Discrete vs. continuum approaches to the assessment of the ecological status in Iberian rivers, does the method matter?

Virgilio Hermoso¹, and Simon Linke¹.

¹ Australian Rivers Institute, Griffith University, Nathan, 4111 Queensland, Australia.

Corresponding author: Virgilio Hermoso.

E-mail: virgilio.hermoso@gmail.com

Running title: Discrete vs. continuum approaches for bioassessment

Abstract

Assessing the ecological status of freshwater ecosystems has become an important topic as a first approach to improve their poor condition. Ecological status is usually measured as the deviation of observed conditions from expected reference conditions in the absence of human disturbances. The ability to produce objective evaluations depends largely on the capacity to correctly define reference conditions. This is usually done following either a type-specific (classification-based) or site-specific approach. Despite the often weak correspondence of classification schemes used in the type-specific approach little attention has been paid to the evaluation of its consequences for defining reference conditions.

Here, we first test the capacity of an environmental classification derived from the implementation of the Water Framework Directive (WFD) to portray freshwater fish assemblage composition in a Mediterranean basin in South Western Iberian Peninsula. We then check the effect of using type-specific vs. site-specific approaches on the performance of different predictive models, as they are a common tool for defining reference conditions elsewhere. We compare predictive models performance using two different classification approaches (*top-down* vs. *bottom-up*) with the same predictive model (RIVPACS) and continuum-like predictive model (ANNA).

The top-down environmental classification did not group reference sites better than random, with high within-class heterogeneity and high between-class homogeneity. Although slightly better, the bottom-up approach did not show high classification performance. This result was not related to the effect weak biological response to environmental gradients, since a true underlying biotic gradient existed (fish assemblages underwent gradual changes through a longitudinal habitat gradient). As a consequence, the classification-based predictive methods failed in fitting the biological data and predicting the expected fish assemblages. Only the continuous predictive approach succeeded and was accurate enough to be used to define reference conditions.

This study shows that the lack of biological significance of classifications may have severe consequences for bioassessment programs. We recommend verifying the accuracy of the classifications before using them as the basis for bioassessment under type-specific approaches. Alternatively, a continuous approach without pre-defined classes should be considered.

KEYWORDS: ANNA, bioassessment, continuous, Mediterranean, reference condition, RIVPACS.

1. Introduction

Assessing the ecological status of freshwater ecosystems has been a key issue in freshwater management for the last decades as a consequence of the poor condition that these ecosystems suffer. Freshwater resources are crucial for human culture, welfare and development, so freshwater environments are submitted to high pressures, overexploitation and transformation rates (Vörösmarty et al., 2010), and freshwater biodiversity is among the most threatened in the world (Dudgeon et al., 2006). In an attempt to address this problem, water policy legislations, such as the Water Clean Act (WCA) in the US or the European Water Framework Directive (WFD) have been developed aiming to protect and improve the condition of aquatic ecosystems (e.g., European Commission, 2000). These laws establish the framework to evaluate the ecological status of freshwater ecosystems, identify the main perturbations responsible for the observed condition, and recommend the most suitable set of actions to lead their recovery. The ecological status is usually defined relative to the deviation of observed status from a reference condition in the absence of human disturbances (reference condition approach, *sensu* Reynoldson, 1997). Reference conditions are established either for a type of freshwater ecosystem (type-specific methods) or determined specifically for each site being evaluated (site-specific methods). Under the type-specific approach, freshwater ecosystems are initially grouped into homogeneous classes, for which a unique reference condition will be described. Different classification methods are used, such as *a priori* top-down environmental classifications (Omernik and Bailey, 1997; Munné and Prat, 2004; Verdonshot and Nijboer, 2004) based on environmental surrogates (e.g., elevation, temperature or rainfall), or *a posteriori* bottom-up biotic classifications (Wright, 1995; Simpson and Norris, 2000; Melcher et al., 2007) based on biological data. On the other hand, site-specific approaches avoid classifications and describe reference conditions for each site under evaluation, by using predictive models for example (e.g., Oberdorff et al., 2002; Hermoso et al., 2010).

The popularity of type-specific approaches is based on the assumption that each subdivision or class of streams has some consistent biological significance (e.g., Frissell et al., 1986). The use of subdivisions helps better understand and cope with the complexity of biological assemblages in the evaluation of the ecological status (Rabeni et al., 2002).

To be useful, any classification (either top-down or bottom-up) should satisfy two main conditions, i) correspond to the distribution of species and hence identify recurring biotic assemblages (Heino and Mykrä, 2006), and ii) portray relationships between biological and physical features including environmental classifiers which reflect ecologically important factors for the biota (Angermeier and Schlosser, 1995; Karr and Chu, 1999). If a classification addresses all of the above, both type I statistical errors (inferring impairment when it does not exist) and type II errors (not detecting impairment when it does exist) could be reduced

(Hawkins et al., 2000a). There is an extensive literature devoted to the evaluation of different classifications systems in terms of how much biological variability they portray (e.g., Hawkins et al., 2000a; Rabeni et al., 2002; Sanchez-Montoya et al., 2007). Some of them report a poor capacity of classes to portray biologically meaningful classes (e.g., classes share most of species), which might limit the precision of bioassessment programs. In particular, weak classifications may difficult the definition of reference conditions, on which the process relies. This could make evaluations fail at determining the degree of impairment and identifying the main perturbations that should be targeted in further rehabilitation programs. However, little efforts have been devoted to the objective assessment of implications of classification weaknesses on bioassessment tools.

