Modeling the impact of landscape types on the distribution of stream fish species

Muriel Gevrey, Frédéric Sans-Piché, Gaël Grenouillet, Loïc Tudesque, and Sovan Lek

Abstract: Modifications of the landscape adjoining streams perturb their local habitat and their biological diversity, but little quantitative information is available on land cover classes that influence the fish species individually. Data collected from 191 sites in the Adour–Garonne Basin (France) were analyzed to assess the effects of land cover on the distribution of fish species. A multimodel approach was carried out to predict fish species using land cover classes and to define the most important classes applying a hierarchical filtering based on artificial neural network method and sensitivity analysis. Firstly, using three single-class models, a selection of the land cover subclasses contributing the most was carried out for each fish species and each class. Secondly, multiclass models were built with all the previously selected subclasses to predict each species (*n*-selected subclass model). Finally, the percentages of contribution for artificial, agricultural, and forest areas obtained for the different model architectures (three class, *n*-selected subclass, and global multiclass models) were compared. The majority of the distribution of fish species was correctly predicted by the single-class models, and different land cover subclasses have been selected depending on the species. Using the *n*-selected subclass models, the predictive performances were globally better than those obtained with other multiclass models.

Résumé : L'habitat local d'un cours d'eau et sa biodiversité sont perturbés par les modifications du paysage adjacent, pourtant il existe peu d'information quantitative sur les classes d'occupation du sol qui influencent les espèces piscicoles individuellement. Des données de 191 sites réparties sur l'ensemble du bassin Adour–Garonne (France) ont été analysées pour évaluer les effets de l'occupation du sol sur la distribution d'espèces de poissons. Une approche « multi-modèles » utilisant la technique des réseaux de neurones artificiels et une analyse de sensibilité a été menée pour déterminer si les espèces étaient efficacement prédites par les classes d'occupation du sol et pour définir celles qui impactaient le plus la distribution des espèces. Premièrement, en utilisant trois modèles à « classe-unique », une sélection des sous-classes d'occupation du sol contribuant le plus a été effectuée pour chaque espèce de poissons et chaque classe. Deuxièmement, des modèles « multi-classes » ont été construits avec toutes les sous-classes sélectionnées précédemment afin de prédire chaque espèce (modèle à *n* sous-classes sélectionnées). Finalement, les pourcentages de contribution des classes territoires artificialisés, territoires agricoles et milieux semi-naturels et forêts ont été comparés pour les différentes architectures de modèles (trois classes, *n* sous-classes sélectionnées et multi-classes global). La distribution de la majorité des espèces a été correctement prédite par les modèles à classe unique et différentes sous-classes d'occupation du sol ont pu être sélectionnées selon l'espèce considérée. En utilisant le modèle à n sous-classes sélectionnées, les performances prédictives étaient globalement meilleures que celles obtenues avec les autres modèles multi-classes.

Introduction

Human actions on the landscape are recognized as being a principal threat to the ecological integrity of river ecosystems, impacting especially on the biota (Strayer et al. 2003; Allan 2004*a*). Even if a large number of studies have shown that the biota can be explained by local variables (e.g., macroinvertebrates (Lancaster and Hildrew 1993), fish (Gore and Nestler 1988)), observing some variations in the habitat along short stretches of a river can give limited conclusions on the impact of these variables on the biota of the whole river (Roth et al. 1996; Sandin and Johnson 2004; Mykra et al. 2007). While Allan and Johnson (1997), using a catchment approach in midwestern USA, concluded that biotic integrity, based on the fish assemblages, was strongly influenced by landscape changes, Townsend et al. (2003) also reviewed studies where large-scale factors such as land use were strongly related to spatial variation in the invertebrate assemblages. Landscapes are altered on multiple scales (Allan 2004*b*), and the ecological consequences must be identified and predicted at these scales to understand the anthropogenic impact on the environment, to manage the natural resources (Fausch et al. 2002).

Streams ecosystems are affected by the human influence on the landscapes, such as deforestation (Jones et al. 1999) and urban (Roy et al. 2003; Walton et al. 2007) or agricultural (Meador and Goldstein 2003) land use. Different studies have tried to understand how landscape variables

Received 9 April 2008. Accepted 21 November 2008. Published on the NRC Research Press Web site at cjfas.nrc.ca on 6 March 2009. J20500

M. Gevrey,[1] F. Sans-Piché, G. Grenouillet, L. Tudesque, and S. Lek. Evolution et Diversité Biologique (EDB), UMR 5174 CNRS – Université Paul Sabatier, 118, route de Narbonne, 31062 Toulouse CEDEX 4, France.

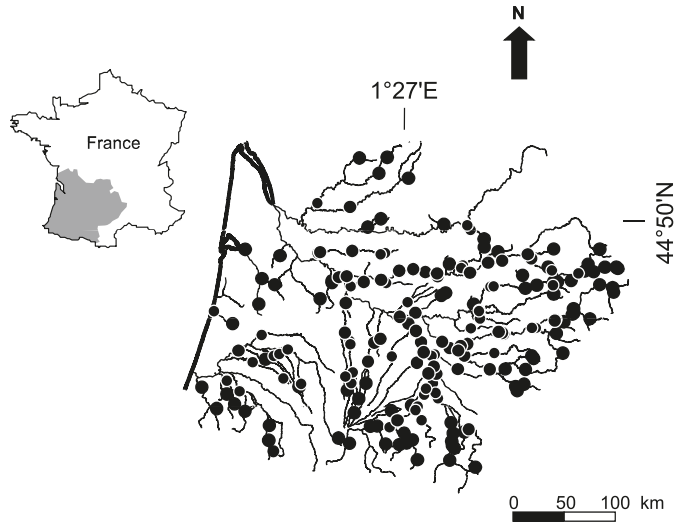
[1]Corresponding author (e-mail: gevrey@cict.fr).

influence the physical, chemical, and biological properties of freshwater systems (reviewed in Allan and Johnson 1997).

Modifications of the landscape do not act directly on fish populations. Their effect on stream morphology, sediment transport, and riparian vegetation will in turn influence the biota present (Jowett et al. 1996). A knowledge of the species distribution and the capacity to predict the occurrence of fish are crucial in managing aquatic biodiversity and assessing the consequences of anthropogenic changes on the river environment (Pont et al. 2005). The development of predictive models to explain biological condition in terms of landscape variables will directly benefit stream conservation and resource management activities (Olden et al. 2006) and will be efficient biomonitoring tools to achieve the high ecological status of rivers requested by the European water framework directive (European Union 2000).

A combination of several aspects in this research area makes this study particularly relevant. (*i*) It is a species-specific approach. The effects on fish community structure of forest, urban, and agricultural land uses have been widely investigated. Forest cover seems to have a positive effect on the fish assemblages (Steedman 1988; Maret et al. 1997; Jones et al. 1999), while agriculture use and urbanization lead to degradation (Allan et al. 1997; Walser and Bart 1999; Onorato et al. 2000). In a previous study, Park et al. (2006) demonstrated the importance of the percentage of agricultural land in the prediction of the fish community. However, Pont et al. (2005) clearly demonstrated that sensitivity to local and regional scale processes is species-specific. Even if the community approach using all species at once has been investigated (Olden et al. 2006) to explore the sensitivity of each species, it has never been compared with a species-by-species approach using the same modeling method demonstrating the superiority of one or the other. Our choice is to use the species-specific approach to better identify the relative influence of land cover on fish species distribution. (*ii*) The approach is not only predictive but also explanatory, aiming at identifying the most important variables driving fish species distribution. This is a response to Allan (2004*b*), who points out that there are limitations in the use of catchment-scale studies to link landscape to stream conditions and that more progress is needed in developing explanations for statistical associations between landscape variables and stream conditions. Land cover categories are used in this study instead of the usual agricultural, forest, and urban land cover as described in Park et al. (2006). (*iii*) It is based on artificial neural network models (ANNs) incorporating ecological knowledge. Firstly, the necessity of models to understand how changes in land cover affect stream ecosystems is now obvious (Strayer et al. 2003). Secondly, ANNs have been chosen as modeling technique. This method is able to identify nonlinear responses (a limitation noticed by Allan (2004*b*) for linking the landscape to the stream condition), does not require any assumptions concerning the statistical distributions of the variables, and often outperforms other techniques when faced with complex problems (Hilbert and Ostendorf 2001; Pearson et al. 2004). Moreover, the standard criticisms, which are that a large quantity of data is required to train and test the networks and that the causal relationships between the dependent and independent variables are not identified, can be avoided by using a special training method called leave-one-out with small data set (Spitz and Lek 1999) and using further analysis such as sensitivity analysis to increase the explanatory power of the approach (Gevrey et al. 2003). Finally, the idea suggested by Olden et al. (2006) of developing more ecologically relevant ANNs better adapted to the problems is used, and an innovative model is presented.

