Energy availability and habitat heterogeneity predict global riverine fish diversity

Jean-François Guégan*, Sovan Lek† & Thierry Oberdorff‡

* ORSTOM, Laboratoire d'Hydrobiologie Marine et Continentale,
UMR CNRS 5556, Station Méditerranéenne de l'Environnement Littoral de
l'Université de Montpellier II, 1 quai de la Daurade, F-34200 Sète, France
† CNRS-UMR 5576, CESAC, Université Paul Sabatier, Bât. IVR3,
118 route de Narbonne, F-31062 Toulouse cedex, France
‡ Muséum National d'Histoire Naturelle, Laboratoire d'Ichtyologie générale et
appliquée, 43 rue Cuvier, F-75231 Paris cedex 05, France

....

Processes governing patterns of richness of riverine fish species at the global level can be modelled using artificial neural network (ANN) procedures. These ANNs are the most recent development in computer-aided identification and are very different from conventional techniques[1,2]. Here we use the potential of ANNs to deal with some of the persistent fuzzy and nonlinear problems that confound classical statistical methods for species diversity prediction. We show that riverine fish diversity patterns on a global scale can be successfully predicted by geographical patterns in local river conditions. Nonlinear relationships, fitted by ANN methods, adequately describe the data, with up to 93 per cent of the total variation in species richness being explained by our results. These findings highlight the dominant effect of energy availability and habitat heterogeneity on patterns of global fish diversity. Our results reinforce the species-energy theory[3] and contrast with those from a recent study on North American mammal species[4], but, more interestingly, they demonstrate the applicability of ANN methods in ecology.

A central issue in macroecology is to determine the forces that shape large-scale patterns of species richness[5,6]. Three main hypotheses have been proposed to explain the spatial variability in species diversity. The first, the species-area hypothesis[7], implies that species richness increases as a power function of surface area; the second, the species-energy hypothesis[3,8], predicts that species variation is correlated with energy availability in the system; the third, the historical hypothesis[9], explains species richness gradients in terms of patterns of recolonization and maturation of ecosystems after glaciation. However, so far none of the three theories has been supported to the exclusion of the other, and many causative factors have been cited even though total available energy has gained currency as a major influencing parameter of species diversity. Here we model processes governing patterns of riverine fish species richness. We use ANNs, known for their capacity to process non-linear relationships between variables[1,2]. The data we present are best explained by the hypothesis that both distribution of available energy and habitat heterogeneity limit fish species richness in rivers on a worldwide scale.

Global-scale patterns of fish species richness in rivers have previously been investigated using linear statistical models[10]. Results suggest that factors related to components of river size (surface area and flow regime) and energy availability (net primary productivity) are most important in predicting fish diversity, whereas the role of other possible factors (such as contemporary climate and/or history) are of only marginal importance. The effect of contemporary available energy has been demonstrated on different groups of organisms[10–14], although some other factors (historical influence, for example) may predict patterns of richness[9,15]. We have reinvestigated previous work[10] employing ANN methods, which do not require a linear relationship between variables and so may be better suited to model nonlinear phenomena (Fig. 1). ANNs are different from multiple linear regressions in that the relationships between independent parameters and fish species richness (SR) are estimated by an iterative trial-and-error procedure. Each influencing parameter
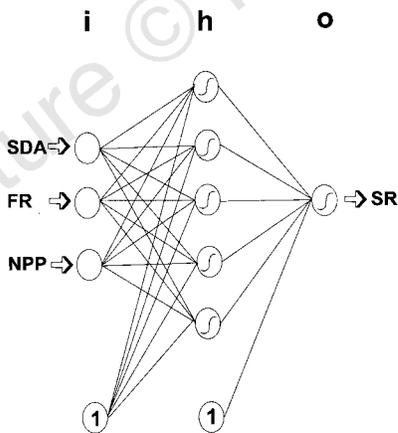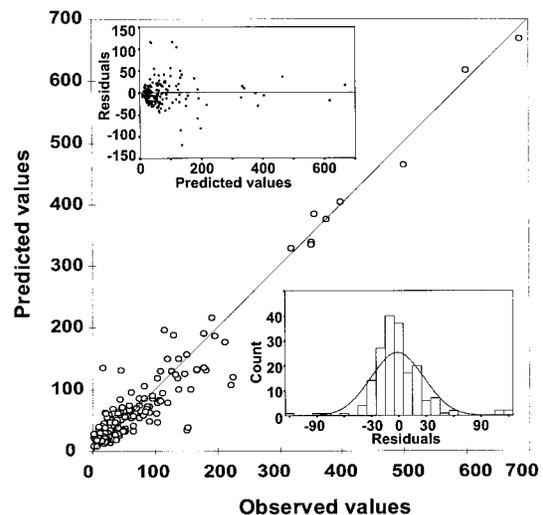

