An expert system based on fish assemblages for evaluating the ecological quality of streams and rivers

Michele Scardi[a],*, Stefano Cataudella[a], Paola Di Dato[b], Eugenio Fresi[a], Lorenzo Tancioni[a]

[a]Department of Biology, 'Tor Vergata' University of Rome, Via della Ricerca Scientifica, 00133 Rome, Italy
[b]Department of Ecology and Sustainable Economic Development, University of Tuscia, Via S. Giovanni Decollato 1, 01100 Viterbo, Italy

ARTICLE INFO

ABSTRACT

Fish species have been often used as indicators of environmental quality in aquatic ecosystems, while biotic indices based on fish have become common tools in ecological monitoring. Nevertheless, such indices are far from perfect, mainly because they are based on assumptions that sometimes are not met and because they cannot be optimized from a computational point of view. As any other method, they rely upon expert judgments for selecting relevant metrics, combining metrics into a score and defining thresholds between ecological status classes in the scoring scale. Provided that no procedure can be entirely objective in evaluating ecological status, as this very concept is inherently subjective, we propose a novel approach in which the unavoidable subjective elements only play a role in the earliest steps, while the subsequent optimization of the evaluation procedure is as objective as possible. An expert system, designed after this concept for Latium (Central Italy) river basins and based on a multilayer perceptron neural network, was developed and implemented into a Graphical User Interface (GUI) in order to make it easily accessible to non-technical users. The neural network reconstructs experts' judgments on the basis of a set of abiotic descriptors and fish assemblage composition, thus providing consensus estimates of ecological status for any river stretch. This approach allows easily the incorporation into the expert system of new data and new expert judgments as soon as they become available. However, the very first version of the expert system is already able to correctly classify 2 out of 3 cases, while the worst classification error does not exceed a single class of ecological status.

© 2007 Elsevier B.V. All rights reserved.

1. Introduction

Fish species have been regarded as very effective biological indicators of environmental quality in different aquatic ecosystems (Fausch et al., 1990; Whitfield, 1996), not only because of their iconic value, but also because of their sensitivity to subtle environmental changes (Karr, 1981). Obviously, fish responses to environmental disturbances, including hydromorphological ones, are different in time and space from those of simpler organisms, as they tend to be integrated over larger intervals (Scardi et al., 2006).

The relevance of fish species as biotic indicators has been explicitly mentioned not only in scientific studies (e.g. Karr and Dudley, 1981; Oberdorff and Hughes, 1992), but also in European and American laws and regulations (European Commission, 1992; European Union, 2000; Kurtz et al., 2001) as well as in those from other countries. In particular, fish are explicitly mentioned among the biological "quality elements" that are to be considered to assess the ecological status of surface waters according to the European Water Framework Directive (WFD) (European Union, 2000).

Several environmental assessment methods based on fish have been developed during the last two decades, mostly

* Corresponding author.
E-mail address: mscardi@mclink.it (M. Scardi).

inspired to the seminal work by Karr (1981), who developed the Index of Biotic Integrity (IBI). The IBI is a multimetric index, as it combines into a single score different variables (i.e. *metrics*) that are thought to respond to environmental disturbance. The original IBI metrics included twelve ecological attributes of fish assemblages related to species richness, assemblage composition and abundance, trophic guilds and fish condition.

The multimetric IBI approach is very flexible, as different metrics can be selected according to regional ecological conditions, and this is certainly the main reason why it has been adapted to a number of countries and river basins, not only in North America (e.g. Karr and Dudley, 1981; Karr et al., 1986; Plafkin et al., 1989; Fausch et al., 1990), but also in Europe (e.g. Hughes and Oberdorff, 1999) and in other continents (e.g. Steedman, 1988; Lyons et al., 1995; Kleynhans, 1999; Harris, 1995; Hugueny et al., 1996; Hay et al., 1996; Kamdem Toham and Teugels, 1999; An et al., 2002).

While many IBI clones and other biotic indices were developed during the last two decades, other methods were also proposed. Although they may differ from the IBI paradigm because of more complex scoring criteria or because they take into account not only the fish assemblage composition, but also some environmental attributes (e.g. geomorphological types), in most cases they are still based on a multimetric approach. The European Fish Index (EFI) is a recent example of such a second generation multimetric index (FAME Consortium, 2004).

Multimetric indices, although very popular, are not the only available solution for evaluating the environmental quality of streams and rivers. In fact, some interesting attempts have been also made by comparing the observed fish assemblages to those modeled on the basis of relevant physical variables, thus obtaining indices based on deviation from expectation of the fish fauna composition (e.g. Oberdorff et al., 2002). This approach involves the selection of reference sites and the application of multivariate analyses, and it seems more accurate than multimetric indices (Reynoldson et al., 1997). Multivariate methods, however, have been developed almost exclusively for benthic macroinvertebrates so far.

While (multimetric) biotic indices have become very common ecological tools, even the most successful ones have been criticized (e.g. Suter, 1993; Karr and Chu, 1999; Norris and Hawkins, 2000). Criticisms stem from very diverse reasons, like, for instance, lack of diagnostic capacity, unpredictable interactions between different metrics (e.g. same final score resulting from different arrangements of the single metrics scores), the inherent circularity of selecting indicators that are supposed to respond to human pressures. Moreover, Reynoldson et al. (1997) pointed out that multimetric indices do not exploit all the collected information, that metrics are often redundant and that errors can be compounded.

