multivariate analysis and follow the concept of Reference Condition. Reynoldson et al. (1997) defined the reference condition as a group of sites in which physical, chemical and biological features are within the range characterized as undisturbed or minimally disturbed. The predictive models measure river health as the alteration of the biological community composition to an expected community under reference conditions. Predictive models are founded on the biological classification of reference sites, based on the similarity between species composition (Reynoldson et al. 1997). According to Zamora-Muñoz and Alba-Tercedor

(1996), the main advantage of multivariate methods is the small reduction of multidimensional data with consequent minimal loss of information, identifying the direction of data variability. The greater disadvantage is the huge effort in the initial construction phase. Nevertheless, this problem can be bypassed with software that integrates model analysis and yields easily understandable results (Feio 2004).

Initially, the predictive models were based on macroinvertebrate communities. The first one, RIVPACS (River InVertebrate Prediction And Classification System – Wright et al. 1993) was developed in Great Britain. It was followed by the BEAST (BEnthic Assessment of Sediment – Reynoldson et al. 1995, 1997) developed in Canada and the AUSRIVAS in Australia (AUstralian RIver Assessment Scheme – Marchant et al. 1997, Simpson and Norris 2000). Later, other predictive models appeared based on fishes (Kennard et al. 2006), diatoms (Chessman et al. 1999, Feio et al. 2007) and macrophytes (Aguiar et al. 2011).

In Portugal, the first predictive models were built initially for the Mondego River basin based on macroinvertebrate communities and following the BEAST approach (Feio et al. 2004, 2006a, 2006b, 2007). Three predictive models were built with 55 reference sites using three levels of taxonomic resolution: 1) the lowest practical taxonomic level (Feio et al. 2007); 2) family level (Feio et al. 2006b) and 3) order level (Feio et al. 2006a). Models performances were tested with 20 test sites that covered a wide range of stream types and all seasons. The best performing model was the one built at the lowest practical taxonomic level.

Since then, other studies addressing predictive models based on different approaches, using different biological communities have been developed (review in Feio & Poquet 2011). Feio et al. (2010) built also predictive models based on functional parameters: one for decomposition (D model), using microbial and total decomposition rates in oak and alder leaves and the other based on biofilm characteristics (B model), using sediment respiration rates, biofilm growth and total chlorophyll *a* of biofilm on natural substrata, the autotrophic index and fungal biomass on conditioned oak leaves. This study showed that functional variables, especially decomposition, can be useful ecological indicators in monitoring programs. Aguiar et al. (2011) built two macrophyte predictive models, one based on RIVPACS and the other based on BEAST approaches. The models were developed for the entire country (Portugal) and the objectives were to test the suitability of two predictive modeling approaches to macrophyte communities as a water-quality assessment tool and compare their performance with other more common approaches. Almeida and Feio (2012) tested the adaptation of the RIVPACS/AUSRIVAS methods to Portuguese rivers through the development of a predictive model based on diatoms (DIATMOD model). Mendes et al. (2012) used two diatom predictive models developed for Portugal: the MoDi based on BEAST and the DIATMOD based on RIVPACS/AUSRIVAS approaches. The goals of this study were to determine the effect of substrate type and the evaluation method on the assessment of water quality.

The RIVPACS, BEAST and AUSRIVAS approaches are within the most commonly used predictive models (Feio and Poquet, 2011). However, the RIVPACS is probably the furthermost popular approach and is now well established in several countries such as U.K., Australia, Canada, Sweden and the Czech Republic (Clarke and Murphy 2006).

1.3.1 RIVPACS approach

The development of the RIVPACS models started in October 1977 with two major goals: 1) development of a biological classification of minimally polluted waters in Great Britain based on macroinvertebrate communities; 2) determine if those communities could be predicted based on physical and chemical features for each site.

In 1986, the first version of RIVPACS was implemented on a microcomputer and made available to water industry biologists throughout Great Britain for testing. By then, RIVPACS I included 370 reference sites which resulted in 30 groups. The classification and predictions were based on species level (Wright 2000). In 1990, the National River Authority (NRA) funded the development of an operational version, the RIVPACS II, for use in the 1990 River Quality Survey. RIVPACS II was used at almost 9000 sites in 1990 River Quality Survey throughout England, Wales and Scotland and on a more experimental basis in Northern Ireland where there were no local references. In this version, further streams were added to give a total of 438 reference sites which resulted in a new classification with 25 groups, as well as other improvements (Wright 2000). The 1995 General Quality Assessment (GQA) required an upgrade of the system, so data collected from sites with high biological quality, sites recommended by local biologists and sites of Northern Ireland were added to develop RIVPACS III. One important modification was also implemented: while RIVPACS II was based on qualitative species data, RIVPACS III used qualitative species data plus family data to characterize each site (Wright 2000). Several other improvements were made to the successive versions of RIVPACS such as standardizing sampling protocols, assessing different taxonomic levels, developing single and combined season models, predictive models using both qualitative and quantitative data, studying alternative procedures for site classification and prediction, assessing the uncertainty of the predictive systems outputs (Feio and Poquet 2011). The RIVPACS III+ represents the major step forward through the incorporation of error terms for the O/E ratios used to assess site quality and provides a mechanism for detecting statistically significant spatial and temporal differences between the macroinvertebrate assemblages of sites (Wright 2000). It is implemented in a software package and it was used under the Water Framework Directive in U.K. (Feio and Poquet 2011).

A RIVPACS predictive model is built in several steps. The first and crucial step is the selection of reference sites (Clarke et al. 2003). The next step is collecting the biological and environmental data. Only environmental features not affected by stressors can be used as predictors. Then, the reference sites are grouped according to their similarity between species composition and by means of ordinations obtained from the correspondence analysis method. Then a Discriminant Function Analysis (DFA), which predicts group membership based on multiple linear regressions, is used to select the potential environmental descriptors that best discriminate the reference groups. To determine the discriminatory power of the DFA, the RIVPACS relied on re-substitution and cross-validation analyses. These analyses provide a percentage value of reference sites correctly located in their original groups. In the cross-validation analysis, one reference site is left aside from the others each time and later it is used to rebuild the DFA model. Then the reference sites are used as test sites and the number of sites attributed to its original group gives the percentage of correct classification. The re-substitution analysis occurs the same way as cross-validation, but the DFA model is not rebuilt for each site. The probability of a new test site belonging to a group can be calculated from the Mahalanobis distance between test site and center of each biological group (Clarke et al. 2003). RIVPACS uses these probabilities of belonging to each reference group to calculate the expected fauna. The final probability of capture (Pc) for a taxon in a test site is calculated as the sum of the probabilities of belonging to each biological reference group, weighted by the frequency of occurrence of that taxon inside each group (Feio and Poquet 2011). The deviation of the observed (O) from the expected taxa (E) is measured by the ratio O/E. Low ratio values (O/E close to 0) means that test sites are strongly impacted by some environmental stressor while ratio values close to 1 means that a site is near to reference (Hawkins et al. 2000). The expected number of taxa (E) is calculated as the sum of individual Pc for all taxa found in a test site. Only taxa with $Pc \geq 0.5$ are usually used to calculate E because rare taxa appear to decrease the performance of models. The standard deviation (SD) of O/E characterizes the magnitude of predictor error and low SDs indicates that the model accounts for much of natural variability and provides good predictions. The final biological evaluation obtained by the RIVPACS approach can vary along an assessment gradient (O/E gradient). Based on the taxa predicted to occur, the RIVPACS approach also produces two biotic indices: the ASPT and BMWP for each test site. For the ASPT the lower 5%, and for the number of taxa and the BMWP the lower 10% of the reference O/E distribution are then used as a threshold to considerer the test sites as impacted by setting the quality classes (below reference) in order to calculate the deviation of a site from the reference condition (Feio and Poquet 2011).

1.3.2 BEAST approach

The BEAST approach was developed in Canada to create criteria for sediment quality of the North American Great Lakes by Reynoldson et al. (1995, 1997). This approach was based on methods developed in the United Kingdom with the main goal to determine predictive associations between the macroinvertebrates and the physic-chemical parameters. During the period from 1991 to 1993, 345 samples were collected from 245 sites and included in the construction of the predictive model (Reynoldson et al. 1995). Later, Rosenberg et al. (2000) also built a BEAST model type based on macroinvertebrates collected from 127 reference sites during autumns of 1994 and 1995. This model was developed for the Fraser River located in North America.

The BEAST models are built in 3 steps: 1) first a PCA (Principal Component Analysis) analysis was used to identify patterns in environmental data through the ordination axes developed from the biological data; 2) determine the environmental variables that best discriminate the biological groups using a DA (Discriminat Analysis); and 3) identify the environmental variables that differ significantly between the biological groups through an ANOVA (Analysis of Variance) (Feio and Poquet 2011).

The major difference between RIVPACS and BEAST approaches (Figure 1) is the assessment of test sites. In BEAST models, the community composition of the test site is compared to the sites included in the biological reference group to which the test site is most likely to belong, based on their environmental characteristics (discriminant predictors). These data are merged in the same matrix, re-ordinated in a MDS-ordination space and plotted, in the original method, in a banding system defined by Gaussian probability ellipses (90, 99, 99.9%). The distance of the test site from the ordination centroid results in the biological assessment. A site located in the first band (inner ellipse) is considered equivalent to the reference, a site located in the second band (90-99%) was potentially different from reference, a site located in the third band (99-99.9%) is different from reference and a site located in the fourth band (beyond 99.9%) is very different from the reference. Thus, the BEAST predictive model gives a direct evaluation of the water status (Reynoldson et al. 1997, Feio and Poquet 2011).

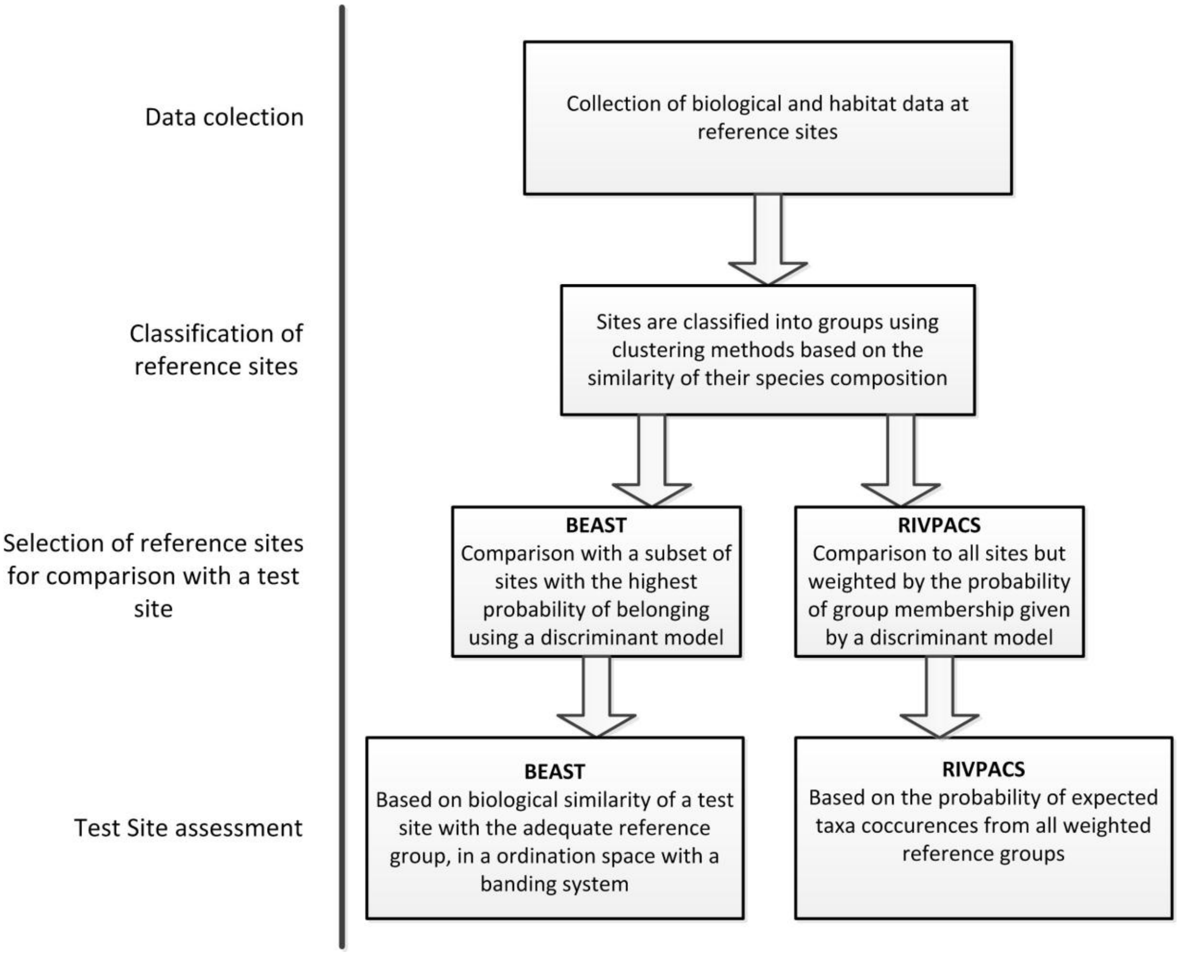

Figure 1 - Flowchart of assessment methods using BEAST and RIVPACS approaches. ***Adapted from Reynoldson et al. (1997).***

1.4 Goals

Independently of the biological assessment methods used, when more than two biological elements are evaluated, there is a need for a global assessment of the studied site. Presently, that is commonly done in Europe by combining the assessments *a posteriori*. This combination is done based on the “one-out all-out” approach, which is a conservative approach that many researchers consider unrealistic.

Therefore, the main goal of this study is to evaluate if a predictive model based on the evaluation of biodiversity using the taxa from two biological elements (macroinvertebrates and diatoms), produces a simpler and simultaneously more holistic and accurate assessment of streams health than individual assessments combined *a posteriori*. For that purpose, we used two different approaches due to their different characteristics: 1) the RIVPACS technique which is based on presence/absence data and only includes frequent taxa and 2) the BEAST methodology, based on abundance data that takes into account the entire community. For comparison, we built six predictive models, three of them based on RIVPACS approach and the other three based on BEAST approach, for continental Portugal: two for diatoms, another two for macroinvertebrates, and the last two integrating diatom and invertebrate assemblages. For all the six predictive models, 23 impacted stream sites affected by mine drainage, organic contamination and industrial effluents were used as test sites. The performance of the combined models was achieved by comparing the assessment of the test sites against the assessments made by the individual ones. Those assessments were also compared with those that would be obtained with the "one-out, all out" approach currently used in Europe.

Chapter 2 – STUDY AREA

The study area (Figure 2) comprises three adjacent river catchments with a total area of 11215 km^2 located in central Portugal: Mondego, Vouga and Lis. This region has a temperate Atlantic climate and highly diverse geological landscapes (Feio et al. 2009a).

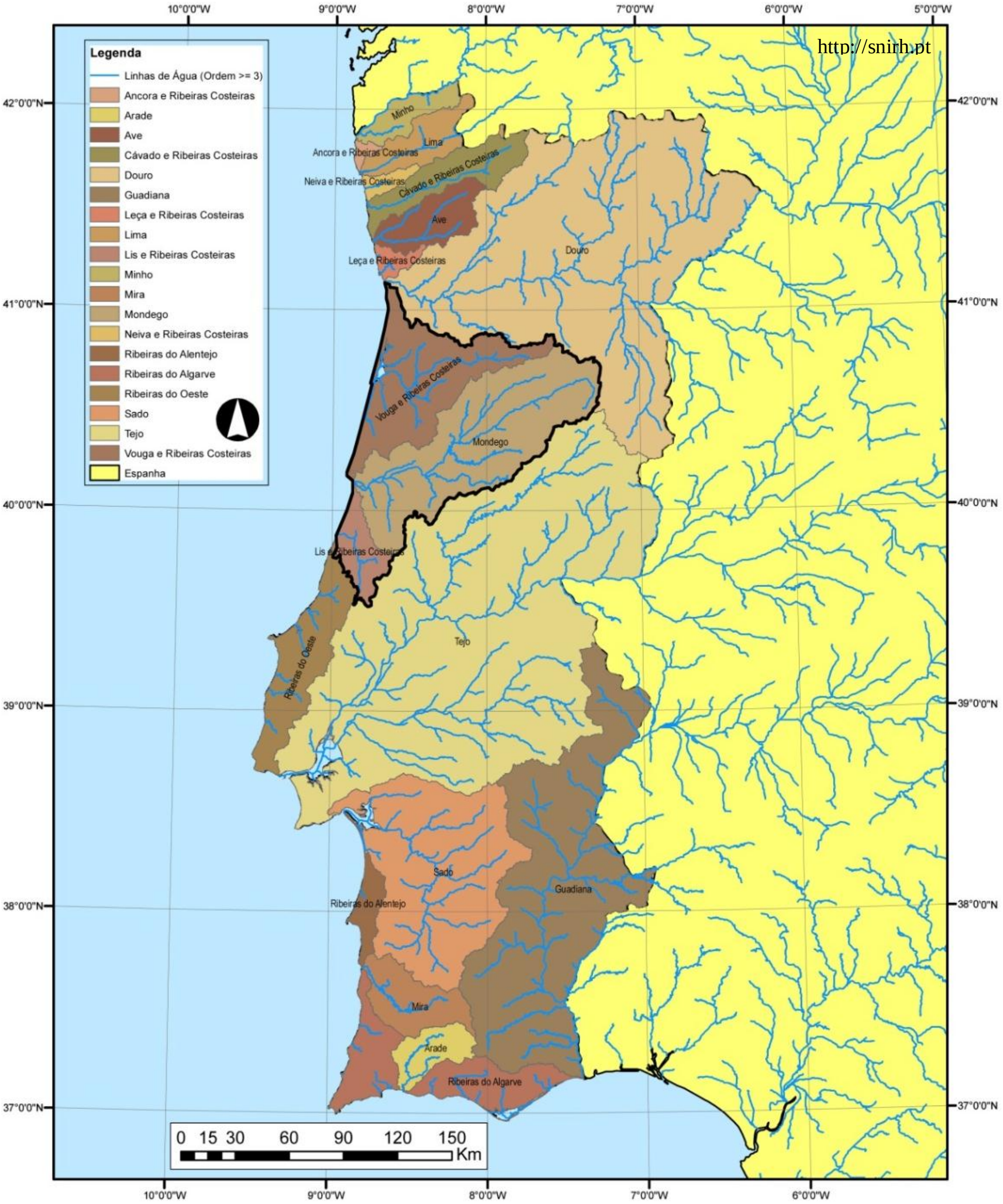

Figure 2 - Hydrological basins of Portugal. The study area is marked with a black outline.

The Mondego is the largest river entirely in national territory located between 39°46’ and 40°48’ N and 7°14 and 8°52’ W, covering an area of 6670 km^2. The river starts to flow in Serra da Estrela at 1547 m of altitude in a small fountain called “Mondeguinho” and runs along 300 km until it reaches the Atlantic Ocean, nearby the city of Figueira da Foz. The main tributaries of this river are Dão on the right bank and Pranto, Arunca, Ceira and Alva rivers on the left one. The basin has an approximately rectangular shape elongated in NE-SW direction. Along the river, three distinct segments can be distinguished: high, medium and low sections. In the high section the river flows through glacial valleys in which the substrate is coarse and mostly granite and schist. In the medium section the river flows in valleys between Serra da Estrela and Coimbra where the Dão, Alva and Ceira rivers converge. The dominant substrates remain the same as in the high section, granite and schist. In the low section the river runs through Coimbra in open valleys to floodplains and the bedrock is limestone with fine sediments (Feio et al. 2007, PBH 1999a). The main anthropogenic pressures felt in the littoral (low section) are agriculture (extensive rice fields) and urban effluents. In the interior the main impacts are the presence of dams and weirs, some milk and cheese industries and mine drainage (Feio et al. 2009a, Feio et al. 2010).

The Vouga river source is located at 930 m of altitude in Serra da Lapa (Chafariz of Lapa), located in Viseu district. The Vouga’s basin is the second largest that runs entirely in Portugal and it is limited at 40°15’ and 40°5’ N and 7°33’ and 8°48’ W. The Vouga river covers a total area of 3706 km^2. This basin is composed by a hydrographic set of rivers that discharge very close to the Vouga’s mouth in Aveiro estuary (Ria) that communicates with the ocean. The main rivers of this set are Águeda, Cértima, Caster, Antuã and Boco rivers and also Corujeira stream in which the substrate is schist and granite. The Vouga river flows along different types of valleys: through an upland until São Pedro do Sul where the basin is elongated-shaped, in a valley with a high slope between São Pedro do Sul and Albergaria-a-Velha, through open valleys until Aveiro and in the estuary (PBH 1999b). The major impacts affecting this basin are the large-scale eucalyptus plantations and paper pulp industries (Feio et al. 2010).

The Lis river is the smallest catchment covering an area of 945 km^2 and it is limited between 39°30′ and 40°00′ N and 8°35′ and 8°00′ W. Lis basin topography is smooth, mostly below 200 m of altitude. The maximum altitude is 562 m in Pedra of Altar. The main water courses are Lis and Lena rivers that run through a limestone massif and an old pine forest. The valleys of Lis and Lena rivers are wide and flat, typical of alluvial floodplains. The Lis valley only narrows when it crosses the anticlinal structure of Leiria and then extends downstream of the confluence of Lena where it forms an alluvial floodplain with 1 km wide. The coastline consists of dunes that include some of the highest in our country (50 m) (Feio et al. 2007, PBH 1999c). The major impacts affecting Lis basin are the dense urbanization and the swine farming (Feio et al. 2010).

2.1 Test site characterization

Twenty-three study sites were sampled in the spring of 2011. Within the 23 sites, 14 were collected in Mondego basin, 7 in Vouga basin and 2 in Lis basin. These sites were selected to cover different levels and types of anthropogenic degradation. The codes attributed to test sites came from a preexisting database in which the letter corresponds to the basin (M to Mondego, V to Vouga and L to Lis) and the numbers corresponds to the sampling order and consequently not sequential.

2.1.1 Botão (M18)

This site is located in Botão stream at 85 m of altitude and belongs to the Mondego basin. The M18 site (Figure 3) is 3 m wide and 24 cm of depth (on average, at sampling location). When sampled, the water was clear and the channel substrate was coble and gravel/pebble. Around the site eucalyptus plantations were present. The riparian vegetation included grasses and alders.

Figure 3 - Sampling site Botão (M18).

2.1.2 Foz do Alva (M55)

This site is located in Alva river at 39 m of altitude and belongs to the Mondego basin. The M55 site (Figure 4) is 12.4 m wide and a depth of about 40 cm. At the time of sampling, the water was clear and the channel substrate was dominanted by cobles. Around the site eucalyptus plantations and acacias were present. The riparian vegetation, included grasses and alders and many acacias.

Figure 4 - Sampling site Foz do Alva (M55).

2.1.3 Lousã-Piscinas (M2002)

This site is located in São João stream at 236 m of altitude and belongs to the Mondego basin. The M2002 site (Figure 5) is 5.10 m wide and a depth of 25 cm. At the time of sampling, the water was clear and the channel substrate was dominated by boulders and stones made of schist. Around the site semi-natural mixed woodland was present.

Figure 5 - Sampling site Lousã-Piscinas (M2002).

2.1.4 Lousã-Fábrica do Papel (M101)

This site is located in São João stream at 176 m of altitude, downstream from M2002 site and belongs to the Mondego basin. The M101 site (Figure 6) is 22 cm deep. At the sampling time, the water was clear and the channel substrate was dominated by coble and gravel/pebble. Part of the channel was obviously realigned. Around the site the land was used for agriculture (orchards) and pasture. The riparian vegetation, when present, included brambles and acacias. The sample location is downstream of a bridge and a weir, as shown in the photo. The major impacts affecting this stream are a landfill and a paper pulp industry.

Figure 6 - Sampling site Lousã-Fábrica do Papel (M101).

2.1.5 **Foz do Ceira (M2001)**

This site is located in Ceira river at 19 m of altitude and belongs to the Mondego basin. The M2001 site (Figure 7) has a 42 cm depth. At the sampling time, the water was clear and the channel substrate was gravel/pebble and sand. The riparian vegetation included herbs and alders and acacias.

Figure 7 - Sampling site Foz do Ceira (M2001).

2.1.6 Lorvão (M108)

This site is located in Lorvão stream at 141 m of altitude and belongs to the Mondego basin. The M108 site (Figure 8) is 1 m wide and is 25 cm deep. At the sampling time, the water was clear and the channel substrate was coble and gravel/pebble. Surrounding the site, on the right bank, the land was used for agriculture (orchards). On the left bank there was a wall, a road and houses. The riparian vegetation was composed by grasses and brambles. The major anthropogenic activity affecting this stream is the housing.

Figure 8 - Sampling site Lorvão (M108).

2.1.7 Casal do Ermio (M109)

This site is located in Ceira river at 66 m of altitude and belongs to the Mondego basin. The M109 site (Figure 9) is 20 m wide and a depth of 36 cm. At the sampling time, the water was clear and the channel substrate was coble. Around the site semi-natural mixed woodland was present. The riparian vegetation was composed by saplings and acacias.

Figure 9 - Sampling site Casal do Ermido (M109).

2.1.8 Nossa Senhora da Piedade de Tábua (M112)

This site is located in Nossa Senhora da Piedade de Tábua stream at 288 m of altitude and belongs to the Mondego basin. The M112 site (Figure 10) is 2.5 m wide and has a depth of 17 cm. At the sampling time, the water was clear and the channel substrate was cobles and boulders. Around the site semi-natural mixed woodland was present. The riparian vegetation was mainly composed by oaks. Futher away the dominant trees were eucalyptus. The forest had suffered a wildfire recently by the time of sampling.

Figure 10 - Sampling site Nossa Senhora da Piedade de Tábua (M112).

2.1.9 Miranda do Corvo 3 (M111)

This site is located in Corvo river at 107 m of altitude and belongs to the Mondego basin. The M111 site (Figure 11) is 3.5 m wide and has 15 cm depth. At the sampling time, the water was moderately clear and the channel substrate was gravel/pebble. The banks were reinforced with wooden fences. The riparian vegetation was composed by occasional trees and grasses. The sample was collected inside a park that is downstream of two bridges (road and train).

Figure 11 - Sampling site Miranda do Corvo 3 (M111).

2.1.10 Miranda do Corvo (M110)

This site is located in Corvo river at 90 m of altitude and belongs to the Mondego basin. The M110 site (Figure 12) is 5 m wide and has 36 cm depth. At the sampling time, the water was moderately clear and the channel substrate was gravel/pebble and sand. The riparian vegetation was composed of some alders and grasses but was discontinued, especially on the right bank. The sample was collected inside a park that is downstream of two bridges (road and train) and downstream of the confluence of Miranda do Corvo 3 sampling point.

Figure 12 - Sampling site Miranda do Corvo (M110).

2.1.11 Covão dos Mendes/Crespos (M43)

This site is located in Crespos site at 75 m of altitude and belongs to the Mondego basin. The M43 site (Figure 13) is 25 cm deep. At the sampling time, the water was turbid and the dominant channel substrate was sand with gravel/pebbles. The riparian vegetation was composed by grasses, brambles and some alders were present. Around the site acacias were present.

Figure 13 - Sampling site Covão dos Mendes/Crespos (M43).

2.1.12 Cunha Baixa (M123)

This site is located in Castelo river stream at 411 m of altitude and belongs to the Mondego basin. The M123 site (Figure 14) is 22 cm deep. At the sampling time, the water was clear and the channel substrate was gravel/pebble. The banks were resectioned and reinforced with brick walls. The riparian vegetation was composed by grasses, brambles and several alders were present. Around the site semi-natural mixed woodland was present and land was used also for agriculture (orchards) and pasture. The sample was collected downstream of extraction of uranium mines.

Figure 14 - Sampling site Cunha Baixa (M123).

2.1.13 Casal da Misarela (M49)

This site is located in Mondego river at 3 m of altitude and belongs to the Mondego basin. The M49 site (Figure 15) has a depth of 45 cm. The site is located downstream from a riverine beach. At the sampling time, the water was clear and the channel substrate was gravel/pebble. Part of the channel was resectioned and reinforced with rip-rap. There were several sandy side bars with vegetation on both sides of the channel. Around the site eucalyptus plantations and acacias were present. The sample was collected downstream of a wooden bridge.

Figure 15 - Sampling Casal da Misarela (M49).

2.1.14 Urgeiriça (M122)

This site is located in Pantanha stream at 329 m of altitude and belongs to the Mondego basin. The M122 site (Figure 16) is 1.50 m wide and 20 cm deep. At the sampling time, the water was turbid and the channel substrate was silt and clay. The banks were reinforced with rip-rap. The riparian vegetation was grasses and brambles. Around the site acacias were present. The major impact affecting the stream is the extraction in a uranium mines.

Figure 16 - Sampling site Urgeiriça (M122).

2.1.15 Mogofores (V78)

This site is located in Cértima river at 27 m of altitude and belongs to the Vouga basin. The V78 site (Figure 17) is 40 cm depth. At the sampling time, the water was turbid and the channel substrate was gravel/pebble. The riparian vegetation was composed by herbs and occasional exotic trees nearby a park. Around the site the land was used for agriculture (orchards) and urban development.

Figure 17 - Sampling site Mogofores (V78).

2.1.16 Vila Verde (V94)

This site is located in Levira river at 25 m of altitude and belongs to the Vouga basin. The V94 site (Figure 18) is 4.70 m wide and 36 cm deep. At the sampling time, the water was turbid and the channel substrate was gravel/pebble and sand. Part of the channel was obviously realigned. The riparian vegetation was composed by herbs and occasional trees. Around the site the land was used for pasture. This stream is affected by industry, mostly ceramics.