Fig. 1. Distribution of the sampling sites (●) along the Adour–Garonne catchment area (inset shows location in southwestern France).



This study focused on the effect of land cover on the distribution of fish species in the Adour–Garonne Basin, France. The main objectives were to (*i*) address the land cover – fish species linkage, (*ii*) examine which land cover features influenced fish species distribution, and (*iii*) provide a methodological framework aiming at selecting the most important land cover features to use as explanatory variables by using a model comparison.

Materials and methods

Site description

A total of 191 stream sites were sampled in the Adour–Garonne Basin (Fig. 1). The Adour–Garonne hydrographic network covers the southwest Atlantic areas of France. It extends over 116 000 km^2 from Charentes and the Massif Central to the Pyrenees, gathering 120 000 km of watercourses, including 68 000 km of permanent rivers running out towards the Atlantic Ocean. The Garonne is the main channel, spanning over 580 km from the Pyrenees to the Gironde estuary on the Atlantic coast. The Adour–Garonne watershed presents a large range of altitudes and geological substrates (Tison et al. 2005). From south to northwest, topography and climate determine three major landscape types: the pyrenean mountains, with a pronounced relief; a large green hill zone of piedmont; the valley of the River Garonne with flooding zones and alluvial terraces (see Park et al. (2006) for further details).

Table 1. List of the 20 species used in the analysis with their Latin and common names as well as the abbreviations used in the paper and their prevalence (ratio of the number of fish species occurrences to the number of sampling sites).

Abbreviation	Latin name	Common name	Prevalence
gog	*Gobio gobio*	Gudgeon	0.79
php	*Phoxinus phoxinus*	Eurasian minnow	0.68
lec	*Leuciscus cephalus*	European chub	0.62
bab	*Barbus barbus*	Barbel	0.61
sat	*Salmo trutta fario*	Brown trout	0.61
rur	*Rutilus rutilus*	Roach	0.53
ana	*Anguilla anguilla*	European eel	0.46
lel	*Leuciscus leuciscus*	Common dace	0.45
ala	*Alburnus alburnus*	Bleak	0.41
pef	*Perca fluviatilis*	European perch	0.31
leg	*Lepomis gibbosus*	Pumpkinseed (common sunfish)	0.29
cht	*Chondrostoma toxostoma*	South-west European nase	0.29
esl	*Esox lucius*	Northern pike	0.24
abb	*Abramis brama*	Common bream	0.23
tit	*Tinca tinca*	Tench	0.23
cog	*Cottus gobio*	Bullhead	0.19
sal	*Sander lucioperca*	Pike-perch	0.16
lap	*Lampetra planeri*	European brook lamprey	0.16
icm	*Ictalurus melas*	Black bullhead	0.14
sce	*Scardinius erythrophthalmus*	Rudd	0.14

Fish species data

The data gathered was from two data sets from the fish database of the Aquatic Environment Team, School of Agronomy at Toulouse (ENSAT), and the French national office for water and aquatic environment (ONEMA). Depending on water depth at the sampling sites, electrofishing surveys were done either by wading in shallow areas or by boat in the deeper reaches. In the case of wider and deeper rivers, gill-netting was used in still waters and both gill- and drift-netting for running waters. This combination of methods allows an effective assessment of fish diversity in rivers (Seegert 2000). To remove sampling bias, only presence–absence data were considered (Hughes and Gammon 1987). Samples from the 191 sites were collected between 1986 and 1996. Details of the fish sampling are given in Park et al. (2006).

In the data set, 34 fish species were identified. Fish species that occurred at only a small percentage of the sites (i.e., <10% of sampling sites) were removed from the analysis (Waite 2000). The final data set contained 20 species (Table 1).

Land cover quantification

A geographical information system (GIS) (Arcview 3.1; ESRI, Redlands, California, USA) was used to quantify the relative percentage (%) of different land cover classes. For each site, the boundaries were determined using a digital elevation model, and then the hydrographic basin areas were obtained. Theses surfaces were overlapped with the Corine Land Cover 2000 layer (Coordination of Information on the Environment), based on automated interpretation of Landsat ETM+ satellite images and data integration with existing digital maps. The land cover data were extracted for each sampling site (www.eea.europa.eu/themes/landuse/clc-download).

The Corine classification system is made up of three hierarchical levels, the first one being composed of five main classes: (*i*) artificial surfaces, (*ii*) agricultural areas, (*iii*) forest and seminatural areas, (*iv*) wetlands, and (*v*) water bodies. The second level contains 11 classes. The third level, which is the most detailed, includes 44 classes. In this study, only the three first level 1 classes (hereafter called ARTI for artificial surfaces, AGRI for agriculture areas, and FOR for forest and seminatural areas) were included in the analysis, as wetlands and water bodies were sparsely represented in the Adour–Garonne Basin (around 0.07% and 0.14%, respectively, on average per site). These three level 1 classes were divided into 10, 8, and 12 level 3 subclasses, respectively, thus providing 30 subclasses (i.e., the third hierarchical level of Corine Land Cover 2000, see Table 2 for details).

ANNs for prediction

The relationship between the occurrence of fish species and land cover classes for the 191 sites was identified using a feed-forward multilayer perceptron trained using the back-propagation error algorithm (BP) (Rumelhart et al. 1986). The BP algorithm is a supervised learning algorithm designed to minimize the mean square error between the predictions computed by the network and the observations.

Network architecture

The network is composed of three layers of neurons: the input layer, the hidden layer, and the output layer. The neurons of each layer are connected to all the neurons of adjacent layers. The neurons receive and send signals through these connections, which are assigned a weight. These weightings modulate the intensity of the signal they transmit. The network's final configuration, resulting in a specific number of neurons and a specific learning rate, is

Table 2. List of the Corine land cover classes, constituting three hierarchical levels.

Corine level 1	Corine level 2	Corine level 3
ARTI	Urban fabric	c111: continuous urban fabric
		c112: discontinuous urban fabric
	Industrial, commercial, transport units	c121: industrial or commercial units
		c122: road and rail networks and associated land
		c124: airports
	Mine, dump, and construction sites	c131: mineral extraction sites
		c132: dump sites
		c133: construction sites
	Artificial, nonagriciultural, vegetated areas	c141: green urban areas
		c142: sport and leisure facilities
AGRI	Arable land	c211: nonirrigated arable land
	Permanent crops	c221: vineyards
		c222: fruit trees and berry plantations
	Pastures	c231: pastures
	Heterogenous agricultural areas	c241: annual crops associated with permanent crops
		c242: complex cultivation patterns
		c243: land principally occupied by agriculture, with major areas of natural vegetation
		c244: agro-forestry areas
FOR	Forest	c311: broad-leaved forest
		c312: coniferous forest
		c313: mixed forest
	Shrub–herbaceous vegetation associations	c321: natural grassland
		c322: moors and heathland
		c323: sclerophyllous vegetation
		c324: transitional woodland–shrub
	Open spaces with little or no vegetation	c331: beaches, dunes, and sand plains
		c332: bare rock
		c333: sparsely vegetated areas
		c334: burnt areas
		c335: glaciers and perpetual snow

Note: ARTI, artificial surfaces; AGRI, agricultural areas; FOR, forest and seminatural areas.

arrived at by testing various possibilities and selecting the one that provides the best compromise between bias and variance (Kohavi 1995).

Algorithm

Two main steps comprise the algorithm: (*i*) the propagation of the signal from the input layer to the output layer, with the objective of computing the observed output and (*ii*) the backpropagation of the signal from the output layer back to the input layer to modify the weighting values. These two steps are repeated or iterated many times in what is called an epoch to minimize the error between the network output and the dependent variable or the desired output until specified end conditions are reached.