The accuracy of methods for evaluating ecological status (as well as any other ecological property) always decreases when they are applied at large spatial scale. This is not surprising, because no model (or method, or index, in this case) can be simple, general and accurate at the same time. In order to be simple, and multimetric indices are designed to be inherently simple, a method can be general, but not accurate, or it can be accurate, but not general.

Biotic indices are not exceptions to this rule, and this is the reason why they usually have to be calibrated at regional scale or at river basin scale in order to provide accurate results. This is the case, for instance, of the original IBI concept, which has been adapted to a number of different (eco)regions and river basins, resulting in a large number of local implementations (Miller et al., 1988). Obviously, all these implementations share the same common rationale, although the metrics that are taken into account change from case to case. On the other hand, indices which are aimed at generality always fail when applied to many different ecoregions. The problems that have been experienced with the EFI in several European countries demonstrate the limited accuracy of a multimetric index that was designed to be both general and simple (e.g. de Sostoa et al., 2004; FAME Consortium, 2004).

The simplicity versus accuracy (or generality) tradeoff, which applies to any model, or concept, or method, is the reason why the multimetric approach, being inherently simple, cannot be accurate at large spatial scale (i.e. when it has to be also general). In particular, the simplicity of all multimetric approaches involves the assumption that each metric is linearly or monotonically related to the ecological status (e.g. Karr, 1981; FAME Consortium, 2004), even though linear or monotonic responses are the exception, not the rule in ecology. For instance, species distributions along environmental gradients are usually unimodal, and very often even more complex because of interspecific competition. However, when only a small portion of an environmental gradient is taken into account, species responses can be regarded as locally linear. Likewise, the assumption of linear or monotonic relationships between metrics and ecological status that supports multimetric indices can be approximately met at small spatial scales (e.g. at the scale of a single river basin), whereas it is most certainly violated at larger scales (e.g. throughout more diverse ecoregions).

The case of species richness, a very common metric (Kennard et al., 2006), is a typical example of such a problem. While in multimetric indices species richness (overall, native or referred to a specific taxonomic group or guild) is assumed to be positively and monotonically related to the ecological status, it is well known that a moderate disturbance usually favors an increase in species richness, and the intermediate disturbance hypothesis is the formal expression of this familiar evidence (Connell, 1978). These effects are often caused by hydromorphological pressures, but other factors, like, for instance, competition between species, may induce even more complex biotic responses.

A major problem with (multimetric) biotic indices is the lack of computational optimization. In fact, they usually take into account only a handful of metrics, thus causing an information loss, while the selected metrics are heuristically processed. For instance, in many cases indices are obtained by summing up scores assigned to each metrics, under the assumption that all metrics have the same weight, that no metrics is redundant and that no interactions between metrics exist. It is obvious that these assumptions are quite simplistic and seldom compliant with reality. As a matter of fact, (multimetric) biotic indices are often conceptually sound as far as the underlying ecological rationale is concerned, but they exploit only part of the available information, and in a suboptimal way.

The lack of computational optimization of multimetric indices, however, may be not really relevant as far as accuracy is concerned. As a matter of fact, a good implementation of a

multimetric index, especially when developed at regional scale, can be enough accurate for most practical purposes and the gain in accuracy that can be obtained using alternate approaches might be negligible. On the contrary, the lack of computational optimization severely affects the development phase of multimetric and other biotic indices, which is time consuming and usually based on trial-and-error procedures, especially when the scoring scale has to be properly gauged. This implies that different developers may produce very different solutions based on the same data set.

In spite of these problems, the index-based approach is very popular and often perceived as an objective procedure. In reality, though, it involves many steps that are based on expert judgment (e.g. metrics selection, data transformations, definition of thresholds in index scoring scale, etc.), because expert judgment is the key for any environmental assessment, evaluation or diagnosis. In fact, the very concept of ecological status (or environmental quality, ecosystem health, etc.) is at the same time very clear in the field and absolutely vague when it has to be translated into a set of rules, or into a concise definition, exactly like the concept of beauty. And, like beauty, ecological status is not an emergent property of ecosystems, nor a property that can be univocally defined (Scrimgeour and Wicklum, 1996): on the contrary, it is based on the personal interpretation of a natural phenomenology. Nevertheless, ecologists who share a common theoretical background (a common ecological aesthetics?) usually agree in ranking a set of sites according to their ecological status.

Although it plays a fundamental role, expert judgment might be controversial in applications that involve not only ranking, but also setting critical boundaries. For instance, experts not always agree with each other in case slightly disturbed sites (i.e. in "good" ecological status according to the European WFD) have to be recognized and separated from moderately impaired sites.

We tried to address this problem by developing an expert system based on a neural network, which was trained to associate expert judgments to environmental and fish assemblage data. In essence, this solution is based on the assumption that the complex biotic relationships that link fish assemblage composition to environmental conditions can be implicitly embedded into a neural network and that such a neural network can be trained to reproduce consensus expert judgment.

In this paper we present the first implementation of such method, based on a regional data set. To this aim, we trained a working demo of the expert system using data from Latium river basins (Central Italy). However, the same method can be easily retrained as soon as other data become available and we are actually planning to update and extend the expert system on a routine basis.

2. Materials and methods

Environmental data and fish assemblage composition were recorded in 62 sites in Latium streams and rivers (Central Italy) during summer 2005. Sampling sites were located in the Tevere river basin as well as in other minor river basins (Sacco, Marta and Mignone). Most sampling sites were wadable, and sampling was performed by means of standard electrofishing gear.