Figure 18 – Sampling site Vila Verde (V94).

2.1.17 São João da Madeira (V125)

This site is located in Ul river at 188 m of altitude and belongs to the Vouga basin. The V125 site (Figure 19) is 2.70 m wide and 15 cm depth. At the sampling time, the water was turbid and the dominant channel substrate was clay. The riparian vegetation, when present, was composed by trees and herbs. The sample was collected inside a park surrounded by alders and placed downstream of an industrial area and a wastewater treatment plant.

Figure 19 - Sampling site São João da Madeira (V125).

2.1.18 Travanca (V124)

This site is located in Travanca river at 100 m of altitude and belongs to the Vouga basin. The V124 site (Figure 20) is 23 cm deep. At the sampling time, the water was moderately clear and the channel substrate was clay. The riparian vegetation was composed by herbs. The sample was collected downstream of an industrial area.

Figure 20 - Sampling site Tranvanca (V124).

2.1.19 Alfusqueiro (V36)

This site is located in Alfusqueiro river at 46 m of altitude and belongs to the Vouga basin. The V124 site (Figure 21) is 16.20 m wide and 26 cm deep. At the sampling time, the water was clear and the channel substrate was composed by cobbles, boulders and bedrock. The riparian vegetation was dominated by acacias and herbs. Around the site eucalyptus plantations were present.

Figure 21 - Sampling site Alfusqueiro (V36).

2.1.20 Carvalhal (V118)

This site is located in Caima river at 46 m of altitude and belongs to the Vouga basin. The V118 site (Figure 22) is 11 m wide and has a mean depth of 35 cm. At the sampling time, the water was turbid and the channel substrate was composed of boulders and cobbles. The riparian vegetation was composed by trees, herbs and brambles in both margins. Around the site some acacias were present. The major impact affecting this stream is a paper pulp industry.

Figure 22 - Sampling site Carvalhal (V118).

2.1.21 Estarreja (V119)

This site is located in Antuã river at 23 m of altitude and belongs to the Vouga basin. The V119 site (Figure 23) is 7.10 m wide and 46 cm deep. At the sampling time, the water was very turbid and the channel substrate was sand. Part of the channel was obviously realigned and reinforced with rip-rap. The riparian vegetation, when present, was composed by brambles and herbs and occasional trees on right bank, in a recreational park area. The major potential impacts affecting this stream are due to chemical industries in the area.

Figure 23 - Sampling site Estarreja (V119).

2.1.22 Colmeias (L42)

This site is located in Agudim stream at 139 m of altitude and belongs to the Lis basin. The V119 site (Figure 24) is 3.90 m wide and 26 cm deep. At the sampling time, the water was very clear and the channel substrate was gravel/pebble and sand. The stream was reinforced with concrete walls. The riparian vegetation, when present, was composed by brambles and herbs. The land around the site was used for agriculture (orchards) and pasture. Evidence of recent weed cutting was noticed.

Figure 24 - Sampling site Colmeias (L42).

2.1.23 Chãs (L120)

This site is located in Milagres stream at 48 m of altitude and belongs to the Vouga basin. The L120 site (Figure 25) is 2.50 m wide and 25 cm deep. At the sampling time, the water was very turbid and the channel substrate was sand. The banks were resectioned. The riparian vegetation was composed by brambles and herbs. The major impact affecting this stream is a swine industry and discharges before sampling was evident.

Figure 25 - Sampling site Chãs (L120).

Chapter 3 – METHODS

3.1 Data-base

Biological data of macroinvertebrate and diatom communities and abiotic data from the 143 reference sites (undisturbed or minimally disturbed) used for model building (calibration and validation sites) belong to a national database held by the Portuguese Water Institute (INAG, I.P.). Data was gathered during a national campaign held during the springs of 2004 and 2005.

Reference sites were selected based on previous knowledge, expert judgment and collected information and finally selected following strict criteria. All the sites shared good chemical quality (low concentrations of nitrate, nitrite, phosphate, ammonia, BOD_5, COD and pH in accordance with lithology of sites), minimal changes in the riparian zone, no signs of recent changes in the channel morphology and all expected habitats present, low levels of urbanization and industrial activities in the catchment area, minimum impacts on the natural hydrological regime and low levels of fine sediment load (Feio et al. 2009a).

The biological and abiotic data of test sites (potentially disturbed; run through the model) was collected in the spring of 2011.

3.2 Biological data

Diatoms were sampled following the recommendations of INAG I.P. (2008a) and Prygiel and Coste (2000). We also followed the recommendation of Kelly et al. (1998) in which the preferred substrate for monitoring streams and rivers is stones and rocks. However, when this substrate was no available, we proceeded to the sampling of epiphytic community and ultimately to the epipsammic community, as a previous study showed that differences in substrates do not interfere with indices classifications (Mendes et al. 2012). The epilithic community (attached to rocks) was sampled by scraping several submerged stones using a toothbrush (Figure 26) in order to complete an area of about 100 cm^2 at well-defined conditions of light, depth and current velocity. The

Figure 26 - Sampling of epilithic diatom community.

cobbles/pebbles were located between 10 and 30 cm depth in unshaded areas whenever possible and with a current velocity varying from 10 to 50 cm s^{-1}. The epiphytic community (Figure 27) was obtained by squeezing the submerged vegetation. The epipsammic community (Figure 28) was collected from the surface layer of riverbed sediment using a pipette with a cut tip. The collected material was preserved with formaldehyde (8 to 10% final concentration) and correctly labeled. In the laboratory, the samples were oxidized to remove the organic material using nitric acid (HNO_3) method. A small homogenized amount of sample (about 2 ml) was put in a centrifuge tube and centrifuged during 5 minutes at 1500 rpm. After this period, the supernatant was removed; distilled water was added and centrifuged again. The number of centrifugations depends on the amount of preservative added. For the oxidation, we added about 4 ml of nitric acid 65% to a 2 ml of sample. An enough amount of potassium dichromate ($K_2Cr_2O_7$)

Figure 27 - Sampling of epiphytic diatom community.

Figure 29 - Sampling of epipsammic diatoms.

was also added until the solutions acquired an orange hue. The samples were left overnight at room temperature to oxidize. After this period, the samples were centrifuged during 5 minutes at 1500 rpm in order to eliminate any nitric acid present. Later, permanent slides were mounted using Naphrax® (refractive index > 1.6). In each sample, about 400 valves were counted and identified to species or infra-specific level mainly using taxonomic guides (Aboal et al. 2003, Krammer 2000, 2002, 2003, Krammer and Lange-Bertalot 1986, 1988, 1991a, 1991b, Levkov 2009, Werum and Lange-Bertalot 2004). For that, we used a light microscope equipped with a 100x immersion objective of NA 1.32.

Macroinvertebrates were sampled following the recommendations of INAG I. P. (2008b). The macroinvertebrates were collected with a hand-net (Figure 29) of 500 µm mesh size. Each sample was composed by 6 sub-samples that were proportionally distributed by the most representative existing habitats (stones, sand and silt, blocks,

Figure 28 - Sampling of macroinvertebrates.

submerged plants, algae) and each defined by an area of 1 m x 0.25 (hand-net width) towards upstream. The organisms with greater capacity of setting were removed manually, especially in areas with current where there are suitable substrates such as blocks and stones. After collected, the composite samples were preserved with formaldehyde. In the laboratory, the invertebrates were separated from the sediment and organic debris and then preserved in ethanol 70%. Identification of macroinvertebrates was made to the lowest taxonomic level possible, mostly to genus level, using a stereo microscope.

3.3 Abiotic data

A total of 10 variables were used to characterize the reference sites and as potential discriminant variables (Discriminant Function Analysis step; see Data analyses) in model building (Table 1).

Table 1 - Potential discriminant variables used to characterize the reference sites and build predictive models.

Potential predictive variables
Latitude (rectangular)
Longitude (rectangular)
Altitude (m)
Distance to source (km)
Typical flow regime (temporary to permanent)
Mean annual temperature (^{o}C)
Mean annual precipitation (mm)
Lithology (category: 1 to 3)
Alkalinity (mg/l CO_3^{2-})
Hardness (mg/l $CaCO_3$)

For all test sites, the variables described in Table 2 were also calculated or measured at each test site. Most of these variables (with the exception of those in italic) describe the

Table 2 - Environmental variables measured or calculated for each test site. In italic are those not used as pressure variables.

Environmental variables
Water width (m)
Water depth (cm)
Water temperature (^{o}C)
BOD$_5$, Biological Oxygen Demand (mgl^{-1})
COD, Chemical Oxygen Demand (mgl^{-1})
Phosphates (mgl^{-1} PO_4^{3-})
Flow velocity (ms^{-1})
N amoniacal (N-NH$_4$; mg Nl^{-1})
Nitrates (mgl^{-1})
Nitrites (mgl^{-1})
Total P (mg Pl-^{1})
Chlorides (mg Cll^{-1})
Sulfates (mg SO$_4$l^{-1})
Silica (mg SiO$_2$l^{-1})
Conductivity (µScm^{-1})
Total suspended solids (mgl^{-1})
pH
Dissolved oxygen (mgl^{-1})
Oxygen saturation (%)
Connectivity (category: 1 to 5)
Hydrological Regime (category: 1 to 5)
Integrity of riparian zone (category: 1 to 5)
Sediment discharge (category: 1 to 5)
Morphological condition (category: 1 to 5)
Acidification and toxicity (category: 1 to 5)
Organic contamination and nutrients enrichment (category: 1 to 5)
HMS (Habitat Modification Score, calculated after field observations according to the River Habitat Survey, E. A. 2003)
HQA (Habitat Quality Assessment, calculated after field observations according to the River Habitat Survey, E. A. 2003)

environmental pressures affecting test sites and are from here on called pressure variables. The River Habitat Survey (RHS) was used for river hydromorphological assessment and physical characterization (E. A. 2003). The RHS data were collected by means of 10 equidistant “spot check” transects of 1 m wide and a sweep-up summary of 500 m. The RHS has two scoring systems, the Habitat Quality Assessment (HQA) and the Habitat Modification Score (HMS). The HQA results from the sum of 10 sub-indices related to the physical habitat diversity, vegetation, river channel and land-use. The HMS quantifies the extent of human intervention through the presence of weirs, bank and channel modifications, etc. Additionally, at each test site 2 L of water for subsequent laboratory analyses of physical-chemical parameters were collected.

For water sampling, polyethylene bottles with screw cap were used. Before sampling, the bottles were washed *in situ* with the stream water to avoid possible contaminations. The bottles were kept in a freezer and sent later to a laboratory for chemical analysis.

The categorized environmental variables (Table 1 and Table 2) were based on the European Project FAME (Schumtz 2004). The FAME variables range from 1 (no obvious deviation from the reference condition, undisturbed/minimally disturbed) to 5 (highly impacted). The hydrological regime refers to the flow pattern and includes all the hydrologic changes. The connectivity is related to the extent of the impoundment and its impact on the migration of the existing organisms. The integrity of the riparian zone refers to cut of vegetation and/or introduction of exotic species. The sediment discharge quantifies the load of sediment in the water column and deposited on the riverbed. The morphological condition relies on the modifications of the riverbed and bank-face of the river. The acidification and toxicity is linked to the symptoms such as clearly unhealthy or even dead organisms or pale. Finally, the organic contamination and nutrients enrichment measured the input of BOD_5 (Biological Oxygen Demand), COD (Chemical Oxygen Demand), NO_3 (nitrates), P_2O_5 (phosphorous) and NH_4 (ammonium) in the water column.

The lithology was based on Atlas Digital do Ambiente (DGA) where the categories correspond to: 1 = sedimentary, 2 = sedimentary + metamorphic and 3 = plutonic rocks.

3.4 Data analysis

3.4.1 Models building

Six predictive models were built in this study. Three of them were built based on RIVPACS approach using the latest improvements on statistical methods, developed by Van Sickle et al. (2005, 2006, 2008). The other three were built based on BEAST approach. The building of the predictive models was done through the web platform AQUAWEB (http://aquaweb.uc.pt/).

Biological and environmental data from 143 reference sites was used. Only the environmental variables not affected by anthropogenic impacts were used in models construction, as recommended by Simpson and Norris (2000). The biological data, initially abundance, was transformed into presence/absence data for the building of the RIVPACS type of models and into relative abundance for the building of the BEAST type of models. In this last case, the data were additionally fourth root transformed in order to reduce the weight of very abundant taxa. Each environmental variable was transformed towards normality.

For the models based on RIVPACS approach, the reference dataset was divided into calibration and validation data. The validation dataset included 10% of the total reference sites spread across the study area and were left out of the model construction in order to validate the model's responses. The calibration dataset was grouped according to the similarity of the biota through a clustering technique (UPGMA - Unweight Pair Group Method with Arithmetic mean) based on Bray-Curtis similarity and with the help of a non-metric multidimensional scaling (MDS). Each group had at least 5 reference sites. Groups with less than 5 sites can be associated with the loss of representative taxa of reference conditions or to underrepresented stream types (Wright et al. 1993). Then, the environmental variables that best discriminated the biological groups, the environmental predictors, were selected using a Discriminant Function Analysis (DFA). Here we applied to our reference data an alternative approach implemented by Van Sickle et al. (2006). This approach explores all possible candidate discriminant function models and ranks all possible models based on the best results of an a F-test and Wilks's lambda test. The F-statistics measures the variation between and within the reference groups. The variation between groups compared with within groups variation increases with the increment in the F-statistics value. The Wilk's values measured the group's separation, small values denotes strong group separation. This approach enables retention of a number of best models of each order (lowest Wilk's values). To determine the discriminatory power of Wilk's lambda we used re-substitution and cross-validation analysis. These analyses provide a percentage value of reference sites correctly alocated to their original groups. A SIMPER analysis (Bray-Curtis similarity measure) was used in addition to characterize groups based on their most representative taxa. This analysis determines which taxa contribute to the within group's similarity. The overall RIVPACS models performance can be achieved by comparison between its predictions of expected and observed taxa. For that purpose, the mean value (MN) and the root mean squared error (RMSE) of O/E were calculated. The MNOE measures the bias and if its value is equal to one, this means that the predictive model is unbiased. Their correspondent standard deviation (SD) indicates the model's precision. A model is considered to be as much precise as lower its SDOE value is. The RMSE combines the bias and variability of prediction errors into a single measure of model performance. Low RMSE values would denote an improvement in the overall model performance. The overall performance of any predictive model at calibration sites can be evaluated by comparison with an upper boundary for a model to be effective. Such a boundary is the RMSE (O/E) value of a null model. This null model is defined as a limit that would be achieved whenever a predictive model fails to explain any of the natural-gradient variation in assemblages. This limit is only achieved if a model fails to account for variability in taxon richness under reference conditions. Thereby, the Replicate Sampling Standard Deviation (RSSD) is the theoretical definition of the lower bound on the SD one should expect for any model and the maximum precision attainable. This score may be compared to the attained precision of the model through its standard deviation, and the distance between these scores indicates how much improvement may be achieved. Together, the null model and the resubstitution-sampling SDs estimate the minimum and maximum precision, respectively, reachable by any predictive model of a given set of reference values (Van Sickle et al. 2005).

Regarding the models based on BEAST approach, the first steps are similar to RIVPACS approach. The reference dataset were also divided into calibration and validation data (10% of the total reference sites). The reference sites were classified into groups based on the biological community structure (UPGMA; Bray-Curtis similarity) and with the help of a MDS analysis. Then, a DFA was applied to select the environmental variables that best discriminated the biological group and to calculate the percentage of reference sites correctly attributed to their original reference group. The F-statistics was calculated to test for significant difference between and within the reference groups as well as the Wilk's values to measure the groups's separation.

3.4.2 Classification system

The classification schemes adopted here consist on a modification to the original methods (RIVPACS and BEAST) in order to make them more compliant with the five-class system defined in the WFD. Therefore, in both cases, the five quality status classes were defined as: high – 1, good – 2, moderate – 3, poor – 4 and bad – 5. The boundary between high and good classes for the RIVPACS models was defined at the 25th percentile of the reference sites O/E_{50} ratios. For the BEAST models the boundaries were defined through 4 Gaussian probability ellipses (75, 90, 99 and 99.9%).

3.4.3 Models validation and testing

Fourteen reference sites not included in the models building were used for the validation of the models. Those sites were used as test sites in order to verify if each model assessed them as reference sites.

For the RIVPACS predictive models we followed the method proposed by Linke et al. (2005) in which a predictive model is considered accurate if the regression line of the reference Observed versus Expected values runs through or close to the origin (between -1,5 and 1,5) and if the slope is close to unity (acceptable range 0,85 – 1,15).

For the BEAST predictive models we tested its performance based on the percentage of reference sites correctly attributed to their original group (Cross-validation and Resubstitution analysis).

The three models were tested by running 23 test sites covering different anthropogenic impact levels. In RIVPACS, the calculations of O/E ratios started with a discriminant analysis and then followed the same steps described for the reference sites. Unlike RIVPACS, which uses probabilities of a site for belonging to each group, the BEAST only assigns a site to its most probable group. Each test site was compared to the reference sites belonging to the most probable group through an MDS ordination to which the probability ellipses were added.

3.4.4 Response to pressures

In order to evaluate the responses to the environmental degradation of the six models a Spearman rank correlation (SYSTAT 10) was applied. The analysis was done between the class assessments attributed by each model to both reference and test sites and pressure variables (see Table 2). To obtain a view of the global degradation a PCA analysis (Primer 6 and Permanova β17) was performed using both reference and test sites. Pressure variables were previously transformed towards normality.

To observe if there was a continuous decrease of the water quality with the increase in quality classes produced by the different models, Box Plots (Minitab 16) were also done between the classes and the pressure variables.

In addition, in order to compare the responses to degradation with those previously obtained, Spearman rank correlations (SYSTAT 10) were also applied between the combined *a posteriori* assessment that would be done according to the “one-out all out” approach (for both type of models) and the pressure variables.

Chapter 4 – RESULTS

4.1 RIVPACS Diatom model - RIV DIAT model

From the Cluster analysis of the calibration dataset based on the presence/absence diatom biological matrix, six groups were elected (Figure 30; Table 12) ranging from 9 to 38 reference sites each.

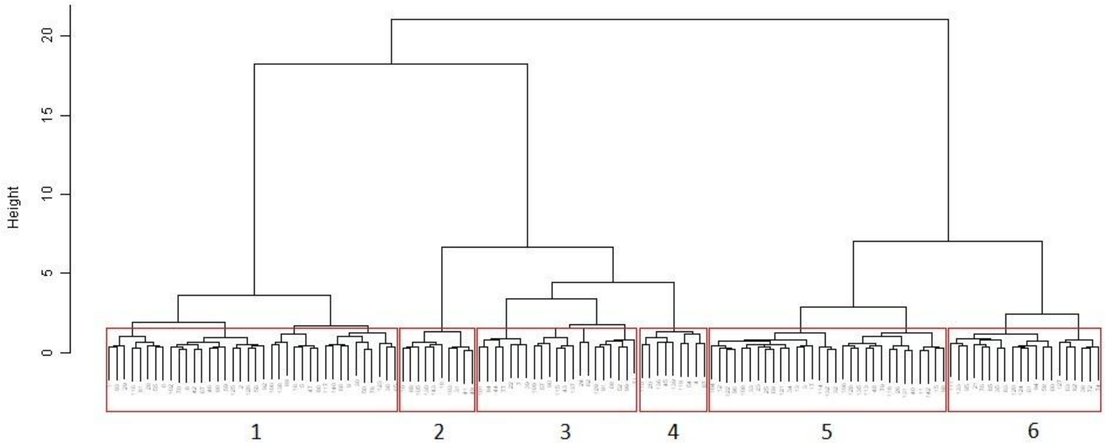

Figure 30 - Cluster of calibration reference sites used for RIV DIAT model construction. Each number corresponds to a group.

The ordination of these six groups in the three axes of the MDS analysis (stress: 0.17) shows little overlap. Groups 3 and 5 are the most scattered, with some samples appearing more dissimilar from the remaining of the group, in the MDS ordination space (Figure 31).

The most contributive taxa to the similarity within each reference group (SIMPER analyses; Bray-Curtis coefficient; presence/absence transformation) are listed in Table 3. The only taxon common to all groups of the RIV DIAT predictive model is *Achnanthidium minutissimum* (ADMI). Other taxa also relatively common include: *Encyonema minutum* (ENMI), *Fragilaria capucina* var. *vaucheriae* (FCVA) and *Gomphonema parvulum* (GPAR). Some taxa were found to be exclusive to each group such as *Navicula gregaria* (NGRE) in group 1, *Nitzschia palea* (NPAL) in group 2, *Gomphonema rhombicum* (GRHB) in group 3, *Navicula cryptotenella* (NCTE) in group 4, *Gomphonema pumilum* (GPUM) in group 5 or *Eolimna minima* (EOMI) in group 6.

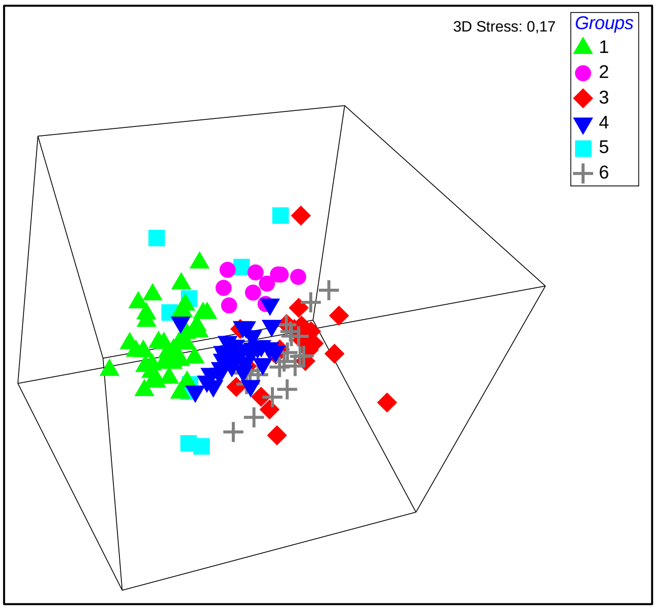

Figure 31 - Three-dimensional MDS ordination of six groups based on diatom references from the calibration dataset of RIV DIAT model.

The environmental variables (Table 2) and the six biological groups were used to run the discriminant function analysis, following the "best-subset" approach proposed by Van Sickle et al. (2006). The best five DF (Discriminante Function) models from each order, up to the 10th order (using 10 environmental predictors), based on their Wilk's lambda and F-statistic values, were retained, resulting in forty-six best models from 1023 possibilities. A final model was chosen, from the combination of the best results of F and Wilks parameters, Standard Deviation, RMSE values and percentage of correct classifications.

The selected final model, classified correctly the calibration sites in 75.2% and 66.7% to their respective classification groups, by Re-substitution and Cross-validation analysis, respectively (Table 12). This model uses nine environmental variables that best discriminate the reference groups: alkalinity, altitude, distance to source, hardness, typical flow regime, latitude, lithology, mean annual precipitation and mean annual temperature.

Table 3 - Most representative taxa of the 6 reference groups of RIV DIAT model, obtained by SIMPER analysis. The diatoms represented were found in 50% or more of the sites.

Groups	Most representative diatoms
1	*Achnanthidium minutissimum*, *Cocconeis placentula* var. *lineata*, *Planothidium frequentissimum*, *Cocconeis euglypta*, *Reimeria sinuata*, *Gomphonema parvulum* var. *parvulum* f. *parvulum*, *Fragilaria capucina* var. *vaucheriae*, *Navicula gregaria*
2	*Achnanthidium minutissimum*, *Encyonema minutum*, *Fragilaria capucina* var. *vaucheriae*, *Gomphonema parvulum* var. *parvulum* f. *parvulum*, *Nitzschia palea*, *Surirella angusta*, *Diatoma mesodon*
3	*Achnanthidium minutissimum*, *Gomphonema rhombicum*, *Achnanthidium minutissimum*, *Encyonema minutum*, *Fragilaria capucina* var. *vaucheriae*, *Gomphonema parvulum* var. *parvulum* f. *parvulum*, *Karayevia oblongella*, *Diatoma mesodon*
4	*Achnanthidium minutissimum*, *Reimeria sinuata*, *Achnanthidium biasolettianum*, *Encyonema minutum*, *Cocconeis placentula* var. *placentula*, *Nitzschia paleacea*, *Fragilaria capucina* var. *vaucheriae*, *Achnanthidium helveticum*, *Cocconeis placentula* var. *lineata*, *Navicula cryptotenella*, *Planothidium frequentissimum*, *Achnanthidium subatomoides*
5	*Achnanthidium minutissimum*, *Gomphonema pumilum*
6	*Achnanthidium minutissimum, Eolimna minima*, *Encyonema minutum*, *Cocconeis placentula* var. *placentula*, *Gomphonema parvulum* var. *parvulum* f. *parvulum*, *Achnanthidium biasolettianum*, *Achnanthidium helveticum*, *Fragilaria capucina* var. *capucina*

The mean values and the correspondent standard deviation of the predictor environmental variables for each reference group are shown in Table 4. Group 1 includes mainly temporary streams in lowland areas where the climate is the driest and hottest with the highest water hardness. The lithology is mostly sedimentary and metamorphic and the alkalinity is high. Group 2 is composed mainly by the smallest permanent rivers located at medium altitude and with low air temperature in the North of Portugal. The mean annual precipitation is the highest and the alkalinity is the lowest. The stream channels are mainly composed of plutonic rocks (granite) and the water has the lowest hardness. Group 3 is characterized by small permanent rivers with low alkalinity and hardness. The sites are located at medium altitude where the temperature is the lowest among all groups and the mean annual precipitation is high. The lithology is mainly sedimentary and metamorphic. Group 4 includes mostly permanent streams located at the highest altitude. They have medium alkalinity, low hardness and channels have sedimentary and metamorphic lithology. The rivers belonging to group 5 are characterized by high air temperature, as well as high water hardness and the highest alkalinity. The majority is permanent streams with low altitude and located in the south of Portugal. The lithology is similar to those belonging to group 4 and the mean annual precipitation shows intermediate values. In group 6, rivers are the largest, permanent, and have medium precipitation and low alkalinity. These sites are located at medium altitude and have the lowest hardness. The lithology is mostly sedimentary and metamorphic.

Table 4 - Environmental characterization of the 6 reference groups of streams used in the RIV DIAT predictive model.

Discriminant variables	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Latitude (rectangular)	38.7 (±1.1)	41.6 (±0.5)	40.3 (±0.7)	41.1 (±0.8)	39.5 (±1.8)	41.4 (±0.3)
Altitude (m)	173.9 (±128.6)	256.9 (±209.1)	311.8 (±283.9)	425.5 (±211.6)	199.4 (±264.6)	408.9 (±207.6)
Distance to source (km)	26.3 (±22.1)	13.8 (±10.6)	16.9 (±22.8)	32.4 (±39.3)	30.6 (±27.5)	37.4 (±32.9)
Typical flow regime (categorical data)	0.3 (±0.5)	0.9 (±0.3)	1.0 (±0.0)	0.9 (±0.2)	0.9 (±0.3)	1.0 (±0.0)
Mean annual temperature (^{o}C)	15.0 (±1.2)	12.3 (±1.3)	12.0 (±2.1)	12.7 (±1.1)	14.3 (±2.0)	12.9 (±1.3)
Mean annual precipitation (mm)	772.3 (±212.8)	1968.1 (±475.5)	1356.8 (±358.3)	788.7 (±354.4)	1079.7 (±607.7)	1171.1 (±477.7)
Lithology (categorical data)	2.0 (±0.5)	2.7 (±0.5)	2.0 (±0.5)	2.3 (±0.5)	2.0 (±0.7)	2.4 (±0.6)
Alkalinity (mgl^{-1})	75.5 (±73.3)	7.4 (±3.9)	15.0 (±21.2)	41.7 (±26.8)	83.3 (±82.7)	24.1 (±18.1)
Hardness (mgl^{-1})	99.5 (±76.0)	14.2 (±13.3)	18.6 (±36.4)	22.5 (±19.0)	92.9 (±78.6)	10.9 (±15.2)

Figure 32 shows that the RMSE (O/E) values calculated for the calibration dataset were located within the upper (SDcal) and the lower (SDRS cal) boundaries for which the model was effective (null model). The RMSE (O/E) calibration dataset was also located between the above mentioned boundaries (SDOEcal – 0.242).

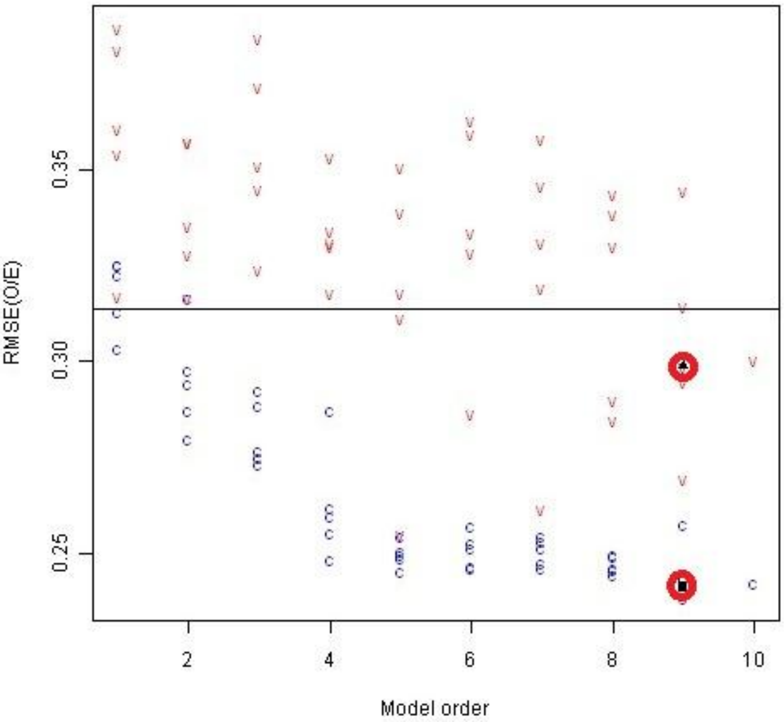

Figure 32 - RMSE (O/E) values for both calibration (c) and validation (v) datasets are shown as well as the corresponding maximum of RMSE for calibration dataset (black line) and the selected model (red circles).