Figure 1 Three-layer feed-forward artificial neural network (ANN) structure used in this work. There are three input neurons (i) for surface of the drainage area (SDA), mean annual flow regime (FR), and net primary productivity (NPP), and only one output neuron (o) which corresponds to fish species richness (SR) in rivers. The hidden layer (h) has five neurons, determined as the optimal configuration that gives lower error during training with minimal computing time. There are two additional bias nodes labelled with a constant input value of 1.0. Initially, the network was trained with a set of 183 rivers and their corresponding parameters for 1,000 iterations. We then examined the capability of the trained network to predict SR with a 'leave-one-out' procedure (see Fig. 4 and statistical analysis).

Figure 2 Prediction of fish species richness (SR) using the 3-5-1 artificial neural network (ANN) model shown in Fig. 1. Scatterplots compare predicted and observed SR values. This relationship is highly significant ($n = 183$, $r = 0.958$, $P < 0.0001$). The diagonal line illustrates points at which the predicted value equals the observed value. Top inset, relationship between the residual values obtained from the ANN model and the predicted values. The horizontal line represents points for which residuals equal zero. The relationship shows no obvious sign of dependence of residuals ($n = 183$, $r = 0.018$, $P = 0.805$), which indicates that the ANN model fits the data well. Lower inset, frequency histogram of residuals with most values centred near zero ($n = 183$, mean = -2.00, s.d., ±28.82).

is assigned with different weights and the combined weighted values are added to predict SR. We tested two aspects of ANN performance. First, we evaluated the ability of the independent parameters (surface of the drainage area (SDA), flow regime (FR), and net primary productivity (NPP)) to predict SR of new samples, and thus for modelling fish species diversity processes on the global scale. Second, we analysed the contribution profile of each predictor as a measure of sensitivity to match data. For this, we used a three-layer feed-forward (3-5-1) neural network: that is, three input neurons corresponding to the three independent parameters, five hidden neurons determined as the optimal configuration (best compromise between bias and variance) and one output neuron for SR, which was trained using the backpropagation algorithm[1] (Fig. 1). In the past decade, ANN models have been widely applied in different research fields[1,2] (physics, chemistry, behavioural sciences) but very few studies have focused on the use of ANNs in theoretical ecology and evolution[2]. We now explore this possibility in the context of conservation ecology.

The 3-5-1 ANN model (Fig. 1) accurately predicted the pattern of observed SR on a global scale (Fig. 2). The contribution profiles of the three predictors for explaining SR estimates are illustrated in Fig. 3. The predictive performance of the ANN model gave significant results with 92.9 per cent of rivers achieving a perfect data fit (Fig. 4), thereby increasing the significance of predictions.

Examination of Fig. 3 shows that for FR and NPP parameters, there is a strong positive effect on richness patterns, with a sigmoidal

contribution between the ability of these two variables to match data and SR values. In contrast, the contribution of the SDA variable (that is, river size) contributes little to variation in global fish diversity, with a contribution profile better fitting a gaussian function where the maximum of sensitivity is achieved for median fish richness values and average river sizes (Fig. 3). This conflicts with previous studies showing the importance of drainage surface area on fish species patterns[15–17]. The small contribution of surface area as a predictor variable indicates the extent to which previous investigations were strongly influenced by log–linear transformations. SDA and FR could be causally linked and these processes are probably acting together, confounding their effects. However, discharge may be a more direct measure of available habitat diversity because it may implicitly integrate a third dimension in river size, the volume of available water for fish communities[16,17].

