



Macroinvertebrate richness patterns in North African streams

Olivier Beauchard, James Gagneur and Sebastien Brosse* *LADYBIO, CNRS-UPS Bat IVR3, Université Paul Sabatier, Toulouse, France*

Abstract

Aim To test the hypothesis that macroscale environmental variables explain local taxonomic richness of stream macroinvertebrates, and then to identify the relationships between these variables and benthic fauna richness in North Africa.

Location North Africa, from West Morocco to East Tunisia.

Methods Using a large-scale data base made of 211 sampling sites gathered from an area of 500,000 km², an artificial neural network model has been built to identify and predict the influence of macroscale environmental variables on local macroinvertebrate richness.

Results The correlation coefficient (r) between observed and estimated taxon richness values was 0.75 ($P < 0.001$), and the model explained more than 55% ($r^2 = 0.563$) of the macroinvertebrate richness variation. Macroinvertebrate richness was, therefore, accurately predicted using only three environmental variables accounting for hydrology (number of rainy days), geographical factors, i.e. connections between European and North African land masses (longitude) and climatic gradient (altitude).

Main conclusions Stream macroinvertebrate richness in North Africa results from a combination of climatic, geographical and hydrological parameters. Although consistent with current biogeographic and ecological concepts mainly derived from European and North American streams, this study underlines the specificity of dry Mediterranean ecosystems. The shape of diversity patterns results from climatic conditions and their associated environmental seasonal dynamics, which screens geographical processes.

Keywords

Morocco, Algeria, Tunisia, benthic fauna, stream macroinvertebrates, Ephemeroptera, Plecoptera, Trichoptera, Coleoptera, richness.

INTRODUCTION

Since the 1980s, more and more studies have focused on the factors influencing biodiversity at different spatial scales and provided evidence that taxonomic richness of aquatic organisms is determined by factors acting at a given scale (Ricklefs & Schluter, 1993; Brosse *et al.*, 2003), which strongly influence the perception of patterns and processes in stream ecosystems (Frissell *et al.*, 1986; Minshall, 1988; Vinson & Hawkins, 1998). Concerning stream macroinvertebrates, this aspect has been documented for local spatial scales, i.e. from site to river basin

(Vinson & Hawkins, 1998; Li *et al.*, 2001), whereas macroscale studies have scarcely been conducted (Huston, 1994; Gaston, 2000). Moreover, in a recent review, Vinson & Hawkins (1998) underlined the need for regional studies on stream macroinvertebrate diversity. Within this framework, increasing interest is given to Mediterranean ecosystems of North Africa that have been considered as biodiversity hotspots (Lavorel & Richardson, 1999; Myers *et al.*, 2000). In the same way, the striking properties of Mediterranean stream ecosystems have recently been underlined by Gasith & Resh (1999). However, up to now, most studies dealing with aquatic fauna have only considered diversity at local scale, leading to a rather good knowledge of benthic fauna of most of the North African river basins from East Tunisia to West Morocco (e.g. Lounaci, 1987; Boumaiza, 1994; Lounaci *et al.*, 2000).

*Correspondence: Sebastien Brosse, LADYBIO, CNRS-UPS, Bat IVR3, Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse cedex 4, France. E-mail: brosse@cict.fr

Although a large amount of information is available, no study has been conducted yet to provide a more general insight into stream macroinvertebrate richness patterns in North Africa. To remedy that deficiency, our aim was to initiate a large scale analysis of stream macroinvertebrate richness in this region. Such a geographical system is determined when integrating sites inside relatively obvious boundaries acting as environmental barriers (Brown, 1995), here made of the Mediterranean Sea and Sahara desert. To understand invertebrate diversity beyond the intrinsic features of drainage basins covering large areas, macro-variables are considered as reliable parameters to identify the regional characteristics (Brown, 1995; Oberdorff *et al.*, 1999), giving rise to habitat properties at local scale (Vinson & Hawkins, 1998).

Among the main global descriptors usually involved, altitude, climate and hydrology are the most commonly used and recognized factors to determine richness and habitat patterns (Frissell *et al.*, 1986; Corkum, 1992; Vinson & Hawkins, 1998). In addition, latitude and longitude have also been employed to identify biomes or historical factors and play a major role in taxon richness patterns (Corkum, 1992; Samways & Osborn, 1998; Lobo & Davis, 1999). These relationships between environmental descriptors and taxon richness are usually derived from statistical and modelling methods based on linear principles (see James & McCulloch, 1990, for a review). However, these approaches cannot overcome some significant biases because of both the complexity and presumed nonlinearity of invertebrate–environment relationships and inherent correlations among variables (Carter *et al.*, 1996). To deal with such difficulties, transformation of nonlinear variables can appreciably improve the results in some situations, but this is far from always being the case (Lek *et al.*, 1996; Brosse *et al.*, 1999). Artificial neural networks (ANNs), on the other hand, are efficient in dealing with systems ruled by complex nonlinear relationships and provide an alternative to traditional statistical methods (Lek *et al.*, 1996; Lek & Guégan, 2000). They have been successfully applied to the prediction of macroinvertebrate taxon number (Walley & Fontama, 1998; Brosse *et al.*, 2003) and to the identification of habitat factors that account for species richness at different spatial scales (Guégan *et al.*, 1998; Brosse *et al.*, 2001, 2003). In addition to the predictive value of the models, the influence of each variable introduced in the modelling procedure can be quantified using specific algorithms (Lek *et al.*, 1996; Dimopoulos *et al.*, 1999).

In this paper, our aim was to identify and understand the relationships between macroscale features and stream macroinvertebrate richness. We used ANN to build a predictive model of macroinvertebrate richness in North African streams based on macroscale environmental variables accounting for climate, hydrology and geographical factors. The results of the models were used to identify influential variables and then to determine the role of these variables on stream insect diversity, leading to the formulation of hypotheses about the biogeography of stream macroinvertebrate richness in the Mediterranean region.

DATA AND METHODS

Study site and data

The study area extends from longitude 8.67° W to 10.75° E and from latitude 30.40° S to 37.25° N; it covers about 500,000 km² through Morocco, Algeria and Tunisia (usually called the Maghreb). This geographic unit is isolated by natural boundaries: the Mediterranean Sea to the north and the Sahara desert to the south (Fig. 1a). In addition to the geographical frontiers of this area, this Mediterranean system constitutes a single biogeographic unit of the Palaearctic region, called the X zone (Illies, 1978).

The macroinvertebrate data come from taxon lists published since the 1980s from Algeria, Morocco and Tunisia covering 318 sampling sites. Incomplete studies, i.e. those dealing with only one species, genus, family or order were removed from the database. In the same way, studies dealing with anthropic disturbance or polluted areas were also dismissed. Therefore, only data dealing with large comprehensive studies of macroinvertebrates were retained to avoid biases in richness assessment (Appendix I). Moreover, only sites where macroinvertebrate fauna has been determined at the same taxonomic level (species or genera) were considered. Although species level is often used for well-known and low diversified taxonomic groups, e.g. mammals or reptiles (Gaston, 2000), a specific determination level is rarely achieved in large scale studies of highly diversified groups such as macroinvertebrates. Moreover, the macroinvertebrate taxonomic resolution required is dependent on the spatial scale selected, species level identification not being necessary beyond the catchment scale (Dolédéc *et al.*, 2000). Genus level, on the other hand, has been proved accurate for the description of community patterns (Gaston & Spicer, 1998; Li *et al.*, 2001); therefore, genera were selected as a compromise between available taxonomic information and number of sites. In addition, the use of genus level has the advantage of avoiding potential biases ascribed to unsure species determination made by various authors. In this way, thirty-nine sites with complete inventories to the genus level for the entire macroinvertebrate fauna were considered (i.e. including Oligocheta, Gastropoda, Ephemeroptera, Odonata, Plecoptera, Heteroptera, Coleoptera, Trichoptera and Diptera). However, aiming to increase the spatial representation of richness patterns in North Africa, additional sites with genus richness of some macroinvertebrate orders may be considered. We, therefore, checked the relationship between overall richness (i.e. all orders considered) and the most represented orders, namely Ephemeroptera (E), Plecoptera (P), Trichoptera (T), Coleoptera (C), EPT and EPTC. A taxonomic group or a combination of taxonomic groups can be considered as a relevant surrogate for overall richness if (i) a strong linear relationship is detected between a richness estimator and overall richness, (ii) the surrogate gives a realistic image of the richness of the taxonomic groups other than those used as estimators, and (iii) the group or combination of groups is present in most sampling sites. The combinations EPT and EPTC were tested as they have already been used as surrogates for overall richness