The slope (0,993) of the O versus E regression of the selected model, the intercept (0,097) and the R^2 (0,842) were all within the range of an accurate model and very close to an ideal model (Linke et al. 2005).

The observed (O) taxa of the references sites are similar to the list of expected taxa (E) resulting in a histogram where the reference sites are mostly close to the unit as shown in Figure 33.

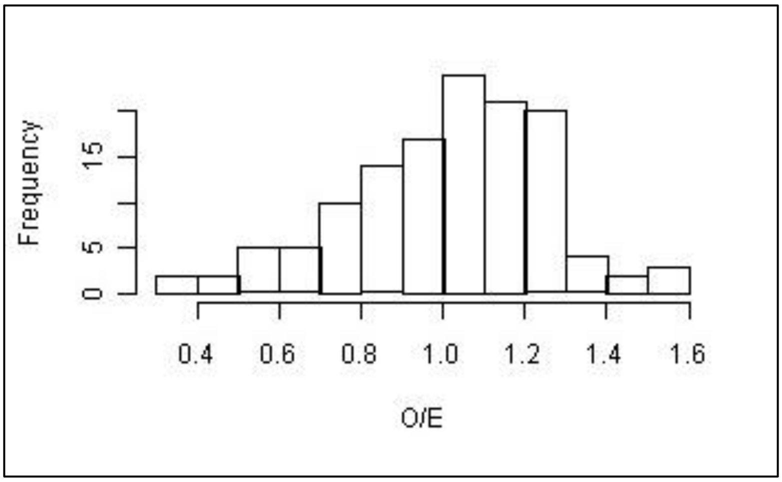

Figure 33 - Distribition of frequencies of the O/E_{50} of all reference sites in the RIV DIAT model.

However, the average Observed/Expected ratio for the validation dataset was 0.81 and about 20% of the sites didn’t achieved at least good quality status (Table 5). The meaning of each site code (numbers ranged from 1 to 143) is presented in Appendix A.

The predictive model based on diatom communities was built using data from 143 reference sites (Appendix A) and a total of 297 different taxa were found in these samples (Appendix B).

Table 5 - RIV DIAT model validation with 14 reference sites not included in the model construction.

Sites	Class	OE50
14	High	0.91
19	High	1.17
27	Good	0.72
37	Good	0.84
51	High	1.22
53	Good	0.83
64	Moderate	0.49
73	Moderate	0.47
75	Moderate	0.44
97	High	1.03
107	Good	0.70
112	Good	0.74
134	High	0.87
141	High	0.96
Average		0.81

4.2 RIVPACS Macroinvertebrate model - RIV INV model

The Cluster analysis of the macroinvertebrate biological reference data resulted in six groups (Figure 34) ranging from 8 to 36 sites. The MDS analysis (stress: 0.17) shows that group 4 is the best defined and distinct but is also the smallest one. Group 6 presents some dispersion and overlapping with the remaining in the 3D view (Figure 35).

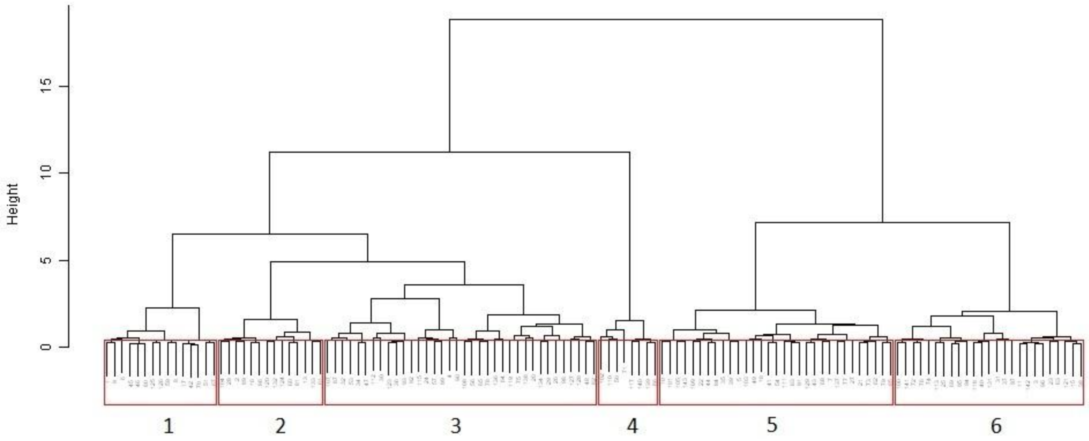

Figure 35 - Cluster of calibration reference sites used for RIV INV model construction. Each number corresponds to a group.

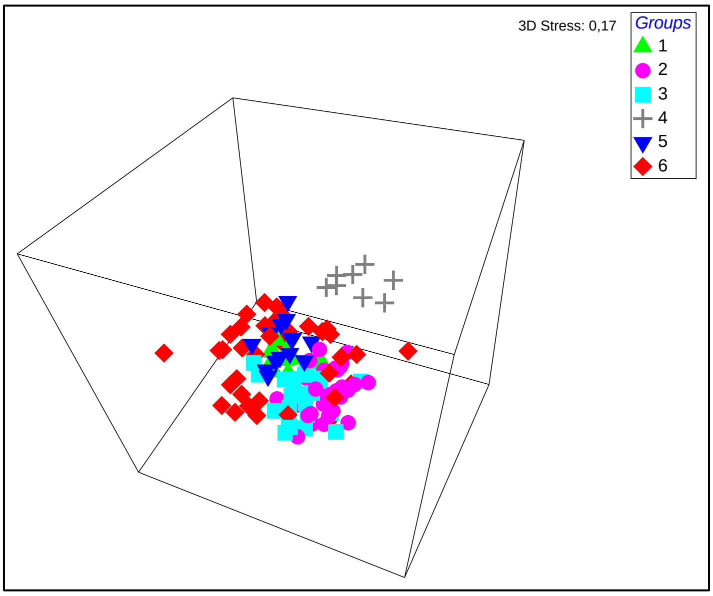

Figure 34 - Three-dimensional MDS ordination of the 6 groups based on macroinvertebrate reference community from the calibration dataset of RIV INV model.

The most representative taxa (SIMPER analysis; Bray-Curtis coefficient; presence/absence transformation) of the six reference groups are presented in Table 6. *Baetis* sp. and *Chironominae* were the taxa shared by all groups. Other taxa were frequently observed such as *Orthocladiinae*, *Oligochaeta*, *Hydropsyche* sp., *Tanypodinae* and *Simulidae*. Only three groups presented exclusive taxa, such as *Limonidae* in group 1, *Tabanidae* in group 4 or *Hydraena* sp. in group 6.

Table 6 - Most representative taxa of the 6 reference groups of RIV INV model, obtained by SIMPER analysis. The invertebrates represented were found in 50% or more of the test sites.

Groups	Most representative invertebrates
1	*Ancylus* sp., *Baetis* sp., Ceratopogoninae, Chironominae, Limonidae, Oligochaeta, Orthocladiinae, *Oulimnius* sp., Simulidae, Tanypodinae, *Caenis* sp., *Hydropsyche* sp.
2	*Baetis* sp., *Ecdyonurus* sp., Chironominae, Orthocladiinae, Tanypodinae, *Hydropsyche* sp., Oligochaeta, Simulidae, *Atherix* sp., *Leuctra* sp., *Limnius* sp.
3	*Baetis* sp., *Caenis* sp., Chironominae, *Hydropsyche* sp., *Leuctra* sp., Orthocladiinae, *Ecdyonurus* sp., Tanypodinae, Oligochaeta, *Bezzia* sp., *Limnius* sp., *Oulimnius* sp., Simulidae, *Polycentropus* sp., *Ancylus* sp., *Atherix* sp., *Siphonoperla* sp., *Habrophlebia* sp.
4	*Ochthebius* sp., *Orthetrum* sp., *Orthotrichia* sp., Tabanidae, *Baetis* sp., Ceratopogoninae, Chironominae
5	Chironominae, *Baetis* sp., Orthocladiinae, Oligochaeta, Tanypodinae, *Caenis* sp., *Hydropsyche* sp., *Leuctra* sp., *Ecdyonurus* sp., *Bezzia* sp., *Oulimnius* sp.
6	*Baetis* sp., Chironominae, *Ecdyonurus* sp., Oligochaeta, Orthocladiinae, Tanypodinae, Simulidae

The “best-subset” approach (Van Sickle et al. 2006) selected 46 models from 1023 possibilities. The selected model classified correctly the calibration dataset to their original classification groups in 64.3% by the Re-substitution analysis and 55.8% by the Cross-validation analysis (Table 12). These groups were best discriminated by nine variables which includes altitude, distance to source, hardness, typical flow regime, latitude, lithology, longitude, mean annual precipitation and mean annual temperature. The mean values and the correspondent standard deviation of the predictor environmental variables for each reference group are shown in Table 7. Group 1 is composed by small streams, mainly temporary, dominated by sedimentary rocks with high hardness and located at low altitude (southern rivers). Precipitation is low and mean annual temperature is high. Group 2 is composed by the smallest permanent rivers with the lowest hardness at moderate altitude. Precipitation is the highest and air temperature the lowest. Group 3 is characterized by permanent streams placed at the highest altitude in the North of Portugal where precipitation is high while air temperature and the water hardness are low.

Table 7 - Environmental characterization of the 6 reference groups of streams used in the RIV INV predictive model.

Discriminant variables	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Latitude (rectangular)	37.7 (±0.5)	41.0 (±0.7)	41.3 (±0.4)	37.6 (±0.3)	40.6 (±1.0)	40.2 (±1.1)
Longitude (rectangular)	-8.2 (±0.6)	-8.0 (±0.4)	-7.3 (±0.5)	-8.2 (±0.4)	-7.6 (±0.7)	-7.7 (±0.6)
Altitude (m)	132.1 (±124.9)	391.3 (±238.2)	401.0 (±201.1)	106.9 (±88.5)	230.9 (±97.3)	301.8 (±279.6)
Distance to source (km)	17.2 (±10.8)	12.8 (±10.0)	23.6 (±27.8)	59.8 (±33.3)	48.5 (±33.6)	29.3 (±34.0)
Typical flow regime (categorical data)	0.2 (±0.4)	1.0 (±0.2)	1.0 (±0.0)	0.4 (±0.5)	0.9 (±0.3)	0.6 (±0.5)
Mean annual temperature (^{o}C)	15.8 (±0.4)	12.0 (±1.4)	12.7 (±0.8)	16.0 (±0.6)	13.9 (±1.2)	13.5 (±2.0)
Mean annual precipitation (mm)	664.5 (±82.4)	1621.0 (±523.0)	1000.4 (±466.9)	613.1 (±82.0)	817.6 (±239.5)	958.7 (±419.8)
Lithology (categorical data)	1.9 (±0.4)	2.3 (±0.6)	2.2 (±0.4)	2.0 (±0.0)	2.1 (±0.5)	2.2 (±0.7)
Hardness (mgl^{-1})	94.4 (±60.5)	11.1 (±9.0)	14.0 (±12.8)	163.0 (±15.8)	36.7 (±34.4)	60.4 (±79.1)

Group 4 includes mainly large temporary watercourses located at the lowest altitude where the climate is the driest (southern rivers). Mean annual temperature and water hardness are the highest. In group 5, the sites are also large and located at medium altitude with medium temperature and precipitation. The majority are permanent rivers with low hardness. The streams belonging to group 6 are mostly permanent and are characterized by medium values of altitude, precipitation, air temperature and water hardness.

Figure 36 shows that the RMSE (O/E) values calculated for the calibration sites were positioned within the boundaries proposed for the concept of null model (Van Sickle et al. 2005) (SDOEnull model=0.249>SDOEcal=0.195>SDRScal=0.118). According to Linke et al. (2005), the model is statistically accurate (slope=1.002; intercept=-0.085 and R^2=0.704).

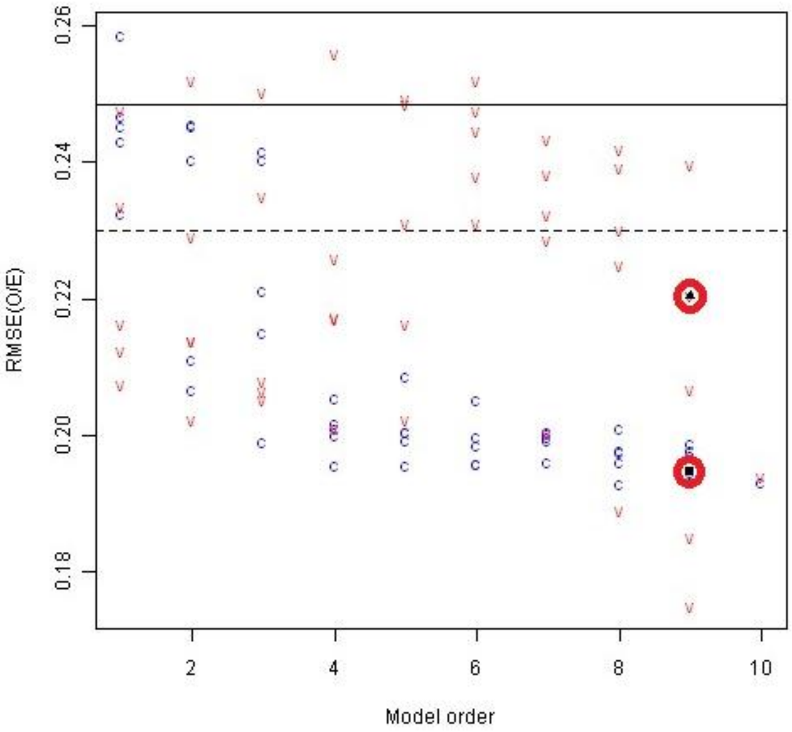

Figure 36 - RMSE (O/E) values for both calibration (c) and validation (v) datasets are shown as well as the corresçponding maximum of RMSE for calibration dataset (black line) and the selected model (red circles).

The ratio O/E taxa of most reference sites are close to 1 as observed in Figure 37 meaning that the observed taxa are equivalent to the expected ones.

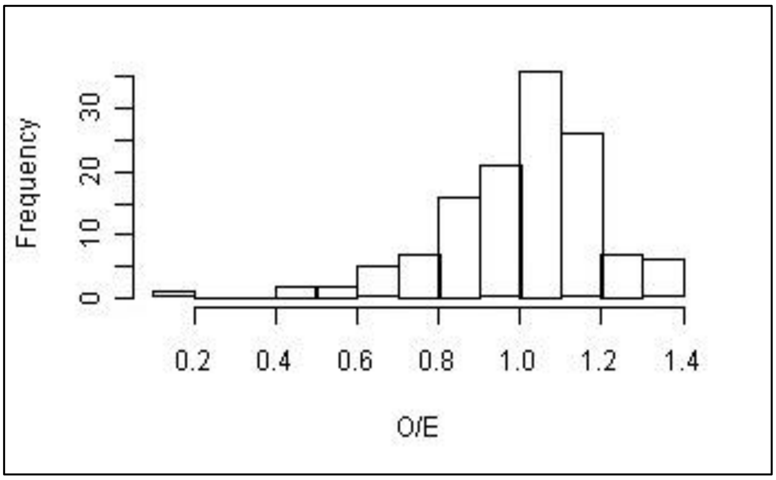

Figure 37 - Distribution of frequencies of the O/E 50 of all reference sites used in the RIV INV model.

Table 8 comprises the average of Observed/Expected ratio for validation reference sites (1.02) and less than 10% didn’t achieve at least the good quality status. The meaning of each site code (numbers ranged from 1 to 143) is presented in Appendix A.

The RIV INV predictive model was also based on the same 143 reference sites as the DIAT model (Appendix A). A total of 301 different taxa were observed in the reference samples (Appendix C).

Table 8 - RIV INV model validation with 14 reference sites not included in the model construction.

Sites	Class	OE50
12	High	1.21
14	High	1.04
19	Poor	0.38
33	Good	0.79
52	High	0.98
55	High	1.08
58	High	1.17
88	High	1.09
106	High	1.00
114	High	1.25
116	High	1.23
122	High	0.90
130	High	1.19
135	High	0.95
Average		1.02

4.3 RIVPACS Macroinvertebrate and Diatom model - RIV INV+DIAT model

The summary of the characteristics of RIV INV+DIAT model are present in Table 12. The cluster analysis lead to the formation of six biological reference groups (Figure 38) ranging from 11 to 28 sites. The MDS analysis (stress: 0.17) shows that group 5 is the most dispersed (Figure 39).

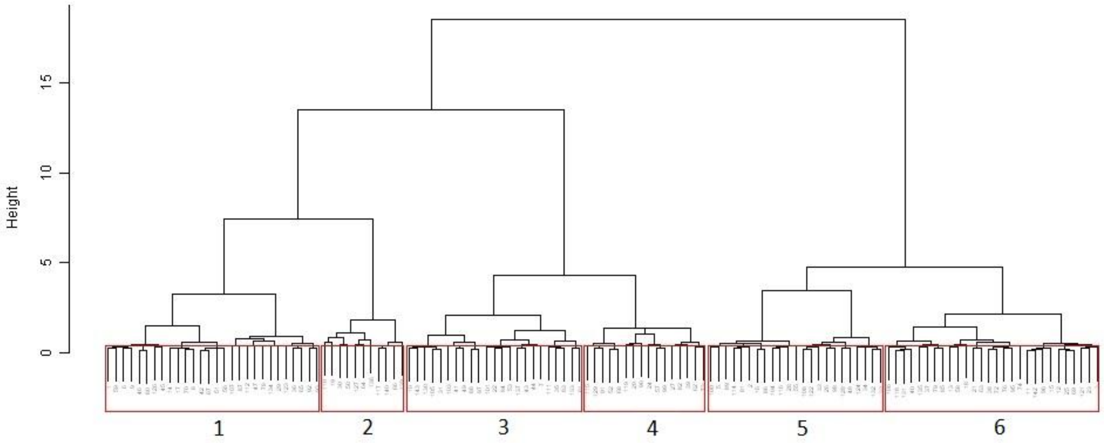

Figure 38 - Cluster of calibration reference sites used for RIV INV+DIAT model construction. Each number corresponds to a group.

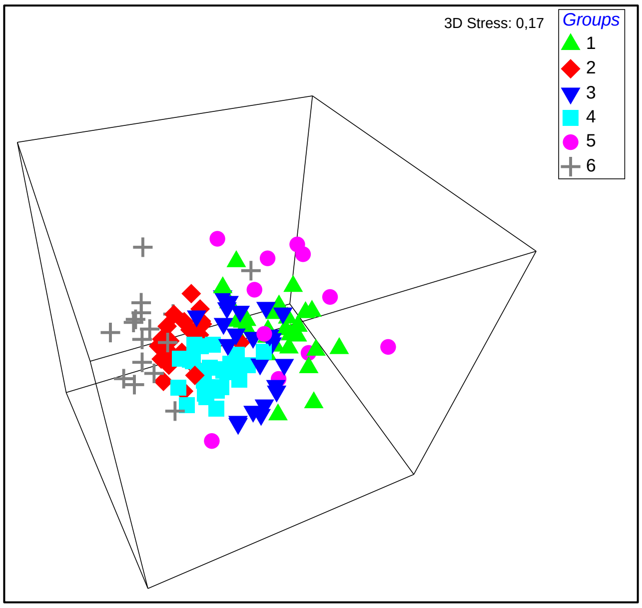

Figure 39 - Three-dimensional MDS ordination of the 6 groups based on the combined biological reference community of macroinvertebrates and diatoms from the calibration dataset of RIV INV+DIAT model.

The most representative taxa (SIMPER analysis; Bray-Curtis coefficient; presence/absence transformation) of the six reference groups are denoted in Table 9. The taxa common to all groups includes two invertebrate taxa, *Baetis* sp. and *Chironominae*, and only one diatom, *Achnanthidium minutissimum* (ADMI). However, other taxa such as Tanypodinae, Orthocladiinae, Oligochaeta, *Hydropsyche* sp., Simulidae, *Gomphonema parvulum* var. *parvulum* f. *parvulum*, *Reimeria sinuata*, *Fragilaria capucina* var. *vaucheriae* and *Encyonema minutum* also contributed to the similarity within the groups. Some taxa were found to be exclusive to each group such as *Reimeria sinuata* (RSIN) in group 1, *Onychogombus* sp. and *Gomphonema rhombicum* (GRHB) in group 2, *Orthetrum* sp. and *Cyclotella meneghiniana* in group 3, *Esolus* sp. in group 5 or *Bezzia* sp. and *Achananthidium helveticum* in group 6.

The combination of the six reference groups and the 10 environmental variables resulted in 46 best models from 1023 possibilities (Van Sickle et al. 2006).

The six reference groups were 61.2% and 55% correctly classified in their respective classification groups by Re-substitution and Cross-validation analysis, respectively (Table 12). The reference groups were best discriminated by the following variables: alkalinity, altitude, distance to source, hardness, typical flow regime, latitude, lithology and mean annual precipitation. Table 10 comprises the average values and the standard deviation of the predictor environmental variables for each reference group. The lithology is the same for all the six groups, sedimentary and metamorphic rocks. Group 1 is composed by streams mainly temporary with high hardness and located at low altitude where the climate is the driest (southern Portugal). Alkalinity is high and hardness is intermediate. Group 2 is characterized by permanent rivers located in the North of Portugal with the lowest alkalinity at medium altitude. Precipitation is high and water hardness is low. Group 3 includes mostly large permanent watercourses placed at medium altitude with medium alkalinity. Both precipitation and hardness are low. Group 4 contains mainly permanent rivers located at the highest altitude. Mean annual precipitation and alkalinity are intermediate while hardness is low. In group 5, the majority are temporary streams at low altitude where the climate is dry and alkalinity and hardness are highest of the dataset (southern Portugal). The watercourses belonging to group 6 are all small and permanent and are characterized by medium altitude and low alkalinity. Water hardness is the lowest while precipitation is highest.

Table 9 - Most representative taxa of the 6 groups of RIV INV+DIAT model, obtained by SIMPER analysis. The invertebrates and diatoms represented were found in 50% or more of the sites.

Groups	Most representative invertebrates and diatoms
1	*Baetis* sp., Chironominae, Orthocladiinae, Tanypodinae, Oligochaeta, Simulidae, *Caenis* sp., *Hydropsyche* sp., *Oulimnius* sp., *Ancylus* sp., Limonidae, *Achnanthidium minutissimum*, *Planothidium frequentissimum*, *Cocconeis euglypta*, *Cocconeis placentula* var. *lineata*, *Reimeria sinuata*
2	*Baetis* sp., *Ecdyonurus* sp., *Hydropsyche* sp., Chironominae, Orthocladiinae, Tanypodinae, Oligochaeta, Simulidae, *Limnius* sp., *Caenis* sp., *Oulimnius* sp., *Leuctra* sp., *Atherix* sp., *Achnanthidium minutissimum*, *Encyonema minutum*, *Gomphonema parvulum* var. *parvulum* f. *parvulum, Gomphonema rhombicum*, *Fragilaria capucina* var. *vaucheriae*
3	*Baetis* sp., Chironominae, Orthocladiinae, Oligochaeta, Tanypodinae, *Ecdyonurus* sp., *Ablabesmyia* sp., *Caenis* sp., *Serratella* sp., Simulidae, *Bezzia* sp., *Hydropsyche* sp., *Arctocorisa* sp., *Achnanthidium minutissimum*, *Coccoenis placentula* var. *lineata*, *Reimeria sinuata*, *Planothidium frequentissimum*, *Navicula cryptocephala*, *Navicula cryptotenella*, *Encyonema minutum*
4	*Baetis* sp., Chironominae, *Hydropsyche* sp., Oligochaeta, Orthocladiinae, Tanypodinae, *Ecdyonurus* sp., *Leuctra* sp., Simulidae, *Hydraena* sp., *Bezzia* sp., *Oulimnius* sp., *Limnius* sp., *Isoperla* sp., *Atherix* sp., *Polycentropus* sp., *Habrophlebia* sp., *Siphonoperla* sp., *Achnanthidium minutissimum*, *Achnanthidium subatomoides*, *Achnanthidium helveticum*, *Cocconeis placentula* var. *lineata*, *Gomphonema parvulum* var. *parvulum* f. *parvulum*, *Reimeria sinuata*, *Achanthidium biasolettianum*, *Encyonema minutum*, *Fragilaria capucina* var. *vaucheriae*
5	Chironominae, *Baetis* sp., *Caenis* sp., *Oulimnius* sp., *Ancylus* sp., *Orthetrum* sp., *Achnanthidium minutissimum, Gomphonema pumilum*, *Fragilaria capucina* var. *vaucheriae*, *Gomphonema parvulum* var. *parvulum* f. *parvulum*, *Cyclotella meneghininana*, *Navicula cryptocephala*
6	*Baetis* sp., Chironominae, Orthocladiinae, Tanypodinae, *Esolus* sp., *Ecdyonurus* sp., *Leuctra* sp., Oligochaeta, *Serratella* sp., *Limnius* sp., *Achnanthidium minutissimum*, *Gomphonema parvulum* var. *parvulum* f. *parvulum*

Table 10 - Environmental characterization of the 6 reference groups of streams used in the RIV INV+DIAT predictive model.

Discriminant variables	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Latitude (rectangular)	38.5 (±1.3)	41.2 (±0.7)	40.5 (±0.7)	41.3 (±0.3)	38.5 (±1.4)	40.6 (±0.6)
Altitude (m)	149.6 (±104.2)	301.9 (±194.7)	361.6 (±196.0)	426.6 (±196.8)	193.4 (±240.7)	406.8 (±343.8)
Distance to source (km)	26.7 (±25.6)	22.1 (±24.5)	40.7 (±38.8)	18.2 (±15.3)	37.7 (±32.0)	11.8 (±10.8)
Typical flow regime (categorical data)	0.3 (±0.4)	1.0 (±0.0)	0.9 (±0.3)	1.0 (±0.2)	0.5 (±0.5)	1.0 (±0.0)
Mean annual precipitation (mm)	705.4 (±174.8)	1507.6 (±562.4)	862.7 (±256.9)	1044.9 (±496.5)	740.8 (±335.6)	1554.5 (±376.4)
Lithology (categorical data)	2.0 (±0.5)	2.3 (±0.6)	2.2 (±0.6)	2.3 (±0.5)	2.0 (±0.4)	2.1 (±0.6)
Alkalinity (mgl^{-1})	62.7 (±55.5)	12.3 (±9.3)	48.9 (±61.4)	31.2 (±20.8)	85.6 (±79.9)	21.0 (±53.4)
Hardness (mgl^{-1})	97.6 (±69.1)	12.0 (±10.6)	26.4 (±29.7)	15.6 (±15.6)	161.3 (±114.3)	9.8 (±7.8)

The selected model is shown in Figure 40 as well as the RMSE values for both validation and calibration data. The RMSE (O/E) calibration dataset located between boundaries of an effective model (Van Sickle et al. 2005) and corresponds to a SDOEcal of 0.164.

The model is statistically accurate (slope=1.024; intercept=-0.008 and R^2=0.728). according to Linke et al. (2005).

The O/E values of most reference sites are close to 1, which means that the observed taxa are almost equal to the expected ones (Figure 41).

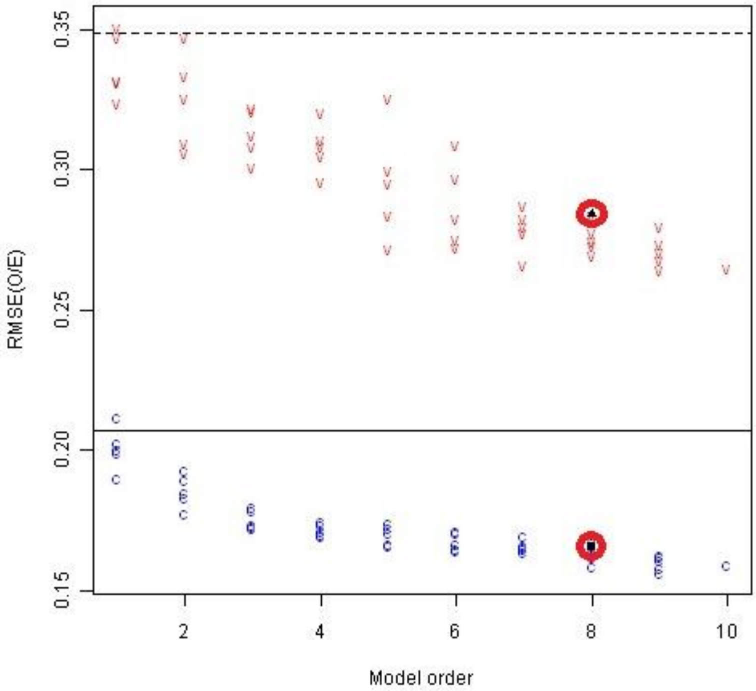

Figure 40 - RMSE (O/E) values for both calibration (c) and validation (v) datasets are shown as well as the corresçponding maximum of RMSE for calibration dataset (black line) and the selected model (red circles).