Interestingly, FR and NPP predictors strongly influence patterns of global-scale SR (Fig. 3). As rivers with high flow regimes may generally contain a greater array of habitat configurations[18], a part of island biogeographic theory[7], local habitat heterogeneity may induce this increase in global SR. Additionally, the influence of net primary productivity, a measure of energy availability, demonstrates the importance of energy input on riverine fish richness patterns on a worldwide scale. This fish species richness–primary productivity function resembles the logistic model obtained for the three species richness–evapotranspiration relationship[14]. Climatic
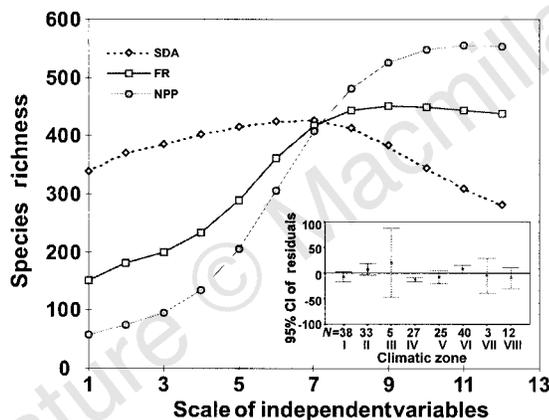
Figure 3 Contribution of the three independent variables (SDA, FR and NPP) used in the 3-5-1 ANN model. Sensitivity profiles explain SR (see Fig. 1 for abbreviations). Contribution of each independent variable to SR estimates is assessed by visual examination of nonlinear profiles: NPP and FR variables have a sigmoid function, with the NPP curve showing a higher range of variation than the FR function; SDA variable better fits a gaussian function. Both NPP (which explained 43% of the total variance) and FR (31%) show a predominant positive effect on SR, whereas SDA (26%) has a weaker effect on SR value. The relative importance of influencing parameters on SR was calculated according to refs 29 and 30. Inset, residual values generated by the ANN model plotted against the 8 categories of climatic zones (I to VIII). Bars represent 95% confidence intervals (CI), and N values (for each categorical climatic zone) indicate the number of rivers analysed. Compared to the zero residual value for which the ANN model perfectly fits the data, we obtained a significant t-test value for climatic zones IV (oceanic areas, overestimated mean value: t-test $= -6.29$, $P < 0.001$) and VI (continental areas, underestimated mean value: t-test $= 2.11$, $P < 0.041$) only. Thus, the additional contribution of climatic topography to global richness patterns appears to be negligible ($r = 0.97$, $P < 0.0001$) in comparison with the total effect of NPP, FR and SDA variables ($r = 0.93$, $P < 0.0001$).
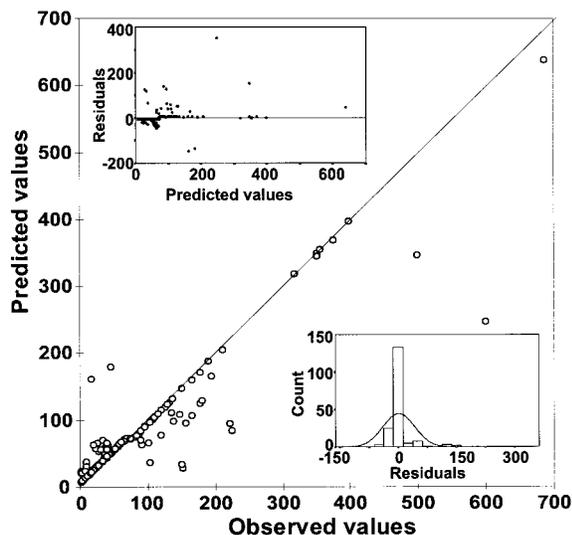
Figure 4 'Leave-one-out' cross-validation test for the 183 rivers analysed in this study. The relationship between predicted and observed SR values is shown; the correlation is highly significant ($n = 183$, $r = 0.929$, $P < 0.0001$) and most points perfectly fit the straight line for which predicted values equal observed ones (many points are superimposed on the figure). Top inset, the relationship shows some dependencies of residuals ($n = 183$, $r = 0.259$, $P < 0.001$) essentially due to some unfitted values in the model. Lower inset, frequency histogram of residuals with most values centred near zero ($n = 183$, mean $= 1.40$, s.d., ± 40.82). Both insets illustrate conventional standard controls on statistics as for insets in Fig. 2.

topography (that is, the presence of a given river in one of the eight conventional climatic zones; see description of data) did not significantly contribute to the explanation of patterns of global-scale fish species richness (Fig. 3, inset), confirming that global fish species variability is not related to topography and latitudinal gradients, as was suggested by the historical hypothesis[9]. Our findings contradict a recent study in which North American mammal species richness was best predicted by a hierarchical sequence of limiting factors: that is, local energy availability was important in comparatively cold regions of high latitudes (Alaska and most of Canada), and topographic heterogeneity best explained species diversity for highly productive regions of the southern part of the continent (Southern Canada and United States)[4]. The main discrepancy between this previous work and our study on fish diversity lies in the different scales of the two analyses and in their geographical scope. Thus, we believe that large-scale fish-richness patterns are best explained by both energy availability and habitat diversity: the more energy available, the more fish species the aquatic environment can support; additionally, for regions with identical energy inputs, habitat heterogeneity may favour coexistence of more fish species.