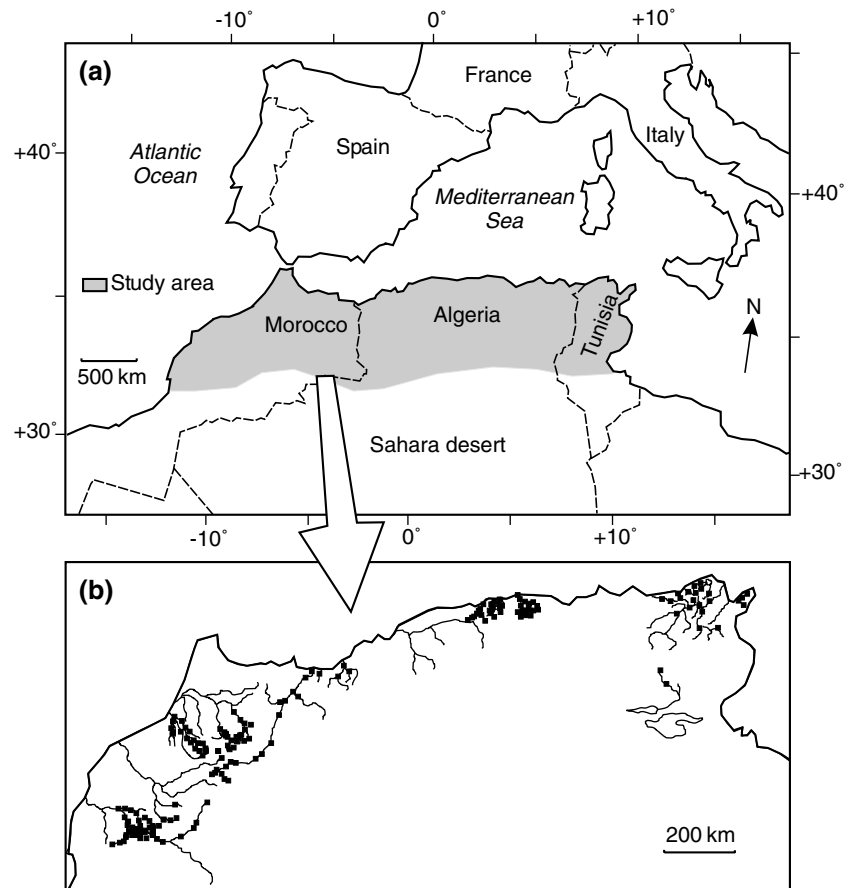


Figure 1 Map of North Africa and South West Europe: (a) location of the study area (shaded); (b) main river basins and the 211 sampling sites covering the northern part of Morocco, Algeria and Tunisia (i.e. the Maghreb). Sahara desert covers the southern part of these countries, and was not considered because of the absence of streams.

(Cayrou *et al.*, 2000; Wright *et al.*, 2000; Li *et al.*, 2001). The most accurate relationships were obtained between overall richness and EPTC (see Results), and consequently, EPTC richness was hereafter considered as a surrogate for overall richness. The resulting EPTC data matrix contained 211 sampling sites and 139 genera, dispersed throughout most drainage basins (Fig. 1b) of biogeographic zone X (from Illies, 1978).

Seven macroscale variables used as richness predictors were derived from site surveys, map records and from a global scale meteorological database (0.5° square grid coverage) provided by Leemans & Cramer (1991), containing long term (29 years) monthly averages of meteorological parameters. The seven macroscale variables used as richness predictors were altitude, coefficient of variation of monthly precipitation (i.e. standard deviation of monthly precipitation/annual average precipitation; calculated from the twelve monthly values), longitude, annual number of rainy days, maximum monthly annual precipitation, distance from the source and surface area of drainage basin. These seven variables were not or were only slightly correlated and are commonly used in large scale ecological studies (Brown, 1995; Vinson & Hawkins, 1998). Although productivity indices are frequently used for this kind of study, we did not use them because the most frequently used productivity

assessment methods (Leith, 1975; Oberdorff *et al.*, 1997) provided contradictory results because of the particular relationships between temperature and precipitation under a Mediterranean climate.

Modelling procedure

In order to establish a link between genus richness and the seven macroscale variables, we used an ANN procedure. The ANN architecture was a layered feed-forward network, in which the nonlinear processing elements (neurones) are arranged in successive layers, with a one-way flow of information (i.e. weights) from input layer to output layer, through a hidden layer (Fig. 2). In ANN, the computational or processing elements are called neurones. Like a natural neurone, they have many inputs but only a single output, which can stimulate other neurones in the network. Neurones from one layer are connected to all neurones in the adjacent layer(s), but no lateral connections within a layer, nor feedback connections, are possible. The number of input and output units depends on the representations of the input and the output objects, respectively. The output neurone computed the value of the dependent variable (EPTC genus richness). As a complement, a 'bias' neurone was added to each computational layer (i.e. hidden and output layers);

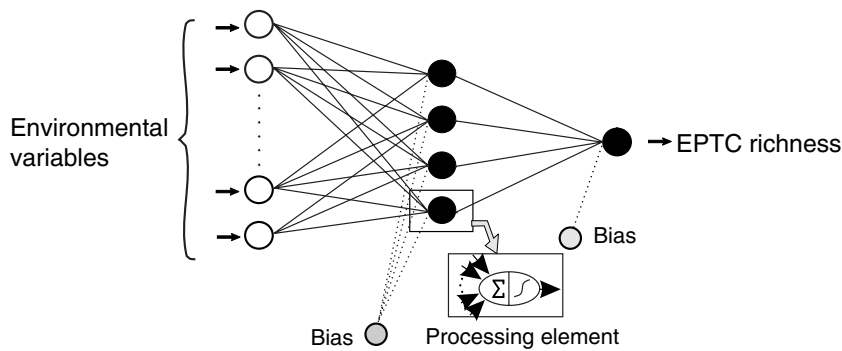


Figure 2 Three-layered feed-forward ANN used to model EPTC richness. There are as many input neurons as explanatory variables, and one output neuron, which correspond to EPTC genus richness. Connections between neurons are shown by solid lines. There are two additional bias nodes labelled with a constant input value of 1.

these two neurones (Fig. 2) had a constant input value of one and were used to lower biases in the modelling procedure (Rumelhart *et al.*, 1986). Training the network consists of using a training data set to adjust the connection weights to minimize the error between observed and predicted values. The weights play an important role in the propagation of the signal through the network. They establish a link between the input variables and their associated output variable and 'contain' the knowledge of the ANN about the problem–solution relationship. This training was performed according to an iterative process called the back-propagation algorithm (Rumelhart *et al.*, 1986). Model reliability was measured using the correlation coefficient (r) between observed and estimated genus richness values, which provides information on the significance of the model.

The modelling was carried out in two steps. First, model training was performed using the whole data matrix. This step was used to estimate the performance of the ANN in learning data. Second, we used the 'leave-one-out' bootstrap cross-validation test (Efron, 1983), a method for estimating generalization error based on 'resampling'. This procedure is a generalization of k -fold cross-validation, where data is divided into k subsets of equal size. The ANN is then trained k times, each time leaving out one of the subsets from training, and using only the omitted subset to test prediction capabilities of the model. If k equals the sample size, this is called 'leave-one-out' cross-validation. Considering our data, each model was trained on 210 sites and tested on the remaining site that was not included in the training set. This procedure was repeated 211 times to train the model using all possible training set combinations and to test model predictions on all the sites. This procedure is appropriate for our data as it should be used when the amount of data is limited and/or when each sample is likely to have 'unique information' (Efron, 1983; Kohavi, 1995); moreover, it has been found to be efficient for ANN modelling (Guégan *et al.*, 1998; Brosse *et al.*, 2001, 2003). This second step allows the prediction capabilities of the network to be assessed.