The main aims of this work are i) to test the capacity of two different classification methods (top-down and bottom-up) to portray biological variability; and ii) to evaluate the accuracy of discrete (type-specific) vs. continuum (site-specific) approaches to identifying reference conditions. We firstly evaluate the strength of a top-down classification derived from the implementation of the WFD´s system B in a Mediterranean basin in the south-western Iberian Peninsula (Guadiana River). We then compare the performance of three different predictive models: two classification-based and a continuum approach and discuss about the causes of different performance values. Finally, we make some recommendations for future applications of these approaches.

2. Methods

2.1 Study area

The Guadiana River basin is located in the South-Western Iberian Peninsula draining a total area of 67,039 km² to the Atlantic Ocean (Fig. 1). It features a typical Mediterranean climate, with high intra and interannual variation in discharge and severe floods and droughts. Potenciano de las Heras (2004) reported annual rainfall values ranging 171-1118 mm in some areas of the Guadiana basin in the period 1850-2004, mostly concentrated in winter. Nearly one half of the basin (49.1%) is currently being used for agriculture. Other common human perturbations are related to river channel modifications due to river channelization and degradation and even complete depletion of the riparian forest and water quality deterioration (Urrea and Sabater, 2009). The Guadiana River basin contains 6 of the 29 environmental types defined following the specifications of system B proposed in the European Water Framework Directive (European Commision, 2003; Ministerio de Medio Ambiente, 2005) (Fig. 1).

2.2 Fish community and environmental data

Fish communities were characterized in 239 localities across the basin (Fig. 1) during spring in 2002, 2005, and 2006. Fish sampling at each location was carried out along 100 m long stream reaches whenever possible using single-pass electrofishing. The sampling stretch covered all habitats available following the recommendations of the FAME Consortium (2004) for European rivers. Sály et al. (2009) and Bertrand et al. (2006) demonstrated this method to be adequate to characterize spatial and temporal trends in fish abundance and composition. Species richness ranged between 1-8 (Mean $\pm$ SE, 3.0 ± 0.1).

Habitat was characterised by 37 environmental variables, covering three different spatial scales: site, reach and basin. All of them were selected from a larger set of environmental variables for their relative independence of human perturbations. Two approaches were used in this characterization: field measures to characterize the habitat at the reach scale and remotely sensed measures to record data from available digital maps (Table 1). All these variables were tested for normality and conveniently transformed when necessary for further analyses.

2.3 Testing the biological significance of environmental classifications

To avoid the pernicious effect of perturbations in our assessment of classification's accuracy, we only considered those sites less affected by human activities (reference sites hereafter). Reference sites are characterised by low urban or agricultural land uses within the whole basin and at the reach scale (500 m around the sampling point). Furthermore, bank and channel structure as well as the riparian zone should be in natural condition (Hermoso et al., 2009). To ensure that reference sites are not impacted by invasive fish, all sites where exotic species accounted for more of 5% of total fish abundance were also discarded (Kennard et al., 2006). We finally considered 57 reference sites distributed across all the environmental classes except *Continental-Mediterranean mineralized rivers* (Fig. 1). This class is exclusively located in the mainstream of the Guadiana River draining through a highly intensive-agriculture transformed landscape. We then tested the biological significance of environmental classifications for the remaining 5 water body types. The inclusion of rare species in the analysis may have negative consequences on classification systems and bioassessment tools (Cao et al., 2001; Van Sickle et al., 2007), so to reduce the noise introduced by very rare species we did not consider species present in less than 5% of localities in the analysis below.

We measured classification strength (CS) using a Multiple Response Permutation Procedure (MRPP) analysis in R statistical software Version 2.13.1 (R Development Core Team 2004). The test calculates a within-class agreement statistic ($A = 1 - \delta / E\delta$) that compares the overall weighted mean of within-class mean dissimilarity (δ in Formula 1) and the average dissimilarity value across classes (mean of all the pairwise comparisons, $E\delta$).

$$\delta = \frac{\sum_{i=1}^n d_i * w_i}{n} \quad (\text{Formula 1})$$

where d_i is the average dissimilarity for class i and w_i the weight applied to class i [$w_i = (s_i * (s_i - 1)) / 2$, with s_i = number of sites within class i]. MRPP also estimates a significance value associated with A , by performing n permutations (9999 in our case) on the dissimilarities across classes. A ranges between 1 and -1, when all assemblages within classes are identical, A reaches its maximum value of 1. If between-class similarity equals that expected by chance $A = 0$, and if between-class similarity is higher than expected by chance $A < 0$. To better understand the origin of the observed A value, we further explored the different components of CS, by checking within-class homogeneity (W) and between-class similarity (B) (Van Sickle and Hughes, 2000).

2.4 Influence of landscape continuum on biota

We used Non-Metric Multidimensional Scaling (MDS) to explore fish assemblage structure based on presence-absence data along the sampling sites. We used presence-absence since it is more suitable for comparisons between streams within a whole basin in a large spatial context, as our case, whilst abundance data is specially recommended for smaller areas where many taxa are widely distributed. Presence-absence data gives more weight to species with a reduced distribution than abundance data (Rabeni et al., 2002). Moreover, population densities are submitted to greater temporal (seasonal and inter-annual) and spatial rates of change than presence-absence data even under natural conditions in Mediterranean and other similar harsh environments (Meffe and Minckley, 1987; Matthews and Marsh-Matthews, 2003; Magalhães et al., 2007). Species presences have been proved to be more persistent than abundance through natural dramatic climatic events (frequent in these kinds of environments) as severe droughts or floods (Magalhães et al., 2007). In this way, we followed the recommendations in Hermoso et al. (2010) to reduce the potential differences in biotic composition across seasons and years, at which the study was carried out. The MDS was performed on a similarity matrix built on presence-absence data, recommended because it ignores joint absences, the predominant case in site x species matrices (Faith et al., 1987).