Table 1 – List of the fish assemblage descriptors: 30 fish species and 2 species richnesses (overall and juveniles only) are included

Species/faunistic variable	
Abramis brama	Leuciscus cephalus
Alburnus alburnus alborella	Leuciscus lucumonis
Alosa fallax	Leuciscus souffia muticellus
Anguilla anguilla	Liza ramada
Barbus plebejus/tyberinus	Mugil cephalus
Carassius carassius	Petromyzon marinus
Chondrostoma genei	Pseudorasbora parva
Cobitis taenia bilineata	Rutilus rubilio
Cyprinus carpio	Rutilus rutilus
Dicentrarchus labrax	Salaria fluviatilis
Esox lucius	Salmo trutta
Gambusia holbrooki	Sander lucioperca
Gasterosteus aculeatus	Scardinius erythrophthalmus
Gobius nigricans	Tinca tinca
Lampetra fluviatilis	Overall species richness
Lampetra planeri	Juveniles only species richness

Species that are not explicitly mentioned are taken into account in species richnesses.

Fish assemblage composition was recorded as the number of fish caught in the sampling stretch for each species, but only binary presence/absence data have been used in our expert system so far, because we aimed at maximum compatibility with previous qualitative data sets. Although 43 species were identified in our samples, only 30 of them, i.e. those which occurred in more than 5% of the records, were explicitly considered (Table 1), whereas the remaining species were only taken into account when computing the overall species richness. A separate species richness value was computed for juveniles (young of the year) only. Information about 27 environmental variables (Table 2) was also recorded during fish sampling. Most of these variables were selected because they had been already considered in previous studies, but some of them might be discarded in future implementations of the expert system on the basis of a sensitivity analysis.

Finally, expert judgments were recorded. They were meant as a global evaluation of the sampling sites ecological status, not just as an evaluation of the fish fauna composition, and they were issued by postdoctoral researchers with more than 10 years of experience in the study area. When possible, more than a single expert operated at each sampling site, thus allowing association of more than a single expert judgment to the same environmental and fish assemblage data. Expert judgments were fuzzy coded, i.e. a membership value was recorded for each ecological status class in the *high*, *good*, *moderate*, *poor* and *bad* range. This way of coding ecological status is very flexible, as it also allows expressing uncertainty. For instance, in case it was not possible to discriminate between *moderate* and *good* ecological status (a very difficult task and a very relevant problem in European countries because of the WFD) a 50% *good* and 50% *moderate* expert judgment was issued.

Although our goal was to develop just a working demo of the expert system, the number of available records was still not large enough to properly train a multilayer perceptron neural network. Therefore, more information was added to the training set by simulating changes to real records that could affect the expert judgment. In other words, the same experts

Table 2 – List of the environmental descriptors: 27 variables are included

Environmental variable	Unit/Notes	
Elevation	m	
Depth	m	
Runs	% of the wetted surface	
Pools	% of the wetted surface	
Riffles	% of the wetted surface	$\sum_{i=1}^{4} p_i = 100\%$
Uniform flux	% of the wetted surface where the water flow is apparently uniform at surface (relevant in larger rivers)	
Wetlands	presence or absence of wetlands connected to the river in normal flow conditions (binary variable)	
Bars or islands	presence or absence (binary variable)	
Boulders	% of the sampling site surface	
Rocks and pebbles	% of the sampling site surface	
Gravel	% of the sampling site surface	$\sum_{i=1}^{5} p_i = 100\%$
Sand	% of the sampling site surface	
Silt and clay	% of the sampling site surface	
Flow velocity	semiquantitative score in the [0,5] range	
Vegetational cover	% of the surface of the sampling site covered by aquatic macrophytes	
Shade	% of the surface of the sampling site shaded at noon	
Anthropic disturbance	semiquantitative score in the [0,4] range	
Dams upstream	distance in km (use 100 in case there are no dams)	
Dams downstream	presence or absence (binary variable)	
Lake upstream	distance in km (use 50 in case there are no lakes)	
Summer water temperature	°C (mid-June to August)	
Turbidity	NTU	
pH		
Conductivity	$\mu S \ cm^{-1}$	
Dissolved oxygen	% saturation	
Basin area	Square root of the basin area in km^2	
Distance from source	km	

Units and notes for each variable are shown in the right column.

who evaluated the ecological status of the sampling sites were requested to think about species whose presence or absence would change their judgment or to environmental variables that, if changed in value, would affect their evaluation.

For instance, given a site in which the fish assemblage included both adults and juveniles of the species that are supposed to be present on the basis of the environmental information, the overall evaluation would certainly become less positive in case less or no juveniles were found. Basically, such *virtual* records were used to add relevant information that was not found in the available real records, but they are obviously based on ecological scenarios that are very likely to occur. In this first attempt we added 157 virtual records to our data base, i.e. we derived one to three virtual records from each real record. We also tried to balance the number of records that were associated to each class of ecological status, adding virtual records to this end. This way no class of ecological status included less than 40 records.

Before training the neural network, the available records were divided into two subsets, for training ($n=150$) and test ($n=69$). Records were assigned to training or test subset on the basis of a stratified procedure. At first several strata were defined according to combinations of ecological status class and sampling site elevation, then records to be assigned to each data subset were randomly selected from each stratum.

The overall number of input variables in our data set was 59, including 27 environmental variables, 30 species and 2 values for species richness (overall and juveniles), whereas the output variables were 5, corresponding to membership values for each class of ecological status defined by the European WFD.