In the present study, we applied a stepwise procedure to identify the most influential variables. We first set up a predictive model of EPTC genus richness using the seven environmental variables. Then, the least influential variable in the model was removed and a new model built with six environmental variables. This stepwise procedure

was repeated to test model reliability with seven to one environmental variables. For each of the six stepwise procedures, model architecture (number of input and hidden neurones, number of iterations) was calibrated using the method illustrated by Lek *et al.* (1996) to provide the best predictive results. Model predictive efficiency was assessed using the correlation coefficient between observed and estimated genus richness in training and testing procedures for each step of the stepwise procedure, so the model providing the best compromise between predictive efficiency and number of input variables could be identified. In this 'best' model, EPTC genus richness was analysed with regard to the relative contributions of the input variables using a specific algorithm: the relative contribution of each variable in the models was obtained by a method based on the response of the partial derivative of the output, called the PAD method (Dimopoulos *et al.*, 1999; Gevrey *et al.*, 2003). This explanatory procedure has been proved to be efficient in various applications of ANN in ecology (Reyjol *et al.*, 2001; Gevrey *et al.*, 2003).

The patterns identified by the PAD method were verified, justified and extended using related variables taken from the genus database and the Leemans & Cramer (1991) climate database and the relationships between each variable and genus richness was illustrated using the 'locally weighted-scatterplot-smoother' usually called 'Lowess' (Cleveland, 1979). In this procedure, each sample is smoothed using a defined proportion of the nearest neighbours to the target point. Optimal fitting is obtained iteratively to minimize the residuals between observed and estimated values. The proportion of samples perfectly fitted by the Lowess smoother was indicated by the f value of each fitted distribution. The parameter f varies between 0 and 1 according to the sensitivity of the analysis and is determined empirically by testing various possibilities and selecting the one that provides the best generalization ability to visualize general data tendencies (Trexler & Travis, 1993). One of the major advantages of this method is that it is able to perform accurate fitting of both linear and nonlinear data, and automatically shows the dependence of the response of the predictor. However, no equation is associated with the Lowess curve because of its nonparametric nature, and therefore only graphical results are obtained. Graphical results could be validated by a nonparametric statistical test

(e.g. Mann–Whitney or Kruskal–Wallis) after sorting environmental values into different classes, but we did not use these methods given that sorting data into classes is not representative of a continuous nature. Moreover, Lowess fitting is commonly used in ecology, and provides reliable descriptive information without further post-analysis (Trexler & Travis 1993; James *et al.* 1996; Brosse *et al.* 1999).

RESULTS

The relationships between genus richness of the four most commonly considered insect orders, i.e. Ephemeroptera, Plecoptera, Trichoptera and Coleoptera, and overall genus richness was verified for each order and two combinations of orders (EPT and EPTC) (Table 1). Considering orders one by one, none provided a reliable representation of the overall richness. Although Coleoptera did exhibit strong relationships with overall richness and overall richness minus Coleoptera, this order was absent from 48% of the sites, and therefore cannot be used as a surrogate for overall richness (Table 1). Both EPT and EPTC showed strongly significant linear relationships with overall macroinvertebrate genus richness ($P < 0.001$; $r = 0.81$ and 0.97 for EPT and EPTC, respectively). However, as shown in Table 1, the use of EPT as a surrogate for overall richness minus EPT may involve drawbacks because of the lower representativity of orders other than EPT. On the contrary, the relationship between EPTC and overall richness provided much more reliable overall richness estimation as both overall richness and overall richness minus EPTC were reliably predicted (Table 1). EPTC can, therefore, be considered as a surrogate for overall richness and hereafter considered as genus richness estimator. This allowed us to consider a final data matrix made of 211 sites, which was employed to set up the ANN models.

The stepwise procedure applied to the ANN models testified that three environmental variables, namely number of annual rainy days, longitude and altitude provided the best compromise between richness predictability and model complexity. The four successively removed environmental variables (maximum monthly annual precipitation, surface

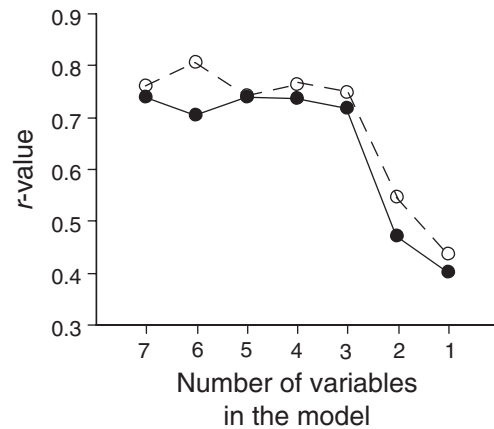


Figure 3 Selection of the input variables in the ANN models. Correlation coefficient between observed and estimated EPTC richness predicted after ANN training (solid line and dark circles) and testing (dashed line and open circles) procedures are represented according to the number of input variables introduced in the model. During this stepwise procedure, the variable contributing least to a model (contribution was estimated using PAD algorithm, see text for details) was removed in the following model. All seven variables were considered in the first model. In the other six models, the variables successively removed were maximum precipitation, surface of drainage basin, coefficient of variation of monthly precipitation, distance from the source, altitude and longitude in the models with six, five, four, three, two, and one input variables, respectively.

area of drainage basin, coefficient of variation of monthly precipitation and distance from the source) weakly contributed to the prediction of EPTC richness as the r -value did not decrease significantly when these variables were removed from the model (Fig. 3). After the ANN training procedure using the three significant environmental variables, the relationship between observed and predicted richness values was highly significant ($r = 0.72$, $P < 0.001$) (Fig. 4a). Predictive values derived from the leave-one-out testing procedure of the ANN also provided a reliable prediction of EPTC richness, with an r -value between observed and predicted richness comparable to that obtained after the training procedure ($r = 0.75$, $P < 0.001$) (Fig. 4b). This ANN model, therefore, explained more than 55% ($r^2 = 0.563$) of the variation of macroinvertebrate richness in the Maghreb. Moreover Fig. 4b illustrates the reliability of the predictive model, as a large proportion of the samples were perfectly predicted and, therefore, well aligned along the diagonal of best prediction of coordinates (1 : 1). Therefore, the use of only three environmental variables provided reliable prediction of EPTC richness as testified by the highly significant relationship between observed (i.e. measured) and estimated (i.e. predicted by the model) values of genus richness provided by the leave-one-out testing procedure.

The results of the PAD algorithm illustrate the relative contribution of the three predictors in the model. Number of rainy days had the strongest influence with a contribution of roughly 62%, the two other variables, namely

Table 1 Linear relationships (correlation coefficient, r) between overall richness, overall richness minus tested group and the four most represented insect orders

Tested group	Overall richness	Overall richness – tested group	Occurrence (%)
E	0.613**	0.558**	94.9
P	0.040 (ns)	0.010 (ns)	64.1
T	0.753**	0.684**	43.6
C	0.964**	0.941**	51.3
EPT	0.805**	0.718**	97.4
EPTC	0.965**	0.913**	97.4

**Highly significant ($P < 0.01$); ns, not significant ($P > 0.05$). The percentage of occurrence of each order or order combination (for EPT and EPTC) in the database is also indicated.