We then checked the relationship between the spatial ordination obtained in MDS and the set of environmental variables describing natural gradients (Table 1). We used an independent multiple regression models for each of the three first ordination axes from MDS to try to describe the biological variation with the mentioned environmental variables. A stepwise approach with entry and removal criterion (set at $r^2 = 0.1$) was followed. All variables were checked for redundancy prior their inclusion in models ($r < 0.5$). If freshwater fish assemblages

were structured along environmental gradients we would expect the multiple regression models to explain a substantial proportion of each MDS's axes.

2.5 Influence of classifications' accuracy on bioassessment

Ecological status is commonly assessed by comparing current assemblages to expected biotic communities in the absence of human perturbation obtained from predictive models (Hawkins et al., 2000b; Oberdorf et al., 2002). RIVPACS models (Wright, 1995; Clarke et al., 2003) are one of the most widespread and accepted methods in bioassessment. Thus the accuracy of the evaluation depends on the ability of the predictive methods to correctly predict the expected fauna to be used as reference in the comparison. Failing at predicting species composition correctly will lead to overestimation or underestimation of the ecological status of a given freshwater ecosystem.

To test the influence of different classification systems on predictive model accuracy, we built three different predictive models using the same data:

- A forced RIVPACS model in which we avoided the initial classification step by using the top-down environmental classification described previously.
- A standard RIVPACS model, in which the classes were derived using a bottom-up approach. We used a UPGMA classification method, based on the Bray-Curtis similarity of fish assemblages (Clarke et al., 2003). The same set of environmental predictor was used in both RIVPACS models, to avoid the effect of predictors on model performance.
- An existing ANNA model (Linke et al., 2005) for the region (see Hermoso et al., 2010), in which sites are treated as a continuum and in which predictions are derived from the environmentally most similar reference sites. ANNA models finds the set of reference sites most environmentally-similar to the target site under evaluation and predicts its community composition based on the community composition of those nearest neighbours (Linke et al., 2005). The contribution of each selected reference site to the final prediction is weighted according to their relative distance to the test site. Then the closer the selected reference site, the higher is its weight on predictions. Our ANNA model used the nearest 6 reference sites to predict each species occurrences.

Since the original number of reference sites we got within the Guadiana River basin was not large enough for model construction and validation, we selected a total of 28 extra reference sites from adjacent basins in the same biogeographical region (Tinto, Odiel and Guadalquivir basins; Fig. 1). All new reference sites belonged to the same environmental classes as the Guadiana sites, so they did not introduce additional variability for classification based

approaches. All sites were surveyed during the same period and using the same standardized sampling approach as explained above. Moreover new sites did not hold any new native species, so we did not include any additional biological variability in the analyses. Reference sites in these new catchments were identified using the same criteria as for the Guadiana River basin, so all of them fulfilled the minimum quality criteria. We used 65 reference sites to build the models and an independent set of 20 reference sites to validate them. We validated the models by plotting the observed vs the expected species richness. As sites used in the validation are reference sites, we anticipated a 1:1 correlation between observed and expected. We checked model fit through the regression line slope, intercept and R^2 (Linke et al. 2005) and the standard deviation of O/E values (Van Sickle et al., 2005) in the validation data set. For the standard deviation, the upper and lower baselines for model performance were given by the SD of the best possible and the null models, as Van Sickle et al. (2005) suggested. The SD of a good model has to improve that displayed by the null model and should lie close to the best possible model. The classification strength of the *a posteriori* classification used in the Standard-RIVPACS model was evaluated similarly to the WFD environmental classes.

In case of a strong concordance of environmental classifications with fish assemblages, we expect the top-down RIVPACS model to perform as strongly as a RIVPACS model based on biota-driven classifications and the very-well performing ANNA model (Hermoso et al., 2010).

3. Results

3.1 Biological significance of environmental classifications

Overall, we found a weak correspondence between the fish assemblages and the environmental classification, indicating low biological significance of this classification. Mean within-class homogeneity was quite similar to between-class similarity (Table 2), so environmental classes were highly overlapped (Fig. 2). MRPP showed that the top-down environmental classification had a between-class similarity higher than expected if localities would have been grouped randomly ($A=-0.03$, $P=764$). The lack of concordance between the top-down classification and fish assemblages was not due the presence of homogeneous fish assemblages in the Guadiana River catchment. Fish species distribution patterns did exist and followed a clear habitat gradient, as indicated by the predictive capacity of the multiple regression models (Fig. 2). This result suggests a transitional replacement in species composition along the natural environmental gradient that was not portrayed by the environmental classification.