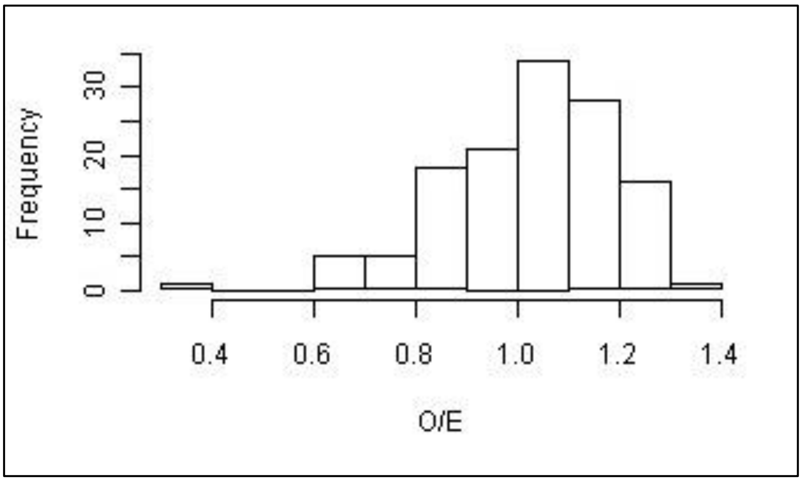

Figure 41 - Distribution of frequencies of the O/E 50 of all reference sites used in the RIV INV+DIAT model.

The mean Observed/Expected ratio for validation of reference sites was 0.86 and almost of 30% of validation reference sites were classified above the good quality status (Table 11). The meaning of each site code (numbers ranged from 1 to 143) is presented in Appendix A.

The RIV INV+DIAT predictive model was based on the same 143 reference sites as the other models (Appendix A). This model results from the combination of the two communities (diatoms and macroinvertebrates) so a total of 598 different taxa were observed in the reference sites (Appendix B and Appendix C).

Table 11 - RIV INV+DIAT model validation with 14 reference sites not included in the model construction.

Sites	Class	OE50
4	Poor	0.33
54	High	1.04
60	Good	0.80
71	Moderate	0.56
75	Moderate	0.63
77	High	1.00
94	High	0.97
102	Moderate	0.60
109	High	0.90
113	High	1.23
120	High	1.03
125	High	1.14
138	Good	0.73
141	High	1.05
Average		0.86

The summary of the characteristics of each RIVPACS type predictive model is shown in Table 12.

Table 12 - Summary of the characteristics of the three RIVPACS predictive models.

	RIV DIAT	RIV INV	RIV DIAT+INV
Number of groups:	6	6	6
F-stat	9.347	7.376	8.063
Wilks	0.070	0.109	0.118
MNOEcal	1.004	0.999	1.026
RMSEcal	0.241	0.195	0.166
SDOE null model	0.315	0.249	0.207
SDOEcal	0.242	0.195	0.164
SDRScal	0.195	0.118	0.102
Correct classification:			
Re-substitution	75.2%	64.3%	61.2%
Cross-validation	66.7%	55.8%	55.0%
Discriminant variables:			
	Alkalinity Altitude Distance to source Hardness Typical flow regime Latitude Lithology Mean annual Precipitation Mean annual Temperature	Altitude Distance to source Hardness Typical flow regime Latitude Lithology Longitude Mean annual Precipitation Mean annual Temperature	Alkalinity Altitude Distance to source Hardness Typical flow regime Latitude Lithology Mean annual Precipitation
OE regression values:			
R^2	0.842	0.704	0.792
Slope	0.993	1.002	1.024
Intersection	0.097	-0.085	0.15
WFD classes (minimum OE values):			
High - Good	0.859	0.897	0.935
Good - Moderate	0.644	0.673	0.701
Moderate - Poor	0.430	0.449	0.467
Poor - Bad	0.215	0.224	0.233

4.4 BEAST Diatom model - BEAST DIAT model

From the Cluster analysis of the calibration dataset based on relative abundance of diatom matrix, 6 groups were selected (Figure 42; Table 22) ranging from 7 to 39 reference sites each.

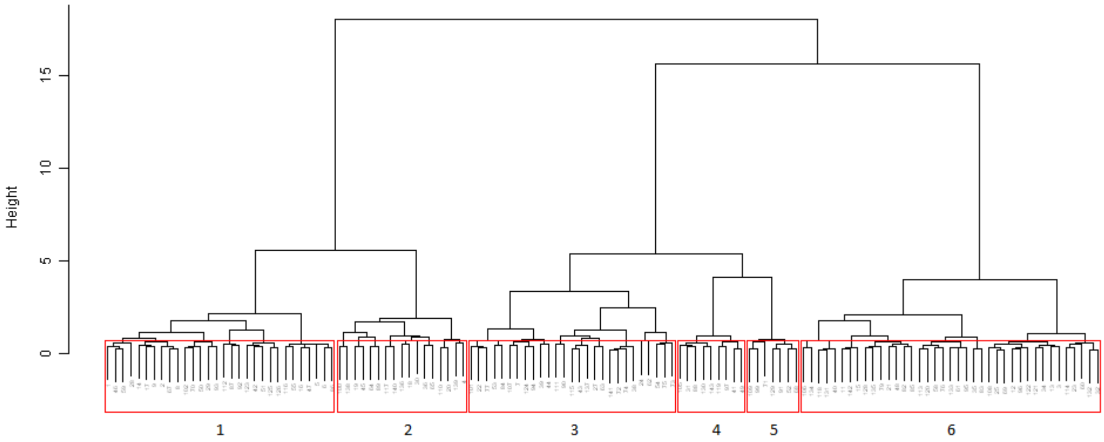

Figure 42 - Cluster of the calibration reference sites used for BEAST DIAT model construction. Each number corresponds to a group.

The spatial ordination of the six groups shows little overlap (MDS analysis; stress: 0.18). Group 3 is the most dispersed while group 5 is the most cohesive (Figure 43).

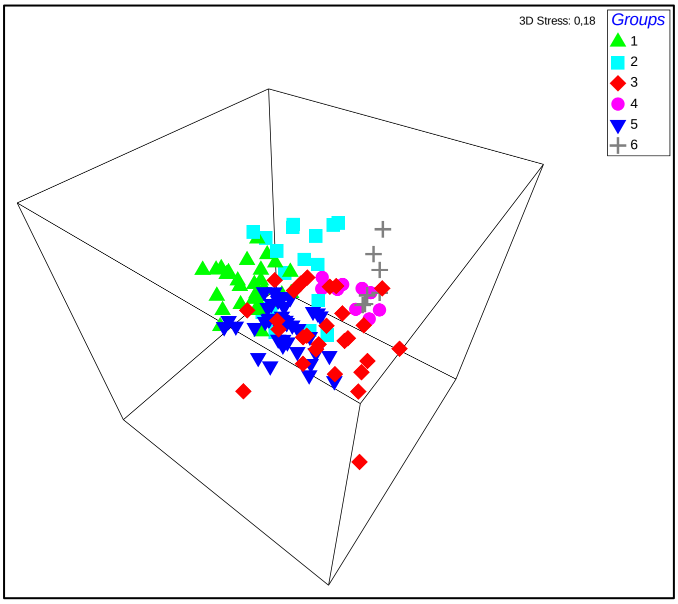

Figure 43 - Three-dimensional MDS ordination of the six groups based on diatom reference community from the calibration dataset of BEAST DIAT model.

The most contributive taxa to the similarity within each reference group (SIMPER analysis; Bray-Curtis coefficient; fourth root transformation) are present in Table 13. The only taxon common to all groups is *Achnanthidium minutissimum* (ADMI). The groups seem well defined with no taxa appearing in more than 2 groups. Some groups include exclusive taxa such as *Eolimna minima* (EOMI) in group 3, *Nitzschia palea* (NPAL) in group 4 or *Achnanthidium biasolettianum* (ADBI) in group 5.

Table 13 - Most representative diatom taxa of the 6 reference groups of BEAST DIAT model, obtained by SIMPER analysis. The diatoms represented were found in 50% or more of the sites.

Groups	Most representative diatoms
1	*Achnanthidium minutissimum, Cocconeis placentula* var. *lineata, Cocconeis euglypta, Planothidium frequentissimum, Reimeria sinuata, Amphora pediculus*
2	*Achnanthidium minutissimum, Gomphonema pumilum, Fragilaria capucina var. vaucheriae*
3	*Achnanthidium minutissimum, Eolimna minima, Cocconeis placentula var. placentula, Gomphonema parvulum var. parvulum f. parvulum, Fragilaria capucina var. capucina, Gomponema pumilum*
4	*Achnanthidium minutissimum*, *Frafilaria capucina var. vaucheriae, Gomphonema parvulum var. parvulum f. parvulum, Nitzschia palea, Surirella angusta, Diatoma mesodon*
5	*Achnanthidium minutissimum*, *Coconeis placentula var. placentula, Achnanthidium biasolettianum, Encyonema minutum, Reimeria sinuata, Gomphonema rhombicum, Achnanthidium helveticum, Achnanthidium subatomoides, Cocconeis placentula* var. *lineata*
6	*Achnanthidium minutissimum, Gomphonema rhombicum*

The environmental variables and the six biological groups were used to run the discriminant function analysis. The best five DF models from each order, up to 10th order (using 10 environmental predictors), based on their Wilk's lambda and F-statistic values, were retained, resulting in forty-six best models from 1023 possibilities. The elected model was selected mainly based on the percentage of correct classifications, Cross-validation (58.9%) and Re-substitution (65.1%) analysis, as well as the discriminant variables (Table 22).

The model uses five environmental variables that best discriminate the reference groups: alkalinity, altitude, hardness, latitude and mean annual precipitation. The mean values and the correspondent standard deviation of the predictor variables for each reference group are shown in Table 14. Group 1 includes sites located in southern Portugal, in lowland areas were the climate is dry, water hardness is medium and alkalinity is high. Lithology is mostly sedimentary and metamorphic and the alkalinity. Group 2 is also from the south and is composed by rivers located at the even lowest altitude where the climate is the driest of the entire dataset. Both alkalinity and hardness are also the highest. Group 3 is characterized by sites with low altitude; water alkalinity and hardness are the lowest of all sites. Mean annual precipitation is medium. Group 4 includes the most northern streams with low altitude, hardness and alkalinity which characterize siliceous river beds. The climate is the moistiest. The rivers belonging to group 5 have low precipitation and water hardness, medium alkalinity and are located at the highest altitude. Group 6 is characterized by low values of altitude, alkalinity and hardness. These sites have medium mean annual precipitation.

Table 14 - Environmental characterization of the 6 reference groups of streams used in the BEAST DIAT predictive model.

Discriminant variables	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Latitude (rectangular)	38.6 (±1.2)	38.8 (±1.2)	40.8 (±0.6)	41.7 (±0.2)	41.3 (±0.3)	40.1 (±1.1)
Altitude (m)	175.8 (±128.1)	157.6 (±194.7)	360.0 (±196.0)	219.1 (±196.8)	417.1 (±240.7)	196.5 (±343.8)
Mean annual precipitation (mm)	779.3 (±207.7)	742.4 (±213.9)	1366.7 (±457.7)	2020.6 (±217.7)	912.0 (±444.0)	1456.2 (±470.6)
Alkalinity (mgl^{-1})	69.3 (±66.5)	82.3 (±87.6)	15.0 (±13.4)	5.9 (±1.8)	38.9 (±26.9)	16.0 (±28.3)
Hardness (mgl^{-1})	90.6 (±61.0)	128.5 (±118.8)	8.8 (±7.0)	10.5 (±12.6)	19.7 (±19.5)	32.6 (±62.6)

Less than 10% of the validation reference sites were classified below the good quality status (Table 15). The meaning of each site code (numbers ranged from 1 to 143) is presented in Appendix A.

Table 15 - BEAST DIAT model validation with 14 reference sites not included in the model construction.

Sites	Class
10	High
26	High
33	High
37	High
56	Good
57	Good
66	Good
78	Good
80	High
81	Moderate
98	High
103	Good
104	Good
127	Good

4.5 BEAST Macroinvertebrate model - BEAST INV model

Six reference groups were obtained based on their macroinvertebrate communities after a Cluster analysis (Figure 44; Table 22) ranging from 8 to 34 reference sites.

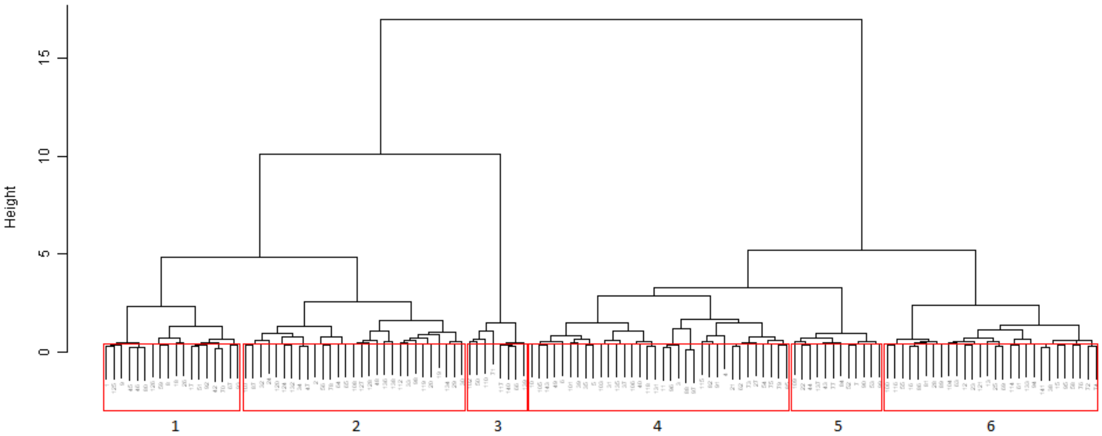

Figure 44 - Cluster of the calibration reference sites used for BEAST INV model construction. Each number corresponds to a group.

The MDS analysis (stress: 0.16) shows most of the reference groups overlapped, with the exception of group 4 (Figure 45).

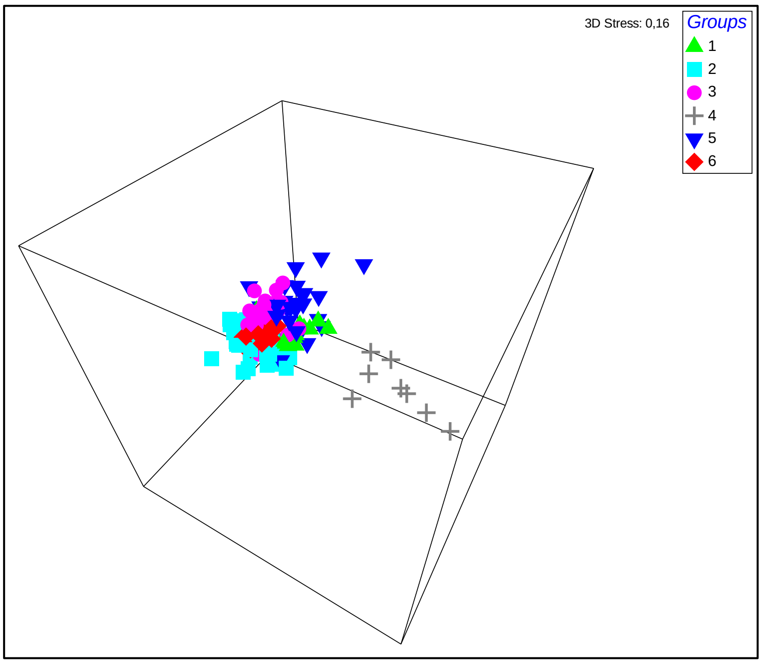

Figure 45 - Three-dimensional MDS ordination of the six groups based on biological reference community from the calibration dataset of BEAST INV model.

The most contributive taxa to the similarity within each reference group (SIMPER analysis; Bray-Curtis coefficient; fourth root transformation) are present in Table 16. No taxa were found to be common to all groups. However, the taxa *Baetis* sp., Orthocladiinae, Tanypodinae and Chironominae were found in all groups except in the fourth. All taxa belonging to group 4 are exclusive of this group. Few other taxa are exclusive to the remaining groups: *Oulimnius* sp. in group 1, *Atherix* sp in group 2 and *Limnius* sp. in group 6. This is in accordance with the results of the MDS shown above.

The selected model presented a percentage of correct classifications of 73.6 and 62% of Re-substitution and Cross-validation, respectively (Table 22).

The model selected eight environmental variables that best discriminate the reference groups: alkalinity, altitude, distance to source, hardness, typical flow regime, latitude, longitude and mean annual temperature. The mean values and the correspondent standard deviation of the predictor variables for each reference group are shown in Table

17. Group 1 is composed by streams mainly temporary with high temperature and located at low altitude. Alkalinity is low and water hardness is medium. The rivers belonging to Group 2 are all small and permanent located at the highest altitude in North of Portugal. These sites are characterized by low values of temperature, alkalinity and hardness. Group 3 includes mainly permanent streams placed at medium altitude where the temperature is also medium while alkalinity and water hardness are low. Group 4 is characterized by the highest values of temperature, alkalinity and hardness. This group includes mainly large temporary watercourses located at the lowest altitude in southern Portugal. In group 5, the sites are located at low altitude with medium temperature, alkalinity and hardness. The majority are permanent rivers. The streams belonging to group 6 are all permanent and are characterized by the lowest values of temperature, alkalinity and hardness. The sites are located at low altitude.

Table 16 - Most representative taxa of the 6 reference groups of BEAST INV model, obtained by SIMPER analysis. The invertebrates represented were found in 50% or more of the sites.

Groups	Most representative invertebrates
1	Chironominae, Orthocladiinae, *Baetis* sp., Simulidae, *Oulimnius* sp., Oligochaeta, Tanypodinae, Limonidae
2	*Baetis* sp., Orthocladiinae, Chironominae, Tanypodinae, Oligochaeta, *Hydropsyche* sp., Simulidae, *Ecdyonurus* sp., *Atherix* sp., *Leuctra* sp.
3	*Baetis* sp., Chironominae, *Leuctra* sp., Orthocladiinae, Oligochaeta, *Hydropsyche* sp., *Caenis* sp., Simulidae, *Ecdyonurus* sp., Tanypodinae, *Serratella* sp.
4	*Orthetrum* sp., *Orthotrichia* sp., Tabanidae, *Ochthebius* sp., *Setodes* sp.
5	Chironominae, Orthocladiinae, *Baetis* sp., Tanypodinae, Oligochaeta, *Caenis* sp.
6	*Leuctra* sp., *Baetis* sp., Tanypodinae, Orthocladiinae, Chironominae, *Ecdyonurus* sp., *Serratella* sp., *Limnius* sp.

Table 17 - Environmental characterization of the 6 reference groups of streams used in the BEAST INV predictive model.

Discriminant variables	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Latitude (rectangular)	38.0 (±1.1)	41.2 (±0.7)	40.9 (±0.8)	37.6 (±0.3)	40.3 (±1.0)	40.4 (±0.2)
Altitude (m)	152.2 (±143.3)	445.4 (±259.3)	320.5 (±166.0)	106.9 (±88.5)	278.6 (±234.3)	211.7 (±174.8)
Distance to source (km)	18.6 (±11.4)	8.5 (±6.3)	34.4 (±30.3)	59.8 (±33.3)	37.6 (±36.9)	24.2 (±27.6)
Typical flow regime (categorical data)	0.3 (±0.4)	1.0 (±0.0)	1.0 (±0.2)	0.4 (±0.5)	0.7 (±0.5)	1.0 (±0.0)
Mean annual temperature (^{o}C)	15.5 (±1.2)	12.5 (±1.6)	13.1 (±1.0)	16.0 (±0.6)	13.7 (±1.4)	12.1 (±1.8)
Alkalinity (mgl^{-1})	44.0 (±38.1)	31.7 (±57.8)	31.3 (±22.6)	94.5 (±20.6)	64.4 (±70.2)	14.8 (±18.8)
Hardness (mgl^{-1})	86.4 (±60.9)	21.3 (±40.9)	19.5 (±14.3)	163.0 (±15.8)	71.4 (±101.4)	7.5 (±6.6)

About 15% of validation reference sites were classified below the good quality status (Table 18). The meaning of each site code (numbers ranged from 1 to 143) is presented in Appendix A.

Table 18 - BEAST INV model validation with 14 reference sites not included in the model construction.

Sites	Class
14	High
36	High
41	Good
57	Good
60	Good
68	Good
83	High
111	Moderate
113	Good
122	Moderate
123	Good
129	High
130	High
142	Good

4.6 BEAST Macroinvertebrate and Diatom model - BEAST INV+DIAT model

The summary of the characteristics of BEAST INV+DIAT model are present in Table 22. The cluster analysis lead to the formation of six biological reference groups (Figure 46) ranging from 6 to 31 sites. The MDS analysis (stress: 0.16) shows that most of the groups are well defined with some overlap except the sixth group that presents some dispersion (Figure 47).

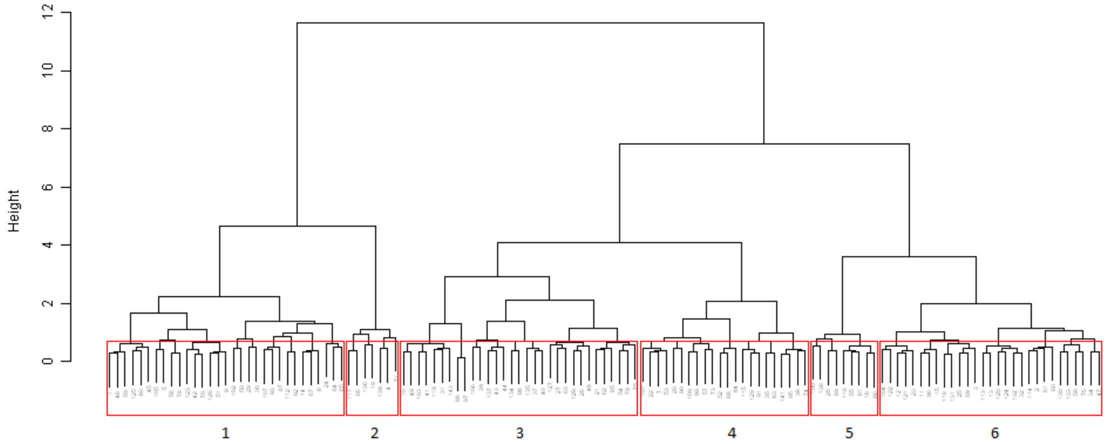

Figure 46 - Cluster of the calibration reference sites used for BEAST INV+DIAT model construction. Each number corresponds to a group.

The most representative taxa (SIMPER analysis; Bray-Curtis coefficient; fourth root transformation) of the six reference groups are denoted in Table 19. The taxa common to all groups includes one invertebrate taxon, Chironominae, and one diatom, *Achnanthidium minutissimum* (ADMI). However, other taxa such as Orthocladiinae, Oligochaeta and *Cocconeis placentula var. lineata* (CPLI) also contributed for the similarity within the groups. Some taxa were found to be exclusive to one group such as *Planothidium frequentissimum* (PLFR) in group 1, *Ephemerella* sp. and *Fragilaria capucina var. vaucheriae* (FCVA) in group 2, *Ablabesmyia* sp. and *Encyonema silesiacum* (ESLE) in group 3, *Serratella* sp. and *Gomphonema rhombicum* (GRHB), *Reimeria sinuata* (RSIN) in group 5 or *Orthetrum* sp. in group 6.

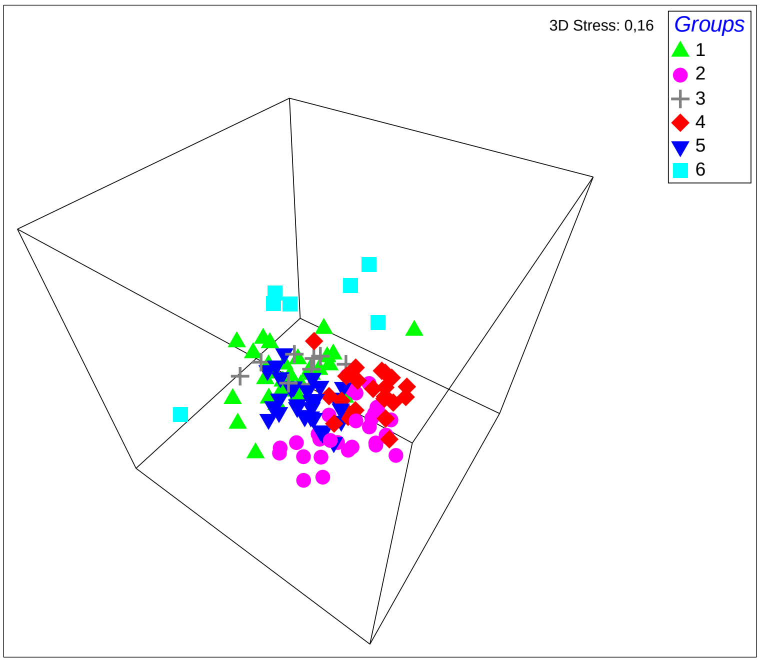

Figure 47 - Three-dimensional MDS ordination of the six groups based on biological reference community from the calibration dataset of BEAST INV+DIAT model.

The selected model classified correctly the calibration dataset to their original classification groups in 72.1% by the Re-substitution analysis and 66.7% by the Cross-validation analysis (Table 22). These groups were best discriminated by eight variables which includes alkalinity, distance to source, hardness, typical flow regime, latitude, lithology, longitude and mean annual temperature. The mean values and the correspondent standard deviation of the predictor environmental variables for each reference group are shown in Table 20. All groups share the same lithology, sedimentary and metamorphic rocks, except the second group which is composed by plutonic rocks. Group 1 is composed by streams mainly temporary with high temperature and alkalinity (southern Portugal). Water hardness is intermediate. Group 2 is characterized by the smallest permanent rivers located in the North of Portugal with low values of temperature, alkalinity and hardness. Group 3 includes mostly permanent watercourses with an intermediate air temperature and low alkalinity and hardness. Group 4 is characterized by permanent rivers and characterized by low values of temperature, alkalinity and hardness. In group 5, all the streams are large and permanent with medium temperature and alkalinity (northern Portugal). Water hardness is low. The watercourses belonging to group 6 are mostly permanent, located in the South of Portugal, and are characterized by the highest values of temperature, alkalinity and hardness.

Table 19 - Most representative taxa of the 6 reference groups of BEAST INV+DIAT model, obtained by SIMPER analysis. The invertebrates and diatoms represented were found in 50% or more of the sites.

Groups	**Most representative diatoms and invertebrates**
1	*Baetis* sp., Chironominae, Orthocladiinae, Simulidae, *Oulimnius* sp., Tanypodinae, Oligochaeta, *Caenis* sp., *Achnanthidium minutissimum, Cocconeis placentula var. lineata, Planothidium frequentissimum, Cocconeis euglypta,*
2	*Baetis* sp., Tanypodinae, Orthocladiinae, Oligochaeta, Chironominae, *Ecdyonurus* sp., *Hydropsyche* sp., *Ephemerella* sp., Simulidae, *Leuctra* sp., *Oulimnius* sp., *Habrophlebia* sp., *Gomphonema parvulum var. parvulum f. parvulum, Eunotia minor, Fragilaria capucina var. vaucheriae, Achnanthidium minutissimum*
3	*Leuctra* sp., *Baetis* sp., Chironominae, *Ablabesmyia* sp., *Hydropsyche* sp., *Chimarra* sp., Oligochaeta, Orthocladiinae, *Caenis* sp., *Polycentropus* sp., *Cocconeis placentula var. lineata, Encyonema silesiacum, Cocconeis euglypta, Achnanthidium minutissimum*
4	*Baetis* sp., Orthocladiinae, Tanypodinae, Chironominae, *Leuctra* sp., *Ecdyonurus* sp., *Serratella* sp., Oligochaeta, *Gomphonema rhombicum, Achnanthidium minutissimum*
5	*Baetis* sp., Chironominae, Tanypodinae, Orthocladiinae, *Caenis* sp., *Hydropsyche* sp., *Ecdyonurus* sp., Oligochaeta, *Leuctra* sp., Simulidae, *Achnanthidium minutissimum, Achnanthidium biasolettianum, Cocconeis placentula var. placentula, Reimeria sinuata, Cocconeis placentula var. lineata, Encyonema minutum*
6	Chironominae, *Caenis* sp., *Orthetrum* sp., *Oulimnius* sp., *Achnanthidium minutissimum, Gonphonema pumilum*

Table 20 - Environmental characterization of the 6 reference groups of streams used in the BEAST INV+DIAT predictive model.