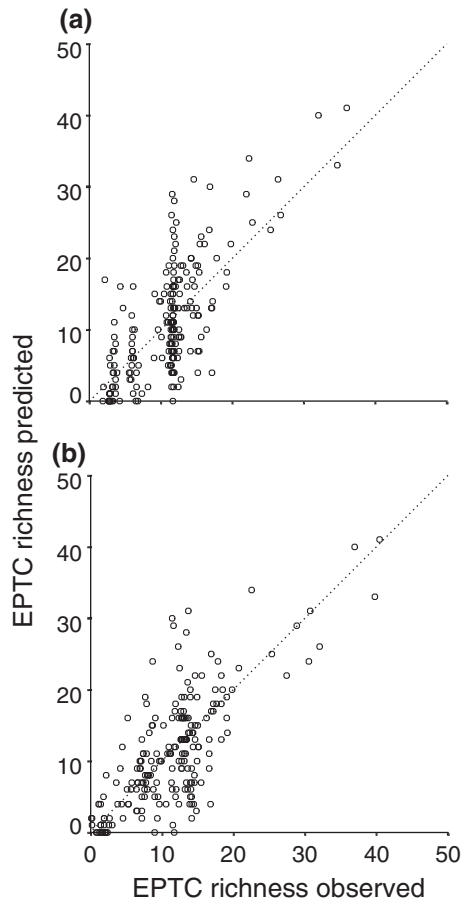


Figure 4 Artificial neural network model prediction of taxon richness using three input variables (annual number of rainy days, longitude and altitude). Scatterplots of predicted values vs. observed values are given for: (a) training procedure; and (b) leave-one-out cross-validation test. The dotted line indicates the perfect fit (1 : 1 ratio).

longitude and altitude accounted for about 20% and 18% of the information given by the model, respectively. The scatterplots of each of these environmental variables vs. genus richness (Fig. 5) illustrate the heterogeneity of the data set. A large range of responses was obtained for the three environmental variables and no *a priori* interpretation of the figures was possible. Data fitting by parametric functions did not reveal any significant tendencies so we used the Lowess nonparametric smoothing method. Using this method, the influence of each of the three environmental variables on genus richness was clearly visualized. Genus richness increased linearly with the annual number of rainy days up to 60 days per year (Fig. 5a). Above this value, richness remained stable and maximal. The Lowess curve representing longitude patterns (Fig. 5b) showed that richness peaked in the western part of the study area, i.e. between -8.7° and -3.2° . In the central part of Maghreb the Lowess analysis predicted a lower diversity, followed by a relative increase in the eastern part; however, because

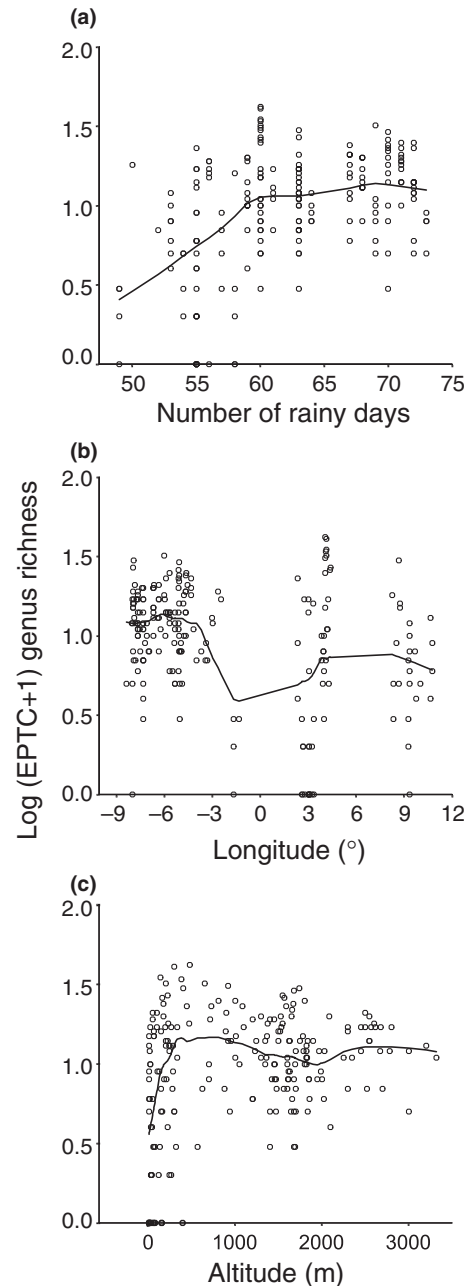


Figure 5 Bivariate plots of the three environmental variables vs. EPTC richness. Lowess curves, represented by a solid line, were used to fit data: (a) number of rainy days, $f = 0.50$; (b) longitude, $f = 0.35$; (c) altitude, $f = 0.30$. The f value indicates the proportion of samples perfectly fitted by the Lowess smoother. f varies between 0 and 1 according to the sensitivity of the analysis and is determined empirically by testing various possibilities and selecting the one that provides the best generalization to visualize data tendencies (see text for details).

of the high heterogeneity of genus richness in this area, the results should be considered with caution. To confirm and provide details on the influence of longitude on genus

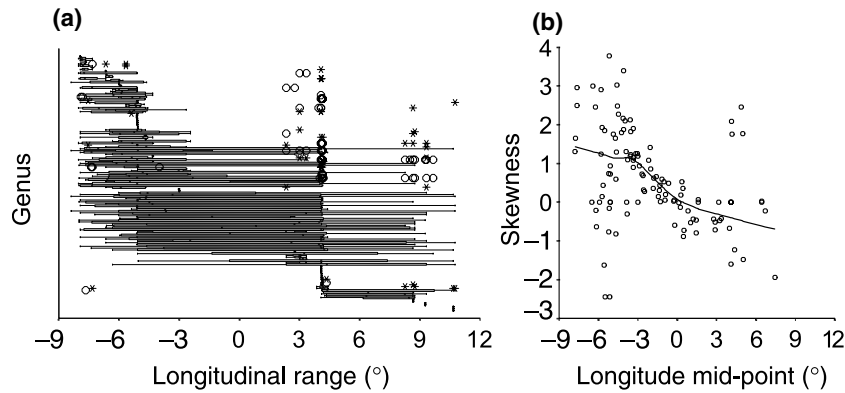


Figure 6 Longitudinal range and symmetry of the spatial distribution of the 139 genera: (a) Boxplots represent the longitudinal range of each genus. Genera were sorted according to their longitude mid-point. Bold lines within boxes represent the median, circles show values between one-and-a-half and three times the box length (outliers) and asterisks represent values above three box-lengths (extremes). (b) Skewness (i.e. distribution symmetry) vs. longitude mid-point of each genus. Skewness values around zero indicate symmetric distribution, positive and negative values reveal distributions skewed to the left and to the right, respectively. Data were smoothed using a Lowess curve to visualize tendencies ($f = 0.60$).

richness patterns, we calculated two descriptors of taxon distribution for each genus: longitude distribution (represented as boxplots, Fig. 6a) and skewness (Fig. 6b). The genera found in the entire longitudinal range exhibited a symmetric distribution (i.e. skewness close to 0), whereas those found in the western and eastern parts had a restricted longitudinal range, skewed to the left (i.e. skewness < 0) and to the right (i.e. skewness > 0), respectively. That analysis allowed visualizing the spatial range of each genera according to its biogeographic distribution from west to east, revealing that genera colonizing the eastern and western sites have a restricted skewed distribution, whereas genera found in the central part of Maghreb were more widespread with a symmetric distribution. Finally, altitude (Fig. 5c) showed a linear increase from sea level to piedmont areas (*ca.* 400 m), and then remained stable and maximal up to 3500 m. These altitude–richness relationships can be related to climatic characteristics such as temperature and precipitation (Table 2). A strong negative linear relationship ($P < 0.01$) was found between maximum, minimum and mean temperatures and altitude. In the same way, precipitation was also related to altitude, but in this case, a strong positive relationship ($P < 0.01$) was only found with minimum monthly precipitation, whereas no linear relationship was detected with maximal monthly precipitation and total annual precipitation ($P > 0.05$).