The best architecture for the multilayer perceptron was selected on the basis of a heuristic test. In fact, we trained neural networks with a number of nodes in the hidden layer ranging from 10 to 50 and then we selected the one which provided the best results. On the basis of this procedure, the final architecture of the multilayer perceptron neural network was set to 59–25–5. Different architectures, however, provided very similar results.

All the environmental variables and the species richness values were normalized into the [0,1] interval. Obviously, binary data for species occurrence did not require normalization. Sigmoid activation functions were used in the hidden layer nodes $[f(x)=1/(1-e^{-x})]$, whereas *softmax* activation functions (Bridle, 1990) were used in the output layer nodes. The *softmax* function scales the neural network outputs so that their sum is 1 and that each output can be regarded as a probability, i.e., in this particular case, as a membership value for each class of ecological status.

The neural network training was performed by means of the most common algorithm, i.e. the error back-propagation (Rumelhart et al., 1986). A constant value was set for both the learning rate (0.9) and the momentum (0.1), while overtraining was avoided using an early stopping strategy. The neural network training was performed according to a "learning per pattern" paradigm. Moreover, training patterns were submitted to the neural network in random order at each learning epoch, thus avoiding that memorization of the submission order could adversely affect the training. Finally, jittering, i.e. addition of a small amount of noise to input patterns at each epoch (Györgyi, 1990), was performed during the training phase. Gaussian noise with $\mu=0$ and $\sigma=0.01$ was used. Jittering helps neural network generalization by providing a virtually unlimited number of artificial training patterns that are closely related, even though not exactly identical, to the original ones.

The confusion matrix obtained from the neural network output using a *winner-takes-all* strategy for the classification of the test records was analyzed by means of the weighted Kappa statistics (Cohen, 1960; Fleiss et al., 1969). Neural network outputs were also analyzed after conversion into a continuous variable by weighted average. In this case a linear correlation coefficient (Pearson, 1896) was computed to evaluate the neural network performance.

Finally, a Graphical User Interface (GUI) was wrapped around the trained neural network, thus providing a user-friendly and interactive access to the expert system (Fig. 1). The GUI was designed in order to make the neural network completely transparent to the user, who can interact with the expert system in real time, observing changes in classification while input data are modified.

3.   Results

In spite of the problems related with the *curse of dimensionality*, the synaptic weights of the trained 59–25–5 multilayer perceptron did not grow too large (88% of the weights were <0.5 and 97% of them <1.0) and the overall response of the neural network to changes in input data was very smooth. Given the very small ratio between the number of training patterns (150) and the number of synaptic weights (1630), a proper training was theoretically impossible, but ecological problems are often exceptions to this rule. In fact, their real dimensionality is usually much smaller than expected, especially due to associations between species whose response to environmental variables is similar and to tight relationships between environmental variables.

For instance, the 30 species that were selected as neural network binary inputs might combine in 2^{30} different ways, i.e. in more than one billion of different fish assemblages. In practice, only 107 (partly) different fish assemblages were found in 219 records. Together with appropriate training strategies (jittering, early stopping, random selection of training patterns at each epoch) these ecological constraints to the theoretical dimensionality of the problem minimized the effects of the *curse* and made it possible to properly train a neural network. An indirect evidence of training stability was also provided by the very small differences that were observed between neural networks with different numbers of nodes in the hidden layer.

The comparison between neural network outputs and expert judgments can be carried out in different ways, depending on the defuzzyfication method. In fact, the *softmax* activation functions of the output layer nodes returned output values that summed up to 1, which could be regarded as memberships for each ecological status class. These membership values, however, were relative to ordered categorical variables and therefore they could be either used to compute a continuous score, or to define the most probable class.

The first solution is the most straightforward, as a continuous ecological status score (ESS), ranging from 1 (best) to 5 (worst), can be easily obtained as a weighted average of the class membership values:

$$ESS = \sum_{k=1}^{5} k \cdot p_k$$

where k is the class rank and p_k is the class membership value.