Model validation

Cross-validation is a method to assess the prediction accuracy on independent data (Verbyla and Litaitis 1989). This technique consists of dividing the data matrix into several parts, one to train the model and a second one to monitor network errors after each training cycle, enabling the training to be stopped when the network begins to be overtrained or over-fit the data (Tarassenko 1998). Finally, the calibrated model is tested on data that are completely unknown to the model.

Model evaluation

In case of presence–absence prediction, the simplest and most widely used measure of the accuracy of the overall classification is the number of correctly classified cases (hereafter called the correct assignment score). Several metrics can be used to test the sensitivity or the specificity of the model (see Joy and Death 2004 for details). Cohen's Kappa statistic of similarity (Cohen 1960) is a commonly used statistic that provides a measure of proportional accuracy, adjusted for chance agreement. This test provides a robust evaluation of a model's performance relatively independent of the species frequency, occurrence, or prevalence (Manel et al. 2001; Liu et al. 2005.)

Adaptations applied in the current study

Several decisions were taken in this study concerning the way to use the method. (*i*) Model architecture: after several tests concerning the changes of predictive power according to the number of hidden neurons, five neurons in the hidden layer were used for all the models. (*ii*) Model validation: be-

cause of the number of sampling sites (191), the matrix has been divided into two parts: two-thirds (127) of the data was used to train the model and one-third (64) was used to validate it. (*iii*) Model evaluation: as mentioned previously, the models can be evaluated in the case of presence–absence prediction using the correct assignment score. However, the output given by the model is a continuous value between 0 and 1. Usually a threshold of 0.5 is chosen. If the predicted value is higher, the species is considered present; otherwise it is considered absent. It has been demonstrated that this threshold could be detrimental to the quality of the results. In one of the rare studies comparing different approaches with determining thresholds, Liu et al. (2005) verified the hypothesis that a good presence–absence prediction would be obtained by taking the prevalence of model-building data as the threshold. Simple and effective, the prevalence approach appeared to be at least as good as the more complicated approaches (e.g., sensitivity-specificity sum maximization approach). Because of the data set partition, we used the training data set prevalence as threshold. Moreover, the correct assignment score is useful to decide if a model is correct or not only with the association of a Kappa test. In this study, the limiting Kappa value was 0.4. (*iv*) Model accuracy: to be certain that the results obtained were not simply due to chance, 500 replicates were run for each model, with each different, randomly chosen sampling sites in the training and in the test data set. This bootstrapping procedure was used to test the confidence of the model's results.

The importance of predictor variables

It is necessary to apply a supplementary method to the trained network to find the most important inputs that explain species presence or absence. There are many ways to perform a sensitivity analysis (see Gevrey et al. 2003 for details), but the connection weight method (Olden and Jackson 2002) was selected in this study to check the contribution of the variables.

The connection weight method was also applied to each of the 500 random models (replicates) used in the prediction step to test the stability of the contributions obtained. The average percentage of the relative contribution of each variable was then recorded to rank the variables by order of importance. Moreover, when a selection of variables was needed, it was decided that only the variables with a contribution higher than 10% were used (Brosse et al. 2003). A Tukey's honestly significant difference (HSD) test for multiple comparisons was used after an analysis of variance (ANOVA). The Tukey's HSD test is a post hoc test designed to perform a pairwise comparison of the means to see where the significant difference is. It was used here to select, among the variables, those for which the contributions could be lower than 10% but not significantly different from those with a contribution higher than 10%.

The modeling process

The total number of land cover subclasses used in this study was 30. All these subclasses could be used in the input of a predictive model; however, the contribution of the input will be very difficult to interpret because of the large number of subclasses that should share 100% of the contribution. The differences of contribution between several subclasses could then be very difficult to define. The aim of this modeling process was to avoid this problem using (*i*) three single-class models for each species corresponding to the three main classes of land cover (ARTI, AGRI, and FOR) to select the most important subclasses in each of these three classes and then (*ii*) building a multiclass model called the "*n*-selected subclass" model using as input the *n*-selected contributing subclasses to predict the species. The process is explained in detail below.

Step 1: subclass selection using single-class models

Each species was predicted using three different single-class models: (*i*) the 10 ARTI subclasses; (*ii*) the 8 AGRI subclasses, and (*iii*) the 12 FOR subclasses (see Table 2 for details). At the end, three models were built for each species. Each model was validated with a score associated to a Kappa value. The connection weights approach was applied to each validated model to select the most important variables (Fig. 2*a*).

Step 2: species prediction using n*-selected subclass models*

The selected variables of the first step were used as the input of an *n*-selected subclass model. Each species was then predicted using this model and the models validated by the score and the Kappa values. The connection weights approach was applied to each validated model to define the importance of each input. The percent contribution for ARTI, AGRI, and FOR was then calculated by summing the contribution of each subclass (Fig. 2*b*).

Multiclass models comparison

To appreciate the results of this *n*-selected subclass model, it was compared with multiclass models that directly use all classes or all subclasses as predictive variables. Two other multiclass models were then used for each species using either the 30 level three subclasses (called 30-subclass model) or the three classes (called 3-class model) at the same time. The predictive abilities of these two approaches were compared with the *n*-selected subclass model using an ANOVA test and a Tukey's HSD test. Applying the connection weight method, the percentage of ARTI, AGRI, and FOR was also calculated for each of the three types of multiclass models as explained in step 2. The percentages obtained were then compared (Fig. 3).

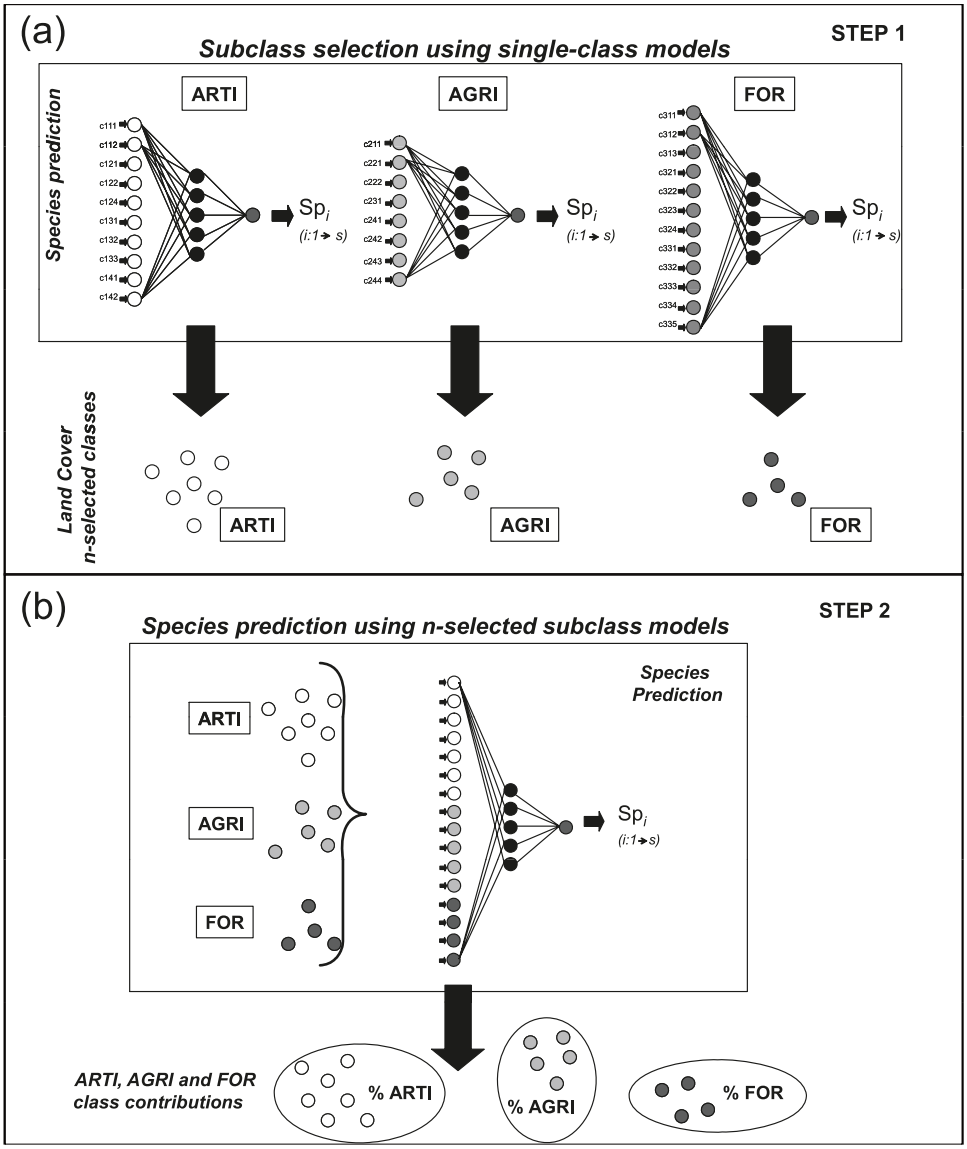
The ANN program written by the authors in Matlab (The MathWorks Inc., Natick, Massachusetts) and several libraries of R statistical program (www.r-project.org/) were used for the analysis.