3.2 Influence of environmental classifications on predictive model accuracy

The top-down environmental classification did not produce a valid RIVPACS model, with almost random predictions (Table 3). Although the standard bottom-up RIVPACS model outperformed the top-down RIVPACS model, it was still very poor as the slope and intercept of the O/E line for the validation data set exceeded the limits proposed by Linke et al. (2005) (Table 3). Both models used a similar number of classes (top-down RIVPACS used the 5 *a priori* environmental classes tested above, and the standard bottom-up RIVPACS used 4 *a posteriori* defined classes). The classes in the bottom-up classification were selected to maximize the compositional differences among classes by the UPGMA. Average species richness/ class was also very similar in both RIVPACS approaches, since it ranged between 2.7-4.2 for the top-down RIVPACS and 2.5-4.6 for the standard bottom-up RIVPACS model. The improvement in model performance when using bottom-up defined classes could be related to the higher classification strength of the bottom-up classification that the standard RIVPACS used ($A=0.19$, $P<0.001$). However, the steep slope of the regression line between observed and expected richness that the standard bottom-up RIVPACS model showed increases the risk of Type I error when assessing the ecological status. For the subset of reference sites used in the calibration, the expected species richness to be found where higher than the observed in field (Table 3). For this reason the O/E score will be below 1 even in reference sites, which will be wrongly labelled as perturbed ($O/E<1$). Moreover, the predictive range of species richness was very narrow and never over four species, which increases the probability of Type II errors as well. RIVPACS approach would be inefficient in detecting species loss in the richest sites and thus potentially fails to detect perturbation.

The continuous ANNA model outperformed the classification-based models. The slope (not different from 1, $b=1.063$, t-test, $p=0.58$), intercept (not different from 0, $a=0.058$, $p=0.95$), R^2 and SD (Table 3) were within the standard range of models (see Linke et al., 2005). This ensures that Type I and Type II errors in the bioassessment are lower than in the other models (Table 3). Furthermore, ANNA was the only model which correctly predicted the presence of all the species at least once.

4. Discussion

The ability of bioassessment programs to objectively evaluate the ecological status of a given system depends largely on how well we can describe expected conditions for each site under evaluation, or in other words, how accurately we can define reference conditions. Biological assemblages are known to change along environmental gradients (Rahel and Hubert, 1991), so reference conditions should ideally account for those changes. Despite the improvement of bottom-up vs top-down classification schemes to portray the patterns in fish assemblages [similar to Padial et al. (2012)], none of the classifications tested in this study were

able to define strong biological classes. Even when the biological data were used to define classes in the bottom-up approach, the classification obtained was characterized by high between-class similarity, and far from an ideal classification for bioassessment purposes (Dodkins et al., 2005). Moreover, classes in the environmental top-down classification were not more biologically homogeneous than a random aggregation. Under this scenario, the observed fish assemblage would be evaluated against a “random-like” set of reference conditions that do not account for the natural variability in assemblage composition. This is a key issue, since the likelihood of committing type I and II errors when using “random-like” reference conditions is very high. Freshwater fish assemblages tend to become richer and more diverse towards low reaches (e.g., Magalhães et al., 2002; Clavero et al., 2005), so the use of reference conditions not better chosen than random will bias the evaluation process. For example, when assessing the status of a headwater reach, the expected assemblage should resemble that of headwaters (low richness and diversity), otherwise the evaluation will be prone to underestimate the true ecological status of this headwater reach (expected richness and diversity will always be higher than achievable for that reach). On the other hand, if evaluating the ecological status of a low reach using headwater fish assemblages as reference conditions, the evaluation will tend to overestimate the true status (even when some perturbation might have happened, species richness and diversity could be higher than expected for a headwater reach). Given the key importance of correctly defining reference conditions and the difficulties of doing it when using classification approaches, we would strongly recommend evaluating the biological significance of any classification (either top-down or bottom-up) used for bioassessment purposes.

Several studies have analysed the concordance between environmental classifications and biological communities. According to these studies, environmental classifications sometimes portrayed the variability of stream assemblages to some degree, although the strength of these classifications was rather weak (Hawkins et al., 2000a; Van Sickle & Hughes, 2000; Heino et al., 2002). Despite the poor capacity of classes to portray biologically homogeneous classes reported in previous studies, the use of these classifications has still been strikingly recommended. We acknowledge that achieving a perfect classification is a rather complex task in nature, but whenever the performance of a given classification is not good enough as to ensure reasonable estimates of reference condition we would recommend avoiding it and using continuous approach instead. These approaches are available elsewhere and are increasingly being used in biological assessments [e.g., Generalised Dissimilarity Modelling (Ferrier, 2002), ANNA (Linke et al., 2005) or Multivariate Adaptive Regression Splines (Leathwick et al., 2005)].

The lack of biological significance of classifications has traditionally been linked to the effect of many taxa varying independently and weakly in response to environmental gradients

(Heino et al., 2003; Snelder et al., 2004; Heino and Mykrä, 2006). The weak biological response to environmental gradients could have been argued as the reason of the poor classification performance reported here. Mediterranean fish assemblages are characterized by low species richness, high endemism and low specialization levels (fish species have evolved in a harsh environment where specialization is not the best strategy). The conjunction of low-species richness assemblages and generalist species could entail weak responses to environmental gradients and then difficulties to obtain strong classifications. However, this seems not to be the reason of the poor performance showed by the two alternative classification approaches we tested here. In our case a response of freshwater fish assemblages to the main environmental gradient existed. We found patterns in species composition that could be related to a longitudinal (upstream-downstream) environmental gradient. The effects of these gradual biological changes in response to habitat gradients on grouping performance have extensively been reported in scientific literature: (i) Van Sickle and Hughes (2000) showed that a classification based on the proximity of sites was stronger than any other a priori classification. (ii) Hawkins and Vinson (2000) showed that classifications in which sites within a class were located relatively close to one another were stronger than classifications in which sites were located in geographically non-contiguous classes. Thus, the inability of both classification systems to account for the observed longitudinal zonation of biotic communities (Hawkins and Vinson, 2000; Snelder et al., 2004) seems to be the main reason of the poor grouping performance that they displayed. This might not be constrained by the peculiarities of the Mediterranean fish assemblages used here. Similar performance levels have been reported in other areas with distinct fish assemblages, such as the Wet-tropics region in Australia (Januchowski et al., 2011), so our recommendations could be extrapolated elsewhere.