In the GUI the ESS can be found in the horizontal bar that is located above the histogram in the lower right corner. The ESS is 1.8 in the screenshot shown in Fig. 1. Next to the ESS another numerical value is also shown, i.e. the Environmental Quality
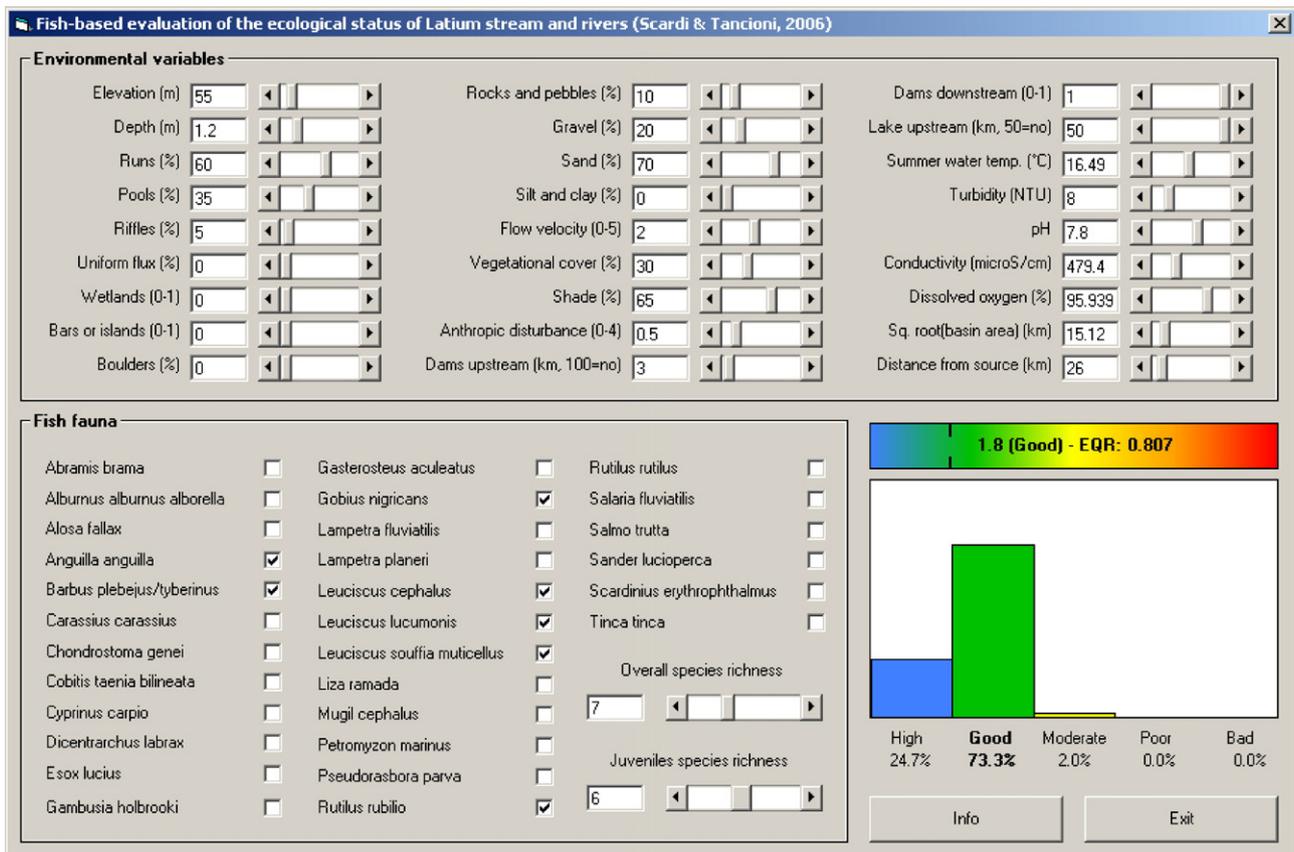

Fig. 1 – The Graphical User Interface (GUI) of the expert system. The classification results are shown in the lower right corner both graphically and alphanumerically: the histogram shows the membership values for each class of ecological status, while the values in the horizontal bar indicate the ESS and the EQR. In the histogram, the ecological status class selected by the *winner-takes-all* criterion has boldface labels. The GUI is fully interactive and the classification is instantly updated at each change in the input data (ecological status classes are color-coded).
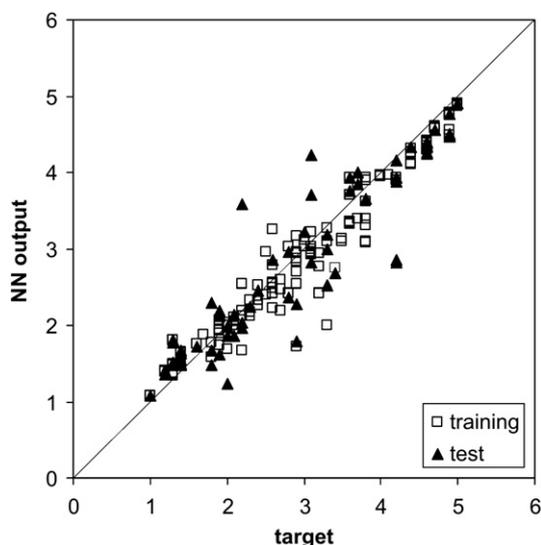
Fig. 2 – Comparison between expert judgments and neural network outputs. Results are shown as ESS, i.e. as weighted averages obtained from class membership values (1 = best ecological status, 5 = worst ecological status). Both training (white squares, $r=0.978^{***}$) and test (black triangles, $r=0.932^{***}$) data are shown.

Ratio (EQR), which is 0.807 in Fig. 1. The EQR is explicitly required by the European Water Framework Directive as the basis for classifying the ecological status. It ranges from 1, in case of an ecological status close to pristine or undisturbed, to 0, in case of a heavily disturbed ecological status. In theory, the EQR is to be computed with respect to some pre-defined reference conditions, but in our expert system the reference conditions are implicitly defined by the expert judgments used for training the neural network, and therefore the EQR can be easily obtained by reversing the ESS and scaling it into the [0,1] interval.

In Fig. 2 the ESS values obtained from the neural network are compared to the ESS values computed on the basis of actual expert judgments. Both the training (white squares) and the test (black triangles) data subsets are shown. Differences between the two data subsets are minimal, and the linear correlation between neural network outputs and known values is very high in both cases ($r=0.978^{***}$ for the training data subset and

Table 3 – Confusion matrix obtained from the *winner-takes-all* classification of the training data subset

Expert judgment	Neural network					Total
	High	Good	Moderate	Poor	Bad	
High	19	21	0	0	0	40
Good	2	40	4	0	0	46
Moderate	0	14	35	2	0	51
Poor	0	0	9	31	0	40
Bad	0	0	0	6	36	42
Total	21	75	48	39	36	219

Correctly Classified Instances (CCI) were 73.5%, while the weighted Kappa statistics was $K_w=0.822$ ($p<0.001$).