Discriminant variables	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Latitude (rectangular)	38.6 (±1.3)	41.3 (±0.5)	39.9 (±0.3)	40.7 (±0.5)	41.2 (±0.6)	38.1 (±0.1)
Longitude (rectangular)	-8.1 (±0.6)	-7.9 (±0.5)	-7.7 (±0.3)	-8.0 (±0.3)	-7.3 (±0.6)	-8.2 (±0.6)
Distance to source (km)	23.1 (±22.6)	8.0 (±5.9)	24.9 (±14.9)	22.3 (±23.1)	47.9 (±40.2)	42.8 (±34.8)
Typical flow regime (categorical data)	0.3 (±0.5)	1.0 (±0.2)	0.8 (±0.4)	1.0 (±0.0)	1.0 (±0.0)	0.6 (±0.5)
Mean annual temperature (^{o}C)	15.2 (±1.1)	12.3 (±1.5)	13.7 (±1.1)	12.1 (±1.9)	13.0 (±1.0)	15.3 (±1.0)
Lithology (categorical data)	1.9 (±0.5)	2.7 (±0.5)	2.1 (±0.6)	2.1 (±0.4)	2.1 (±0.4)	1.9 (±0.4)
Alkalinity (mgl^{-1})	77.7 (±75.9)	14.0 (±9.6)	16.7 (±3.7)	19.7 (±45.1)	47.0 (±32.7)	88.8 (±63.9)
Hardness (mgl^{-1})	104.5 (±74.9)	8.8 (±7.6)	23.3 (±8.6)	8.3 (±7.0)	27.5 (±27.6)	193.6 (±109.7)

About 15% of the validation sites were classified above the good quality status (Table 21). The meaning of each site code (numbers ranged from 1 to 143) is presented in Appendix A.

The summary of the characteristics of each BEAST type predictive model is shown in Table 22.

Table 21 - BEAST INV+DIAT model validation with 14 reference sites not included in the model construction.

Sites	Class
6	High
17	High
18	Good
33	Good
36	High
72	High
77	Good
82	Good
94	Moderate
108	Good
110	High
111	Moderate
140	High
142	Good

Table 22 - Summary of the characteristics of the three BEAST predictive models.

	BEAST DIAT	BEAST INV	BEAST DIAT+INV
Number of groups:	6	6	6
F-stat	14.433	9.809	10.147
Wilks	0.11	0.083	0.077
Correct classification:			
Re-substitution	65.1%	73.6%	72.1%
Cross-validation	58.9%	62%	66.7%
Discriminant variables:			
	Alkalinity	Alkalinity	Alkalinity
	Altitude	Altitude	Distance to source
	Hardness	Distance to source	Hardness
	Latitude	Hardness	Typical flow regime
	Mean annual Precipitation	Typical flow regime	Latitude
		Latitude	Lithology
		Longitude	Longitude
		Mean annual Temperature	Mean annual Temperature

4.7 Assessment of test sites

The test sites were classified from poor to high quality status by RIV DIAT model (Table 23): 47.8% in high, 17.4% in good, 26.1% in moderate and 8.7% in poor quality classes. This model attributed mostly high quality to the test sites and only two of them were classified as poor (M55 and M112). RIV INV predictive model classified the test sites from bad to high (Table 23): 17.4% in high, 34.8% in good, 34.8% in moderate and 13% in poor quality classes. At last, the combined model, RIV INV+DIAT; attributed classes from poor to high like the RIV DIAT model (Table 23): 13% in high, 43.5% in good, 30.5% in moderate and 13% in poor quality classes.

Table 23 – Water quality classes attributed to test sites by the three predictive models, RIV DIAT, RIV INV and RIV INV+DIAT.

Sites	RIV DIAT	RIV INV	RIV INV+DIAT
M18	High	Moderate	Good
V118	Good	High	Good
M49	Good	Moderate	Good
M109	Moderate	High	High
L120	Moderate	Moderate	Moderate
L42	High	Good	Moderate
M43	Moderate	Good	Good
M123	Good	Moderate	Moderate
M55	Poor	Good	Good
M2001	High	Good	Good
M108	High	Good	Good
M101	High	High	High
M2002	Good	Good	Good
M111	High	Good	Good
M110	High	Good	Good
V78	High	Moderate	Moderate
M112	Moderate	Moderate	Moderate
V125	Moderate	Poor	Poor
V124	Moderate	Poor	Poor
M122	Poor	Moderate	Moderate
V94	High	Moderate	Moderate
V119	High	Poor	Poor
V36	High	High	High

Regarding the BEAST models, the test sites were classified from bad to high quality status by BEAST DIAT model (Table 24): 8.7% in high, 13% in good, 56.5% in moderate, 17.4% in poor and 4.4% in bad quality classes. Most of the test sites were classified as moderate quality and only one of them was classified as bad (M112). BEAST INV predictive model classified the test sites from bad to good status (Table 24): 4.4% in good, 21.7% in moderate, 30.4% in poor and 43.5% in bad quality classes. Finally, the combined model BEAST INV+DIAT attributed classes from bad to good (Table 24): 8.7% in good, 52.2% in moderate, 26.1% in poor and 13% in bad classes.

Table 24 - Classes attributed to test sites by the three predictive models, BEAST DIAT, BEAST INV and BEAST INV+DIAT.

Sites	BEAST DIAT	BEAST INV	BEAST INV+DIAT
M18	Moderate	Bad	Poor
V118	Good	Poor	Moderate
M49	Good	Poor	Poor
M109	Moderate	Good	Moderate
L120	Poor	Bad	Poor
L42	Moderate	Bad	Poor
M43	Moderate	Poor	Moderate
M123	Poor	Bad	Poor
M55	Moderate	Moderate	Moderate
M2001	Moderate	Moderate	Moderate
M108	Moderate	Poor	Moderate
M101	Moderate	Moderate	Moderate
M2002	Moderate	Moderate	Moderate
M111	Poor	Bad	Poor
M110	Poor	Poor	Bad
V78	Moderate	Poor	Good
M112	Bad	Poor	Moderate
V125	High	Bad	Moderate
V124	Moderate	Bad	Bad
M122	Good	Bad	Good
V94	Moderate	Bad	Moderate
V119	Moderate	Bad	Bad
V36	High	Moderate	Moderate

In tables 25 and 26 the assessment of the combined models against the assessment made by the worst classification by individual models (which is common practice in the context of the WFD) is shown.

Table 25 – Classification of test sites according to what is used in the context of the WFD (worst classification obtained by the individual diatom or the macroinvertebrate RIVPACS models) and RIV INV+DIAT model.

Sites	WFD	RIV INV+DIAT
M18	Moderate	Good
V118	Good	Good
M49	Moderate	Good
M109	Moderate	High
L120	Moderate	Moderate
L42	Good	Moderate
M43	Moderate	Good
M123	Moderate	Moderate
M55	Poor	Good
M2001	Good	Good
M108	Good	Good
M101	High	High
M2002	Good	Good
M111	Good	Good
M110	Good	Good
V78	Moderate	Moderate
M112	Moderate	Moderate
V125	Poor	Poor
V124	Poor	Poor
M122	Poor	Moderate
V94	Moderate	Moderate
V119	Poor	Poor
V36	High	High

The RIV INV+DIAT model classified about 70% of the test sites with the same quality class as the WFD approach. Most of the remaining sites had a classification of higher quality in the combined model than with the WFD approach. In the case of BEAST, the combined model and WFD only share about 40% of the test sites with equal classification. As RIV INV+DIAT, the BEAST INV+DIAT model also attributed higher quality status to the sites comparing to WFD classifications.

Table 26 - Classification of test sites according to what is used in the context of the WFD (worst classification obtained by the individual diatom or the macroinvertebrate BEAST models) and BEAST INV+DIAT model.

Sites	WFD	BEAST INV+DIAT
M18	Bad	Poor
V118	Poor	Moderate
M49	Poor	Poor
M109	Moderate	Moderate
L120	Bad	Poor
L42	Bad	Poor
M43	Poor	Moderate
M123	Bad	Poor
M55	Moderate	Moderate
M2001	Moderate	Moderate
M108	Poor	Moderate
M101	Moderate	Moderate
M2002	Moderate	Moderate
M111	Bad	Poor
M110	Poor	Bad
V78	Poor	Good
M112	Bad	Moderate
V125	Bad	Moderate
V124	Bad	Bad
M122	Bad	Good
V94	Bad	Moderate
V119	Bad	Bad
V36	Moderate	Moderate

Spearman rank correlation coefficients between the quality classes of the RIVPACS type of models and the pressure variables for both reference and test sites showed significant correlations (*P*<0,05) between RIV DIAT and pH, BOD_5, nitrites, ammonia, sediment discharge, hydrological regime and HMS (Table 27). The RIV INV predictive model was correlated with dissolved O_2, conductivity, nitrates, nitrites, ammonia, total phosphorous, sediment discharge, hydrological regime, morphological condition, HQA and HMS (Table 27). Finally, the combined predictive model RIV INV+DIAT showed significant correlation with dissolved O_2, BOD_5, nitrates, nitrites, ammonia, total phosphorous, riparian zone, sediment discharge, hydrological regime, morphological condition, HQA and HMS (Table 27).

Regarding the BEAST type of models, significant correlations were found (P<0,05) between BEAST DIAT model and dissolved O_2, nitrites, ammonia, total phosphorous, riparian zone, sediment discharge, hydrological regime, morphological condition, connectivity and HMS (Table 28). The BEAST INV predictive model was correlated with dissolved O_2, nitrates, nitrites, ammonia, riparian zone, total phosphorous, sediment discharge, hydrological regime, acidification and toxicity, morphological condition, connectivity, HQA and HMS (Table 28). At last, the combined predictive model BEAST INV+DIAT showed significant correlation with dissolved O_2, nitrates, nitrites, ammonia, total phosphorous, riparian zone, sediment discharge, hydrological regime, morphological condition, connectivity, HQA and HMS (Table 28).

Regarding RIV INV+DIAT method (Table 29), Spearman rank correlation coefficients showed significant correlations with dissolved oxygen, BOD_5, nitrates, nitrites, ammonia, total phosphorous, riparian zone, sediment discharge, hydrological regime, morphological condition, HQA and HMS pressure variables. The WFD method concerning the RIVPCAS models was correlated with dissolved oxygen, BOD_5, nitrates, nitrites, ammonia, total phosphorous, sediment discharge, hydrological regime, HQA and HMS. In this case, most of those variables were better correlated with the RIV INV+DIAT, the combined model than with the individual models.

Table 27 - Spearman rank correlation (rs) and respective *P* value between quality classes of the three RIVPACS type of predictive models and pressure variables for reference and test sites.

Pressure variables	RIV DIAT	RIV INV	RIV INV+DIAT
O_2 diss	rs=-0.126 *P*=0.111	rs=-0.248 *P*<0.05	rs=-0.301 *P*<0.05
pH	rs=-0.180 *P*<0.05	rs=0.006 *P*=0.943	rs=-0.058 *P*=0.469
Conductivity	rs=-0.044 *P*=0.582	rs=0.185 *P*<0.05	rs=0.126 *P*=0.112
BOD_5	rs=0.182 *P*<0.05	rs=0.032 *P*=0.688	rs=0.156 *P*<0.05
COD	rs=0.031 *P*=0.697	rs=0.073 *P*=0.356	rs=0.099 *P*=0.213
Nitrates	rs=0.096 *P*=0.225	rs=0.220 *P*<0.05	rs=0.333 *P*<0.05
Nitrites	rs=0.186 *P*<0.05	rs=0.423 *P*<0.05	rs=0.414 *P*<0.05
Ammonia	rs=0.285 *P*<0.05	rs=0.482 *P*<0.05	rs=0.504 *P*<0.05
Total P	rs=0.127 *P*=0.108	rs=0.383 *P*<0.05	rs=0.402 *P*<0.05
Rip zone	rs=0.039 *P*=0.622	rs=0.130 *P*=0.101	rs=0.279 *P*<0.05
Sediments	rs=0.221 *P*<0.05	rs=0.377 *P*<0.05	rs=0.376 *P*<0.05
Hyd reg	rs=0.195 *P*<0.05	rs=0.159 *P*<0.05	rs=0.233 *P*<0.05
Acid e Tox	rs=0.139 *P*=0.080	rs=0.078 *P*=0.326	rs=0.174 *P*=0.082
Morph cond	rs=-0.036 *P*=0.655	rs=0.224 *P*<0.05	rs=0.268 *P*<0.05
Org cont	rs=-0.016 *P*=0.842	rs=-0.004 *P*=0.956	rs=-0.094 *P*=0.236
Connectivity	rs=0.020 *P*=0.798	rs=0.018 *P*=0.818	rs=0.111 *P*=0.163
HQA	rs=-0.116 *P*=0.143	rs=-0.295 *P*<0.05	rs=-0.261 *P*<0.05
HMS	rs=0.175 *P*<0.05	rs=0.349 *P*<0.05	rs=0.392 *P*<0.05

Table 28 - Spearman rank correlation (rs) and respective *P* value between quality classes of the three BEAST type of predictive models and pressure variables for reference and test sites.

Pressure variables	BEAST DIAT	BEAST INV	BEAST INV+DIAT
O_2 diss	rs=-0.267 $P<0.05$	rs=-0.251 $P<0.05$	rs=-0.219 $P<0.05$
pH	rs=0.083 $P=0.297$	rs=0.058 $P=0.462$	rs=0.048 $P=0.543$
Conductivity	rs=0.094 $P=0.237$	rs=0.129 $P=0.102$	rs=0.041 $P=0.608$
BOD_5	rs=0.031 $P=0.696$	rs=0.007 $P=0.928$	rs=-0.020 $P=0.806$
COD	rs=0.011 $P=0.891$	rs=0.039 $P=0.622$	rs=-0.020 $P=0.800$
Nitrates	rs=0.129 $P=0.104$	rs=0.217 $P<0.05$	rs=0.163 $P<0.05$
Nitrites	rs=0.632 $P<0.05$	rs=0.666 $P<0.05$	rs=0.617 $P<0.05$
Ammonia	rs=0.666 $P<0.05$	rs=0.742 $P<0.05$	rs=0.659 $P<0.05$
Total P	rs=0.611 $P<0.05$	rs=0.662 $P<0.05$	rs=0.552 $P<0.05$
Rip zone	rs=0.310 $P<0.05$	rs=0.384 $P<0.05$	rs=0.404 $P<0.05$
Sediments	rs=0.257 $P<0.05$	rs=0.488 $P<0.05$	rs=0.346 $P<0.05$
Hyd reg	rs=0.361 $P<0.05$	rs=0.405 $P<0.05$	rs=0.385 $P<0.05$
Acid e Tox	rs=0.102 $P=0.200$	rs=0.183 $P<0.05$	rs=0.060 $P=0.451$
Morph cond	rs=0.455 $P<0.05$	rs=0.586 $P<0.05$	rs=0.571 $P<0.05$
Org cont	rs=-0.023 $P=0.772$	rs=0.000 $P=0.997$	rs=-0.017 $P=0.831$
Connectivity	rs=0.204 $P<0.05$	rs=0.194 $P<0.05$	rs=0.197 $P<0.05$
HQA	rs=-0.133 $P=0.093$	rs=-0.293 $P<0.05$	rs=-0.240 $P<0.05$
HMS	rs=0.462 $P<0.05$	rs=0.656 $P<0.05$	rs=0.616 $P<0.05$

Spearman rank correlation coefficients between the classifications according to the approach in practice in Europe and the classifications of the BEAST INV+DIAT model and the pressure variables (Table 30) showed that both methods were significantly correlated with the same 12 variables: dissolved oxygen, nitrates, nitrites, ammonia, total phosphorous, riparian zone, sediment discharge, hydrological regime, morphological condition, connectivity, HQA and HMS. Only one variable, acidification and toxicity, was significantly correlated with the WFD method. Nevertheless, most of those variables showed a higher correlation with WFD approach; only two of them (riparian zone and morphological condition) were better correlated with the BEAST combined method.

Table 31 summarizes the Spearman rank correlation coefficients between the classifications attributed by all the approaches applied in this study and the pressure variables. The BEAST INV model showed to be highly correlated with a greater number of pressure variables among all the approaches. On the other hand, by comparison, all the RIVPACS approaches showed to be highly correlated with just one (RIV DIAT and RIV INV) or two (RIV INV+DIAT and RIV WFD) pressure variables.

Table 29 - Spearman rank correlation (rs) and respective *P* value between quality classes of the RIV INV+DIAT and WFD and pressure variables for reference and test sites. In bold the highest significant correlations between methods, for each variable are highlighted.

Pressure variables	WFD	RIV INV+DIAT
O_2 diss	rs=-0.199 $P<0.05$	**rs=-0.301** $P<0.05$
pH	rs=-0.113 $P=0.154$	rs=-0.058 $P=0.469$
Conductivity	rs=0.073 $P=0.360$	rs=0.126 $P=0.112$
BOD_5	**rs=0.167** $P<0.05$	rs=0.156 $P<0.05$
COD	rs=0.074 $P=0.352$	rs=0.099 $P=0.213$
Nitrates	rs=0.210 $P<0.05$	**rs=0.333** $P<0.05$
Nitrites	rs=0.348 $P<0.05$	**rs=0.414** $P<0.05$
Ammonia	rs=0.407 $P<0.05$	**rs=0.504** $P<0.05$
Total P	rs=0.318 $P<0.05$	**rs=0.402** $P<0.05$
Rip zone	rs=0.140 $P=0.076$	rs=0.279 $P<0.05$
Sediments	rs=0.323 $P<0.05$	**rs=0.376** $P<0.05$
Hyd reg	rs=0.156 $P<0.05$	**rs=0.233** $P<0.05$
Acid e Tox	rs=0.068 $P=0.390$	rs=0.174 $P=0.082$
Morph cond	rs=0.119 $P=0.132$	**rs=0.268** $P<0.05$
Org cont	rs=-0.013 $P=0.866$	rs=-0.094 $P=0.236$
Connectivity	rs=0.018 $P=0.817$	rs=0.111 $P=0.163$
HQA	**rs=-0.304** $P<0.05$	rs=-0.261 $P<0.05$
HMS	rs=0.251 $P<0.05$	**rs=0.392** $P<0.05$

Table 30 - Spearman rank correlation (rs) and respective *P* value between quality classes of the BEAST INV+DIAT and WFD and pressure variables for reference and test sites. In bold the highest significant correlations between methods, for each variable are highlighted.

Pressure variables	WFD	BEAST INV+DIAT
O_2 diss	**rs=-0.266** *P*<0.05	rs=-0.219 *P*<0.05
pH	rs=0.060 *P*=0.452	rs=0.048 *P*=0.543
Conductivity	rs=0.124 *P*=0.116	rs=0.041 *P*=0.608
BOD_5	rs=0.010 *P*=0.900	rs=-0.020 *P*=0.806
COD	rs=0.019 *P*=0.807	rs=-0.020 *P*=0.800
Nitrates	**rs=0.189** *P*<0.05	rs=0.163 *P*<0.05
Nitrites	**rs=0.648** *P*<0.05	rs=0.617 *P*<0.05
Ammonia	**rs=0.720** *P*<0.05	rs=0.659 *P*<0.05
Total P	**rs=0.644** *P*<0.05	rs=0.552 *P*<0.05
Rip zone	rs=0.354 *P*<0.05	**rs=0.404** *P*<0.05
Sediments	**rs=0.480** *P*<0.05	rs=0.346 *P*<0.05
Hyd reg	**rs=0.428** *P*<0.05	rs=0.385 *P*<0.05
Acid e Tox	**rs=0.207** *P*<0.05	rs=0.060 *P*=0.451
Morph cond	rs=0.551 *P*<0.05	**rs=0.571** *P*<0.05
Org cont	rs=-0.004 *P*=0.965	rs=-0.017 *P*=0.831
Connectivity	rs=0.197 *P*<0.05	rs=0.197 *P*<0.05
HQA	**rs=-0.259** *P*<0.05	rs=-0.240 *P*<0.05
HMS	**rs=0.638** *P*<0.05	rs=0.616 *P*<0.05

Table 31 – Spearman rank correlation (rs) and respective *P* value between classes of all the approaches and pressure variables for reference and test sites. The red rectangles show the highest and the grey ones the second highest significant correlations between methods, for each variable.

Pressure variables	RIV DIAT	RIV INV	RIV INV+DIAT	RIV WFD	BEAST DIAT	BEAST INV	BEAST INV+DIAT	BEAST WFD
O_2 diss	rs=-0.126 P=0.111	rs=-0.248 P<0.05	**rs=-0.301** P<0.05	rs=-0.199 P<0.05	**rs=-0.267** P<0.05	rs=-0.251 P<0.05	rs=-0.219 P<0.05	rs=-0.266 P<0.05
***P*H**	rs=-0.180 P<0.05	rs=0.006 P=0.943	rs=-0.058 P=0.469	rs=-0.113 P=0.154	rs=0.083 P=0.297	rs=0.058 P=0.462	rs=0.048 P=0.543	rs=0.060 P=0.452
Conductivity	rs=-0.044 P=0.582	**rs=0.185** P<0.05	rs=0.126 P=0.112	rs=0.073 P=0.360	rs=0.094 P=0.237	rs=0.129 P=0.102	rs=0.041 P=0.608	rs=0.124 P=0.116
BOD$_5$	**rs=0.182** P<0.05	rs=0.032 P=0.688	rs=0.156 P<0.05	**rs=0.167** P<0.05	rs=0.031 P=0.696	rs=0.007 P=0.928	rs=-0.020 P=0.806	rs=0.010 P=0.900
COD	rs=0.031 P=0.697	rs=0.073 P=0.356	rs=0.099 P=0.213	rs=0.074 P=0.352	rs=0.011 P=0.891	rs=0.039 P=0.622	rs=-0.020 P=0.800	rs=0.019 P=0.807
Nitrates	rs=0.096 P=0.225	rs=0.220 P<0.05	**rs=0.333** P<0.05	rs=0.210 P<0.05	rs=0.129 P=0.104	**rs=0.217** P<0.05	rs=0.163 P<0.05	rs=0.189 P<0.05
Nitrites	rs=0.186 P<0.05	rs=0.423 P<0.05	rs=0.414 P<0.05	rs=0.348 P<0.05	rs=0.632 P<0.05	**rs=0.666** P<0.05	rs=0.617 P<0.05	**rs=0.648** P<0.05
Ammonia	rs=0.285 P<0.05	rs=0.482 P<0.05	rs=0.504 P<0.05	rs=0.407 P<0.05	rs=0.666 P<0.05	**rs=0.742** P<0.05	rs=0.659 P<0.05	**rs=0.720** P<0.05
Total *P*	rs=0.127 P=0.108	rs=0.383 P<0.05	rs=0.402 P<0.05	rs=0.318 P<0.05	rs=0.611 P<0.05	**rs=0.662** P<0.05	rs=0.552 P<0.05	**rs=0.644** P<0.05
Ri*P* zone	rs=0.039 P=0.622	rs=0.130 P=0.101	rs=0.279 P<0.05	rs=0.140 P=0.076	rs=0.310 P<0.05	**rs=0.384** P<0.05	**rs=0.404** P<0.05	rs=0.354 P<0.05
Sediments	rs=0.221 P<0.05	rs=0.377 P<0.05	rs=0.376 P<0.05	rs=0.323 P<0.05	rs=0.257 P<0.05	**rs=0.488** P<0.05	rs=0.346 P<0.05	**rs=0.480** P<0.05
Hyd reg	rs=0.195 P<0.05	rs=0.159 P<0.05	rs=0.233 P<0.05	rs=0.156 P<0.05	rs=0.361 P<0.05	**rs=0.405** P<0.05	rs=0.385 P<0.05	**rs=0.428** P<0.05
Acid e Tox	rs=0.139 P=0.080	rs=0.078 P=0.326	rs=0.174 P=0.082	rs=0.068 P=0.390	rs=0.102 P=0.200	**rs=0.183** P<0.05	rs=0.060 P=0.451	**rs=0.207** P<0.05
Mor*Ph* cond	rs=-0.036 P=0.655	rs=0.224 P<0.05	rs=0.268 P<0.05	rs=0.119 P=0.132	rs=0.455 P<0.05	**rs=0.586** P<0.05	**rs=0.571** P<0.05	rs=0.551 P<0.05
Org cont	rs=-0.016 P=0.842	rs=-0.004 P=0.956	rs=-0.094 P=0.236	rs=-0.013 P=0.866	rs=-0.023 P=0.772	rs=0.000 P=0.997	rs=-0.017 P=0.831	rs=-0.004 P=0.965
Connectivity	rs=0.020 P=0.798	rs=0.018 P=0.818	rs=0.111 P=0.163	rs=0.018 P=0.817	**rs=0.204** P<0.05	rs=0.194 P<0.05	**rs=0.197** P<0.05	**rs=0.197** P<0.05
HQA	rs=-0.116 P=0.143	rs=-0.295 P<0.05	rs=-0.261 P<0.05	**rs=-0.304** P<0.05	rs=-0.133 P=0.093	**rs=-0.293** P<0.05	rs=-0.240 P<0.05	rs=-0.259 P<0.05
HMS	rs=0.175 P<0.05	rs=0.349 P<0.05	rs=0.392 P<0.05	rs=0.251 P<0.05	rs=0.462 P<0.05	**rs=0.656** P<0.05	rs=0.616 P<0.05	**rs=0.638** P<0.05

The Box Plots analysis shows the evolution of the biological classifications (from high to bad) of all models with the environmental degradation (Figure 48 – Figure 53). Only some examples of the pressure variables significantly correlated with each predictive model are presented.

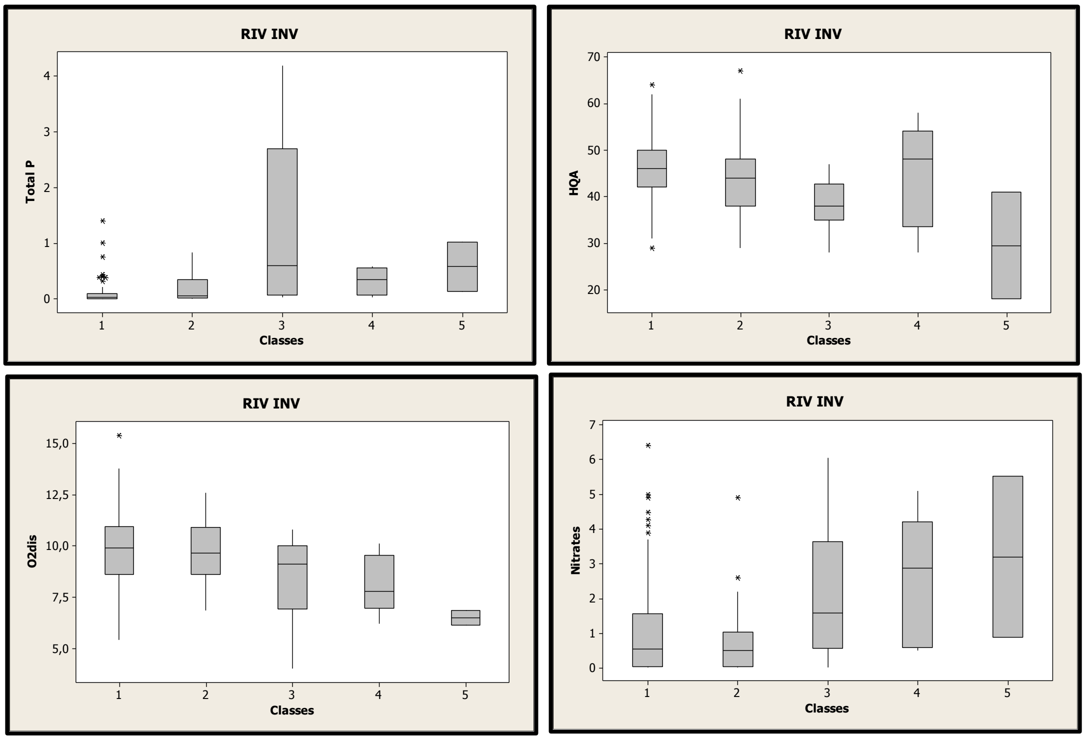

Figure 48 - Box plots representing the examples of clear relationship between increasing pressure level (for total phosphorous, HQA, dissolved O_2 and nitrates variables) and the classification attributed by RIV INV model.

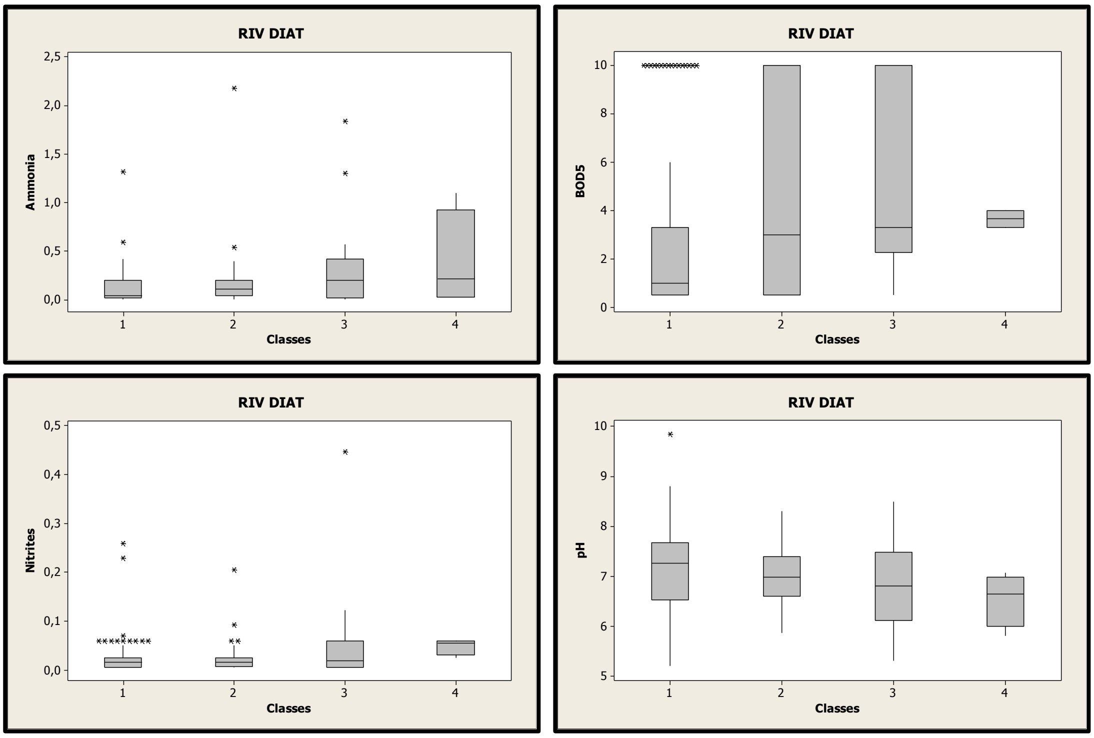

Figure 49 - Box plots representing the examples of clear relationship between increasing pressure level (for ammonia, BOD$_5$, nitrites and pH variables) and the classification attributed by RIV DIAT model.