Table 2 Linear correlation coefficient (r) between altitude and climatic factors

	Temperature			Precipitation		
	Minimum	Maximum	Mean	Minimum	Maximum	Total
Altitude	-0.763**	-0.573**	-0.732**	0.673**	-0.119 (ns)	0.135 (ns)

Minimum, maximum and mean monthly annual temperatures, and minimum, maximum monthly precipitation and total annual precipitation are considered; (ns) not significant, $P > 0.05$; **Highly significant, $P < 0.01$.

DISCUSSION

EPTC vs. overall richness

Although EPT (Ephemeroptera, Plecoptera and Trichoptera) have been considered as sufficiently representative of the overall fauna for water quality assessment procedures such as RIVPACS (Wright *et al.*, 2000) or AUSTRIVAS (Smith *et al.*, 1999), it remains a poor estimator of the richness of the other orders (Table 1). Better results arise when adding Coleoptera to the three previous orders, considering both overall richness and richness of the other orders (overall richness – EPTC). This can be explained by the fact that Ephemeroptera and Trichoptera are generally found in the entire river basins, whereas Plecoptera are usually confined to headwaters (Williams & Feltmate, 1992; Allan, 1995). This explains why the EPT richness is biased at some sites. In contrast, this drawback is balanced by taking into account Coleoptera, which are present over whole river basins, but which are more diverse in downstream sites (Williams & Feltmate, 1992). As a consequence, EPTC can be considered as a relevant descriptor of overall taxon richness, in accordance with Cayrou *et al.* (2000). Moreover, the reliable assessment of overall macroinvertebrate richness using a restricted number of taxa (EPTC) confirms the efficiency of such an approach in biodiversity analysis and that this method can be employed for richness studies at different spatial scales, from unique drainage basin (Cayrou *et al.*,

2000) to macroscale analysis with a large range of stream conditions.

Model reliability

The main processes that determine taxon richness were adequately approximated by ANN, which are known to be able to model nonlinear systems in ecology without transformation of the data (Lek *et al.*, 1996; Scardi, 1996). In the present study, the network was able to reliably recognize (i.e. training procedure) and predict new data (i.e. leave-one-out testing procedure) with a good performance (Fig. 4b). The quality of the models, estimated using the correlation coefficient (r) between observed and estimated genus richness, was similar for training and testing procedures, indicating that the model is fully able to predict taxon richness from new and independent samples and that the data matrix adequately represents the large range of environmental situations encountered in North Africa. Indeed, the model never failed to predict any sample, and even though some discrepancies were found between observed and estimated values, high richness sites were never predicted as low richness areas and inversely, areas with low richness were rarely predicted as diversified. This result first indicates that our model is able to produce relevant results and to deal with the complexity of the natural environment using only three macroscale environmental variables. Second, the absence of unexpected results (i.e. model failure) indicates the relevance of the data matrix. Indeed, if some faunistic lists had been poorly set up or if disturbed sites had been considered (i.e. subjected to high anthropic disturbance), the model would not be able to predict relevant richness, as the three considered macrovariables do not account for anthropic disturbances or unrealistic faunistic lists. Such discrepancy between observed and estimated values are commonly used as a river health assessment method (Wright *et al.*, 2000; Oberdorff *et al.*, 2001) or as an estimator of the relevance of the information introduced in the modelling procedure (Brosse & Lek, 2000). Therefore, it can be concluded that the data matrix is relevant to our study, and that the three environmental predictors provide sufficient information to predict taxon richness. It should also be noticed that the stepwise procedure applied to ANN models showed that genus richness in North Africa can be relevantly predicted using only three environmental variables showing that taxon richness can be predicted using a few macroscale features integrating the main environmental characteristics, as previously shown for the prediction of global and regional riverine fish richness (Guégan *et al.* 1998; Oberdorff *et al.* 1999).

Relationships between taxon richness and environmental variables

The ANN model identifies three influential environmental variables to the prediction of genus richness. From an ecological point of view, these relationships between environmental variables and taxon richness should be considered

with caution, as these analyses, based on correlation, do not necessarily involve relevant ecological processes (Begon *et al.*, 1996). However, the only way to establish reliable causal relationships is to use experimental designs, which is not possible in the case of large scale biogeographical studies. To deal with this possible drawback, the results of our macroscale analysis are justified using previous experimental and theoretical studies conducted at a lower spatial scale, which provided causal relationships between environmental variables and taxon richness (see Frissell *et al.*, 1986; Vinson & Hawkins, 1998, for a review). Such procedures are commonly and efficiently used in large scale ecological studies (Brown, 1995; Begon *et al.*, 1996), such as fish or macroinvertebrate richness approaches (Guégan *et al.*, 1998; Vinson & Hawkins, 1998). The three environmental variables considered are known to be influential on benthic fauna and their influence on macroinvertebrate richness and life histories has previously been underlined in experimental studies (see references below). Therefore, the relationships between these variables and taxon richness identified in this paper account for environmental processes influencing stream macroinvertebrate richness patterns in North Africa.

The number of rainy days was identified as the most influential variable, as it accounted for more than 60% of the information in the model. This variable is the only pluviometric parameter retained in the model and represents the annual number of rainy events. Rainfall constitutes the major source of water for stream ecosystems (Minshall, 1988; Resh *et al.*, 1988; Allan, 1995) and the frequency of rainy events determines the amount of water for streams within watersheds (Abdul & Gillham, 1989; Groisman *et al.*, 2001). Streamflow patterns are strongly dependent on rainfall, and therefore act on the dynamics of macroinvertebrate fauna (Flecker & Feifarek, 1994). Such a tendency occurs in Mediterranean ecosystems with a dry climate, so rainy events play a key role in aquatic ecosystem functioning (Gasith & Resh, 1999). In the present macroscale context, the number of rainy days influences streamflow constancy. The selection of the number of rainy days (i.e. constancy of water supply) in the model, rather than maximal precipitation (i.e. amount of water supply), underlines that taxonomic richness is more influenced by water supply frequency than water amount. Surprisingly, the coefficient of variation of precipitation was not selected as influential, whereas this variable is commonly used to characterize habitat disturbance in large scale studies (Ricklefs, 1980; Hugues & James, 1989; Poff, 1996). The low influence of this variable, calculated from monthly precipitation and, therefore, representing seasonal dynamics (periodic succession of winter rainfall events and dry summer season in North Africa), may be related to its inability to integrate the temporal dimension of disturbing rainy events (i.e. flash floods), which are known to be unpredictable in Mediterranean ecosystems (Gagneur, 1994; Gasith & Resh, 1999). Although flash floods influence macroinvertebrate richness (Gagneur, 1994; Gasith & Resh, 1999), no relationship has been detected between number of rainy days and the coefficient of varia-

tion of precipitation ($r = -0.136$; $P = 0.049$). Thus, even if stormy disturbing events exist, and considering the huge contribution of the number of rainy days in the model, these observations favour a time sequence effect such as seasonal intermittence. In addition, a highly significant negative correlation between number of rainy days and maximum temperature ($r = -0.712$; $P < 0.001$) suggests a climatic interaction responsible for limited rainfall events (Arora & Boer, 2001) giving rise to harsh summer conditions for the fauna. In the same way, a highly significant negative correlation occurs with the coefficient of variation of the number of rainy days ($r = -0.490$; $P < 0.001$), underlying an increasing stability of water supply with an increasing number of annual rainy days. The number of rainy days, therefore, accounts for the temporal persistence of suitable habitats for macroinvertebrates. In this way, an increase of taxonomic richness can be ascribed to an increase of environmental stability as suggested by the habitat templet concept (Southwood, 1977). According to these considerations, the Lowess analysis of the number of rainy days (Fig. 5a) suggests that a drying summer period, below sixty rainy days per year, reduces taxon richness. Taxon richness, therefore, increases inversely to stream intermittence. Above sixty rainy days, drying did not occur, explaining a higher environmental stability and hence higher taxon richness. In these perennial streams, richness is, therefore, hardly affected by the number of rainy days.