As expected, the absence of biological concordance of the top-down and bottom-up classifications had immediate consequences on model performance. The top-down RIVPACS model did not predict species richness at new sites better than randomly and would thus not be suitable for bioassessment. This is not a surprising conclusion as the high within-class heterogeneity displayed by the top-down environmental classes did not allow the RIVPACS model to match particular fish assemblage compositions to specific environmental conditions. Given the proven weaknesses of the top-down classification derived from the application of the WFD's system B that we tested here and its potential negative effects when identifying reference conditions, we recommend avoiding its use for fish-based bioassessment.

The bottom-up classification approach (standard RIVPACS model) resulted in better model performance than the top-down classification due to the improvement in the classification strength, and the ability to discriminate between different fish communities. However, both type-specific predictive models were outperformed by the continuous approach

through the ANNA model. Given the clear biotic pattern we found, assemblages would not be expected to show much tendency to cluster into discrete classes, and any classification of species into a few classes would thus only allow a coarse estimation of the actual fauna expected to occur at a site (Hawkins and Vinson, 2000). This is expected to happen even when using bottom-up classifications as the standard RIVPACS did. Predictions of expected conditions derived from this bottom-up classification entailed a high risk of type I and II statistical errors. On the other hand, the ANNA model was an average/good model. As expected a model that accounted for the continuous nature of streams would better fit the biological data and would deliver more accurate predictions - as the ANNA model did.

The implementation of bioassessment programs for the protection and restoration of biological integrity poses a great advance in the management of worldwide freshwater ecosystems. However, two key lessons might be learnt from this study:

1. Environmental classifications often have no biological meaning. Then bioassessment methods that classify sites into a few discrete classes are likely to be limited in how precisely they can specify expected reference conditions, and thus the degree of biological impairment that they can detect.
2. Continuous approaches are sometimes more accurate than classifications. They avoid the effects of imprecise classifications and better portray the continuous nature of rivers.

So when stream classifications affect the precision of bioassessment as severely as our analysis indicates, it is highly recommended to use methods that do not rely on site classification for specifying reference conditions (Hawkins and Vinson, 2002). Instead of that we recommend using alternative methods which avoid artificial classifications and recognise streams as biological continua through site-specific measures. These methods allow more refined predictions of reference conditions and more objective assessment of ecological status hence.

5. Acknowledgement

The fieldwork was funded by the Confederación Hidrográfica del Guadiana. V.H. held a pre-doctoral fellowship financed by the Spanish Ministry of Education (AP-2004-1414). S.L.'s postdoctoral fellowship was funded by eWater CRC.

References

- Angermeier, P.L., Schlosser, I.J., 1995. Conserving aquatic biodiversity: beyond species and populations. *Am Fish Soc Symp* 17, 402-414.
- Bertrand, K.N., Gido, K.B., Guy, C.S., 2006. An evaluation of single-pass versus multiple-pass backpack electrofishing to estimate trends in species abundance and richness in prairie streams. *Trans Kans Acad Sci* 109, 131–138.
- Cao, Y., Larsen, D.P., Thorne, St-J., 2001. Rare species in multivariate analysis for bioassessment: some considerations. *J N Am Benthol Soc* 20, 144-153.
- Clarke, R.T., Wright, J.F., Furse, M.T., 2003. RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. *Ecol Model* 160, 219-233.
- Clavero, M., Blanco-Garrido, F., Prenda, J., 2005. Fish-habitat relationship and fish conservation in small coastal streams in southern Spain. *Aquat Conserv* 15, 415-426.
- Dodkins, I., Rippey, B., Harrington, T.J., Bradley, C., Ni Chathain, B., Kelly-Queen, M., McGarrigle, M., Hodge, S., Trigg, D., 2005. Developing an optimal river typology for biological elements within the Water Framework Directive. *Water Res* 39, 3479-3486.
- Dudgeon, D., Arthington, A.H., Gessner, M.O, Kawabata, Z., Knowler, D., Lévêque, C., Naiman, R.J., Prieur-Richard, A.-H., Soto, D., Stiassny, M.L.J. & Sullivan, C.A. (2006): Freshwater biodiversity: importance, threats, status and conservation challenges. *Biol Rev* 81, 163-182.
- European Commission, 2000. Directive 2000/60/EC of the European Council and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. *Official Journal of the European Communities*, L327 1-72.
- European Commission, 2003. Common Implementation Strategy for the Water Framework Directive (2000/60/EC). Guidance document 10: River and Lakes- Typology, reference conditions and classification systems. REFCOND, Luxemburg.
- Faith, D.P., Minchin P.R., Belbin, L., 1987. Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* 88,135–141.
- FAME Consortium. (2004) Manual for application of the European Fish index—EFI. Version 1.1, January, 2005, pp. 66.
- Ferrier, S., 2002. Mapping spatial pattern in biodiversity for regional conservation planning: Where to from here? *Syst Biol* 51, 331–363.

418 Frissell, C.A., Liss, W.J., Warren, C.E., Hurley, M.D., 1986. A hierarchical framework for
 419 stream habitat classification: viewing streams in a watershed context. *Environ Manage* 10,
 420 199-214.

421 Hawkins, C.P., Norris, R.H., Gerritsen, J., Huges, R.M., Jackson, S.K., Johnson, R.K.,
 422 Stevenson R. J., 2000a. Evaluation of the use of landscape classifications for the prediction
 423 of freshwater biota: synthesis and recommendations. *J N Am Benthol Soc* 19, 541-556.