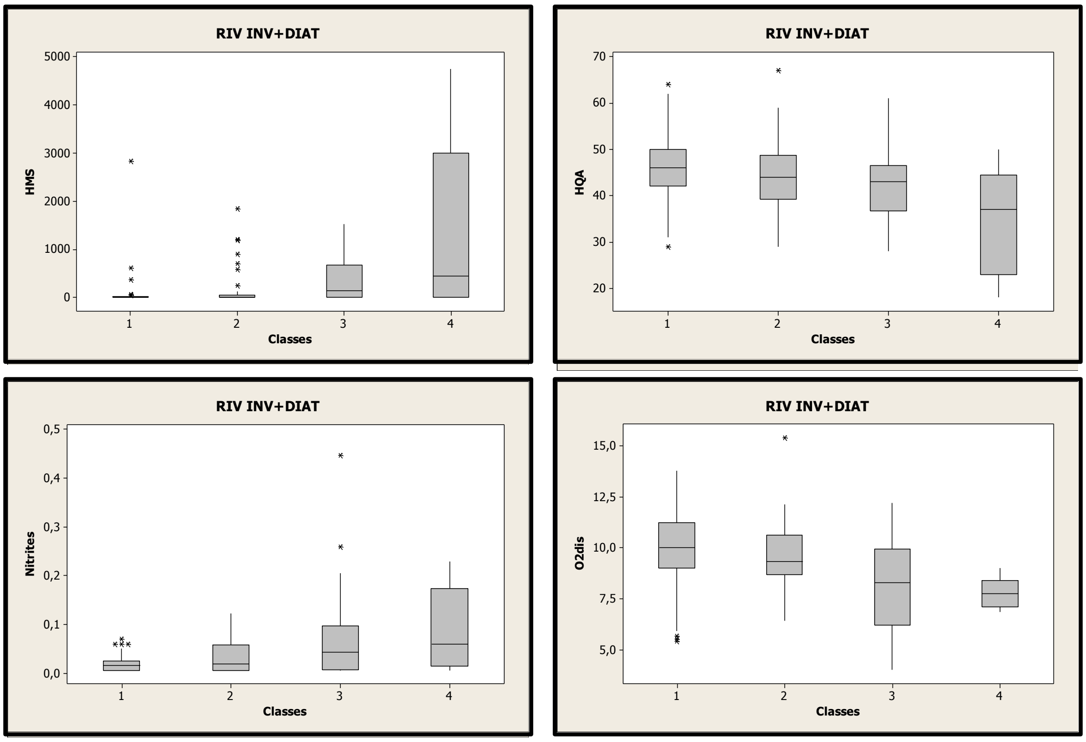

Figure 50 - Box plots representing the examples of clear relationship between increasing pressure level (for HMS, HQA, dissolved O$_2$ and nitrites variables) and the classification attributed by RIV INV+DIAT model.

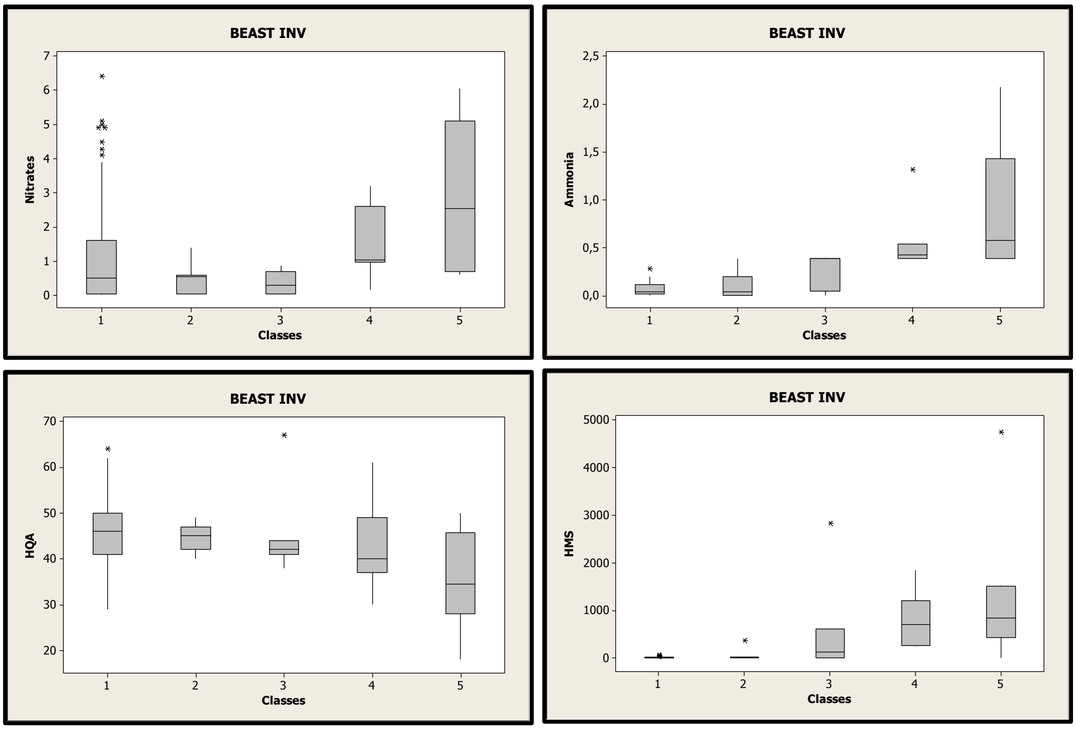

Figure 52 - Box plots representing the examples of clear relationship between increasing pressure level (for HQA, HMS, ammonia and nitrates variables) and the classification attributed by BEAST INV model.

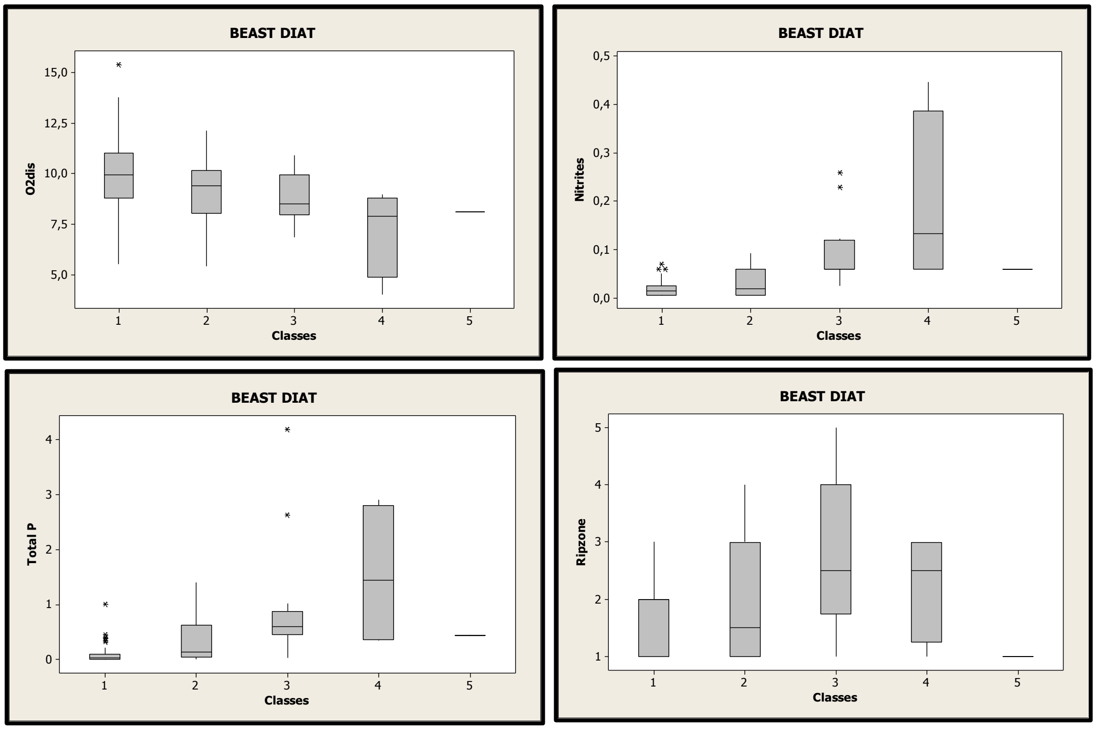

Figure 51 - Box plots representing the examples of clear relationship between increasing pressure level (for riparian zone, total phosphorous, dissolved O_2 and nitrites variables) and the classification attributed by BEAST DIAT model.

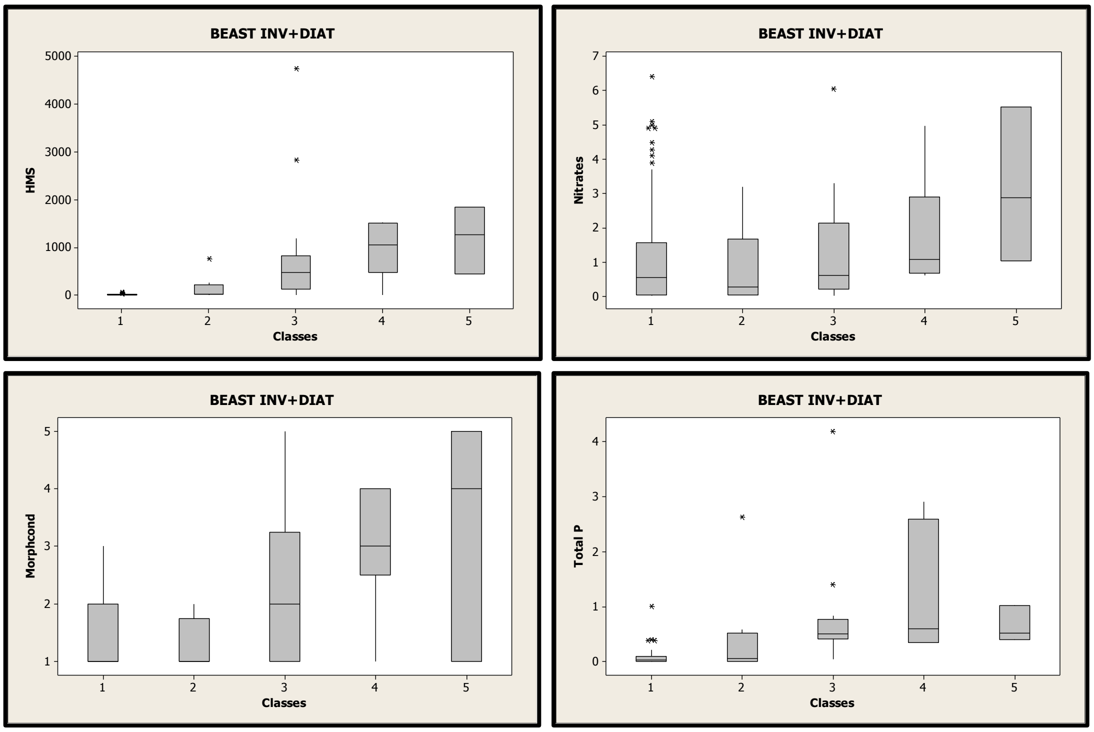

Figure 53 - Box plots representing the examples of clear relationship between increasing pressure level (for total phosphorous, morphological condition, HMS and nitrates variables) and the classification attributed by BEAST INV+DIAT model.

PCA axis 1 explains 45.5% of the total variation and translates, therefore, a general abiotic degradation gradient mainly influenced by the variables total phosphorous, ammonia and HMS (Figure 54). The reference and test sites formed two distinct and well defined groups. On the left side are the sites classified as reference and on the right side of the diagram the potentially impacted sites (test sites).

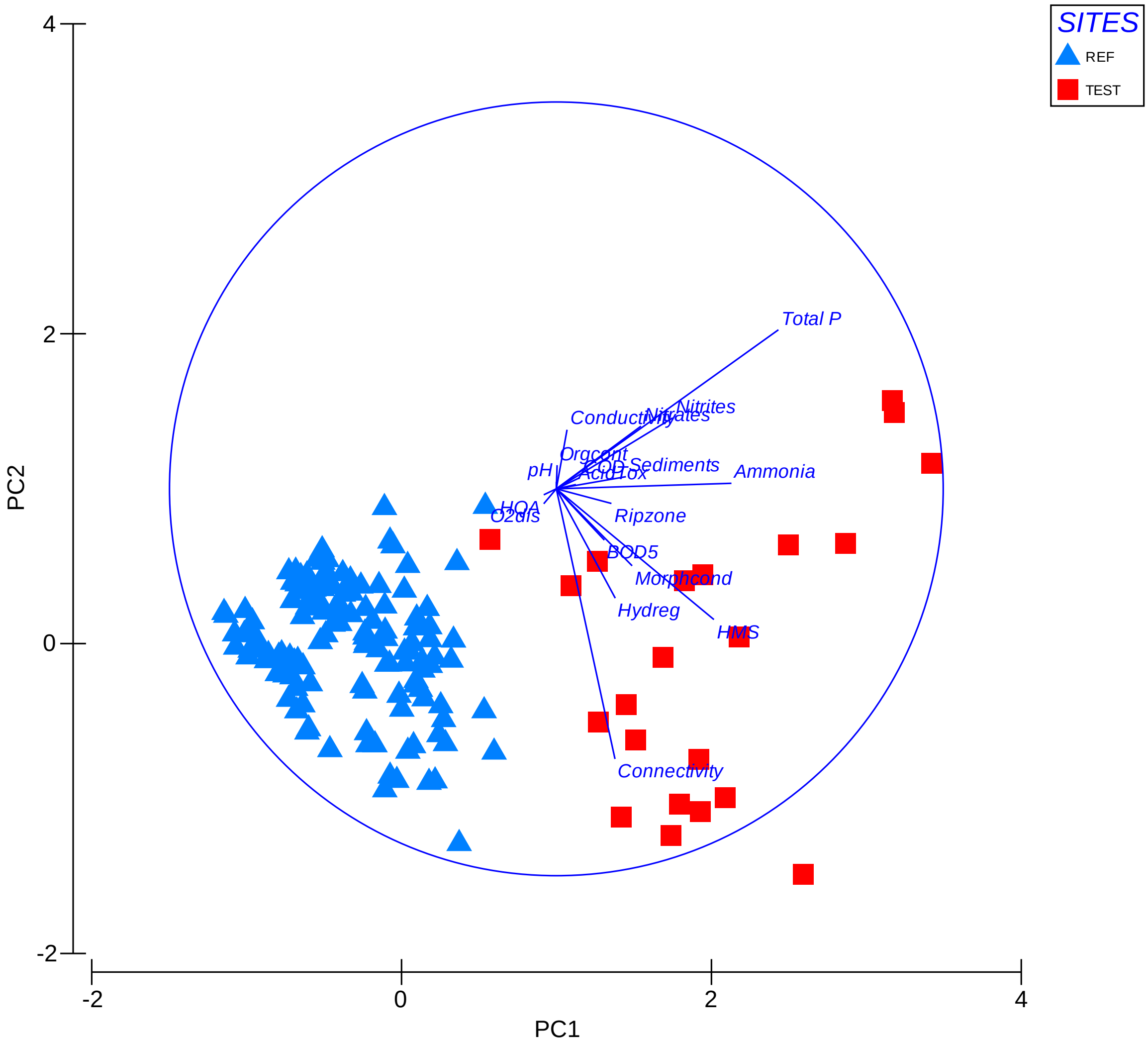

Figure 54 - Principal Component Analysis of all sites based on disturbance variables along axes 1 and 2.

Spearman rank correlation coefficient between the SCORE 1 of the PCA analysis and the models quality classes showed that the BEAST approach is more correlated to global degradation than the RIVPACS approach. The BEAST INV model responds better to general degradation and it’s followed by the BEAST combined model (BEAST INV+DIAT).

Table 32 - Spearman rank correlation (rs) and respective *P* value between quality classes of the three RIVPACS and BEAST predictive models and SCORE 1 of PCA analysis. The red rectangles signed the highest and the gray ones the second highest significant correlations between methods.

	RIV DIAT	RIV INV	RIV INV+DIAT	BEAST DIAT	BEAST INV	BEAST INV+DIAT
SCORE 1	rs=0.257 *P*<0.05	rs=0.465 *P*<0.05	rs=0.534 *P*<0.05	rs=0.737 *P*<0.05	rs=0.835 *P*<0.05	rs=0.766 *P*<0.05

Chapter 5 – DISCUSSION and CONCLUSIONS

The predictor variables selected for the RIV DIAT model, such as alkalinity and hardness, imply that diatoms are strongly influenced by water chemistry. The influence of chemistry on diatom distribution is well known and has been reported in other studies (Passy et al. 2004, Soininen and Könönen 2004, Feio et al. 2007, Carlisle et al. 2008, Almeida et al. 2012). According to Stevenson (1997), climate and geology are also determinant environmental variables which affect the spatial distribution of benthic algae. In RIV INV predictive model, the predictor variables selected were similar to those selected in RIV DIAT. This was also previously described by Carlisle et al. (2008), where both predictive models based on diatoms and invertebrates used climatic and physical/chemical variables. The importance of variables such as latitude, longitude and altitude implies that temperature is a determinant factor of invertebrates composition which is in accordance with Hawkins et al. (2000). Surprisingly, hardness was also one of the predictors selected. Variables related with water chemistry are usually associated to diatoms as referred above. Nevertheless, Hawkins et al. (2000) already suggested that the ionic composition of water is an important determinant of biotic structure. As expected, the discriminant variables selected by RIV INV+DIAT appear to be a combination of those described for both communities, except for mean annual temperature.

The discriminant variables selected for the BEAST DIAT and BEAST INV were almost the same selected in RIV DIAT and RIV INV, but in the last ones more variables were used. The only exception was the variable alkalinity which was selected by BEAST INV model, but not by RIV INV. According to Egglishaw (1968), this variable can be important for invertebrates, as waters with high alkalinity can support more invertebrates by inducing a quicker turnover of the organic matter. In BEAST INV+DIAT predictive model, the predictor variables selected were similar to those selected in RIV INV+DIAT. As in RIV INV+DIAT, the discriminant variables selected by BEAST INV+DIAT, appear to be a combination of those described for both communities.

The three RIVPACS type models (RIV DIAT, RIV INV and RIN INV+DIAT) were considered good models regarding the Observed/Expected regression, which is used to evaluate the accuracy of predictive models (Linke et al. 2005). The precision of RIV DIAT model (SDOEcal=0.242) was better than the null model (SDOEcal null model=0.315) but the difference between predictive model SD and the sampling error (SDRScal=0.195) suggests that there is some potential for further improvement of the model. This model presents also the highest value of RMSE which also means that it requires some model improvement. In fact, the RIV DIAT model didn't perform well both in the internal (sites initially set aside) and in the external validation. Two reference sites (M109 and M112; i.e. good quality) used as test sites were evaluated as moderate quality by RIV DIAT model. Looking at the community composition there was a dominance of ubiquitous sensitive species such as *Achnanthidium minutissimum*, *Cocconeis pseudolineata*, *Fragilaria bidens*, *Stauroneis thermicola* and a few typical species of clean waters were present (*Gomphonema rhombicum*, *Nitzschia dissipata var. dissipata*), which indicates that this site is of good quality. However, only about 30% and 50% of the expected taxa were observed for M109 (O/E=0.582) and M112 (O/E=0.613) test sites, respectively. The RIV DIAT model attributed an average OE value of 0.81 to the validation reference sites. In comparison, the RIV INV model performed better on the validation reference sites with an OE average of 1.02 meaning that the observed taxa are very close to the expected taxa. Additionally, the difference between the SD of the null model and the predictive model (0.054) is smaller, hence this model accounts for slight variability in O/E across reference sites. The difference between predictive model SD and the sampling error (0.077) is also small suggesting slight potential for improving the model. Regarding the external validation, this model shows better performance than the other two, evaluating correctly most of the reference sites used as test sites. Concerning the validation reference sites, the RIV INV+DIAT model didn't perform so well, with an OE average of 0.86 and about 30% of the validation reference sites not achieving at least the good quality status. However, this model could be considered the best of the three because it presents the lowest values of the standard deviation (SDOEnull model=0.207; SDOEcal=0.164; SDRScal=0.102).

The three BEAST models showed good performance regarding the assessment of the validation sites. BEAST INV and BEAST INV+DIAT models classified only about 15% of the sites above the good status while less than 10% of the validation sites were classified above good status in BEAST DIAT model. The percentage of reference sites correctly attributed to their respective reference group was acceptable for the BEAST DIAT (65.1% and 58.9%) and good for BEAST INV (73.6% and 62%) and for BEAST INV+DIAT (72.1% and 66.7%). Other studies also find, after the discriminant analysis, lower classifications for diatoms than for invertebrates (Chessman et al. 1999, Mazor et al. 2006). The variables used as predictors are usually more related to macroinvertebrates resulting in a higher accuracy of BEAST INV model. Further works should re-think those variables by including, for example, others more relevant for diatoms. This hypothesis was already reffered by Chessman et al. (1999), who additionally proposed that possibly the short life cycles of diatoms make them intrinsically less predictable when comparing to macroinvertebrates. The BEAST INV and BEAST INV+DIAT models were very similar, showing better performance than BEAST DIAT model.

The models were tested with a set of sites with different levels of degradation. The RIV DIAT model responded well to degradation such as pH, BOD_5, nitrites, ammonia, sediments discharge, hydrological regime and HMS while BEAST DIAT responded well to dissolved O_2, nitrites, ammonia, total phosphorous, riparian zone, sediment discharge, hydrological regime, morphological condition, connectivity and HMS variables. The fact that diatoms respond well to alterations in nutrients is well described in the literature (Growns 1999, Almeida and Gil 2001, Potapova and Charles 2005, Tison et al. 2007, Tornés et al. 2007). The RIV DIAT was also sensitive to HMS (Habitat Modification Score) which is related with the human intervention through the presence of weirs, banks and channel modifications. Johnson and Hering (2009) already reported that diatom communities, in lowland streams, show a strong response to habitat degradation. The response of diatom assemblages to morphological alterations such as sediments discharge, connectivity was also found by other authors (Soininen 2004, Feio et al. 2009a, Almeida and Feio 2012) which can lead to changes in flow regimes and current velocity. The RIV INV model responded well to changes in dissolved O_2, conductivity, nitrates, nitrites, ammonia, total phosphorous, sediments discharge, hydrological regime, morphological condition, HQA and HMS while the BEAST INV model was significantly correlated with dissolved O_2, nitrates, nitrites, ammonia, riparian zone, total phosphorous, sediment discharge, hydrological regime, acidification and toxicity, morphological condition, connectivity, HQA and HMS. The macroinvertebrate assemblages detected modifications in water chemistry but also on habitat impairment as reported previously by Hawkins et al. (2000) and Feio et al. (2006b). Finally, the combined model RIV INV+DIAT showed to be the most sensitive to the environmental degradation such as: dissolved O_2, BOD_5, nitrates, nitrites, ammonia, total phosphorous, riparian zone, sediments discharge, hydrological regime, morphological condition, HQA and HMS while the combined BEAST INV+DIAT was correlated with dissolved O_2, nitrates, nitrites, ammonia, total phosphorous, riparian zone, sediment discharge, hydrological regime, morphological condition, connectivity, HQA and HMS. Once again, the models based on the two communities seem to combine both sensitivity of diatoms (e.g. nitrites, ammonia) and macroinvertebrates (e.g. HQA, HMS) to the pressure variables. This suggests that the model does not mask the responses of the two communities, but combines them instead. However, in the BEAST models case, the BEAST INV model responds better to the environmental degradation than the BEAST INV+DIAT, being significantly correlated with a greater number of environmental variables. Both diatom models, RIV DIAT and BEAST DIAT showed to be less sensitive to environmental degradation than the invertebrate models. This fact can be associated to misidentification and some taxonomic problems related with some species which are difficult to distinguish on the light microscope. , This problem will not be solved based only on a morphological and structural analysis of the cell wall of diatoms. Mann et al. 2010 suggests that molecular methods such as DNA barcoding, could resolve most of the taxa issues.

Most of the box plots showed a continuous increment in the concentration of nutrients and habitat degradation with the increase of degradation classes meaning that those variables are well reflected in the classifications of both methods. Other variables showed an unclear pattern which means that they also contribute to the environmental degradation but not in a continuous way.

According to WFD, the ecological quality status of rivers should result of the assessments based on several biological elements (aquatic flora, fish and invertebrates).

In fact, the use of multiple biological assemblages should provide a holistic perspective of ecosystem's health, while covering different structural elements and trophic levels. In a conservative way, it has been interpreted under the context of the WFD that the final classification of the ecological status corresponds to that obtained for the biological element with the worst result. In this study, the a priori combination using the RIVPACS approach (RIV INV+DIAT) gave the same result as the *a posteriori* combination by the worse result for most of the test sites (≈70%). However, the disagreements result, that according to the Spearman rank correlation coefficients, the combined method (RIV INV+DIAT) responds better to the environmental degradation than the "one-out all-out" *a posteriori*. On the contrary, the opposite happened with the BEAST approach: with the combined model (BEAST INV+DIAT) only 40% of the test sites obtained the same final quality status than the one-out, all-out approach and the latter st was shown to give better responses to environmental degradation (Spearman rank correlations).

So, should multiple biological elements' assessments be integrated into a single method resulting directly in a single value, or should they be applied independently and then combined into a single value? If we can use a single method, which approach (RIVPACS or BEAST) should we choose?

First of all, our results indicate that the type of data, presence/absence or relative abundance is probably influencing the results, even though there are also statistical differences between the two types of models used (RIVPACS/BEAST). Most of the diatom assessment tools are based on abundance such as the indices (IPS) but also some predictive models (BEAST), nevertheless, the use of presence/absence data is not usual. Regarding the invertebrates, the use of that type of data is more common (predictive models – RIVPCAS; indices - IBMWP) and those assessment tools have been successfully used worldwide (Wright et al. 1993, Abbasi and Abbasi 2011). Though, studies using only diatom binary data work well (Feio et al. 2009b, Almeida et al. 2012), when assessment methods are compared using both types of data, the abundance works better (Chessman et al. 1999, Mendes et al. 2012).

The evaluation of the test sites by the two approaches was different and in general the BEAST models attributed more severe classifications. About 50% of the test sites were classified by BEAST approach at least to classes above when comparing to RIVPACS one, for both individual models. In the invertebrate models, these differences can be explained by the great abundance of taxa belonging to groups characteristic of polluted water such as Gastropoda, Chironomidae, Oligochaeta, Turbellaria and Hirudinae over the presence of the more sensitive taxa. For example, the L120 test site was classified with moderate quality class by RIV INV and as bad quality class by BEAST INV. This site has a total of about 3500 individuals in which more than half belong to the Oligochaeta family, more specifically Tubificidae organisms that are known to inhabit poor oxygenated and rich in organic matter waters and be the last to disappear from contaminated waters (Mosleh et al. 2006). Indeed, the major impact of the L120 test site is a pig farm. However, the presence of other sensitive taxa belonging to Ephemeroptera and Trichoptera and few taxa of more tolerant groups justifies the quality class attributed by the RIV INV model. Regarding the diatom models, the differences in the assessment of most of test sites can be explained the same way. The RIVPACS approach attributed more severe classifications than BEAST only to a few test sites. In these cases, the dominance of taxa was not so evident but more typical taxa of clean waters were noted. For example, the M122 test site, in spite of the presence of some sensitive taxa such as *Achnanthidium minutissimum*, *Eunotia exigua*, *Encyonema minutum*, *Karayevia oblongella*, *Nitzschia dissipata var. dissipata*, *Nitzschia hantzschiana*, *Meridion circulare* var. *circulare*, *Navicula cryptocephala*, *Navicula tenelloides* and *Pinnularia subcapitata* var. *subcapitata*, it was classified with poor quality class by RIV DIAT. However, BEAST DIAT model classified this site with good quality class. The unconformity in the assessment can be explained by the dominance in abundance of those sensitive taxa which in fact account for about 1/3 of the total number of individuals of the site. Nevertheless, it’s difficult to conclude that one assessment method is better than the other because both rely on different assumptions. Mazor et al. (2006) found that RIVPACS is more sensitive to species loss while BEAST is more sensitive to changes in the structure of communities without loss of common taxa. Although BEAST INV+DIAT and RIV INV+DIAT showed significant correlation with the same number of pressure variables, in general, BEAST INV+DIAT showed stronger association with those. In other words, the BEAST INV+DIAT model was here more sensitive to environmental degradation. Chessman et al. 1999 also found in his work higher correlations for predictions based on abundance against the binary (presence/absence data). Though, the RIVPACS approach is less sensitive than the BEAST, it seems to work better when we join the two biological elements, and better than the individual methods and better than the approach proposed by WFD. Answering our questions, we can integrate reliably the assessments of diatoms and macroinvertebrates into a single method without loss of information, if we use presence/absence type of data, and for that we recommend the combined model RIV INV+DIAT. But if we use abundance type of data (BEAST model), the “one-out all-out” *a posteriori* approach is recommended, or in other words, the assessment should be done individually for both diatom and invertebrate communities and combined *a posteriori*.