Results

Step 1: subclass selection using single-class models

The selection of the most powerful model was done using the score and the Kappa value from the 500 replicates obtained per species (the entire results for this step are summarized in Fig. 4). The score values ranged from 0.61 to 0.91. These values were validated by the Kappa values. Among the 20 selected species, two could not be used for the rest of the study owing to too low Kappa values: Eurasian minnow (*Phoxinus phoxinus*) and south-west European

Fig. 2. In the modeling process shown, *s* represents the number of modeled species using artificial neural network symbolized by circles (neurons) and lines (connection weights): (*a*) Step 1: three single-class models are used to predict the fish species (Sp_i) using either ARTI, AGRI, or FOR; the land cover subclasses and the most contributing subclasses for each single-class model are selected using a contribution method. (*b*) Step 2: the fish species are predicted using the selected subclasses of the previous steps (by the *n*-selected subclass model that is a multiclass model) and by applying a contribution method; the contributions in percentages of the classes ARTI, AGRI, and FOR are computed.



nase (*Chondrostoma toxostoma*). Moreover, AGRI and FOR models for common dace (*Leuciscus leuciscus*) and ARTI for bullhead (*Cottus gobio*) were not good enough to be used later on, as the confidence interval of the Kappa values was below the threshold of 0.4. Except for four species, each of the classes was able to predict the species occurrence using the subclasses.

The connection weight method associated to the back-propagation algorithm allowed the selection of the most important land cover subclasses for each single-class model. All classes that contributed more than 10% were considered and validated by a Tukey's HSD test (the results are summarized for all species in Table 3).

Among the 18 fish species that responded to land cover subclasses, bullhead responded to subclasses from two land cover classes (AGRI and FOR) and common dace only from one (ARTI). The others species were all influenced by at least one subclass of each land cover class. All the subclasses considered in this study were at least selected for one species. Among the land cover subclasses selected, those from the AGRI class were on average the most frequently selected ($\bar{x} = 4.3$, $\sigma = 1.9$) following by ARTI ($\bar{x} = 3.94$, $\sigma = 0.75$) subclasses and FOR ($\bar{x} = 2.88$, $\sigma = 1.11$).

In particular, agro-forestry areas (c244), vineyards (c221), fruit trees (c222), and complex cultivation patterns (c242), all subclasses of AGRI, were selected for 14, 13, 10, and 10 fish species, respectively. The most selected ARTI subclasses were dump sites (c132) selected for 12 fish species, followed by road and rail networks (c122) and sport and lei-

Fig. 3. Schematic drawing of the three different types of multiclass models comparison (same symbols as Fig. 2): the *n*-selected subclass model (upper right of panel); the 30-subclass model, which uses all the subclasses available as input, and the 3-class model, which directly uses the three classes as input.

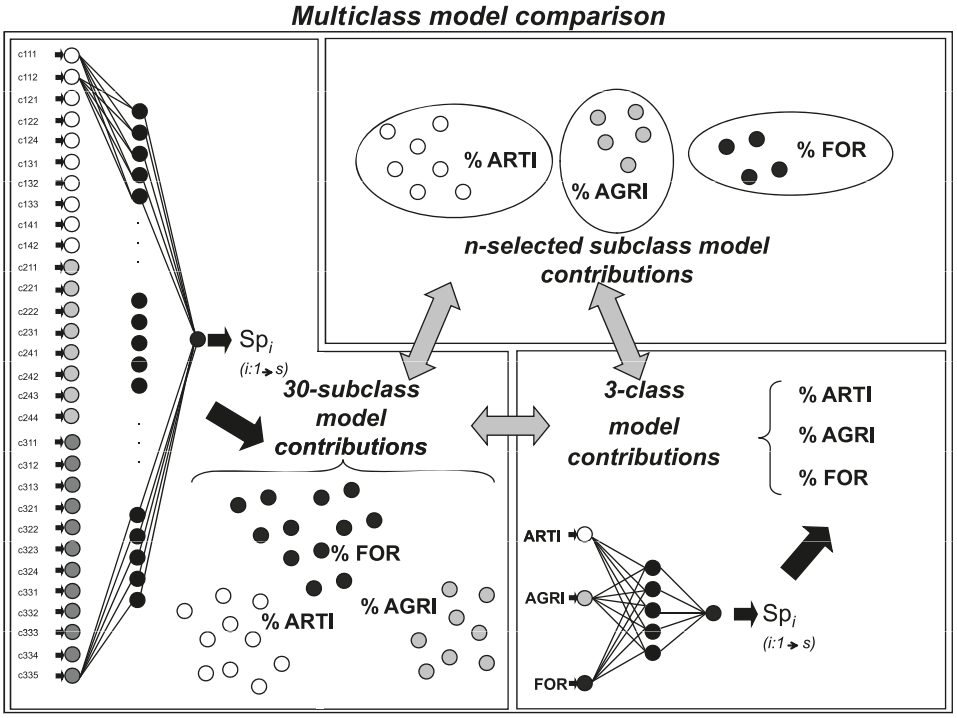


Fig. 4. Cleveland dot plots of the scores and the Kappa values (mean ± standard deviation) obtained using the test data set in the single-class models of (*a*) ARTI, (*b*) AGRI, and (*c*) FOR, results obtained from step 1 (see Fig. 2).

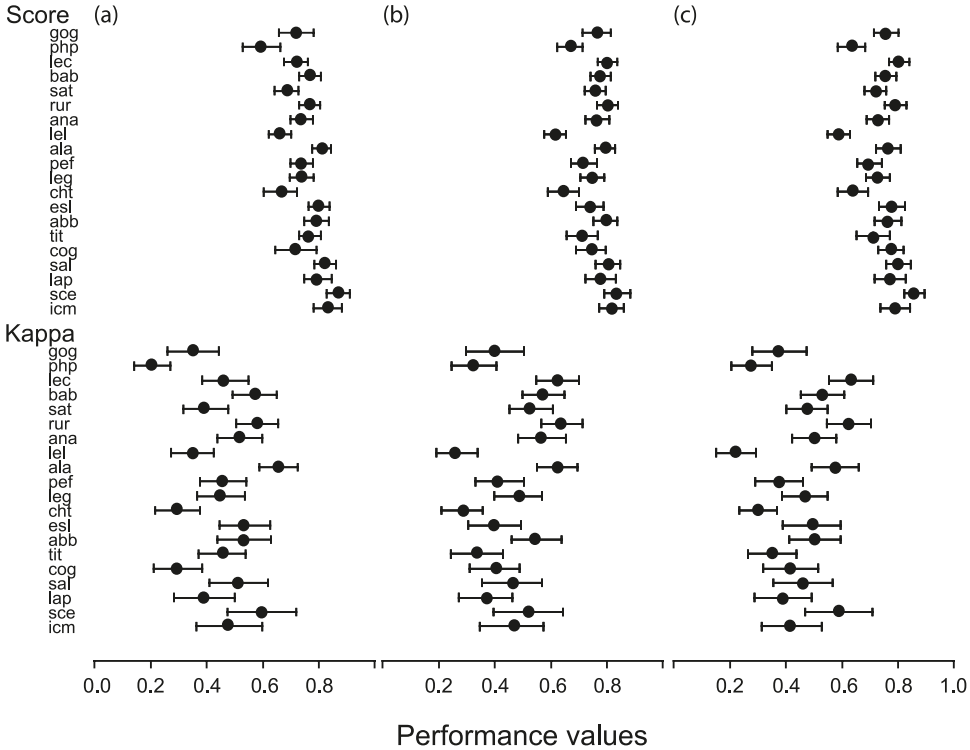


Table 3. Results of the selection of the land cover classes (contribution higher than 10%) for the three single-class models and each species.