424 Hawkins, C.P., Norris, R.H., Hogue, J.N., Feminella, J.W., 2000b. Development and use of
 425 predictive models for assessing the biological integrity of streams. *Ecol Appl* 10, 1456-
 426 1477.

427 Hawkins, C.P., Vinson, M.R., 2000. Weak correspondence between landscape classifications
 428 and stream invertebrate assemblages: implications for bioassessment. *J N Am Benthol Soc*
 429 19, 501-517.

430 Heino, J., Muotka, T., Mykrä, H., Paavola, R., Hämäläinen, H., Koskenniemi, E., 2003.
 431 Defining macroinvertebrate assemblage types of headwater streams: implications for
 432 bioassessment and conservation. *Ecol Appl* 13, 842-852.

433 Heino, J., Muotka, T., Paavola, R., Hämäläinen, H., Koskenniemi, E., 2002. Correspondence
 434 between regional delineations and spatial patterns in macroinvertebrates assemblages of
 435 boreal headwaters streams. *J N Am Benthol Soc* 21, 397-413.

436 Heino, J., Mykrä, H., 2006. Assessing physical surrogates for biodiversity: Do tributary and
 437 stream type classifications reflect macroinvertebrate assemblage diversity in running
 438 waters? *Biol Conserv* 129, 418-426.

439 Hermoso, V., Clavero, M., Blanco-Garrido, F., Prenda, P., 2010. Assessing the ecological status
 440 in species-poor systems: a fish-based index for Mediterranean Rivers (Guadiana River, SW
 441 Spain). *Ecol Indic* 10, 1552-1561.

442 Hermoso, V., Linke, S., Prenda, J., 2009. Identifying priority sites for the conservation of
 443 freshwater fish biodiversity in a Mediterranean basin with high degree of threatened
 444 endemics. *Hydrobiologia* 623, 127-140.

445 Januchowski-Hartley, S.R., Hermoso, V., Pressey, R.L., Linke, S., Kool, J., Pearson, R.G.,
 446 Pusey, B.J., and VanDerWal, J. 2011. Coarse-filter surrogates do not represent freshwater
 447 fish diversity at a regional scale in Queensland, Australia. *Biol Conserv* 144, 2499-2511.

448 Karr, J. R., Chu, E.W., 1999. *Restoring Life in Running Waters: Better Biological Monitoring*.
 449 Island press, Covelo.

450 Kennard, M.J., Harch, B.D., Pusey, B.J., Arthington, A.H., 2006. Accurately defining the
451 reference condition for summary biotic metrics: a comparison of four approaches.
452 *Hydrobiologia* 572, 151-170.

453 Leathwick, J.R., Rowe, D., Richardson, J., Elith, J., Hastie, T., 2005. Using multivariate adaptive
454 regression splines to predict the distribution of New Zealand's freshwater diadromous fish.
455 *Freshwater Biol* 50, 2034-2052.

456 Linke, S., Norris, R.H., Faith, D.P., Stockwell, D., 2005. ANNA: A new prediction method for
457 bioassessment programs. *Freshwater Biol* 50, 147-158.

458 Magalhães M.F., Batalha, D.C., Collares-Pereira M. J., 2002. Gradients in stream fish
459 assemblages across a Mediterranean landscape: contributions of environmental factors and
460 spatial structure. *Freshwater Biol* 47, 1015-1031.

461 Magalhães, M.F., Beja, P., Schlosser, I.J., Collares-Pereira, M.J., 2007. Effects of multi-year
462 droughts on fish assemblages of seasonally drying Mediterranean streams. *Freshwater Biol*
463 52, 1494-1510.

464 Meffe, G.C., Minckley, W.L., 1987. Persistence and stability of fish and macroinvertebrate
465 assemblages in a repeatedly disturbed Sonora desert stream. *Am MidlNat* 117, 177-191.

466 Melcher, A., Schmutz, S., Haidvolg, G., Moder, K., 2007. Spatially based methods to assess the
467 ecological status of European fish assemblage types. *Fisheries Manag Ecol* 14, 453-463.

468 Ministerio de Medio Ambiente, 2005. Caracterización de los tipos de ríos y lagos. Análisis de
469 las características de las demarcaciones. Directiva 2000/60/CE. Centro de Estudios
470 Hidrográficos del CEDEX. Madrid, Spain.

471 Munné, A., Prat, N., 2004. Defining river types in a Mediterranean area: a methodology for the
472 implementation of the EU Water Framework Directive. *Environ Manage*, 34 711-729.

473 Munné, A., Prat, N., Solá, C., Bonada, N., Rieradevall, M. 2003. A simple field method
474 for assessing the ecological quality of riparian habitat in rivers and streams: QBR
475 index. *Aquat Conserv* 13, 147-163.

476 Oberdorf, T., Pont, D., Hugueny, B., Porcher, J-P., 2002. Development and validation of a fish-
477 based index for the assessment of river health in France. *Freshwater Biol* 47, 1720-1734.

478 Omernik, J.M., Bailey, R.G., 1997. Distinguishing between watersheds and ecoregions. *J Am*
479 *Water Resour As* 33, 935-949.

480 Padial, A.A., Siqueira, T., Heino, J., Vieira, L.C.G., Bonecker, C.C., Lansac-Tôha, F.A.,
481 Rodrigues, L.C., Takeda, A.M., Train, S., Velho, L.F.M., Bini, L.M. 2012. Relationships

482 between multiple biological groups and classification schemes in a Neotropical floodplain.
 483 *Ecol Indic* 13, 55–65.