Table 4 – Confusion matrix obtained from the *winner-takes-all* classification of the test data subset

Expert judgment	Neural network					Total
	High	Good	Moderate	Poor	Bad	
High	5	7	0	0	0	12
Good	2	15	1	0	0	18
Moderate	0	5	6	2	0	13
Poor	0	0	2	11	0	13
Bad	0	0	0	4	9	13
Total	7	27	9	17	9	69

Correctly Classified Instances (CCI) were 66.7%, while the weighted Kappa statistics was $K_w=0.775$ ($p<0.001$).

$r=0.932^{***}$ for the test data subset). While extreme ESS values are very accurately estimated by the neural network, a few larger errors are associated to intermediate ESS values. This is not surprising, as the characterization of intermediate classes of ecological status is inherently more controversial.

Even though the ESS assessment seems very accurate, a classification into discrete levels of ecological status is required in most practical applications. Such a classification can be obtained from the neural network output according to different criteria. The most straightforward solution is to round off the ESS to the closest integer value, but this solution is also potentially biased, because the ESS might be influenced by an asymmetrical distribution of the neural network outputs around the most probable ecological status class. In this case, even small membership values for extreme classes might bias the overall estimate of the ESS.

A safer alternate solution is the classical *winner-takes-all* strategy. In other words, each pattern is classified according to the highest class membership value in the neural network output. This strategy seems more robust in case of asymmetrical class membership distributions and it also has the advantage of being more intuitive than others. From a theoretical point of view, it can be affected by bimodal (or trimodal) distributions of class membership values, but such particular cases were never observed while classifying real records.

Confusion matrices based on the *winner-takes-all* strategy were obtained for both the training and the test data subset (Tables 3 and 4). Data in the test subset were never used in the training phase, and therefore they could be safely used to assess the ability of the neural network to reproduce the way expert judgments are issued. The percentage of Correctly Classified Instances (CCI) in the test data subset (66.7%) is not very different from the one in the training data subset (73.5%). Basically, if only test data were taken into account, 2 out of 3 cases were exactly classified, while the remaining cases were misclassified by only one class of ecological status.

The agreement between expert judgments and neural network outputs was tested by means of the weighted Kappa statistics. The weighted version of the Kappa statistics was selected because of the ordered categorical nature of the classification and the following weighting scheme was adopted:

$$w_{i,j} = \frac{|i-j|}{r-1}$$

where $w_{i,j}$ is the weight for the *j*th element of the *i*th row of the confusion matrix and *r* is the rank of the confusion matrix.

The weighted Kappa statistics for the confusion matrix obtained from the test data subset was highly significant (K_w=0.775, p<0.001) and the overall agreement between expert judgments and neural network outputs could be considered as "good" according to Landis and Koch (1977). The same statistics for the confusion matrix obtained from the training data subset was only slightly higher (K_w=0.822, p<0.001).

It is worth noticing that the weighted Kappa statistics for the confusion matrix based on the alternate classification criterion for the ecological status, i.e. on the rounding off of the ESS to the closest integer value, was lower, although still highly significant (K_w=0.715, p<0.001) for the test data subset.

4. Discussion and conclusions

Relationships between environmental variables and fish assemblage composition in streams and rivers are quite tight, and they have been already successfully modeled using statistical approaches (e.g. Oberdorff et al., 2001) as well as neural networks (e.g. Boet and Fuhs, 2000; Joy and Death, 2002; Mastrorillo et al., 1997; Olden and Jackson, 2001; Scardi et al., 2004, 2005). The expert system we developed, however, goes a step farther, because it leverages an Artificial Intelligence approach in the strictest sense, i.e. for reproducing as closely as possible the behavior of human experts.

Conventional strategies for evaluating the ecological status of streams and rivers on the basis of fish fauna include biotic (multimetric) indices and comparisons between observed and expected fauna (either modeled or found in reference sites). Although useful in many cases, especially when applied at regional scale, these methods are not computationally optimized and, although often perceived as objective, unavoidably rely upon choices that are as subjective as the very concept of ecological status (or environmental quality, etc.). Nevertheless, recent developments in environmental laws and regulations of most countries demand ecological status classification procedures to be applied in routine environmental monitoring activities.

Although we are fully aware that such requests imply an oversimplification of the underlying ecological problems, our expert system is aimed at obtaining the best classification procedure by focusing on expert judgments only at the earliest step, accepting their inherent subjectivity and then processing all the relevant biotic and abiotic information as objectively as possible. In short, we aim at mimicking as closely as possible the way human experts issue a consensus expert judgment.

In fact, the multilayer perceptron neural network we trained provided very good results, accurately reproducing expert judgments, even though it was only aimed at demonstrating the feasibility of an Artificial Intelligence approach and therefore the training data set was not nearly as large as needed. In spite of these limits, the neural network classification of test cases closely matched the expert judgments, with a limited number of misclassifications (1 out of 3 cases), which were never worse than a single ecological status class. Moreover, the similarity in terms of CCI and weighted Kappa statistics between the results obtained with training and test data subsets showed that the neural network was properly generalized.