Species	ARTI	AGRI	FOR
gog	c121, c131, c112	c242, c243, c244, c222, c211	c323
php	—	—	—
lec	c121, c132, c142, c111	c242, c211, c222	c321, c312, c323, c324
bab	c111, c121, c124, c132, c141	c244, c242, c222, c221	c323, c324, c312, c321
sat	c112, c111, c132, c121	c211, c221	c321, c331
rur	c111, c132, c141	c241, c211, c221, c244	c321, c322, c323, c311
ana	c121, c122, c112, c111	c241, c222, c221	c334, c331, c324
lel	c132, c122, c142, c121	—	—
ala	c111, c122, c141, c132	c241, c244, c221, c222, c242	c321
pef	c141, c142, c132	c221, c244, c242	c323
leg	c141, c142, c122	c221, c244, c211	c321, c323, c334
cht	—	—	—
esl	c132, c142, c124, c112, c133	c222, c241, c244, c242	c335, c323, c334
abb	c122, c111, c141, c132	c244, c221, c222	c335, c323, c312
tit	c122, c142, c132, c133	c221, c243, c211, c222, c242, c244, c241, c231	c335, c312, c334
cog	—	c244, c211	c332, c321, c333
sal	c132, c142, c122, c133, c124	c243, c222, c244, c221, c242, c211, c231, c241	c323, c335
lap	c141, c133, c142	c231, c243, c244, c221, c211, c222, c241, c242	c333, c335, c313, c332
icm	c131, c142, c124, c141, c122	c221, c241, c244	c321, c323, c335, c334
sce	c122, c112, c131, c132	c241, c242, c243, c221, c244	c323, c313, c334, c324

Note: Dashes in cells mean that the models were not able to correctly predict the species. Species abbreviations can be found in Table 1.

sure facilities (c142), both selected for nine fish species. FOR models showed that sclerophyllous vegetation (c323) was the most frequently selected subclass (for 11 fish species). The subclasses that were selected only by one species are all from the FOR class (c311, broad-leaved forest; and c322, moors and heathland), and it was also for this class that some species had only one subclass selected (c323, sclerophyllous vegetation for gudgeon (*Gobio gobio*) and European perch (*Perca fluviatilis*); and c321, natural grassland for bleak (*Alburnus alburnus*).

Step 2: species prediction using *n*-selected subclass models

Using the subclasses selected for each species in the single-class models, new models (*n*-selected subclass models) have been built to predict the species occurrences with an input of these selected subclasses. Models were validated, as previously, using the score values associated to the Kappa test (Fig. 5 and Table 4). The score values ranged from 0.67 to 0.9, while the Kappa were all higher than 0.4, except for common dace.

A contribution analysis using the connection weights method has been applied. Then, the contribution values of the subclasses belonging to the same classes have been summed to obtain a percentage of contribution for ARTI, AGRI, and FOR. These results are presented as part of the multiclass comparison (Fig. 6). FOR was the most contributing class for only one species (brown trout, *Salmo trutta fario*). For the other species, the most contributing class was either AGRI or ARTI. For example, rudd (*Scardinius erythrophthalmus*), common sunfish (i.e., pumpkinseed, *Lepomis gibbosus*), bleak, European eel (*Anguilla anguilla*), pikeperch (*Sander lucioperca*), and barbell (*Barbus barbus*) were more sensitive to ARTI, while European chub (*Leuciscus cephalus*), roach (*Rutilus rutilus*), black bullhead (*Ictalurus melas*), gudgeon, European perch, northern pike (*Esox lucius*), tench (*Tinca tinca*), bullhead, and European brook lamprey (*Lampetra planeri*) were more sensitive to AGRI.

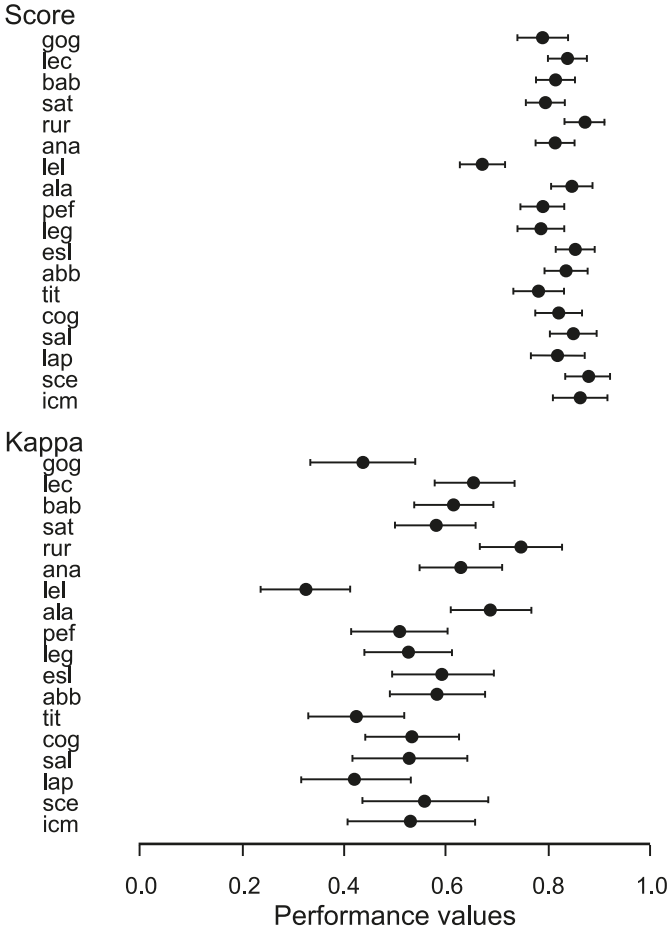
Comparison of multiclass models

The 18 species have also been modeled, using as explanatory variables all the land cover subclasses (30-subclass model) or the three land cover classes (3-class model). The predictive power of these models has been validated and compared with the *n*-selected subclass model using the Kappa value. For all species, there was an effect of the type of model on the Kappa value ($P < 0.0001$, Table 4).

Except for barbel and rudd, all the other fish species had Kappa values that increased from the 3-class model to the *n*-selected subclass model. However, for barbell there was no significant difference between the *n*-selected subclass model and the 30-subclass model. Moreover, while for bullhead and brook lamprey there was no significant difference between the Kappa values obtained with the 3-class model and the 30-subclass model, and for gudgeon there was a better Kappa value obtained using the 3-class model instead of the 30-subclass model, for all the other species, even if the best result was obtained with the *n*-selected subclass model, the 30-subclass model gave better results than the 3-class models.

Using the connection weight method, the contribution of each explanatory variable has been calculated. As carried out in step 2 for the 30-sublass model, the contribution values of the subclasses belonging to the same classes have been summed to obtain a percentage of the contribution for ARTI, AGRI, and FOR and compared between all the multiclass models (Fig. 6). While the most contributing class is for the most part either AGRI or ARTI according to the considered species in the *n*-selected subclass model, for the 3-class model, most of the species are sensitive to ARTI. For the 30-subclass model, the highest contributions are mixed

Fig. 5. Cleveland dot plots of the scores and the Kappa values (mean ± standard deviation) obtained using the test data set in the *n*-selected subclass models, results obtained from step 2 (see Fig. 2).



between ARTI and FOR depending on the species. However, some examples in the case of the 3-class and the 30-subclass models were low in terms of predictive ability (Kappa lower than 0.4).