484 Potenciano de las Heras, A. 2004. Estudio de las inundaciones históricas del Río Amarguillo
 485 (Toledo). Doctorate Thesis (in Spanish). Universidad Complutense de Madrid, Madrid.
 486 ISBN 84-669-2587-2.

487 R Development Core Team, 2004. A language and environment for statistical computing. R
 488 Foundation for Statistical Computing, Viena, Austria. ISBN 3-900051-07-0, URL
 489 <http://www.R-project.org>.

490 Rabeni, C.F., Doysi, K.E., Galat, D.L., 2002. Testing the biological basis of a stream habitat
 491 classification using benthic invertebrates. *Ecol Appl* 12, 782-796.

492 Rahel, F.J., Hubert, W.A., 1991. Fish assemblages and habitat gradients in a rocky-mountain
 493 great-plains stream-biotic zonation and additive patterns of community change. *T Am Fish*
 494 *Soc* 120, 319-332.

495 Reynoldson, T.B., Norris, R.H., Resh, V.H., Day, K.E., Rosenberg, D.M., 1997. The reference
 496 condition: a comparison of multimetric and multivariate approaches to assess water-quality
 497 impairment using benthic macroinvertebrates. *J N Am Benthol Soc* 16, 833-852.

498 Sály, P., Erós, T., Takács, P., Specziár, A., Kiss, I., B író, P., 2009. Assemblage level monitoring
 499 of stream fishes: The relative efficiency of single-pass vs. double-pass electrofishing. *Fish*
 500 *Res* 99, 226–233.

501 Sánchez-Montoya, M. M., Punt í, T., Suárez, M.L., Vidal-Abarca, M.R., Rieradevall, M.,
 502 Poquet, J.M., Zamora-Muñoz, C., Robles, S., Á lvarez, M., Alba-Tercedor, J., Toro, M.,
 503 Pujante, A.M., Munné, A., Prat, N., 2007. Concordance between ecotypes and
 504 macroinvertebrate assemblages in Mediterranean streams. *Freshwater Biol* 52, 2240-2255.

505 Simpson, J.C., Norris, R.H., 2000. Biological assessment of river quality: development of
 506 AUSRIVAS models and outputs. In: Wright, J.F., Sutcliffe, D.W., Furse, M.T. (Eds.),
 507 Assessing the biological quality of fresh waters; RIVPACS and other techniques.
 508 Freshwater Biological Association, Ambleside, Cumbria, U.K, pp. 125–142.

509 Snelder, T. H., Cattaneo, F, Suren, A.M., Biggs, B., 2004. Is the River Environment
 510 Classification an improved landscape-scale classification of rivers? *J N Am Benthol Soc* 23,
 511 580-598.

512 Urrea, G., Sabater, S., 2009. Epilithic diatom assemblages and their relationship to environmental
 513 characteristics in an agricultural watershed (Guadiana River, SW Spain). *Ecol Indic* 9, 693-
 514 703.

515 Van Sickle, J., Hawkins, C.P., Larsen, D.P., Herlihy, A.T., 2005. A null model for the expected
516 macroinvertebrate assemblage in streams. *J N Am Benthol Soc* 24, 178-191.

517 Van Sickle, J., Hughes, R.M., 2000. Classification strengths of ecoregions, catchments and
518 geographic clusters for aquatic vertebrates in Oregon. *J N Am Benthol Soc* 19, 370-384.

519 Van Sickle, J., Larsen, D.P., Hawkins, C.P., 2007. Exclusion of rare taxa affects performance of
520 the O/E index in bioassessment. *J N Am Benthol Soc* 26, 319-331.

521 Verdonschot P.F.M., Nijboer R.C., 2004. Testing the European stream typology of Water
522 Framework Directive for macroinvertebrates. *Hydrobiologia* 516, 37–55.

523 Vörösmarty C. J., P. B. McIntyre, M. O. Gessner, D. Dudgeon, A. Prusevich, P. Green, S.
524 Glidden, S. E. Bunn, C. A. Sullivan, C. Reidy Liermann, and & P. M. Davies. 2010.
525 Global threats to human water security and river biodiversity. *Nature*, 467: 555-561.

526 Welborn, G.A., Skelly, D.K., Werner, E.E., 1996. Mechanisms creating community structure
527 across a freshwater habitat gradient. *Annu RevEcol Syst* 27, 337-363.

528 Wright, J.F., 1995. Development and use of a system for predicting the macroinvertebrate fauna
529 in flowing waters. *Aust J Ecol* 20, 1881-197.

530

Table 1. Set of environmental variables used to describe environmental characteristics. * Denotes those variables possibly affected by human perturbation and not included as predictors in ANNA and RIVPACS models, but used to identify reference sites following the method proposed in Hermoso et al. (2009). Reference sites covered a wide range of environmental conditions as mean values in non-perturbed variables show.