The software implementation of the expert system is not a minor feature in our opinion, and the GUI that makes our neural network transparent to users also plays a major role as far as the acceptance of the expert system is concerned. Most users are already acquainted with biotic indices, while they are not comfortable with a "black box" approach, even though they realize that it makes sense from a theoretical point of view. Interacting with the expert system through the GUI, on the other hand, is very intuitive. For instance, changing input values by means of sliders, which are not strictly needed nor convenient for routine data input, help users to learn how the system reacts to changes in diagnostic variables. In other words, sliders help users to understand the way of reasoning of the expert system and to recognize similarities with their own point of view.

A recurrent criticism to an Artificial Intelligence approach to the evaluation of ecological status is that a lot of data are needed and that other methods are therefore more feasible in data-limited situations. Although the need for data is absolutely true, it is certainly false that more conventional approaches need less information to be correctly developed. In fact, collecting enough relevant information is the basis for any evaluation procedure, from the simplest, i.e. human expert judgment, to the most complex ones. And biotic indices cannot be regarded as exceptions to this rule.

However, information is available in many different forms, and not only field data can be considered as useful information. Knowledge is equally (and probably even more) important, and it can be easily elicited from experts and converted into virtual data records, which may contribute relevant information to the training data set for a neural network. In practice, simulating changes in real records that might affect the expert judgment is a very effective way for eliciting experts' knowledge and transferring it into a neural network with no filters or reinterpretations. Of course, the role of such virtual records will become less important as soon as more data will be available, but they will always play a useful role in transferring knowledge to the expert system.

In conclusion, a reductionist, index-based approach to a problem as complex as the evaluation of ecological status of streams and rivers (assuming that this very concept really makes sense) cannot be fully successful, except in particular cases, like single river basins or very homogeneous ecoregions. As a matter of fact, given the elusive nature of the very concept of ecological status (or environmental quality, health, etc.), the only sensible way to evaluate the overall state of an ecosystem is by expert judgment, which can integrate all the available information and provide a competent diagnosis. Of course, expert judgment is subjective by definition and, due to our limited grasp of ecological processes, potentially biased. Therefore, an expert system based on a typical Artificial Intelligence approach may assist human experts by defining a consensus reference or, if no human experts are available, it may successfully surrogate them.

REFERENCES

An, K.-G., Park, S.-S., Shin, J.-Y., 2002. An evaluation of a river health using the index of biological integrity along with relations to chemical and habitat conditions. Environ. Int. 28, 411–420.

Boet, P., Fuhs, T., 2000. Predicting presence of fish species in the Seine river basin using artificial neuronal networks. In: Lek, S., Gueguan, J.-F. (Eds.), Artificial neuronal networks: application to ecology and evolution. . Environ. Sci. Springer-Verlag, pp. 187–201.

Bridle, J.S., 1990. Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In: Fogleman Soulie, F., Herault, J. (Eds.), Neurocomputing: algorithms, architectures and applications. Springer-Verlag, Berlin, pp. 227–236.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20, 27–46.

Connell, J.H., 1978. Diversity in tropical rainforests and coral reefs. Science 199, 1302–1310.

de Sostoa, A., Caiola, N., Casals, F., 2004. A new IBI (IBICAT) for local application of the E.U. water framework directive. In: Garcia de Jalón, D., Vizcaíno, P. (Eds.), Aquatic Habitats: Analysis and Restoration, IAHR, Madrid, pp. 187–191.

European Commission, 1992. Council Directive 92/43/ECC of 21st May 1992 on the conservation of natural habitats and of wild fauna and flora. Official Journal L206. European Commission, Brussels.

European Union, 2000. Parliament and Council Directive 2000/60/EC of 23rd October 2000. Establishing a framework for community action in the field of water policy. Official Journal PE-CONS 3639/1/00 REV 1, 2000. European Union, Brussels.

FAME Consortium, 2004. Manual for the application of the European Fish Index – EFI. A fish-based method to assess the ecological status of European rivers in support of the Water Framework Directive. Version 1.1, January 2005. Available online at http://fame.boku.ac.at/downloads/manual_Version_Februar2005.pdf.

Fausch, K.D., Lyons, J., Karr, J.R., Angermeier, P.L., 1990. Fish communities as indicators of environmental degradation. Am. Fish. Soc. Symp. 8, 123–144.

Fleiss, J.L., Cohen, J., Everitt, B.S., 1969. Large sample standard errors of kappa and weighted kappa. Psychol. Bull. 72, 323–327.

Györgyi, G., 1990. Inference of a rule by a neural network with thermal noise. Phys. Rev. Lett. 64, 2957–2960.

Harris, J.H., 1995. The use of fish in ecological assessments. Aust. J. Ecol. 20 (1), 65–80.

Hay, C.J., Van Zyl, B.J., Steyn, G.J., 1996. A quantitative assessment of the biotic integrity of the Okavango River, Namibia, based on fish. Water S.A. 22, 263–284.

Hughes, R.M., Oberdorff, T., 1999. Application of IBI concepts and metrics to waters outside the United States and Canada. In: Simon, T.P. (Ed.), Assessing the Sustainability and Biological Integrity of Water Resources using Fish Communities. CRC Press, LLC, pp. 79–93.

Hugueny, B., Camara, S., Samoura, B., Magassouba, M., 1996. Applying an index of biotic integrity based on fish assemblages in a West African river. Hydrobiologia 331, 71–78.

Joy, M.K., Death, R.G., 2002. Predictive modelling of freshwater fish as a biomonitoring tool in New Zealand. Freshw. Biol. 47, 2261–2275.