Discussion

The present study was designed to investigate the responses of riverine fish species to landscape characteristics. Landscape analysis represents an exciting research area for understanding aquatic ecosystems (Wiley et al. 1997). Two main targets were addressed: (*i*) evaluating the global accuracy of our proposed modeling process to predict occurrence patterns of fish species and (*ii*) comparing species responses with landscape types.

Firstly, the models developed, using a screening procedure, have performed well in identifying the impact of landscapes on fish species in the Adour–Garonne Basin, with a mean Kappa value of 0.57 calculated between the observed and simulated distribution of the species modeled. The survey involved 191 sampling sites in the Adour–Garonne Basin, where 18 of the 20 selected species were finally correctly predicted by our modeling processes. Using land cover, our models failed to predict the spatial distribution of only two fish species (minnow and south-west European nase). An important result of this study is the improvement of the prediction using the *n*-selected model instead of the multiclass models using either the three classes or the 30 subclasses as predictive variables. The screening stage of the subclasses was successful in increasing the predictive capacity of the models.

Table 4. Mean values of Kappa from 500 replicates for the three models.

	Model*				
Species	1	2	3	*F*	*P* (>*F*)
gog	0.38b	0.33a	0.44c	171	<0.001
lec	0.59a	0.64b	0.66c	94	<0.001
bab	0.61b	0.58a	0.62b	36	<0.001
sat	0.44a	0.52b	0.58c	360	<0.001
rur	0.57a	0.66b	0.75c	602	<0.001
ana	0.38a	0.58b	0.63c	1219	<0.001
lel	0.17a	0.28b	0.33c	524	<0.001
ala	0.49a	0.61b	0.69c	746	<0.001
pef	0.33a	0.38b	0.51c	514	<0.001
leg	0.32a	0.50b	0.53c	908	<0.001
esl	0.37a	0.55b	0.59c	743	<0.001
abb	0.47a	0.53b	0.58c	181	<0.001
tit	0.29a	0.41b	0.42c	280	<0.001
cog	0.33a	0.34a	0.53b	719	<0.001
sal	0.42a	0.47b	0.53c	122	<0.001
lap	0.35a	0.37a	0.42b	56	<0.001
icm	0.37a	0.45b	0.53c	260	<0.001
sce	0.55b	0.53a	0.56b	10	<0.001

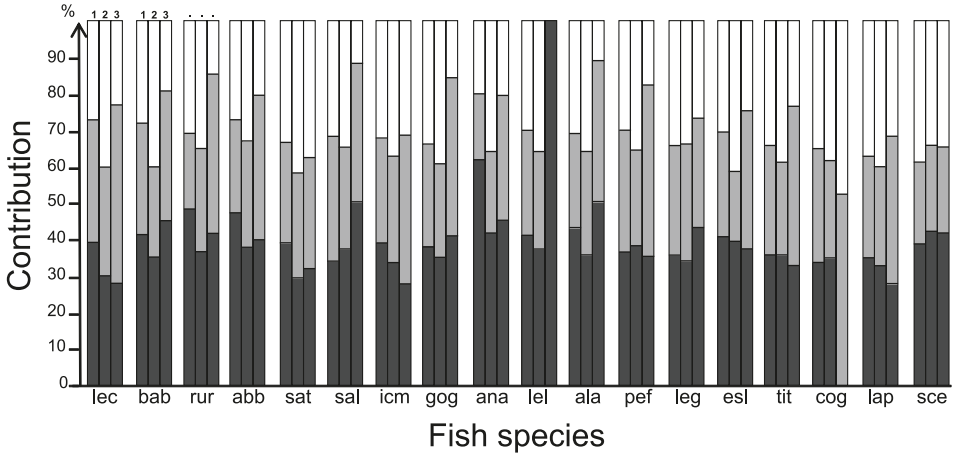
Note: Species abbreviations can be found in Table 1. The *P* values correspond to the results of the analysis of variance (ANOVA) test to determine an effect of the type of model on the model quality. A *P* value lower than 0.05 indicates a difference between the three Kappa means and then an effect of the type of model. Mean Kappa values with a common letter are not significantly different at $P = 0.05$ (Tukey's honestly significant difference test).

*1, the 3-class model; 2, the 30-subclass model; 3, the *n*-selected subclass model.

Secondly, each species was explained differently by the three land cover types. More precisely, the first step of our modeling process, i.e., the prediction of each species using the land cover subclasses of either ARTI, AGRI, or FOR, highlighted the contribution of different land cover subclasses according to the species considered. For example gudgeon, European chub, barbell, brown trout, and roach seem to be sensitive to forest and associations of vegetation, while northern pike, pike-perch, and black bullhead are sensitive to associations of vegetation and open spaces with little or no vegetation. Gudgeon, bleak, and European perch are only sensitive to associations of vegetation. Tench, pike-perch, and brook lamprey are the only species sensitive to pastures, but they are also the three species explained by all the AGRI subclasses, whereas brown trout is only explained by two of them.

This result is a key finding of this study. Landscape influence is species-specific for the distribution of stream fish in the Adour–Garonne Basin. The contrasting sensitivity of the 20 selected species to the land cover descriptors leads us to reject the hypothesis that all the species of the assemblage found at a site have common responses to the environmental constraints. Our study confirms the work of Pont et al.

Fig. 6. Bar plot of the mean percentage (from 500 replicates) of the contribution of the three classes (bottom to top): AGRI (dark gray), ARTI (light gray), and FOR (open). For each species (abbreviations can be found in Table 1), three bars are represented corresponding to the three multiclass models compared: 1, the 3-class model; 2, the 30-subclass model; and 3, the *n*-selected subclass model. These results are obtained from step 3 (see Fig. 2).



(2005), who showed that all species do not have the same perception of the heterogeneity of their habitat. They clearly illustrated that the sensitivity to local and regional processes is species-specific. Therefore, landscape changes have been often studied with respect to the community whatever the organisms studied: diatoms (Pan et al. 2004), macroinvertebrates (Roy et al. 2003; Sandin and Johnson 2004; Mykra et al. 2007), or fish (Jones et al. 1999; Bisson et al. 2002; Roy et al. 2007). Park et al. (2006) have studied the same data as here at the community scale. They have shown that three clusters representing three typical fish assemblages are related for two of them with the percentage of agricultural land. Moreover, they have taken into account not only land cover variables but also topographical factors. Surface area and distance from source seemed to play an important role for these assemblage predictions. Using these other factors, the real impact of land cover seems to be transparent. However, the role of these topographical factors would already be included in the land cover variables. To quantify the unique effect of land cover, a solution could be to first predict the land cover by the topographical factors and use the residuals of these models as the land cover devoid of any topographical variables (Olden et al. 2006). For the same basin, the Adour–Garonne in France, another study has recently been published on macroinvertebrates and their relations with land cover and physical variables (Compin and Cereghino 2007). The authors showed that changes in land cover type led to a change in the patterns of macroinvertebrate functional feeding groups. In their study, they also used ANNs as modeling techniques: the self-organizing map (Kohonen 1982, 2001). Unfortunately, the results shown were only descriptive, and the authors did not address any prediction analysis of their macroinvertebrate functional feeding groups to really explain their relationship with land cover.

ANNs are frequently used in ecology (Joy and Death 2004; Worner and Gevrey 2006; Lencioni et al. 2007). In our study, not only the prediction of species using land cover variables was tested but also the role of each of these variables to explain the species distribution. To test the real impact of land cover only, no other types of variables have been used. Moreover, knowing the complexity needed to determine the impact of each explanatory variable when their number is too large (Zurada et al. 1994), a special modeling process going through a screening step has been proposed. This procedure aimed to exploit the potential of neural networks to incorporate prior ecological knowledge into a model structure to develop more ecologically relevant ANNs as suggested by Olden et al. (2006). A classical use of the method was not powerful enough, as demonstrated here by comparing the predictive power of the developed model after screening and the classical model that used as input all the explanatory variables available at different levels of detail (3 classes or 30 subclasses). The performance clearly increased using the model created.