The longitudinal distribution of the fauna provides information about the spatial patterns of taxon richness (Fig. 5b). The spatial distribution of the 139 genera shows a west-east geographical organization of the fauna. Large scale spatial richness patterns are commonly related to latitudinal gradient as a surrogate for the environmental context (Begon *et al.*, 1996; Gaston & Spicer, 1998; Lyons & Willig, 1999; Gaston, 2000). However, given the limited latitudinal range of the Maghreb, latitude cannot be used as a relevant macro-scale variable. In contrast, spatial patterns can be illustrated using a longitudinal gradient between the same climatic and biogeographical zones (Illies, 1978). Within a regional biogeographical context, the Maghreb can be considered as an island for aquatic invertebrates that are isolated by natural boundaries from streams belonging to adjacent biogeographic zones: the Mediterranean Sea to the North and the Sahara desert to the South (Fig. 1a). Macroinvertebrates from the Maghreb are strongly related to European fauna (i.e. more than 80% of the Maghreb genera are known from Europe), whereas Afrotropical invertebrates are poorly represented in these two land masses (Illies, 1978). According to the island biogeography concept (McArthur & Wilson, 1967), which explains diversity gradients by distances between land masses and colonization capabilities, we can hypothesize that most population exchanges are likely to occur between Europe and the Maghreb, via Spain and Morocco in the western part of Maghreb, and Italy and Tunisia in the eastern part (Fig. 1a). According to Fig. 5b, richness is maximal in the western part where the sea channel is narrow, allowing easy and frequent

population exchanges. A similar process is identified in the eastern part, but is less marked because of the increased distance between the land masses. This assumption is supported by study of the spatial range and distribution shape of each genus (Fig. 6). According to Tilman & Kareiva (1997), in one biogeographic unit, taxa should be distributed symmetrically. This was verified for the genera found in the entire longitudinal range (Fig. 6). However, taxa with limited distribution in the Maghreb were mainly located in the western and eastern parts. Western taxa exhibited a distribution skewed to the left (i.e. skewness < 0), whereas eastern taxa had a distribution skewed to the right (i.e. skewness > 0). As the theoretical distribution of a given taxon is symmetrical (Tilman & Kareiva, 1997), the complementary part of the distribution range of taxa restricted to the western and eastern sites is found in Spain and Italy (Illies, 1978), respectively. This is supported by previous observations on macroinvertebrate population exchanges between Europe and the Maghreb using these dispersion routes, as shown by Lounaci (1987). In contrast, genera having a symmetrical distribution in Maghreb are represented by widespread North African genera, and abundant genera in both Europe and Maghreb (Illies, 1978). This justifies past and present biogeographical linkages between Europe and Maghreb, confirming that longitude accounts for geographical patterns, within the context of insular biogeography (McArthur & Wilson, 1967). These results parallel recent theoretical considerations concerning population exchanges between large land masses (Lomolino, 2000), explaining the low genus endemism in North Africa. Within this framework, one last question that remains concerns the lower richness in the central areas (central and western Algeria). This can be explained by river scarcity in North Africa lowering population exchanges between Western and Eastern areas, the relative aridity, therefore, limiting macroinvertebrate dispersal, as suggested by Bilton *et al.* (2001).

Altitude is a general integrative factor known to affect fauna through thermal processes (Gaston & Spicer, 1998), and stream insect taxon richness has already been related to that thermal gradient (Vinson & Hawkins, 1998). The relationship between altitude and taxon richness is generally assumed to be a decreasing one because of the evident negative relationship between temperature and altitude (Begon *et al.*, 1996; Gaston & Spicer, 1998). These conclusions, drawn from temperate biomes contrast with our results which show a steep diversity increase from sea level to piedmont areas (Fig. 5c). Such a tendency has already been reported (Dodds & Hisaw, 1925; Tate & Heiny, 1995; Lomolino, 2001), and is generally explained by a higher structural homogeneity of lowland rivers, leading to a reduction of habitat diversity. However, homogeneity may not be sufficient to reduce macroinvertebrate richness down to close-to-zero values as observed in this study. In the case of Mediterranean rivers, in addition to an assumed low habitat heterogeneity, the sharp decline in richness can be attributed to water supply, as a highly significant positive relationship has been detected between altitude and

precipitation minima, and altitude and temperature (non-linear smoothing methods were tried but did not provide significantly better results) (Table 2). Lowland areas are therefore subjected to seasonal droughts or at least to drastic reductions of stream flow (Boumaiza & Thomas, 1986; Gagneur, 1994), with high water temperatures (> 30 °C) (Lounaci *et al.*, 2000), which acts as a huge restricting factor for fauna (Williams & Feltnate, 1992). This major disturbance strongly affects the stream fauna (Vinson & Hawkins, 1998; Gasith & Resh, 1999), which is, overall, rather similar to the one encountered in temperate Europe (Illies, 1978), leading to reduced taxa richness in the lowland sites, as previously observed in the Sebaou drainage basin (North East Algeria) (Lounaci *et al.*, 2000). However, the amplitude of water level fluctuation varies according to river size, as annual drought is not observed for the largest rivers. Therefore, the high heterogeneity of taxon richness in the low-altitude sites accounts for both large perennial rivers and small temporary coastal streams. In the highest mountain sites, i.e. above 2000 m, our results are also not consistent with previous studies. The predictable steep diversity decrease in high mountains known from temperate environments (Ward, 1994) was not observed. Once again, this can be ascribed to the particular climatic characteristics of North Africa, leading to a relatively mild climate, even in the highest mountain sites, compared to equivalent altitudes in temperate biomes.

CONCLUSIONS

The explanation and the prediction of stream macroinvertebrate richness patterns in North Africa have been efficiently achieved using three environmental variables known, from previous local and theoretical studies, to play a major role (Frissell *et al.*, 1986; Minshall, 1988; Vinson & Hawkins, 1998). This shows that overall trends in macroinvertebrate distribution and richness on a large geographical scale can be relevantly assessed using a few pertinent environmental variables providing insights into stream ecology and biogeography in North Africa. However, two of these environmental variables (altitude and number of rainy days), illustrate the particular climatic characteristics of North Africa. Taxon richness was highly dependent on stream flow constancy, ruling habitat dynamics and persistence of fauna. The third variable (longitude) provided insights into geographical patterns, i.e. connections between land masses, within the island biogeography context. Macroinvertebrate richness, therefore, results from a combination of climatic and hydrological parameters which screen geographic processes. This study underlines the specificity of Mediterranean ecosystems and illustrates the concomitant action of both geographical and environmental factors in the generation of aquatic insect diversity. From a more theoretical point of view our results justify and add weight to current biogeographic and ecological concepts based on the influence of environmental and geographical gradients on taxon richness, mainly derived from European and North American streams.