The advantage of ANNs over conventional models stems from their ability directly to take into account nonlinear relationships, a common stumbling block when dealing with ecological systems[19–21], and so provide a more precise idea of the relationship between any influencing parameter and its dependent factor. Therefore, they are powerful models for forecasting purposes. A previous study using logarithmic transformation of variables succeeded in explaining up to 78% of the total variation in richness[10], whereas the ANN method achieved a much higher level (~93%) with only three environmental parameters. Here, we show that large-scale species richness in riverine fish varies among regions as a nonlinear function of contemporary available energy in the system and local habitat heterogeneity in rivers. Characteristics of the ANN model parameters are practical for predicting and assessing global trends in biodiversity loss and habitat fragmentation. Important pervasive forms of environmental degradation due to human activities usually include source pollution, altered hydrological regimes (by dams, diversions and withdrawals) and habitat destruction[22,23]. Consequences of such degradation are that many aquatic species are now threatened with extinction[24]. To protect and maintain aquatic (and terrestrial) biodiversity, an understanding of the relationships between species and ecological processes that shape the entire ecosystem is essential[25,26]. The development of artificial neural network models is a task of major importance in view of projections of global environmental change and the need for water-resource management. □

Methods

Description of data. Data employed in this study[10] are based on a subsample of 183 plots (from a total of 292 rivers) for which all parameter values were available. We omitted the Amazon River basin, for which the richness value of ~2000 known fish species is subject to considerable error, and which, because of its extremely high value, may considerably bias the type of statistics used. We selected the most recent references and adjusted species number to account for extinction and introduction wherever possible. Only riverine fish species were included in the analyses and secondary or migratory euryhaline fishes were systematically withdrawn. Values for species richness (SR) refer to the total number of riverine fish species collected from the entire drainage basin, which corresponds to the current community richness per river. Three independent parameters were used as the best predictors of SR. There were: total surface of the drainage area (SDA) (generally taken from the literature, in km^2); mean annual flow regime (FR) at the river mouth (also taken from the literature, in $m^3 s^{-1}$, data were not available for all rivers, so we used only 183 rivers from the entire data set of 292); and net terrestrial primary productivity (NPP), which refers to the rate of energy flow through the plants of the region where a given river is located. We then tested the influence of contemporary climate

topography on SR values by analysing residual variations of richness obtained from the ANN model across the 8 conventional climatic zones (I to VIII). The rivers fall into the following zones: I, equatorial zone with very high annual precipitation; II, tropical summer-rainfall zone, with heavy rains in the summer and extreme drought during the cooler season; III, subtropical dry-zone of deserts, with very low rainfall; IV, Mediterranean transition zone with winter rainfall; V, warm-temperate climate zone with high humidity in summer; VI, temperate climate zone, with moderate humidity; VII, arid temperate climate zone of continental regions, with low rainfall; VIII, cold–temperate or boreal climate zone, with high precipitation.

Statistical analysis. ANN models are known for their capacity to process nonlinear relationships. We used one of the principles of ANNs, (the backpropagation algorithm[27]), and a 'leave-one-out' cross-validation test (where each river sample is left out of the model formulation in turn and predicted once) to determine its performance (Fig. 4). This procedure is appropriate when the data set is quite small and/or when each sample is likely to have 'unique information' that is relevant to the regression model[27,28], as is frequently found in ecology. We used a typical three-layer feed-forward (3-5-1) ANN (Fig. 1). To determine the relative importance of the three input parameters, we used the procedure for partitioning the connection weights of the ANN model[21,29,30].