The notions of spatial scales and the climate integration as explanatory variables have not yet been tackled here and could be discussed as two possibilities for future consideration. As a response to the difficulty in evaluating the impact of land cover on stream biodiversity using the appropriate scales, a new study comparing the results of the models according to the scales used has been done to predict the distribution of fish species in the Adour–Garonne Basin (G. Grenouillet, M. Gevrey, L. Tudesque, and S. Lek, unpublished data). The problem of the impact of landscape change operating via multiple processes at different scales has been already studied (Roy et al. 2007). It should be interesting now to determine which spatial scale is the most efficient one.

In addition to its influence, land use interacts with other anthropogenic drivers that affect the health of stream ecosystems, including climate change (Meyer et al. 1999), invasive species (Scott and Helfman 2001), and dams (Nilsson and Berggren 2000). Although species occurrences were well predicted in our study using only land cover variables, a number of other factors influence fish species distribution. To improve the prediction results, other types of variables may be included, such as climatic variables. As suggested by Pearson et al. (2002), a challenge for future research will be to develop an integrated approach, incorporating factors such as land use and climatic changes. A few modeling

studies have addressed the interactions between climate and land cover. The addition of land cover variables to pure bioclimatic models did not improve their predictive accuracy in a study by Thuiller et al. (2004). Luoto et al. (2007) found that the determinants of species distributions were hierarchically structured, with the climatic variables playing a role at a large scale and land cover at a finer resolution as mentioned by Pearson et al. (2004). The complexity of the existing links between all the drivers impacting on the streams and their biodiversity, in particular climate and landscape changes, merit a complete study in the future.

In conclusion, the application of the model to 20 fish species of this French river system has enabled the methodology to be tested and conclusions drawn regarding the relative roles of land cover on the species distribution. Further applications of this approach associating climate variables should help advance our understanding of the combined effects of these changes on the distribution of species.

Acknowledgements

This work was supported by the EU FP6 Integrated Project Euro-limpacs (GOCE-CT-2003-505540). We thank Mr. John Woodley for reading our manuscript and for his precious English corrections.

References

Allan, J.D. 2004*a*. Landscapes and riverscapes: the influence of land use on stream ecosystems. Annu. Rev. Ecol. Evol. Syst. **35**: 257–284. doi:10.1146/annurev.ecolsys.35.120202.110122.

Allan, J.D. 2004*b*. Influence of land use and landscape setting on the ecological status of rivers. Limnetica, **23**(3–4): 187–198.

Allan, J.D., and Johnson, L.B. 1997. Catchment-scale analysis of aquatic ecosystems. Freshw. Biol. **37**(1): 107–111. doi:10.1046/j.1365-2427.1997.00155.x.

Allan, J.D., Erickson, D.L., and Fay, J. 1997. The influence of catchment land use on stream integrity across multiple spatial scales. Freshw. Biol. **37**(1): 149–161. doi:10.1046/j.1365-2427.1997.d01-546.x.

Bisson, P.A., Raphael, M.G., Foster, A.D., and Jones, L.L.C. 2002. Influence of site and landscape features on vertebrate assemblages in small streams. *In* Proceedings from the Wood Compatibility Initiative Workshop, Washington, USA, 4–7 December 2001. *Edited by* A.C. Johnson, R.W. Haynes, and R A. Monserud. US Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, Ore. Gen. Tech. Rep. PNW-GTR-563. pp. 61–72.

Brosse, S., Arbuckle, C.J., and Townsend, C.R. 2003. Habitat scale and biodiversity: influence of catchment, stream reach and bedform scales on local invertebrate diversity. Biodivers. Conserv. **12**(10): 2057–2075. doi:10.1023/A:1024107915183.

Cohen, J. 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. **20**: 37–46. doi:10.1177/001316446002000104.

Compin, A., and Cereghino, R. 2007. Spatial patterns of macroinvertebrate functional feeding groups in streams in relation to physical variables and land-cover in southwestern France. Landsc. Ecol. **22**(8): 1215–1225. doi:10.1007/s10980-007-9101-y.

European Union. 2000. Directive of the European Parliament and the Council Establishing a Framework for Community Action in the Field of Water Policy. Legislative acts and other instruments. ENV221 CODEC 513. European Union, Brussels, Belgium.

Fausch, K.D., Torgersen, C.E., Baxter, C.V., and Li, H.W. 2002. Landscapes to riverscapes: bridging the gap between research and conservation of stream fishes. Bioscience, **52**(6): 483–498. doi:10.1641/0006-3568(2002)052[0483:LTRBTG]2.0.CO;2.

Gevrey, M., Dimopoulos, L., and Lek, S. 2003. Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol. Model. **160**(3): 249–264. doi:10.1016/S0304-3800(02)00257-0.

Gore, J.A., and Nestler, J.M. 1988. Instream flow studies in perspective. Regul. Rivers Res. Manage. **2**: 93–101. doi:10.1002/rrr.3450020204.

Hilbert, D.W., and Ostendorf, B. 2001. The utility of artificial neural networks for modelling the distribution of vegetation in past, present and future climates. Ecol. Model. **146**(1–3): 311–327. doi:10.1016/S0304-3800(01)00323-4.

Hughes, R.M., and Gammon, J.R. 1987. Longitudinal changes in fish assemblages and water quality in the Willamette River, Oregon. Trans. Am. Fish. Soc. **116**: 196–209. doi:10.1577/1548-8659(1987)116<196:LCIFAA>2.0.CO;2.

Jones, E.B.D., Helfman, G.S., Harper, J.O., and Bolstad, P.V. 1999. Effects of riparian forest removal on fish assemblages in southern Appalachian streams. Conserv. Biol. **13**(6): 1454–1465. doi:10.1046/j.1523-1739.1999.98172.x.

Jowett, I.G., Richardson, J., and McDowall, R.M. 1996. Relative effects of in-stream habitat and land use on fish distribution and abundance in tributaries of the Grey River, New Zealand. N.Z. J. Mar. Freshw. Res. **30**(4): 463–475.

Joy, M.K., and Death, R.G. 2004. Predictive modelling and spatial mapping of freshwater fish and decapod assemblages using GIS and neural networks. Freshw. Biol. **49**(8): 1036–1052. doi:10.1111/j.1365-2427.2004.01248.x.

Kohavi, R. 1995 A study of crossvalidation and bootstrap for estimation and model selection. Morgan Kaufmann Publishers, Montréal, Que., Canada.

Kohonen, T. 1982. Self-organized formation of topologically correct feature maps. Biol. Cybern. **43**: 59–69. doi:10.1007/BF00337288.

Kohonen, T. 2001 Self-organizing maps. Springer, Berlin, Germany.

Lancaster, J., and Hildrew, A.G. 1993. Flow refugia and the microdistribution of lotic macroinvertebrates. J. N. Am. Benthol. Soc. **12**: 385–393. doi:10.2307/1467619.

Lencioni, V., Maiolini, B., Marziali, L., Lek, S., and Rossaro, B. 2007. Macroinvertebrate assemblages in glacial stream systems: a comparison of linear multivariate methods with artificial neural networks. Ecol. Model. **203**(1–2): 119–131. doi:10.1016/j.ecolmodel.2006.04.028.

Liu, C.R., Berry, P.M., Dawson, T.P., and Pearson, R.G. 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecography, **28**(3): 385–393. doi:10.1111/j.0906-7590.2005.03957.x.

Luoto, M., Virkkala, R., and Heikkinen, R.K. 2007. The role of land cover in bioclimatic models depends on spatial resolution. Glob. Ecol. Biogeogr. **16**(1): 34–42. doi:10.1111/j.1466-8238.2006.00262.x.

Manel, S., Williams, H.C., and Ormerod, S.J. 2001. Evaluating presence–absence models in ecology: the need to account for prevalence. J. Appl. Ecol. **38**: 921–931. doi:10.1046/j.1365-2664.2001.00647.x.

Maret, T.R., Robinson, C.T., and Minshall, G.W. 1997. Fish assemblages and environmental correlates in least-disturbed streams of the upper Snake River basin. Trans. Am. Fish. Soc. **126**(2): 200–216. doi:10.1577/1548-8659(1997)126<0200:FAAECI>2.3.CO;2.

Meador, M.R., and Goldstein, R.M. 2003. Assessing water quality at large geographic scales: relations among land use, water phy-

sicochemistry, riparian condition, and fish community structure. Environ. Manage. **31**(4): 504–517. doi:10.1007/s00267-002-2805-5. PMID:12677296.