Kamdem Toham, A., Teugels, G.G., 1999. First data on an Index of Biotic Integrity (IBI) based on fish assemblages for the assessment of the impact of deforestation in a tropical West African river system. Hydrobiologia 397, 29–38.

Karr, J.R., 1981. Assessment of biotic integrity using fish communities. Fisheries 6, 21–27.

Karr, J.R., Dudley, D.R., 1981. Ecological perspective on water quality goals. Environ. Manag. 5, 55–68.

Karr, J.R., Chu, E.W., 1999. Restoring life in running waters: better biological monitoring. Island Press, Washington D.C.

Karr, J.R., Fausch, F.D., Angermeier, P.L., Yant, P.R., Schlosser, I.J., 1986. Assessing biological integrity in running waters: a method and its rationale. Illinois Natural History Survey, vol. 5. Special Publication. Illinois Natural History Survey, Champaign, IL, USA.

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

Kleynhans, C.J., 1999. The development of a fish index to assess the biological integrity of South African rivers. Water S.A. 25 (3), 265–278.

Kurtz, J.A., Jackson, L.E., Fisher, W.S., 2001. Strategies for evaluating indicators based on guidelines from the Environmental Protection Agency's Office of Research and Development. Ecol. Indic. 1, 49–60.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33, 159–174.

Lyons, J., Navarro-Pérez, S., Cochran, P.A., Santana, E., Guzmán-Arroyo, M., 1995. Index of biotic integrity based on fish assemblages for the conservation of streams and rivers in westcentral Mexico. Conserv. Biol. 9, 569–584.

Mastrorillo, S., Lek, S., Dauba, F., Belaud, A., 1997. The use of artificial neural networks to predict the presence of small-bodied fish in a river. Freshw. Biol. 38, 237–246.

Miller, D.L., Hughes, R.M., Karr, J.R., Leonard, P.L., Moyle, P.B., Schrader, L.H., Thompson, B.A., Daniels, R.A., Fausch, K.D., Fitzhugh, G.A., Gammon, J.R., Halliwell, D.B., Angermeier, P.L., Orth, D.J., 1988. Regional applications of an Index of Biotic Integrity for use in water resource management. Fisheries 13, 12–20.

Norris, R.H., Hawkins, C.P., 2000. Monitoring river health. Hydrobiologia 435, 5–17.

Oberdorff, T., Hughes, R.M., 1992. Modification of an index of biotic integrity based on fish assemblages to characterize rivers of the Seine Basin, France. Hydrobiologia 228, 117–130.

Oberdorff, T., Pont, D., Hugueny, B., Chessel, D., 2001. A probabilistic model characterizing fish assemblages of French rivers: a framework for environmental assessment. Freshw. Biol. 46, 399–415.

Oberdorff, T., Pont, D., Hugueny, B., Porcher, J.P., 2002. Development and validation of a fish-based index for the assessment of 'river health' in France. Freshw. Biol. 47, 1720–1734.

Olden, J.D., Jackson, D.A., 2001. Fish-habitat relationships in lakes: gaining predictive and explanatory insight by using artificial neural networks. Trans. Am. Fish. Soc. 130, 878–897.

Pearson, K., 1896. Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia. Philos. Trans. R. Soc. Lond., A 187, 253–318.

Plafkin, L.P., Barbour, M.T., Porter, K.D., Gross, S.K., Hughes, R.M., 1989. Rapid bioassessment protocols for use in streams and rivers: benthic macroinvertebrates and fish. U.S. Environmental Protection Agency, Washington, DC.

Reynoldson, T.B., Norris, R.H., Resh, V.H., Day, K.E., Rosenberg, D.M., 1997. The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. J. North Am. Benthol. Soc. 16 (4), 833–852.

Rumelhart, D.E., Hinton, G.E., Williams, G.E., 1986. Learning representations by back-propagating errors. Nature 323, 533–536.

Scardi, M., Cataudella, S., Ciccotti, E., Di Dato, P., Maio, G., Marconato, E., Salviati, S., Tancioni, L., Turin, P., Zanetti, M., 2004. Previsione della composizione della fauna ittica mediante reti neurali artificiali. Biol. Ambient. 18, 1–8.

Scardi, M., Cataudella, S., Ciccotti, E., Di Dato, P., Maio, G., Marconato, E., Salviati, S., Tancioni, L., Turin, P., Zanetti, M., 2005. Optimisation of artificial neural networks for predicting fish assemblages in rivers. In: Lek, S., Scardi, M., Verdonschot, P.F., Descy, J.P., Park, Y.S. (Eds.), Modelling community structure in freshwater ecosystems. Springer-Verlag, Berlin, pp. 114–129.

Scardi, M., Tancioni, L., Cataudella, S., 2006. Monitoring methods based on fish. In: Ziglio, G., Siligardi, M., Flaim, G. (Eds.), Biological monitoring of rivers. Wiley, pp. 135–153.

Scrimgeour, C.J., Wicklum, D., 1996. Aquatic ecosystem health and integrity: problems and potential solutions. J. North Am. Benthol. Soc. 15, 254–261.

Steedman, R.J., 1988. Modification and assessment of an index of biotic integrity to quantify stream quality in southern Ontario. Can. J. Fish. aquat. Sci. 45, 492–501.

Suter II, G.W., 1993. A critique of ecosystem health concepts and indexes. Environ. Toxicol. Chem. 12, 1533–1539.

Whitfield, A.K., 1996. Fishes and the environmental status of South African estuaries. Fisheries Manag. Ecol. 3, 45–57.