Scale	Variable	Method	Mean Reference	Mean Perturbed	Range
Site	Water depth (cm)	<i>In situ</i>	40.2	40.0	7.0-200
	Refuge cover availability (m ² of shelter/river width)	<i>In situ</i>	5.0	5.9	0.0-60.6
	Elevation (m) 1	GIS	393.0	392.7	7.1-974.9
	Relative position (dist. to the most headwater point/total length of the stream) 2	GIS	0.4	0.5	0.04-1.00
	Stream order (Strahler) 2	GIS	2.1	2.1	1.0-6.0
	Distance to headwater (km) 2	GIS	97.9	63.2	3.6-1,036.1
	Distance to Guadiana River (km) 2	GIS	71.7	52.7	0.0-196.0
	River width (m) *	<i>In situ</i>	12.6	10.4	1.4-123.0
	Substrate coarseness (Wentworth scale) *	<i>In situ</i>	6.0	4.9	1.0-9.0
	Riparian Quality Index (QBR, Munne et al. [2003]) *	<i>In situ</i>	79.8	54.3	0-100
	NH ₄ ⁺ (mg/L) *	<i>In situ</i>	0.36	1.78	0.02-51.60
	NO ₂ ⁻ (mg/L) *	<i>In situ</i>	0.04	0.13	0.01-2.00
	NO ₃ ⁻ (mg/L) *	<i>In situ</i>	2.74	4.63	0.50-55.90
	PO ₄ ³⁻ (mg/L) *	<i>In situ</i>	0.50	1.20	0.05-23.20
	SO ₄ ²⁻ (mg/L) *	<i>In situ</i>	98.1	123.8	10.0-2380.0
	Cl ⁻ (mg/L) *	<i>In situ</i>	35.5	72.7	2.0-834.0
	Water temperature (°C) *	<i>In situ</i>	19.6	20.9	9.4-32.6
	Conductivity (µS/cm) *	<i>In situ</i>	471.8	684.9	38.0-3230.0
	pH *	<i>In situ</i>	7.7	7.8	2.21-10.63
	Annual precipitation (mm/m ²) 3	GIS	675.7	554.0	370.2-1114.5
	Solar radiation (10 KJ/m ² *day*µm) 3	GIS	2033.8	2034.5	1646-2227
	Average annual air temperature (°C) 3	GIS	15.7	15.8	13.0-18.0
	% Exotic fish abundance*	<i>In situ</i>	1.0	50.0	0-100
Reach (500 m)	Slope (‰) 1	GIS	8.9	4.6	0.00-58.03
	Sinuosity 2	GIS	1.2	1.2	1.00-2.79
	Land uses 4				
	Urban/Industrial (%) *	GIS	5.0	10.0	0.0-36.0
	Intensive agriculture (%) *	GIS	9.0	36.0	0.0-100.0
	Extensive agriculture (%) *	GIS	3.0	8.0	0.0-100.0
Basin	Natural (%) *	GIS	86.0	54.3	0.0-100.0
	Basin area (Drainage surface upstream of each site, 10 ³ km ²) 1	GIS	4.1	2.3	0.9-5919.1
	Gravelius index 1	GIS	1.6	1.7	1.14-2.68
	Land uses 4				
	Urban/Industrial (%) *	GIS	4.1	4.6	0.0-6.7
	Intensive agriculture (%) *	GIS	13.1	27.0	0.0-97.0
	Extensive agriculture (%) *	GIS	7.8	12.5	0.0-89.1
	Natural (%) *	GIS	78.5	59.6	0.9-100.0
	Reservoir (%) *	GIS	0.15	0.43	0.0-21.2
	Population density (inhabitant/km ²)*5	GIS	15.6	23.6	0.0-459.3

Data sources

- 1 Digital Elevation Model (90 m resolution). Confederación Hidrográfica del Guadiana.
- 2 Stream network provided by the Confederación Hidrográfica del Guadiana.
- 3 Atlas Climático Digital de la Península Ibérica (Ninyerola et al., 2005)
- 4 CORINE Land-Cover 1:100.000. Confederación Hidrográfica del Guadiana.

5 Instituto Nacional de Estadística, available at www.ine.es (May 2006).

Table 2. Results of the biological significance of environmental classification analysis. It is shown the partial within-class similarity for each water body type considered (W_i) (all except Continental-Mediterranean mineralized rivers with no reference locality), mean all within-class (W) and between-class (B) similarity.

Water body type	W_i	B
Main rivers in Mediterranean environments	0.71	0.48
Low Mediterranean siliceous mountains rivers	0.50	
Guadiana and Tajo siliceous plain rivers	0.57	
Rivers of “La Mancha”	0.85	
Siliceous rivers of Sierra Morena Low-Foothills	0.47	
W	0.56	

Table 3. Parameters describing the O/E regression line for the validation data set (n=20 sites) for the three different predictive approaches. The upper and lower baselines for model performance measured as SD of the best possible model and the null model *sensu* Van Sickle et al. (2005) is also given between parentheses. Since the validation data set was the same for all the models, these values are shared by all of them. * Denotes parameters for which no significant differences were found ($p < 0.05$) respect the expected value for a good fit (1 and 0 for slope and intercept respectively).

Predictive method	Slope	Intercept	R ²	SD (0.38-0.45)
Top-down RIVPACS	0.28	1.37	0.05	0.41
Bottom-up RIVPACS	1.36	-0.89	0.22	0.40
ANNA	1.06*	0.06*	0.43	0.39

Figure 1. Location of the Guadiana River basin and the extent of the six top-down environmental classes defined using the classification system B specified in the Water Framework Directive.

Figure 2. Multi Dimensional Scaling (MDS) ordination plots for the 57 reference sites in the Guadiana River basin. Different environmental classes are denoted as follow: Main rivers in Mediterranean environments, open squares; Low Mediterranean siliceous mountains rivers, open circles; Guadiana and Tajo siliceous plain rivers, black diamonds; Rivers of “La Mancha, black circles; Siliceous rivers of Sierra Morena Low-Foothills, black triangles. The centroid of each class (average within-class MDS scores) are represented with larger symbols for each class. Bars show within-class Standard Deviation. Environmental variables selected in each multiple regression model are shown on each MDS axis. Model fit (R^2) is also given for each model.