Beyond the aim of this study, its not possible to ignore that, the best of all methods in terms of response to environmental degradation was the BEAST model with invertebrates only, which raises back the question of the need for an additional biological element and on the other hand shows the importance of the use of abundance data in biomonitoring with this biological element, as demanded by the WFD. However, diatoms and macroinvertebrates showed to provide complementary information, besides that, the position of diatoms in the foodchain (primary producers) and their short life cycle allows them to detect and respond to rapid environmental changes that macroinvertebrates can’t.

REFERENCES

Abbasi T, Abbasi SA. 2011. Water quality indices based on bioassessment: The biotic indices. Journal of Water & Health 9: 330-348.

Aboal M, Álvarez-Cobelas M, Cambra J, Ector L. 2003. Diatom Monographs: Floristic list of the non marine diatoms (Bacillariophyceae) of Iberian Peninsula, Balearic Islands and Canary Islands. Ruggell: A.R.G. Gantner Verlag K.G.

Aguiar FC, Feio MJ, Ferreira MT. 2011. Choosing the best method for stream bioassessment using macrophyte communities: Indices and predictive models. Ecological Indicators 11: 379-388.

Alba-Tercedor J, Sánchez-Ortega A. 1988. Un Método Rápido y Simple para Evaluar la Calidad Biológica de las Aguas Corrientes Basado en el de Hellawel (1978). Limnética 4: 51-56.

Alba-Tercedor J, et al. 2002. Caracterización del estado ecológico de ríos mediterráneos ibéricos mediante el índice IBMWP (antes BMWP'). Limnética 21(3-4): 175-185.

Almeida SFP, Gil MCP. 2001. Ecology of freshwater diatoms from the central region of Prtugal. Criptogamie, Algologie 22: 109-126.

Almeida SFP, Feio MJ. 2012. DIATMOD: diatom predictive model for quality assessment of Portuguese running waters. Hydrobiologia.

Andrén C, Jarlman A. 2008. Benthic diatoms as indicators of acidity in streams. Fundamental and Applied Limnology 173: 237-253.

Atazadeh I, Sharifi M, Kelly M. 2007. Evaluation of the Trophic Diatom Index for assessing water quality in River Gharasou, western Iran. Hydrobiologia 589: 165-173.

Azrina MZ, Yap CK, Rahim Ismail A, Ismail A, Tan SG. 2006. Anthropogenic impacts on the distribution and biodiversity of benthic macroinvertebrates and water quality of the Langat River, Peninsular Malaysia. Ecotoxicology and Environmental Safety 64: 337-347.

Barbosa FAR, Callisto M, Galdean N. 2001. The diversity of benthic macroinvertebrates as an indicator of water quality and ecosystem health: a case study for Brazil. Aquatic Ecosystem Health & Management 4: 51-59.

Barbour M, Gerritsen J, Snyder B, Stribling J. 1999. Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish. Washington, D.C.: U.S. Environmental Protection Agency; Office of Water. Report no.

Bellinger B, Cocquyt C, O'Reilly C. 2006. Benthic diatoms as indicators of eutrophication in tropical streams. Hydrobiologia 573: 75-87.

Bere T, Tundisi J. 2011a. Applicability of borrowed diatom-based water quality assessment indices in streams around São Carlos-SP, Brazil. Hydrobiologia 673: 179-192.

Besse-Lototskaya A, Verdonschot PFM, Coste M, Van de Vijver B. 2011. Evaluation of European diatom trophic indices. Ecological Indicators 11: 456-467.

Bieger L, Carvalho A, Strieder M, Maltchik L, Stenert C. 2010. Are the streams of the Sinos River basin of good water quality? Aquatic macroinvertebrates may answer the question. Braz. J. Biol. 70(4): 1207-1215.

Bold HC, Wynne MJ. 1985. Introduction to the Algae. New Jersey: Prentice-Hall, Inc.

Butcher R. 1947. Studies in the Ecology of Rivers: VII. The Algae of Organically Enriched Waters. Journal of Ecology 35(1-2): 186-191.

Cairns Jr. J, Pratt JR. 1993. A History of Biological Monitoring Using Benthic Macroinvertebrates. Pages 10-27 in Rosenberg DM, Resh VH, eds. Freshwater biomonitoring and benthic macroinvertebrates. New York: Chapman & Hall.

Carlisle DM, Hawkins CP, Meador MR, Potapova M, Falcone J. 2008. Biological assessments of Appalachian streams based on predictive models for fish, macroinvertebrate, and diatom assemblages. Journal of the North American Benthological Society 27: 16-37.

Charles DF. 1996. Use of algae for monitoring rivers in the United States. Some examples. Pages 109-118 in Whitton B, Rott E, eds. Use of Algae for Monitoring Rivers, vol. I.

Chessman B, Growns I, Curreys J, Plunkett-Cole N. 1999. The influence of taxonomic level on the per-formance of a predictive model for water quality assessment. Freshwater Biology 41: 317-331.

Chesters R. 1980. Biological Monitoring Working Party. The 1978 National Testing Excercise, Technical Memorandum 19. Water Data Unit. Reading. United Kingdom.

Clarke RT, Murphy JF. 2006. Effects of locally rare ta xa on the precision and sensitivity of RIVPACS bioassessment of freshwaters. Freshwater Biology 51(10): 1924-1940.

Clarke RT, Wright JF, Furse MT. 2003. RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. Ecological Modelling 160: 219-233.

Cooper SR. 1995. Chesapeake Bay Watershed Historical Land Use: Impact on Water Quality and Diatom Communities. Ecological Applications 5: 703-723.

Cummins K. 1992. Invertebrates. Pages 234-250 in Calow P, Petts GE, eds. The Rivers Handbook: Hydrological and Ecological Principles, vol. 1. Oxford: Blackwell Science.

Czerniawska-Kusza I. 2005. Comparing modified biological monitoring working party score system and several biological indices based on macroinvertebrates for water-quality assessment. Limnologica - Ecology and Management of Inland Waters 35: 169-176.

Davis WS. 1995. Biological Assessment and Criteria: Building on the Past. in Davis WS, Simon TP, eds. Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making. Boca Raton: Lewis Publishers.

De Paw N, Vanhooren G. 1983. Method for biological quality assessment of watercourses in Belgium. Hydrobiologia 100: 153-168.

Descy J-P. 1979. A new approach to water quality estimation using diatoms. Nova Hedwigia 119: 119-128.

EA. 2003. River Habitat Survey in Britain and Ireland: Field Survey Guidance Manual. River Habitat Survey: version 2003. Environmental Agency, U.K., 136 pp.

EC Parliament and Council. 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.

Egglishaw HJ. 1968. The quantitative relationship between bottom fauna and plant detritus in streams of different calcium concentrations. Journal of Applied Ecology 5: 731-740.

Feio MJ. 2004. Macroinvertebrates in the Mondego river basin bioassessment.University of Coimbra.

Feio MJ, Reynoldson TB, Graça MA. 2006a. Effect of Seasonal Changes on Predictive Model Assessments of Streams Water Quality with Macroinvertebrates. International Review of Hydrobiology 91: 509-520.

Feio MJ, Reynoldson TB, Gra, a MA. 2006b. The influence of taxonomic level on the performance of a predictive model for water quality assessment. Canadian Journal of Fisheries and Aquatic Sciences 63: 367-376.

Feio MJ, Almeida SFP, Craveiro SC, Calado AJ. 2007. Diatoms and macroinvertebrates provide consistent and complementary information on environmental quality. Fundamental and Applied Limnology / Archiv für Hydrobiologie 169: 247-258.

Feio MJ, Norris RH, Graça MAS, Nichols S. 2009a. Water quality assessment of Portuguese streams: Regional or national predictive models? Ecological Indicators 9: 791-806.

Feio MJ, Almeida SFP, Craveiro SC, Calado AJ. 2009b. A comparison between biotic indices and predictive models in stream water quality assessment based on benthic diatom communities. Ecological Indicators 9: 497-507.

Feio MJ, Alves T, Boavida M, Medeiros A, Graça MAS. 2010. Functional indicators of stream health: a river-basin approach. Freshwater Biology 55: 1050-1065.

Feio MJ, Poquet JM. 2011. Predictive Models for Freshwater Biological Assessment: Statistical Approaches, Biological Elements and the Iberian Peninsula Experience: A Review. International Review of Hydrobiology 96: 321-346.

Figueroa R, Valdovinos C, Araya E, Parra O. 2003. Macroinvertebrados bentónicos como indicadores de calidad de agua de ríos del sur de Chile. Revista Chilena de Historia Natural 76: 275-285.

Growns I. 1999. Is genus or species identification of periphytic diatoms required to determine the impacts of river regulation? J. Appl. Phycol. 11: 273-283.

Harding W, Archibald C, Taylor J. 2005. The relevance of diatoms for water quality assessment in South Africa: A position paper. Water SA 31: 41-46.

Hawkins CP, Norris RH, Hogue JN, Feminella JW. 2000. Development and Evaluation of Predictive Models for Measuring the Biological Integrity of Streams. Ecological Applications 10(5): 1456-1477.

Hill MO. 1979. TWISPAN – a FORTRAN program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes. Ecology and Systematics: Cornell Univ., Ithaca, New York.

INAG IP. 2008a. Protocolo de amostragem e análise para o FITOBENTOS - DIATOMÁCEAS. Ministério do Ambiente, do Ordenamento do Território e do Desenvolvimento Regional. Institudo da Água, I.P.

INAG IP. 2008b. Protocolo de amostragem e análise para os MACROINVERTEBRADOS BENTÓNICOS. Ministério do Ambiente, do Ordenamento do Território e do Desenvolvimento Regional. Instituto da Água, I.P.

INAG IP. 2009. Critérios para a Classificação do Estado das Massas de Água Superficiais: Rios e Albufeiras. Ministério do Ambiente, do Ordenamento do Território e do Desenvolvimento Regional. Instituto da Água, I. P.

Johnson RK, Hering D. 2009. Response of taxonomic groups in streams to gradients in resource and habitat characteristics. J. Appl. Phycol. 46: 175-186.

Jones V. 2007. DIATOM INTRODUCTION. Pages 476-484 in Editor-in-Chief: Scott AE, ed. Encyclopedia of Quaternary Science. Oxford: Elsevier.

Kelly MG, Whitton B. 1995. The Trophic Diatom Index: a new index for monitoring eutrophication in rivers. Journal of Applied Phycology 7: 433-444.

Kennard MJ, Pusey BJ, Arthington AH, Harch BD, Mackay SJ. 2006. Development and Application of a Predictive Model of Freshwater Fish Assemblage Composition to Evaluate River Health in Eastern Australia. Hydrobiologia 572: 33-57.

Kenney MA, Sutton-Grier AE, Smith RF, Gresens SE. 2009. Benthic macroinvertebrates as indicators of water quality: The intersection of science and policy. Terrestrial Arthropod Reviews 2: 99-128.

Kireta AR, Reavie ED, Sgro GV, Angradi TR, Bolgrien DW, Hill BH, Jicha TM. 2012. Planktonic and periphytic diatoms as indicators of stress on great rivers of the United States: Testing water quality and disturbance models. Ecological Indicators 13: 222-231.

Kolkwitz R, Marsson M. 1908. Ökologie der pflanzlichen Saprobien. Berichte der Deutschen Botanischen Gesellschaft 26A: 505-519.

Kolkwitz R, Marsson M. 1909. Ökologie der tierischen Saprobien. Beiträge zur Lehre von des biologischen Gewasserbeurteilung. Internationale Revue der Gesamten Hydrobiologie und Hydrographie 2: 126-152.

Krammer K. 2000. Diatoms of Europe: Diatoms of the European Inland Waters and Comparable Habitats. Ruggell: A.R.G. Gantner Verlag K.G.

Krammer K. 2002. Diatoms of Europe: Diatoms of the European Inland Waters and Comparable Habitats. Ruggell: A.R.G. Gantner Verlag K.G.

Krammer K. 2003. Diatoms of Europe: Diatoms of the European Inland Waters and Comparable Habitats. Ruggell: A.R.G. Gantner Verlag K.G.

Krammer K, Lange-Bertalot H. 1986. Süßwasserflora von Mitteleuropa. Baccillariophyceae 1 Teil: Naviculaceae, Band 2/1. Stuttgart: Gustav Fischer Verlag.

Krammer K, Lange-Bertalot H. . 1988. Süßwasserflora von Mitteleuropa. Baccillariophyceae 2 Teil: Bacillariaceae, Epithemiaceae, Surirellaceae, Band 2/2. . Stuttgart: Gustav Fischer Verlag.

Krammer K, Lange-Bertalot H. . 1991a. Süßwasserflora von Mitteleuropa. Baccillariophyceae 3 Teil: Centrales, Fragilariaceae, Eunotiaceae, Band 2/3. Stuttgart: Gustav Fischer Verlag.

Krammer K, Lange-Bertalot H. . 1991b. Süßwasserflora von Mitteleuropa. Baccillariophyceae 4 Teil: Achnanthaceae, Kritische Ergänzungen zu Navicula (Lineolatae) und Gomphonema. Band 2/4. Stuttgart: Gustav Fischer Verlag.

Lange-Bertalot H. 2001. Diatoms of Europe: Diatoms of the European Inland Waters and Comparable Habitats. Ruggell: A.R.G Gantner Verlag K.G.

Lange-Bertalot H, Cavacini P, Tagliaventi N, Alfinito S. 2003. Diatoms of Sardinia: Rare and 76 new species in rock pools and other ephemeral waters. Ruggell: A.R.G. Gantner Verlag K.G.

Lee RE. 1980. Phycology. New York: Cambride University Press.

Lenoir A, Coste M. 1996. Development of a practical diatom index of overall water quality applicable to the French national water Board network. in Whitton B, Rott E, eds. Use of Algae for Monitoring Rivers II. Innsbruck: Rott E, Publisher.

Levkov Z. 2009. Diatoms of Europe: Diatoms of the European Inland Waters and Comparable Habitats. Liechtenstein: A.R.G. Gantner Verlag K.G.

Linke S, Norris RH, Faith DP, Stockwell D. 2005. ANNA: A new prediction method for bioassessment programs. Freshwater Biology 50: 147-158.

Lowe RL, Pan Y. 1996. Benthic algal comunnities and biological monitors. Pages 705-739 in Stevenson RJ, Bothwell M, Lowe RL, eds. Algal Ecology: Freshwater Benthic Ecosystems. San Diego, CA: Academic Press.

Mann DG, Sato S, Trobajo R, Vanormelingen P, Souffreau C. 2010. DNA barconding for species identification and discovery in diatoms. Criptogamie, Algologie 31(4): 557-577.

Marchant R, Hirst A, Norris RH, Butcher R, Metzeling L, Tiller D. 1997. Classifica-tion and prediction of macroinvertebrate assemblages from running waters in Victoria, Australia. J. N. Am. Benthol. Soc. 16: 664-681.

Matthwes R, Buikema A, Cairns Jr. J. 1982. Biological monitoring part IIA: Receiving system functional methods relationships, and indices. Water Research 16: 129-139.

Mazor RD, Reynoldson TB, Rosenberg DM, Resh VH. 2006. Effects of biotic assemblage, classification, and assessment method on bioassessment performance. Canadian Journal of Fisheries and Aquatic Sciences 63: 394-411.

Mendes T, Almeida SFP, Feio MJ. 2012. Assessment of rivers using diatoms: effect of substrate and evaluation method. Fundamental and Applied Limnology / Archiv für Hydrobiologie 179: 267-279.

Milner AM, Oswood MW. 2000. Urbanization gradients in streams of Anchorage, Alaska: a comparison of multivariate and multimetric approaches to classification. Hydrobiologia 422-423: 209-223.

Mosleh YY, Paris-Palacios S, Biagianti-Risbourg S. 2006. Metallothioneins induction and antioxidative response in aquatic worms Tubifex tubifex (Oligochaeta, Tubificidae) exposed to copper. Chemosphere 64: 121-128.

Muralidharan M, Selvakumar C, Sundar S, Raja M. 2010. Macroinvertebrates as Potential Indicators of Environmental Quality. IJBT 1: 23-28.

Passy SI, Bode RW, Carlson DM, Novak MA. 2004. Comparative Environmental Assessment in the Studies of Benthic Diatom, Macroinvertebrate, and Fish Communities. International Review of Hydrobiology 89: 121-138.

Patrick R. 1949. A proposed biological measure of stream conditions based on a survey of the Conestoga Basin, Lancaster County, Pennsylvania. Proceedings of the Academy of Natural Sciences of Philadelphia 101: 277-341.

Patrick R. . 1973. Use of algae, especially diatoms, in assessment of water quality. ASTM STP 528: 76-95.

Patrick R, Strawbridge D. 1963. Variation in the structure of natural diatom communities. The American Naturalist 97: 51-57.

Patrick R, Hohn MH, Wallace JH. 1954. A new method for determining the patter of the diatom flora. Notulae Naturae 259: 1-12.

PBH. 1999a. Plano de Bacia Hidrográfica do Rio Mondego: Síntese da análise e diagnóstico da situação actual. Sumário Executivo I: 1-182.

PBH. 1999b. Plano de Bacia Hidrográfica do Rio Vouga: Síntese da análise e diagnóstico da situação actual. Sumário Executivo I: 1-204.

PBH. 1999c. Plano de Bacia Hidrográfica do Rio Lis: Síntese da análise e diagnóstico da situação actual. Sumário Executivo I: 1-203.

Perry J, Vanderklein E. 1996. Water Quality: Management of a Natural Resource. United States of America: Blackwell Science.

Potapova M, Charles DF. 2005. Choice of substrate in algae-based water-quality assessment. Journal of the North American Benthological Society 24: 415-427.

Prygiel J, Coste M. 2000. Guide Méthodologique pour la mise em ouvre del'Indice Biologique Diatomées. Agences de L'eau, Ministère de L'Aménagement du Territoire et de L'Environnement i Cemagref, Bordeus.

Reynoldson TB, Bailey RC, Day KE, Rosenberg DM. 1995. Biological guidelines for freshwater sediment based on BEnthic Assessment of SedimenT (the BEAST) using a multivariate approach for predicting biological state. Austr. J. Ecol. 20: 198-219.

Reynoldson TB, Norris RH, Resh VH, Day KE, Rosenberg DM. 1997. The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. J. N. Am. Benthol. Soc. 16(4): 833-852.

Rimet F, Cauchie H-M, Hoffmann L, Ector L. 2005. Response of diatom indices to simulated water quality improvements in a river. Journal of Applied Phycology 17: 119-128.

Rosen BH. 1995. Use of periphyton in the development of biocriteria. Pages 209-215 in Davis WS, Simon TP, eds. Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making. Boca Raton, FL: Lewis Publishers.

Rosenberg DM, Resh VH. 1993. Introduction to Freshwater Biomonitoring and Benthic Macroinvertebrates. Pages 1-9 in Rosenberg DM, Resh VH, eds. Freshwater Biomonitoring and Benthic Macroinvertebrates. New York: Chapman & Hall.

Rott E, Hofmann G, Pall K, Pfister P, Pipp E. 1997. Indikationslisten für Aufwuchsalgen in österreichischen Fliessgewässern. Teil 1: Saprobielle Indikation. Wien, Austria: Bundesministerium für Land- und Forstwirtschaft, Wasserwirtschaftskataster.

Simpson JC, Norris RH. 2000. Biological assessment of river quality: development of AUS-RIVAS models and outputs. Pages 125-142 in Wright JF, Sutcliffe DW, Furse MT, eds. Assess-ing the biological quality of fresh waters: RIVPACS and other techniques. Ambleside. UK: Freshwater Biological Association.

Slàdecek V. 1973. System of water quality from the biological point of view. Archiv für Hydrobiologie und Ergebnisse Limnologie 7: 1-218.

Soininen J. 2004. Assessing the current related heterogeneity and diversity patterns of benthic diatom communities in a turbid and a clear water river. Aquatic Ecology 38: 495-501.

Soininen J, Könönen K. 2004. Comparative study of monitoring South-Finnish rivers and streams using macroinvertebrate and benthic diatom community structure. Aquatic Ecology 38: 63-75.

Solak CN, Acs E. 2011. Water Quality Monitoring in European and Turkish Rivers Using Diatoms. Turkish Journal of Fisheries and Aquatic Sciences 11: 329-337.

Stevenson RJ. 1997. Scale-dependent determinants and consequences of benthic algal heterogeneity. Journal of the North American Benthological Society 16: 1267-1276.

Stevenson RJ. 1998. Diatom indicators of stream and wetland stressors in a risk management framework. Environmental Monitoring and Assessment 51: 107-118.

Stevenson RJ, Lowe RL. 1986. Sampling and interpretation of algal patterns for water quality assessment. ASTM STP 894: 118-149.

Stevenson RJ, Pan Y. 1999. Assessing environmental conditions in rivers and streams with diatoms. Pages 11-40 in Stoermer EF, Smol JP, eds. The Diatoms: Applications for the Environmental and Earth Sciences. United Kingdom: Cambridge University Press.

Stoddard JL, Larsen DP, Hawkins CP, Johnson RK, Norris RH. 2006. Setting expectations fot he ecological condition of streams: the concept of Reference Condiction. Ecological Applications 16: 1267-1276.

Stoermer EF, Smol JP. 1999. The diatoms: Applications for the Environmental and Earth Sciences. United Kigdom: Cambridge University Press.

Tison J, Coste Y-S, Wasson J-G, Rimet F, Ector L, Delmas F. 2007. Predicting diatom reference communities at the French hydrosystem scale: A first step towards the definition of the good ecological status. Ecol. Model. 203: 99-108.

Tornés E, Cambra J, Gomà J, Ortiz JD, Sabater S. 2007. Indicator taxa of benthic diatom communities: a case study in Mediterranean streams. Ann. Limnol. Int. J. Lim. 43: 1-11.

Werum M, Lange-Bertalot H. 2004. Iconographia Diatomologica: Annoted Diatom Micrographs. Ruggell: A.R.G. Gantner Verlag K.G.

Van Sickle J, Huff DD, Hawkins CP. 2006. Selecting discriminant function models for predicting the expected richness of aquatic macroinvertebrates. Freshwater Biology 51: 359-372.

Van Sickle J, Hawkins CP, Larsen DP, Herlihy AT. 2005. A null model for the expected macroinvertebrate assemblage in streams. J. N. Am. Benthol. Soc. 24(1): 178-191.

Whitton B, Kelly M. 1995. Use of algae and other plants for monitoring rivers. Austr. J. Ecol. 20: 45-56.

Whitton B, Rott E, Friedrich G. 1991. Use of Algae for Monitoring Rivers. Institut für Botanik, Universität Innsbruck, Innsbruck, Austria: E. Rott, Publisher.

Woodiwiss F. 1964. The biological system of stream classification used by Trent River Board. Chem. Ind. 83: 443-447.

Wright JF, Furse MT, Armitage PD. 1993. RIVPACS – a technique for evaluating the biological quality of rivers in the U.K. Europ. Water Poll. Contr. 3: 15-25.

Wright JF. 1995. Development and use of a system for predicting the macroinvertebrate fauna in flowing waters. Austr. J. Ecol. 20: 181-197.

Wright JF. . 2000. An Introduction to RIVPACS. Pages 2-24 in Wright JF, Sutcliffe DW, Furse MT, eds. Assessing the Biological Quality of Fresh Waters: RIVPACS and Other Techniques. Ambleside, Cumbria, UK: Freshwater Biologica Association.

Wright JF, Furse MT, Armitage PD. 1993. RIVPACS – a technique for evaluating the biological quality of rivers in the U.K. Europ. Water Poll. Contr. 3: 15-25.

Zamora-Muñoz C, Alba-Tercedor J. 1996. Bioassessment of organically polluted Spanish rivers, using a biotic index and multivariate methods. J. N. Am. Benthol. Soc. 15(3): 332-352.

Zelinka M, Marvan P. 1961. Zur Präzisierung der biolo -gischen Klassifikation der Reinheit fließender Gewässer. Arch. Hydrobiol. 57: 389 – 407.

Appendix A – List of reference sites used for the models construction

Code	Site	Watercourse	Hydrological basin
1	Afluente do Torgal	Ribeira da Capelinha	Mira
2	Agroal	Rio Nabão	Tejo
3	Aguieiras (Sanceriz)	Aguieiras	Douro
4	Alcaria	Rio Alcaide	Lis
5	Aldeia_freiras	Ribeira Pera	Tejo
6	Alegrete	Ribeira de Arronches	Guadiana
7	Alvoco das Várzeas	Ribeira do Alvoco	Mondego
8	Ameixial	Ribeira do Vascãozinho	Guadiana
9	Azenha	Ribeira das Alfambras	Ribeiras do Algarve
10	Ázere (rio Ázere)	Rio Ázere	Lima
11	Azibo (Bragada)	Azibo	Douro
12	Azibo 1 (Balsamão)	Azibo	Douro
13	Azibo 2 (Foz do Azibo)	Azibo	Douro
14	Azinhal de Mouros	Ribeira do Vascão	Guadiana
15	Baceiro (Parâmio)	Baceiro	Douro
16	Barbaído	Rio Tripeiro	Tejo
17	Barranco	Ribeira do Vascão	Guadiana
18	Bazágueda	Rio Bazágueda	Tejo
19	Besteiros Jusante	Ribeira de Seixe	Ribeiras do Algarve
20	Boeiro	Ribeira Serta	Tejo
21	Busteliberne (Bucos Além Rio)	Busteliberne	Douro
22	Cabreira	Rio Ceira	Mondego
23	Calvo (Santa Valha)	Calvo	Douro
24	Canadas/M.te Redondo	Ribeira de Fonte Cova	Lis
25	Candedo (Malhadais)	Candedo	Douro
26	Caravelas	Caravelas	Douro
27	Carrazedo	Ribeira da Alombada	Vouga
28	Carregueira	Ribeira da Carregueira	Tejo
29	Casal_aboboreiras	Ribeira Lousa	Tejo
30	Casal_alecrim	Ribeira Algaz	Tejo
31	Cavacadouro	Rio Homem	Cávado
32	Côa 1 (Cinco Vilas)	Côa	Douro
33	Côa 1 (Seixo do Côa)	Côa	Douro
34	Côa 2 (Azevo)	Côa	Douro
35	Corgo (Cor2)	Corgo	Douro
36	Corte do Pinto	Barranco dos Alcaides	Guadiana
37	Curros (Cu20)	Curros	Douro
38	Eiriz (Eiriz)	Eiriz	Douro
39	Espinhal	Ribeira da Azenha	Mondego
40	Estevais	Estevais	Douro
41	Férrea	Rio Peneda	Lima
42	Ficalho	Ribeira do Vidigão	Guadiana
43	Folgosinho	Ribeira do Freixo	Mondego
44	Folques	Ribeira de Folques	Mondego
45	Foz de Besteiros	Ribeira de Seixe	Ribeiras do Algarve
46	Foz do Carvalhoso	Ribeira de Seixe	Ribeiras do Algarve
47	Foz do Cobrão	Rio Ocreza	Tejo

48	Freixal (Pega/Monte Vasco)	Freixal	Douro
49	Froufe	Rio Froufe	Lima
50	Gomes Aires Montante	Rio Mira	Mira
51	Grândola	Ribeira de Grândola	Sado
52	Guístola	Rio Agadão	Vouga
53	Laborins	Rio Alva	Mondego
54	Lamas de Mouro	Rio Mouro	Minho
55	Lavacolhos	Ribeira de Ximassas	Tejo
56	Lentiscais	Ribeira da Farropinha	Tejo
57	Loriga	Ribeira da Nave	Mondego
58	Louredo (Agunchos)	Louredo	Douro
59	Luzianes	Ribeira do Monte Novo	Mira
60	Maçãs (Junqueira)	Maçãs	Douro
61	Macedo (Ma 20)	Macedo	Douro
62	Manhouce	Ribeira de Manhouce	Vouga
63	Monim (Vilar de Maçada)	Monim	Douro
64	Monte da Fazenda	Ribeira de Erra	Tejo
65	Monte dos Arneiros	Ribeira dos Arneiros	Tejo
66	Monte dos Corvos	Ribeira do Vascão	Guadiana
67	Morenos	Barranco da Corte	Ribeiras do Algarve
68	Mosteirinho	Rio Agadão	Vouga
69	Mosteiro (Malhadais)	Mosteiro	Douro
70	Murtigão	Ribeira do Murtigão	Guadiana
71	Odemira	Rio Mira	Mira
72	Olo (Canadelo/Fridão)	Olo	Douro
73	Olo (Lamas de Olo)	Olo	Douro
74	Olo (Tejão)	Olo	Douro
75	Outeiro das Cabras	Rio Âncora	Lima
76	Paiva (Folgosa)	Paiva	Douro
77	Parada	Rio da Serra	Vouga
78	Pavia	Ribeira do Freixo	Tejo
79	Peio	Peio	Douro
80	Pêro Negro	Ribeira da Cerca	Ribeiras do Algarve
81	Peroviseu	Ribeira da Meimoa	Tejo
82	Pinhão (Barrela)	Pinhão	Douro
83	Pinhão (Pin1)	Pinhão	Douro
84	Piscinas da Lousã	Ribeira de S. João	Mondego
85	Poldras	Poldras	Douro
86	Pomar	Ribeira de Alvito	Tejo
87	Pombal-sul	Rio Arunca	Mondego
88	Ponte do Pingue	Ribeira de Moreira	Ave
89	Portela	Ribeira da Foz	Tejo
90	Praia do Vau	Rio Teixeira	Vouga
91	Praia Fluvial de S.João do Monte	Rio Águeda	Vouga
92	Pulo do Lobo	Ribeira de Limas	Guadiana
93	Queimado	Ribeira do Pardiela	Guadiana
94	Rabaçal (Ra 70)	Rabaçal	Douro
95	Rabaçal (Vale do Armeiro)	Rabaçal	Douro
96	Rabo do Burro (Soeima)	Rabo do Burro	Douro
97	Real	Ribeira de Docim	Ave
98	Redonda	Rio Águeda	Vouga
99	Redondo	Rio Águeda	Vouga