ACKNOWLEDGMENTS

S. Lek, A. Lounaci and A.G.B. Thomas provided helpful comments during the preparation of this manuscript.

REFERENCES

- Abdul, A.S. & Gillham, R.W. (1989) Field studies of the effects of the capillary fringe on streamflow generation. *Journal of Hydrology*, **112**, 1–18.
- Allan, J.D. (1995) *Stream ecology: structure and function of running waters*. Chapman & Hall, London.
- Arora, V.K. & Boer, G.J. (2001) Effects of simulated climate change on the hydrology of major river basins. *Journal of Geophysical Research Atmospheres*, **106**, 3335–3348.
- Begon, M., Harper, J.L. & Townsend, C.R. (1996) *Ecology: individuals, population, and communities*, 3rd edn. Blackwell Science, Oxford.
- Bilton, D.T., Freeland, J.R. & Okamura, B. (2001) Dispersal in freshwater invertebrates. *Annual Review of Ecology and Systematics*, **32**, 159–181.
- Boumaiza, M. (1994) *Recherche sur les eaux courantes de Tunisie. Faunistique, Ecologie et Biogéographie*. PhD Dissertation, University of Tunis, Tunisia.
- Boumaiza, M. & Thomas, A.G.B. (1986) Répartition et écologie des Ephéméroptères de Tunisie (1^{ère} partie) (Insecta, Ephemeroptera). *Archives de l'Institut Pasteur de Tunis*, **63**, 567–599.
- Brosse S. & Lek S. (2000) Modelling roach (*Rutilus rutilus*) microhabitat using linear and nonlinear techniques. *Freshwater Biology*, **44**, 441–452.
- Brosse, S., Lek, S. & Dauba, F. (1999) Predicting fish distribution in a mesotrophic lake by hydroacoustic survey and artificial neural networks. *Limnology & Oceanography*, **44**, 1293–1303.
- Brosse, S., Lek, S. & Townsend, C.R. (2001) Abundance, diversity, and structure of freshwater invertebrates and fish communities: an artificial neural network approach. *New Zealand Journal of Marine and Freshwater Research*, **35**, 135–145.
- Brosse, S., Arbuckle, C.J. & Townsend, C.R. (2003) Habitat scale and biodiversity: influence of catchment, stream reach and bedform scales on local invertebrate diversity. *Biodiversity and Conservation*, **12**, 2057–2075.
- Brown, J.H. (1995) *Macroecology*. University of Chicago Press, Chicago.
- Carter, J.L., Fend, S.V. & Kennelly, S.S. (1996) The relationships among three habitat scales and stream benthic invertebrates community structure. *Freshwater Biology*, **35**, 109–124.
- Cayrou, J., Compin, A., Giani, N. & Cereghino, R. (2000) Associations spécifiques chez les macroinvertébrés benthiques et leur utilisation pour la typologie de cours d'eau. Cas du réseau hydrographique Adour-Garonne. *Annales de limnologie*, **36**, 189–202.
- Cleveland, W.S. (1979) Robust locally-weighted regression and scatterplot smoothing. *Journal of the American Statistical Association*, **74**, 829–836.
- Corkum, L.D. (1992) Spatial distributional patterns of macroinvertebrates along rivers within and among biomes. *Hydrobiologia*, **239**, 101–114.

- Dimopoulos, I., Chronopoulos, J., Chronopoulou-Sereli, A. & Lek, S. (1999) Neural network models to study relationships between lead concentration in grasses and permanent urban descriptors in Athens city (Greece). *Ecological Modelling*, **120**, 157–165.
- Dodds, G.S. & Hisaw, F.L. (1925) Ecological studies on aquatic insects, IV. Altitudinal range and zonation of mayflies, stoneflies and caddisflies in the Colorado Rockies. *Ecology*, **6**, 380–390.
- Dolédéc, S., Olivier, J.M. & Stutzner, B. (2000) Accurate description of taxa and their biological traits in stream invertebrate communities: effects of taxonomic and spatial resolution. *Archiv für Hydrobiologie*, **148**, 25–53.
- Efron, B. (1983) Estimating the error rate of a prediction rule: some improvements on cross-validation. *Journal of the American Statistical Association*, **78**, 316–331.
- Flecker, A.S. & Feifarek, B. (1994) Disturbance and the temporal variability of invertebrate assemblages in two Andean streams. *Freshwater Biology*, **31**, 131–142.
- Frissell, C.A., Liss, W.J., Warren, C.E. & Hurley, M.D. (1986) A hierarchical framework for stream habitat classification: viewing streams in a watershed context. *Environmental Management*, **10**, 199–214.
- Gagneur, J. (1994) Flash floods and drying up as major disturbance upon benthic communities in North-African wadis. *Verhandlungen der Internationale Vereinigung für theoretische und angewandte Limnologie*, **25**, 1807–1811.
- Gasith, A. & Resh, V.H. (1999) Streams in Mediterranean climate regions: abiotic influences and biotic responses to predictable seasonal events. *Annual Review of Ecology and Systematics*, **30**, 51–81.
- Gaston, K.J. (2000) Global patterns in biodiversity. *Nature*, **405**, 220–227.
- Gaston, K.J. & Spicer, J.I. (1998) *Biodiversity: an introduction*. Blackwell Science, Oxford.
- Gevrey, M., Dimopoulos, I. & Lek, S. (2003) Review and comparison of methods to study the contribution of variables in Artificial Neural Networks models. *Ecological Modelling*, **160**, 249–264.
- Groisman, P.Y., Knight, R.W. & Karl, T.R. (2001) Heavy precipitation and high streamflow in the contiguous United States: trends in the twentieth century. *Bulletin of the American Meteorological Society*, **82**, 219–246.
- Guégan, J.F., Lek, S. & Oberdorff, T. (1998) Energy availability and habitat heterogeneity predict global riverine fish diversity. *Nature*, **391**, 382–384.
- Hugues, J.M.R. & James, B. (1989) A hydrological regionalization of streams in Victoria, Australia, with implications for stream ecology. *Australian Journal of Marine and Freshwater Research*, **40**, 303–326.
- Huston, M.A. (1994) *Species diversity: the coexistence of species on changing landscapes*. Cambridge University Press, Cambridge.
- Illies, J. (1978) *Limnofauna Europaea*. Gustav Fischer Verlag, New York.
- James, F.C. & McCulloch, C.E. (1990) Multivariate analysis in ecology and systematics: panacea or Pandora's box? *Annual Review of Ecology and Systematics*, **21**, 129–166.
- James, F.C., McCulloch C.E. & Wiedenfeld D.A. (1996) New approaches to the analysis of population trends in land birds. *Ecology*, **77**, 13–27.
- Kohavi, R. (1995) A study of the cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1137–1143, Montreal, Canada.
- Lavorel, S. & Richardson, D.M. (1999) Diversity, stability and conservation of Mediterranean-type ecosystems in a changing world: an introduction. *Diversity and Distributions*, **5**, 1–2.
- Leemans, R. & Cramer, W. (1991) The IIASA database for mean monthly values of temperature, precipitation and cloudiness of a global terrestrial grid. *International Institute for Applied System Analysis (IIASA)*, RR-91–18.
- Leith, H. (1975) Modelling primary productivity of the world. *Primary productivity of the Biosphere* (ed. by H. Leith and R.H. Whittaker), pp. 237–263. Springer Verlag, New York.
- Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J. & Aulagnier, S. (1996) Application of neural networks to modelling non-linear relationships in ecology. *Ecological Modelling*, **90**, 39–52.
- Lek, S. & Guégan, J.F. (2000) *Artificial neuronal networks, applications to ecology and evolution*. Springer Verlag, Berlin.
- Li, J., Herlihy, A., Gerth, W., Kaufman, P., Gregory, S., Hurquhart, S. & Larsen, D.P. (2001) Variability in stream macroinvertebrates at multiple spatial scales. *Freshwater Biology*, **46**, 87–97.
- Lobo, J.M. & Davis, A.L.V. (1999) An intercontinental comparison of dung beetle diversity between two Mediterranean-climatic regions: local versus regional and historical influences. *Diversity and Distributions*, **5**, 91–103.
- Lomolino, M.V. (2000) A call for a new paradigm of island biogeography. *Global Ecology & Biogeography*, **9**, 1–6.
- Lomolino, M.V. (2001) Elevation gradients of species-density: historical and prospective views. *Global Ecology & Biogeography*, **10**, 3–13.
- Lounaci, A. (1987) *Recherches hydrobiologiques sur les peuplements d'invertébrés benthiques du bassin de l'oued Aïssi (Grande Kabylie)*. PhD Dissertation, University of Algiers, Algeria.
- Lounaci, A., Brosse, S., Ait Mouloud, S., Lounaci-Daoudi, D., Mebarki, N. & Thomas, A. (2000) Current knowledge of benthic invertebrate diversity in an Algerian stream: a species check-list of the Sébaou River basin (Tizi-Ouzou). *Bulletin de la Société d'Histoire Naturelle de Toulouse*, **136**, 43–55.
- Lyons, S.K. & Willig, M.R. (1999) A hemispheric assessment of scale dependence in latitudinal gradients of species richness. *Ecology*, **80**, 2483–2491.
- McArthur, R.H. & Wilson, E.O. (1967) *The theory of island biogeography*. Princeton University Press, Princeton.
- Minshall, G.W. (1988) Stream ecosystem theory: a global perspective. *Journal of the North American Benthological Society*, **7**, 263–288.
- Myers, N., Mittermeier, R.A., Mittermeier, C.G., Da Fonseca, G.A.B. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.
- Oberdorff, T., Hugueny, B. & Guégan, J.F. (1997) Is there an influence of historical events on contemporary fish species