1. Murray, A. F. *Application of Neural Networks* (Kluwer Academic, Boston, 1995).
2. Edwards, M. & Morse, D. R. The potential for computer-aided identification in biodiversity research. *TREE* **10,** 153–158 (1995).
3. Wright, D. H. Species energy theory: an extension of species-area theory. *Oikos* **41,** 495–506 (1983).
4. Kerr, J. T. & Packer, L. Habitat heterogeneity as a determinant of mammal species richness in high-energy regions. *Nature* **385,** 252–254 (1997).
5. Brown, J. H. *Macroecology* (Chicago University Press, Chicago, 1995).
6. Rosenzweig, M. L. *Species Diversity in Space and Time* (Cambridge University Press, Cambridge, 1995).
7. MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography* (Princeton University Press, Princeton, NJ, 1967).
8. Wright, D. H., Currie, D. J. & Maurer, B. A. *Species Diversity in Ecological Communities* (eds Ricklefs, R. E. & Schluter, D.) 66–74 (University of Chicago Press, Chicago, 1993).
9. Whittaker, R. H. Evolution of species diversity in land communities. *Evol. Biol.* **10,** 1–67 (1977).
10. Oberdorff, T., Guégan, J.-F. & Hugueny, B. Global scale patterns of fish species richness in rivers. *Ecography* **18,** 345–352 (1995).
11. Turner, J. R. G., Lennon, J. J. & Lawrenson, J. A. British bird species distributions and the energy theory. *Nature* **335,** 539–541 (1988).
12. Turner, J. R. G., Gatehouse, C. M. & Corey, C. A. Does solar energy control organic diversity? Butterflies, moths and British climate. *Oikos* **48,** 195–205 (1987).
13. Currie, D. J. Energy and large scale patterns of animal and plant species richness. *Am. Nat.* **137,** 27–49 (1991).
14. Currie, D. J. & Paquin, V. Large-scale biogeographical patterns of species richness of trees. *Nature* **329,** 326–327 (1987).
15. Oberdorff, T., Hugueny, B. & Guégan, J.-F. Is there an influence of historical events on contemporary fish species richness in rivers? Comparisons between Western Europe and North America. *J. Biogeogr.* **24,** 461–467 (1997).
16. Welcomme, R. L. Status of fisheries in South American rivers. *Interciencia* **15,** 337–345 (1990).
17. Hugueny, B. West African rivers as biogeographic islands: species richness of fish communities. *Oecologia* **79,** 235–243 (1989).
18. Hugueny, B. Geographical range of west African freshwater fishes: role of biological characteristics and stochastic processes. *Acta Oecologica* **11,** 351–375 (1990).
19. Scardi, M. Artificial neural networks as empirical models for estimating phytoplankton production. *Marine Ecol. Progr. Ser.* **139,** 289–299 (1996).
20. Ehrman, J. M., Clair, T. A. & Bouchard, A. Using neural networks to predict pH changes in acidified Eastern Canadian lakes. *Artif. Intell. Appl.* **10,** 1–8 (1996).
21. Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J. & Aulanier, S. Application of neural networks to modelling non-linear relationships in ecology. *Ecol. Model.* **90,** 39–52 (1996).
22. Barinaga, M. A recipe for river recovery? *Science* **273,** 1648–1650 (1996).
23. Naiman, R. J., Magnuson, J. J., McKnight, D. M., Stanford, J. A. & Karr, J. R. Freshwater ecosystems and their management: a national initiative. *Science* **270,** 584–585 (1995).
24. Ludwig, D., Hilborn, R. & Walters, C. Uncertainty, resource exploitation, and conservation: lessons from history. *Science* **260,** 17 and 36 (1993).
25. Johnson, B. L., Richardson, W. B. & Naimo, T. J. Past, present, and future concepts in large river ecology. *BioScience* **45,** 134–141 (1995).
26. Beattie, M. An ecosystem approach to fish and wildlife conservation. *Ecol. Appl.* **6,** 696–698 (1996).
27. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating error. *Nature* **323,** 533–536 (1986).
28. Kohavi, R. A study of cross-validation and bootstrap for estimation and model selection. *Proc. 14th Int. Joint Conf. on Artif. Intell.* 1137–1143 (Kaufman, Montreal, 1995).
29. Garson, G. D. Interpreting neural-network connection weights. *Artif. Intell. Exp.* **6,** 47–51 (1991).
30. Goh, A. T. C. Back-propagation neural networks for modelling complex systems. *Artif. Intell. Engng* **9,** 143–151 (1995).

Correspondence and requests for materials should be addressed to J.F.G. (e-mail: guegan@univ-montp2.fr).