100	Relvas	Ribeira do Paúl	Tejo
101	Ribeira de Fráguas	Azenha da Costa Má	Vouga
102	Ribeira de Grandola	Ribeira de Grandola	Sado
103	Ribeiro de Baixo	Rio Castro Laboreiro	Lima
104	Róios (Qtª do Vale da Cal)	Ribª Vilariça	Douro
105	Rubiães	Coura	Minho
106	Russilhão	Russilhão	Douro
107	S. João do Monte (Mondego)	Rio Mondego	Mondego
108	S. Pedro (Minas Stº Adrião)	S. Pedro	Douro
109	S.João do Monte (Vouga)	Rio Águeda	Vouga
110	Sabóia	Rio Mira	Mira
111	Sabor (Sab1)	Sabor	Douro
112	Sabor (Sab4)	Sabor	Douro
113	Sabor 1 (Foz do Azibo)	Sabor	Douro
114	Sabor 2 (Felgar)	Sabor	Douro
115	Sabugueiro	ribeira da Fervença	Mondego
116	Salgueiro	Ribeira da Meimoa	Tejo
117	Santa Cruz	Ribeira do Vascão	Guadiana
118	Santa Marinha	Santa Marinha (Mós)	Douro
119	Segude	Rio Mouro	Minho
120	Tâmega3 (Agunchos)	Tâmega	Douro
121	Tedo (Stª Leocádia)	Tedo	Douro
122	Teja (Vesúvio)	Teja	Douro
123	Terges	Rio Terges	Guadiana
124	Tinhela (Martim)	Tinhela	Douro
125	Torgal Jusante	Ribeira do Torgal	Mira
126	Torgal Montante	Ribeira do Torgal	Mira
127	Torno (Povoação)	Torno	Douro
128	Torto (A-do-Bispo)	Torto	Douro
129	Tourigo	ribeira de Marruge	Mondego
130	Tregosa (rio Neiva)	Rio Neiva	Lima
131	Trovisco	Trovisco	Douro
132	Tuela (Guribanes)	Tuela	Douro
133	Tuela (Tue3)	Tuela	Douro
134	Uceira	Uceira	Douro
135	Urtigosa	Ortigosa	Douro
136	Vale da Azinheira	Ribeira de Urtiga	Tejo
137	Vale de Azares	Ribeira da Cabeça Alta	Mondego
138	Vale de Ferradas	Ribeira de Vale Ferradas	Tejo
139	Varzea de Romba	Ribeira de Odelouca	Ribeiras do Algarve
140	Vascão Jusante	Ribeira do Vascão	Guadiana
141	Vidoeiro (Ermida)	Vidoeiro	Douro
142	Vilalva (Serapicos)	Vilalva	Douro
143	Vilar da Veiga (rio Gerês)	Rio Gerês	Cávado

Appendix B – List of observed diatoms

Codes	Taxa
AAMB	*Aulacoseira ambigua* (Grun.) Simonsen
ABRT	*Achnanthidium bioretii* (Germain) Edlund
ACHS	*Achnanthes* species
ADBI	*Achnanthidium biasolettianum* (Grunow in Cl. & Grun.) Lange-Bertalot
ADEG	*Achnanthidium exiguum* (Grunow) Czarnecki
ADHE	*Achnanthidium helveticum* (Hustedt) Monnier Lange-Bertalot & Ector
ADMI	*Achnanthidium minutissimum* (Kützing) Czarnecki
ADMS	*Adlafia minuscula* (Grunow) Lange-Bertalot
ADSH	*Achnanthidium subhudsonis* (Hustedt) H. Kobayasi
ADSO	*Achnanthidium subatomoides* (Hustedt) Monnier, Lange-Bertalot et Ector
ADSU	*Achnanthidium subatomus* (Hustedt) Lange-Bertalot
AEEL	*Achnanthidium exiguum* (Grunow) Czarnecki var. *elliptica* Hustedt
AEXI	*Achnanthes exilis* Kützing
AFOR	*Asterionella formosa* Hassall
AINA	*Amphora inariensis* Krammer
ALIB	*Amphora libyca* Ehr.
ALTE	*Aulacoseira lacustris* f. *tenuior* (Grunow) Houk, Klee & Passauer
AMPS	*Amphora* species
ANMN	*Actinocyclus normanii* (Greg. ex Grev.) Hustedt morphotype *normanii*
APED	*Amphora pediculus* (Kützing) Grunow
APEL	*Amphipleura pellucida* Kützing
AUDI	*Aulacoseira distans* (Ehr.) Simonsen
AUGR	*Aulacoseira granulata* (Ehr.) Simonsen
AUSU	*Aulacoseira subarctica* (O.Muller) Haworth
AVEN	*Amphora veneta* Kützing
BBRE	*Brachysira brebissonii* Ross in Hartley ssp. *brebissonii*
BPAX	*Bacillaria paxillifera* (O. F. Müller) Hendey var. *paxillifera*
CAEX	*Cymbella excisa* Kützing var. *excisa*
CAFF	*Cymbella affinis* Kützing var. *affinis*
CASP	*Cymbella aspera* (Ehrenberg) H.Peragallo
CATO	*Cyclotella atomus* Hustedt
CBAC	*Caloneis bacillum* (Grunow) Cleve
CBAM	*Cymbopleura amphicephala* Krammer
CBNA	*Cymbopleura naviculiformis* (Auerswald) Krammer var. *naviculiformis*
CCIS	*Cymbella cistula* (Ehrenberg) Kirchner
CDUB	*Cyclostephanos dubius* (Fricke) Round
CEUG	*Cocconeis euglypta* Ehrenberg
CHAL	*Craticula halophila* (Grunow ex Van Heurck) Mann
CHEV	*Chamaepinnularia evanida* (Hustedt) Lange-Bertalot
CLAN	*Cymbella lanceolata* (Agardh ?) Agardh var. *lanceolata*

CMEN	*Cyclotella meneghiniana* Kützing
CMLF	*Craticula molestiformis* (Hustedt) Lange-Bertalot
CMNO	*Craticula minusculoides* (Hustedt) Lange-Bertalot
COCE	*Cyclotella ocellata* Pantocsek
COPL	*Cocconeis pseudolineata* (Geitler) Lange-Bertalot
CPED	*Cocconeis pediculus* Ehrenberg
CPLA	*Cocconeis placentula* Ehrenberg var. *placentula*
CPLI	*Cocconeis placentula Ehrenberg* var. *lineata* (Ehr.)Van Heurck
CPLK	*Cocconeis placentula* Ehrenberg var. *klinoraphis* Geitler
CRAC	*Craticula accomoda* (Hustedt) Mann
CSBM	*Craticula submolesta* (Hustedt) Lange-Bertalot
CSIL	*Caloneis silicula* (Ehr.) Cleve
CSMO	*Cymbella simonsenii* Krammer
CTGL	*Cymbella turgidula* Grunow 1875 in A.Schmidt & al. var. *turgidula*
CTPU	*Ctenophora pulchella* (Ralfs ex Kütz.) Williams et Round
CTUM	*Cymbella tumida* (Brebisson)Van Heurck
CYMS	*Cymbella* species
DCOF	*Diadesmis confervacea* K³tzing
DCOT	*Diadesmis contenta* (Grunow ex V. Heurck) Mann
DELL	*Diploneis elliptica* (Kützing) Cleve
DKUE	*Denticula kuetzingii* Grunow var. *kuetzingii*
DMES	*Diatoma mesodon* (Ehrenberg) Kützing
DOBL	*Diploneis oblongella* (Naegeli) Cleve-Euler
DOVA	*Diploneis ovalis* (Hilse) Cleve
DPAR	*Diploneis parma* Cleve
DPST	*Discotella pseudostelligera* (Husdted) Houk et Klee
DSUB	*Denticula subtilis* Grunow
DVUL	*Diatoma vulgaris* Bory 1824
EADN	*Epithemia adnata* (Kützing) Brebisson
EARC	*Eunotia arcus* Ehrenberg var. *arcus*
EARL	*Eunotia arculus* (Grunow) Lange-Bertalot & Nörpel
EBIL	*Eunotia bilunaris* (Ehr.) Mills var. *bilunaris*
EBLL	*Eunotia bilunaris* var. *linearis* (Okuno) Lange-Bertalot & Nörpel-Schempp
ECAE	*Encyonema caespitosum* Kützing
ECES	*Encyonopsis cesatii* (Rabenhorst) Krammer
EETE	*Eunotia exigua* (Breb.) Rabenhorst var. *tenella* (Grunow) Nörpel et Alles
EEXI	*Eunotia exigua* (Brebisson ex Kützing) Rabenhorst
EFAB	*Eunotia faba* Grunow
EGLA	*Eunotia glacialis* Meister
EIMP	*Eunotia implicata* Nörpel, Lange-Bertalot & Alles
EINC	*Eunotia incisa* Gregory var. *incisa*
EMIN	*Eunotia minor* (Kützing) Grunow in Van Heurck
EMUC	*Eunotia mucophila* (Lange-Bert. & Nörpel Schempp) Lange-Bertalot
ENAE	*Eunotia naegeli* Migula

ENCM	*Encyonopsis microcephala* (Grunow) Krammer
ENME	*Encyonema mesianum* (Cholnoky) D. G. Mann
ENMI	*Encyonema minutum* (Hilse in Rabh.) D. G. Mann
ENNG	*Encyonema neogracile* Krammer
EPEC	*Eunotia pectinalis* (Dyllwyn) Rabenhorst var. *pectinalis*
EPTR	*Eunotia paludosa* Grunow var. *trinacria* (Krasske) Nörpel et Alles
EPUN	*Eunotia pectinalis* (Kutz.) Rabenhorst var.undulata (Ralfs) Rabenhorst
ESBM	*Eolimna subminuscula* (Manguin) Moser Lange-Bertalot & Metzeltin
ESLE	*Encyonema silesiacum* (Bleisch in Rabh.) D. G. Mann
ESOL	*Eunotia soleirolii* (Kützing) Rabenhorst
ESOR	*Epithemia sorex* Kützing
ESUB	*Eunotia subarcuatoides* Alles Nörpel & Lange-Bertalot
ESUD	*Eunotia sudetica* O. Muller
ETOR	*Eunotia torula* Hohn
EUIN	*Eunotia intermedia* (Krasske ex Hustedt) Nörpel & Lange-Bertalot
EUNS	*Eunotia* species
EUPA	*Eunotia paludosa* Grunow in Van Heurck var. *paludosa*
EVEN	*Eunotia veneris* (Kützing) De Toni
FARC	*Fragilaria arcus* (Ehrenberg) Cleve var. *acus*
FAUT	*Fragilaria austriaca* (Grunow) Lange-Bertalot
FBID	*Fragilaria bidens* Heiberg
FCAP	*Fragilaria capucina* Desmazieres var. *capucina*
FCDI	*Fragilaria capucina* Desmazieres var. *distans* (Grunow) Lange-Bertalot
FCPL	*Fragilaria capitellata* (Grunow in Van Heurck) J. B. Peterson
FCRO	*Fragilaria crotonensis* Kitton
FCRS	*Frustulia crassinervia* (Breb.) Lange-Bertalot et Krammer
FCVA	*Fragilaria capucina* Desmazieres var. *vaucheriae* (Kützing) Lange-Bertalot
FERI	*Frustulia erifuga* Lange-Bertalot & Krammer
FGRA	*Fragilaria gracilis* Østrup
FHEL	*Fallacia helensis* (Schutz.) D. G. Mann
FMES	*Fragilaria mesolepta* Rabenhorst
FPYG	*Fallacia pygmaea* (Kützing) Stickle & Mann *ssp. pygmaea* Lange-Bertalot
FRHO	*Frustulia rhomboides* (Ehr.) De Toni
FRUM	*Fragilaria rumpens* (Kütz.) G. W. F. Carlson
FSAP	*Fistulifera saprophila* (Lange-Bertalot & Bonik) Lange-Bertalot
FTEN	*Fragilaria tenera* (W.Smith) Lange-Bertalot
FVIR	*Fragilaria virescens* Ralfs
FVUL	*Frustulia vulgaris* (Thwaites) De Toni
GACT	*Gomphonema acutiusculum* (O.Muller) Cleve-Euler
GACU	*Gomphonema acuminatum* Ehrenberg
GAFF	*Gomphonema affine* Kützing
GANG	*Gomphonema angustatum* (Kützing) Rabenhorst
GANT	*Gomphonema angustum* Agardh
GAUG	*Gomphonema augur* Ehrenberg

GCLA	*Gomphonema clavatum* Ehr.
GDEC	*Geissleria decussis* (Ostrup) Lange-Bertalot & Metzeltin
GEXL	*Gomphonema exilissimum* (Grun.) Lange-Bertalot & Reichardt
GGRA	*Gomphonema gracile* Ehrenberg
GMIC	*Gomphonema micropus* Kützing var. *micropus*
GMIN	*Gomphonema minutum* (Ag.) Agardh f. *minutum*
GNOD	*Gyrosigma nodiferum* (Grunow) Reimer
GOLI	*Gomphonema olivaceum* (Hornemann) Brébisson var. *olivaceum*
GOMS	*Gomphonema* species
GPAR	*Gomphonema parvulum* (Kützing) Kützing var. *parvulum* f. *parvulum*
GPSA	*Gomphonema pseudoaugur* Lange-Bertalot
GPUM	*Gomphonema pumilum* (Grunow) Reichardt & Lange-Bertalot
GPVL	*Gomphonema parvulius* Lange-Bertalot & Reichardt
GRHB	*Gomphonema rhombicum* M. Schmidt
GSHO	*Geissleria schoenfeldii* (Hustedt) Lange-Bertalot & Metzeltin
GTRU	*Gomphonema truncatum* Ehr.
GUTA	*Gomphonema utae* Lange-Bertalot & Reichardt
GYAC	*Gyrosigma acuminatum* (Kützing) Rabenhorst
HAMP	*Hantzschia amphioxys* (Ehr.) Grunow in Cleve et Grunow 1880
HCAP	*Hippodonta capitata* (Ehr.) Lange-Bert. Metzeltin & Witkowski
KALA	*Karayevia laterostrata* (Hustdet) Bukhtiyarova
KAPL	*Karayevia ploenensis* (Hustedt) Bukhtiyarova
KASU	*Karayevia suchlandtii* (Hustedt) Bukhtiyarova
KBOT	*Karayevia bottnica* (P. T. Cleve) Lange-Bertalot
KCLE	*Karayevia clevei* (Grunow) Bukhtiyarova
KOBG	*Karayevia oblongella* (Oestrup) M. Aboal
KOSU	*Kobayasiella subtilissima* (Cleve) Lange-Bertalot
LGOE	*Luticola goeppertiana* (Bleisch in Rabenhorst) D. G. Mann
LHUN	*Lemnicola hungarica* (Grunow) Round & Basson
LVEN	*Luticola ventricosa* (Kützing) D. G. Mann
MBAL	*Mastogloia baltica* Grunow
MCCO	*Meridion circulare* (Greville) Agardh var. *constrictum* (Ralfs) Van Heurck
MCIR	*Meridion circulare* (Greville) C. A. Agardh var. *circulare*
MPMI	*Mayamaea permitis* (Hustedt) Bruder & Medin
MVAR	*Melosira varians* Agardh
NAAN	*Navicula angusta* Grunow
NACI	*Nitzschia acicularis* (Kützing) W. M. Smith
NAGN	*Nitzschia agnita* Hustedt
NAMM	*Navicula ammophila* Grunow
NAMP	*Nitzschia amphibia* Grunow f. *amphibia*
NASP	*Navicula* species
NBRE	*Nitzschia brevissima* Grunow
NCAR	*Navicula cari* Ehrenberg
NCPL	*Nitzschia capitellata* Hustedt in A. Schmidt & al.

NCPR	*Navicula capitatoradiata* Germain
NCRY	*Navicula cryptocephala* Kützing
NCTE	*Navicula cryptotenella* Lange-Bertalot
NCTO	*Navicula cryptotenelloides* Lange-Bertalot
NDIS	*Nitzschia dissipata* (Kützing) Grunow var. *dissipata*
NDPV	*Naviculadicta pseudoventralis* (Hustedt) Lange-Bertalot
NDSS	*Neidium densestriatum* (Ostrup) Krammer
NDUB	*Nitzschia dubia* W.M.Smith
NEAF	*Neidium affine* (Ehrenberg) Pfitzer
NEAM	*Neidium ampliatum* (Ehrenberg) Krammer
NFIL	*Nitzschia filiformis* (W. M. Smith) Van Heurck var. *filiformis*
NFON	*Nitzschia fonticola* Grunow in Cleve et Müller
NGRE	*Navicula gregaria* Donkin
NHAN	*Nitzschia hantzschiana* Rabenhorst
NHMS	*Navicula heimansii* Van Dam et Kooyman
NICN	*Nitzschia incognita* Legler et Krasske
NIFR	*Nitzschia frustulum* (Kützing) Grunow var. *frustulum*
NIGR	*Nitzschia gracilis* Hantzsch
NIME	*Nitzschia media* Hantzsch.
NINC	*Nitzschia inconspicua* Grunow
NIPM	*Nitzschia perminuta* (Grunow) M.Peragallo
NISO	*Nitzschia solita* Hustedt
NIVA	*Nitzschia valdestriata* Aleem & Hustedt
NLAN	*Navicula lanceolata* (Agardh) Ehrenberg
NLIN	*Nitzschia linearis* (Agardh) W. M. Smith var. *linearis*
NLST	*Navicula leptostriata* Jorgensen
NLSU	*Nitzschia linearis* (Agardh) W. M. Smith var. *subtilis* (Grunow) Hustedt
NMEN	*Navicula menisculus* Schumann var. *menisculus*
NMIC	*Nitzschia microcephala* Grunow in Cleve & Moller
NNOT	*Navicula notha* Wallace
NPAD	*Nitzschia palea* (Kützing) W.Smith var. *debilis* (Kützing) Grunow in Cl. & Grun
NPAE	*Nitzschia paleacea* (Grunow) Grunow in van Heurck
NPAL	*Nitzschia palea* (Kützing) W.Smith
NPHY	*Navicula phyllepta* Kützing
NPML	*Nitzschia pumila* Hustedt
NPSA	*Navicula pseudoarvensis* Hustedt
NPSL	*Navicula pseudolanceolata* Lange-Bertalot
NRAD	*Navicula radiosa* Kützing
NRCH	*Navicula reichardtiana* Lange-Bertalot var. *reichardtiana*
NREC	*Nitzschia recta* Hantzsch in Rabenhorst
NRFA	*Navicula radiosafallax* Lange-Bertalot
NRHY	*Navicula rhynchocephala* Kützing
NROS	*Navicula rostellata* Kützing
NSHR	*Navicula schroeteri* Meister var. *schroeteri*

NSIG	*Nitzschia sigma* (Kützing) W. M. Smith
NSUA	*Nitzschia subacicularis* Hustedt in A. Schmidt et al.
NTAB	*Nitzschia tabellaria* (Grun.) Grun. in Cl. & Grun.
NTEN	*Navicula tenelloides* Hustedt
NTPT	*Navicula tripunctata* (O.F.Müller) Bory
NTUB	*Nitzschia tubicola* Grunow
NULA	*Nupela lapidosa* (Lange-Bertalot) Lange-Bertalot var. *lapidosa*
NVDL	*Naviculadicta laterostrata* Hustedt
NVDS	*Navicula* (dicta) seminulum (Grunow) Lange-Bertalot
NVEN	*Navicula veneta* Kützing
NZCD	*Nitzschia acicularioides* Hustedt
NZLT	*Nitzschia linearis* (Agardh) W. M. Smith var. *tenuis* (W. Smith) Grunow
NZSS	*Nitzschia* species
PBOR	*Pinnularia borealis* Ehrenberg var. *borealis*
PCHL	*Psammothidium chlidanos* (Hohn & Hellerman) Lange-Bertalot
PCLT	*Placoneis clementis* (Grun.) Cox
PDAU	*Planothidium daui* (Foged) Lange-Bertalot
PDIS	*Planothidium distinctum* (Messikommer) Lange-Bertalot
PDVG	*Pinnularia divergentissima* (Grunow) Cleve var. *divergentissima*
PFIB	*Peronia fibula* (Breb. ex Kutz.) Ross
PGIB	*Pinnularia gibba* Ehrenberg
PINS	*Pinnularia* species
PINT	*Pinnularia interrupta* W.M.Smith
PLEN	*Planothidium engelbrechtii* (Choln.) Round & Bukhtiyarova
PLFR	*Planothidium frequentissimum* (Lange-Bertalot) Lange-Bertalot
PLUN	*Pinnularia lundii* Hustedt var. *lundii*
PMAC	*Pinnularia macilenta* Ehrenberg
PMIC	*Pinnularia microstauron* (Ehr.) Cleve var. *microstauron*
PPRS	*Pseudostaurosira parasitica* (W. Smith) Morales
PPSB	*Parlibellus proctratus* var. *subcapitatus* (Wislouch & Poretzky) M. Aboal
PPSC	*Pseudostaurosira parasitica* var. *subconstricta* (Grunow) Morales
PRAD	*Puncticulata radiosa* (Lemmermann) Håkansson
PRST	*Planothidium rostratum* (Oestrup) Lange-Bertalot
PSAC	*Psammothidium sacculum* (Carter) Bukhtiyarova et Round
PSCA	*Pinnularia subcapitata* Gregory var. *subcapitata*
PSSE	*Pseudostaurosira elliptica* (Schumann) Edlund, Morales & Spaulding
PTCO	*Platessa conspicua* (A. Mayer) Lange-Bertalot
PTDE	*Planothidium delicatulum* (Kütz.) Round & Bukhtiyarova
PTEL	*Planothidium ellipticum* (Cl.) Round & Bukhtiyarova
PTHA	*Planothidium hauckianum* (Grun.) Round & Bukhtiyarova
PTLA	*Planothidium lanceolatum* (Brebisson ex Kützing) Lange-Bertalot
PVIR	*Pinnularia viridis* (Nitzsch) Ehrenberg var. *viridis* morphotype 1
RABB	*Rhoicosphenia abbreviata* (C. Agardh) Lange-Bertalot
RGIB	*Rhopalodia gibba* (Ehr.) O.Muller var. *gibba*

RSIN	*Reimeria sinuata* (Gregory) Kociolek & Stoermer
RUNI	*Reimeria uniseriata* Sala Guerrero & Ferrario
SANG	*Surirella angusta* Kützing
SBIS	*Surirella biseriata* Brebisson in Brébisson & Godey
SBKU	*Surirella brebissonii* var. *kuetzingii* Krammer et Lange-Bertalot
SBRE	*Surirella brebissonii* Krammer & Lange-Bertalot var. *brebissonii*
SBRV	*Staurosira brevistriata* (Grunow) Grunow
SCBI	*Staurosira construens* (Ehr.) var. *binodis* (Ehr.) Hamilton
SCON	*Staurosira construens* Ehrenberg
SCPM	*Staurosira construens* Ehr. var. *pumila* (Grunow in Van Heurck) Kingston
SEBA	*Sellaphora bacillum* (Ehrenberg) D. G. Mann
SEMN	*Eolimna minima* (Grunow) Mann
SLIN	*Surirella linearis* W. M. Smith
SPHO	*Stauroneis phoenicenteron* (Nitzsch.) Ehrenberg
SPRO	*Stauroneis producta* Grunow
SPUP	*Sellaphora pupa* (Kützing) Mereschkowsky
SRBA	*Surirella roba* Leclercq
SSMI	*Stauroneis smithii* Grunow
SSMU	*Staurosira mutabilis* (Wm Smith) Grunow
SSVE	*Staurosira venter* (Ehr.) Cleve & Moeller
STAN	*Stauroneis anceps* Ehrenberg
STAS	*Stauroneis* species
STDE	*Stenopterobia delicatissima* (Lewis) Brebisson ex Van Heurck
STHE	*Stauroneis thermicola* (Petersen) Lund
STKR	*Stauroneis kriegeri* Patrick
STLE	*Stauroneis legume* (Ehrenberg) Kützing
SUCO	*Surirella constricta* W. Smith
TAPI	*Tryblionella apiculata* Gregory
TFAS	*Tabularia fasciculata* (Agardh) Williams et Round
TFEN	*Tabellaria fenestrate* (Lyngbye) Kützing
TFLO	*Tabellaria flocculosa* (Roth) Kützing
TLEV	*Tryblionella levidensis* W. M. Smith
TVIS	*Thalassiosira visurgis* Hustedt
UBIC	*Ulnaria biceps* (Kützing) Compère
UDEA	*Ulnaria delicatissima* var. *angustissima* (Grunow) Aboal & Silva
UUAC	*Ulnaria ulna* (Nitzsch.) Compère var. *acus* (Kütz.) Lange-Bertalot
UULN	*Ulnaria ulna* (Nitzsch.) Compère

Appendix C – List of observed macroinvertebrates

Taxa
Acaridida
Acentrella sp.
Adicella sp.
Aeshna sp.
Agabus sp.
Agapetus sp.
Allogamus sp.
Amphinemura sp.
Anacaena sp.
Anax sp.
Ancylus sp.
Anisops sp.
Anthomyiidae
Aphelocheirus sp.
Aquarius sp.
Asellus sp.
Atherix sp.
Atrichops sp.
Atyaephyra sp.
Aulonogyrus sp.
Baetidae
Baetis sp.
Batracobdella sp.
Beraea sp.
Berosus sp.
BlephDaricera sp.
Boyeria sp.
Branchiura sp.
Caenis sp.
Calamoceras sp.
Calopteryx sp.
Ceraclea sp.
Ceratopogoninae
Chaetarthria sp
Cheumatopsyche sp.
Chimarra sp.
Chironomidae
Chironomini
Chloroperla sp.
Chrysomelidae
Clinocerinae
Cloeon sp.
Coelambus sp.
Colymbetinae
Copelatus sp.
Corbicula sp.
Cordulegaster sp.
Cordulia sp.
Culicinae
Curculionidae
Cyphon sp.
Deronectes sp.
Diamesinae
Dina sp.
Diytiscus sp.
Dolichopodidae
Dryops sp.
Dugesia sp.
Dupophilus sp.
Ecdyonurus sp.
Eiseniella sp.
Elmis sp.
Enchytraeidae
Enochrus sp.
Epeorus sp.
Ephemera sp.
Ephemerella sp.
Epitheca sp.
Erpobdella sp.
Esolus sp.
Gerris sp.
Glossiphonia sp.
Glossiphoniidae
Glossosoma sp.
Gomphus sp.
Graptodytes sp.
Gyrinus sp.
Habroleptoides sp.
Habrophlebia sp.
Haementeria sp.
Haemopis sp.

Haliplus sp.
Haplotaxidae
Haplotaxis sp.
Haselus sp.
Helopdella sp.
Helophorus sp.
Hemerodromiinae
Heptagenia sp.
Heptageniidae
Hexatomini
Holocentropus sp.
Hydaticus sp.
Hydraena sp.
Hydrocyphon sp.
Hydrometra sp.
Hydroporus sp.
Hydropsyche sp.
Hydroptila sp.
Hydroptilidae
Isoperla sp.
Lacasia sp.
Laccobius sp.
Laccophilus sp.
Lepidostoma sp.
Leptophlebiidae
Lestes sp.
Leuctra sp.
Leuctridae
Libellula sp.
Limnaea sp.
Limnephilus sp.
Limnius sp.
Limoniini
Lumbricidae
Lumbriculidae
Lype sp.
Lythoglyphus sp.
Macronychus sp.
Metalype sp.
Microvelia sp.
Naididae
Nais sp.
Naucoris sp.
Nemoura sp.
Nepa sp.
Noterus sp.
Notonecta sp.
Notonectidae
Oecetis sp.
Oligoneuriella sp.
Onychogomphus sp.
Ophidonais sp.
Orectochilus sp.
Oreodytes sp.
Orthetrum sp.
Orthocladiinae
Ostracoda
Oulimnius sp.
Oxyethira sp.
Paraleptophlebia sp.
Paranais sp.
Pediciini
Perla sp.
Perlodidae
Philopotamus sp.
Physa sp.
Pisidium sp.
Planorbarius sp.
Platynemis sp.
Plectrocnemia sp.
Polycelis sp.
Polycentropus sp.
Pomatinus sp.
Porhydrus sp.
Porhydrus sp.
Potamopyrgus sp.
Proasellus sp.
Procambarus sp.
Prosimuliini
Protonemura sp.
Pseudoneureclipsinae
Psychodidae
Psychomyia sp.
Radix sp.
Rhagionidae
Rhithrogena sp.
Rhyacophyla sp.
Scarodytes sp.

Scirtidae
Sericostoma sp.
Serratella sp.
Setodes sp.
Sialis sp.
Sigara sp.
Simuliidae
Simuliini
Slavina sp.
Sphaerium sp.
Stenelmis sp.
Stictonectes sp.
Stratiomydae
Stylaria sp.
Suphrodytes sp.
Sympecma sp.
Synagapetus sp.
Tabanidae
Tanypodinae
Tanytarsiini
Thraulus sp.
Thremma sp.
Tipula sp.
Tipulidae
Triaenodes sp.
Trocheta sp.
Tubificidae
Uncinais sp.
Velia sp.
Wormaldia sp.
Xanthoperla sp.