- richness in rivers? Comparison between Western Europe and North America. *Journal of Biogeography*, **24**, 461–469.
- Oberdorff, T., Lek, S. & Guegan, J.F. (1999) Patterns of endemism in riverine fish of the Northern hemisphere. *Ecology Letters*, **2**, 75–81.
- Oberdorff, T., Pont, D., Hugueny, B. & Chessel, D. (2001) A probabilistic model characterizing fish assemblages of French rivers: a framework for environmental assessment. *Freshwater Biology*, **46**, 399–417.
- Poff, N.L. (1996) A hydrogeography of unregulated streams in the United States and an examination of scale-dependence in some hydrological descriptors. *Freshwater Biology*, **36**, 71–91.
- Resh, V.H., Brown, A.V., Covich, A.P., Gurtz, M.E., Li, H.W., Minshall, W., Reice, S.R., Sheldon, A.L., Wallace, J.B. & Wissmar, R.C. (1988) The role of disturbance in stream ecology. *Journal of the North American Benthological Society*, **7**, 433–455.
- Reyjol, Y., Lim, P., Belaud, A. & Lek, S. (2001) Modelling of microhabitat used by fish in natural and regulated flows in the river Garonne (France). *Ecological Modelling*, **146**, 131–142.
- Ricklefs, R.E. (1980) Latitudinal patterns in seasonality and year-to-year variation in climate. *Caribbean Journal of Science*, **15**, 183–197.
- Ricklefs, R.E. & Schluter, D. (1993) *Species diversity in ecological communities*. University of Chicago Press, Chicago.
- Rumelhart, D.E., Hinton, G.E. & Williams, R.J. (1986) Learning representations by back-propagating error. *Nature*, **323**, 533–536.
- Samways, M.J. & Osborn, R. (1998) Divergence in transoceanic circumtropical dragonfly on a remote island. *Journal of Biogeography*, **25**, 935–946.
- Scardi, M. (1996) Artificial neural networks as empirical models for estimating phytoplankton production. *Marine Ecology Progress Series*, **139**, 289–299.
- Smith, M.J., Kay, J.R., Edward, D.H.D., Simpson, J.C., Pender, A.M., Corle, D.J., Honwitz, P.H.J., Davis, J.A., Yung, F.H., Norvis, R.H. & Horsle, S.A. (1999) AUSRIVAS: using macroinvertebrates to assess ecological conditions of rivers in western Australia. *Freshwater Biology*, **41**, 269–282.
- Southwood, T.R.E. (1977) Habitat, the templet for ecological strategies? *Journal of Animal Ecology*, **46**, 337–365.
- Tate, C.M. & Heiny, J.M. (1995) The ordination of benthic invertebrate communities in the South Platte River basin in relation to environmental factors. *Freshwater Biology*, **33**, 439–454.
- Tilman, D. & Kareiva, P. (1997) *Spatial ecology. The role of space in population dynamics and interspecific interactions*. Princeton University Press, Princeton, USA.
- Trexler, J.C. & Travis J. (1993) Non-traditional regression analyses. *Ecology*, **74**, 1629–1637.
- Vinson, M.R. & Hawkins, C.P. (1998) Biodiversity of stream insects: variation at local, basin and regional scales. *Annual Review of Entomology*, **43**, 271–293.
- Walley, W.J. & Fontama, V.N. (1998) Neural network predictors of average score per taxon and number of families at unpolluted river sites in Great Britain. *Water Research*, **32**, 613–622.
- Ward, J.V. (1994) Ecology of Alpine streams. *Freshwater Biology*, **32**, 277–294.
- Williams, D.D. & Feltmate, B.W. (1992) *Aquatic insects*. CAB International.
- Wright, J.F., Sutcliffe, D.W. & Furse, M.T. (2000) *Assessing the biological quality of freshwaters: RIVPACS and other techniques*. Freshwater Biological Association, Ambleside.

BIOSKETCHES

Olivier Beauchard is a post-graduate student in Ecology at the University of Toulouse. His research is focused on the spatial distribution patterns of stream macroinvertebrates in North Africa. He is also interested in data mining and spatial modelling.

James Gagneur is an Associate Professor of Biology at the University of Toulouse. His research interests include biology and ecology of aquatic insects, mainly in North Africa where he worked for many years. He is also interested in conservation ecology of streams and water quality management.

Sebastien Brosse is an Associate Professor of Biology at the University of Toulouse. He works on the spatial distribution and diversity patterns of freshwater fish and invertebrates. He is also interested in nonlinear statistical modelling, and particularly in the application of artificial neural network models to ecology.

Appendix I Bibliographic sources used to set up the macroinvertebrate database

Author	Year of publication	Reference	Study area
Arab	1989	Master's thesis, Algiers (Algeria)	Central Algeria
Berrahou <i>et al.</i>	2001	<i>Ann Limnol</i>	East Morocco
Boumaïza	1994	PhD, Tunis, Tunisia	Tunisia
Boumaïza & Clergue-Gazeau	1986	<i>Ann Limnol</i>	Tunisia
Boumaïza & Laville	1988	<i>Ann Limnol</i>	Tunisia
Boumaïza & Thomas	1995	<i>Bull Soc Hist Nat Toulouse</i>	Tunisia
Boumaïza & Thomas	1986	<i>Arch Inst Pasteur Tunis</i>	Tunisia
Boumaïza <i>et al.</i>	1986	<i>Ann Limnol</i>	Tunisia
Bouzidi	1989	PhD, Aix-Marseille (France)	Central Morocco
Dakki	1979	PhD, Aix-Marseille (France)	North Morocco
Dakki	1985	<i>Bull Ecol</i>	North Morocco
Djeridane & Salhi	1983	Master's thesis, Algiers (Algeria)	Central Algeria
El Agbani <i>et al.</i>	1992	<i>Bull Ecol</i>	North-west Morocco
Gagneur	1981–1984	Unpubl. data	West Algeria
Gagneur & Aliane	1991	<i>Overview and strategies of Ephemeroptera and Plecoptera</i> , Alba-Tercedor J. & Sanchez-Ortega A. Sandhill Crane Press	West Algeria
Gagneur & Clergue-Gazeau	1988	<i>Ann Limnol</i>	West Algeria
Gagneur & Thomas	1988	<i>Bull Soc Hist Nat Toulouse</i>	West Algeria
Gagneur <i>et al.</i>	1985	<i>Bull Soc Hist Nat Toulouse</i>	West Algeria
Gagneur <i>et al.</i>	1986	<i>Bull Soc Hist Nat Toulouse</i>	West Algeria
Giudicelli & Thiery	1985	<i>Bull Zool Mus</i>	West Morocco
Kraïem	1986	<i>Bull Mens Soc Linn Lyon</i>	North-east Tunisia
Kraïem	1987	<i>Arch Inst Pasteur Tunis</i>	North-east Tunisia
Lounaci <i>et al.</i>	2000	<i>Bull Soc Hist Nat Toulouse</i>	East Algeria
Mouehli <i>et al.</i>	1998	<i>Bull Soc Hist Nat Toulouse</i>	Tunisia
Pihan & Mohati	1983	<i>Bull Fac Sci Marrakech</i>	West Morocco
Pihan <i>et al.</i>	1987	<i>Rev Fac Sci Marrakech</i>	West Morocco
Touabay <i>et al.</i>	2001	<i>Ann Limnol</i>	Central Morocco
Vaillant & Gagneur	1998	<i>Ann Soc Entomol</i>	West Algeria
Zouakh	1995	Master's thesis, Algiers (Algeria)	Central Algeria