Meyer, J.L., Sale, M.J., Mulholland, P.J., and Poff, N.L. 1999. Impacts of climate change on aquatic ecosystem functioning and health. J. Am. Water Resour. Assoc. **35**(6): 1373–1386. doi:10.1111/j.1752-1688.1999.tb04222.x.

Mykra, H., Heino, J., and Muotka, T. 2007. Scale-related patterns in the spatial and environmental components of stream macroinvertebrate assemblage variation. Glob. Ecol. Biogeogr. Lett. **16**(2): 149–159. doi:10.1111/j.1466-8238.2006.00272.x.

Nilsson, C., and Berggren, K. 2000. Alteration of riparian ecosystems caused by river regulation. Bioscience, **50**(9): 783–792. doi:10.1641/0006-3568(2000)050[0783:AORECB]2.0.CO;2.

Olden, J.D., and Jackson, D.A. 2002. Illuminating the "Black box": a randomization approach for understanding variable contributions in artificial neural networks. Ecol. Model. **154**: 135–150. doi:10.1016/S0304-3800(02)00064-9.

Olden, J.D., Poff, N.L., and Bledsoe, B.P. 2006. Incorporating ecological knowledge into ecoinformatics: an example of modeling hierarchically structured aquatic communities with neural networks. Ecol. Inform. **1**(1): 33–42. doi:10.1016/j.ecoinf.2005.08.003.

Onorato, D., Angus, R.A., and Marion, K.R. 2000. Historical changes in the ichthyofaunal assemblages of the upper Cahaba River in Alabama associated with extensive urban development in the watershed. J. Freshwat. Ecol. **15**(1): 47–63.

Pan, Y.D., Herlihy, A., Kaufmann, P., Wigington, J., van Sickle, J., and Moser, T. 2004. Linkages among land-use, water quality, physical habitat conditions and lotic diatom assemblages: a multi-spatial scale assessment. Hydrobiologia, **515**(1–3): 59–73. doi:10.1023/B:HYDR.0000027318.11417.e7.

Park, Y.-S., Grenouillet, G., Esperance, B., and Lek, S. 2006. Stream fish assemblages and basin land cover in a river network. Sci. Total Environ. **365**(1–3): 140–153. doi:10.1016/j.scitotenv.2006.02.046. PMID:16626786.

Pearson, R.G., Dawson, T.P., Berry, P.M., and Harrison, P.A. 2002. SPECIES: a spatial evaluation of climate impact on the envelope of species. Ecol. Model. **154**(3): 289–300. doi:10.1016/S0304-3800(02)00056-X.

Pearson, R.G., Dawson, T.P., and Liu, C. 2004. Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. Ecography, **27**(3): 285–298. doi:10.1111/j.0906-7590.2004.03740.x.

Pont, D., Hugueny, B., and Oberdorff, T. 2005. Modelling habitat requirement of European fishes: do species have similar responses to local and regional environmental constraints? Can. J. Fish. Aquat. Sci. **62**(1): 163–173. doi:10.1139/f04-183.

Roth, N.E., Allan, J.D., and Erickson, D.L. 1996. Landscape influences on stream biotic integrity assessed at multiple spatial scales. Landsc. Ecol. **11**(3): 141–156. doi:10.1007/BF02447513.

Roy, A.H., Rosemond, A.D., Paul, M.J., Leigh, D.S., and Wallace, J.B. 2003. Stream macroinvertebrate response to catchment urbanisation (Georgia, USA). Freshw. Biol. **48**(2): 329–346. doi:10.1046/j.1365-2427.2003.00979.x.

Roy, A.H., Freeman, B.J., and Freeman, M.C. 2007. Riparian influences on stream fish assemblage structure in urbanizing streams. Landsc. Ecol. **22**(3): 385–402. doi:10.1007/s10980-006-9034-x.

Rumelhart, D.E., Hinton, G.E., and Williams, R.J. 1986. Learning internal representations by error propagation. *In* Parallel distributed processing: explorations in the microstructure of cognition. Vol. 1. *Edited by* D.E. Rumelhart. MIT Press, Cambridge, Mass., USA. pp. 318–62.

Sandin, L., and Johnson, R.K. 2004. Local, landscape and regional factors structuring benthic macroinvertebrate assemblages in Swedish streams. Landsc. Ecol. **19**(5): 501–514. doi:10.1023/B:LAND.0000036116.44231.1c.

Scott, M.C., and Helfman, G.S. 2001. Native invasions, homogenization, and the mismeasure of integrity of fish assemblages. Fisheries, **26**(11): 6–15. doi:10.1577/1548-8446(2001)026<0006:NIHATM>2.0.CO;2.

Seegert, G. 2000. Considerations regarding development of index of biotic integrity metrics for large rivers. Environ. Sci. Policy, **3**(Suppl. 1): 99–106. doi:10.1016/S1462-9011(00)00031-9.

Spitz, F., and Lek, S. 1999. Environmental impact prediction using neural network modelling. An example in wildlife damage. J. Appl. Ecol. **36**(2): 317–326. doi:10.1046/j.1365-2664.1999.00400.x.

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**(3): 492–501. doi:10.1139/f88-059.

Strayer, D.L., Beighley, R.E., Thompson, L.C., Brooks, S., Nilsson, C., Pinay, G., and Naiman, R.J. 2003. Effects of land cover on stream ecosystems: roles of empirical models and scaling issues. Ecosystems (N.Y.), **6**(5): 407–423. doi:10.1007/s10021-002-0170-0.

Tarassenko, L. 1998. A guide to neural computing applications. Arnold, London and New York.

Thuiller, W., Araujo, M.B., and Lavorel, S. 2004. Do we need land-cover data to model species distributions in Europe? J. Biogeogr. **31**(3): 353–361.

Tison, J., Park, Y.S., Coste, M., Wasson, J.G., Ector, L., Rimet, F., and Delmas, F. 2005. Typology of diatom communities and the influence of hydro-ecoregions: a study on the French hydrosystem scale. Water Res. **39**(14): 3177–3188. doi:10.1016/j.watres.2005.05.029. PMID:16002120.

Townsend, C.R., Doledec, S., Norris, R., Peacock, K., and Arbuckle, C. 2003. The influence of scale and geography on relationships between stream community composition and landscape variables: description and prediction. Freshw. Biol. **48**(5): 768–785. doi:10.1046/j.1365-2427.2003.01043.x.

Verbyla, D.L., and Litaitis, J.A. 1989. Resampling methods for evaluating classification accuracy of wildlife habitat models. Environ. Manage. **13**: 783–787. doi:10.1007/BF01868317.

Waite, S. 2000 Statistical ecology in practice: a guide to analysing environmental and ecological field data. Prentice Hall, Essex, UK.

Walser, C.A., and Bart, H.L. 1999. Influence of agriculture on instream habitat and fish community structure in Piedmont watersheds of the Chattahoochee River System. Ecol. Freshw. Fish, **8**(4): 237–246. doi:10.1111/j.1600-0633.1999.tb00075.x.

Walton, B.M., Salling, M., Wyles, J., and Wolin, J. 2007. Biological integrity in urban streams: toward resolving multiple dimensions of urbanization. Landsc. Urban Plan. **79**(1): 110–123. doi:10.1016/j.landurbplan.2005.10.004.

Wiley, M.J., Kohler, S.L., and Seelbach, P.W. 1997. Reconciling landscape and local views of aquatic communities: lessons from Michigan trout streams. Freshw. Biol. **37**(1): 133–148. doi:10.1046/j.1365-2427.1997.00152.x.

Worner, S.P., and Gevrey, M. 2006. Modelling global insect pest species assemblages to determine risk of invasion. J. Appl. Ecol. **43**(5): 858–867. doi:10.1111/j.1365-2664.2006.01202.x.

Zurada, J.M., Malinowski, A., and Cloete, I. 1994 Sensitivity analysis for minimization of input data dimension for feedforward neural network. *In* Proceedings of IEEE International Symposium on Circuits and Systems, London, UK, 28 May – 2 June 1994. Vol. 6. IEEE, London, UK. pp. 447–450.