Les réseaux de neurones pour prédire la biodiversité des poissons en eau courante

Philippe Boët, T. Fuhs

Les réseaux de neurones pour prédire la biodiversité des poissons en eau courante

Philippe Boët et Thierry Fuhs

Tous les pêcheurs le savent et cela ne nous surprendra pas : les belles rivières font les beaux poissons. Les communautés piscicoles s'organisent en effet autour de trois espaces fondamentaux, celui de la reproduction, de l'alimentation ou de l'abri, auxquels sont associées des fonctions vitales. Cette organisation est étroitement dépendante des multiples paramètres de l'environnement qui s'expriment à différentes échelles de temps et d'espace. Les communautés piscicoles peuvent ainsi être considérées comme une expression de « l'état de santé » des écosystèmes aquatiques (Fausch, 1990).

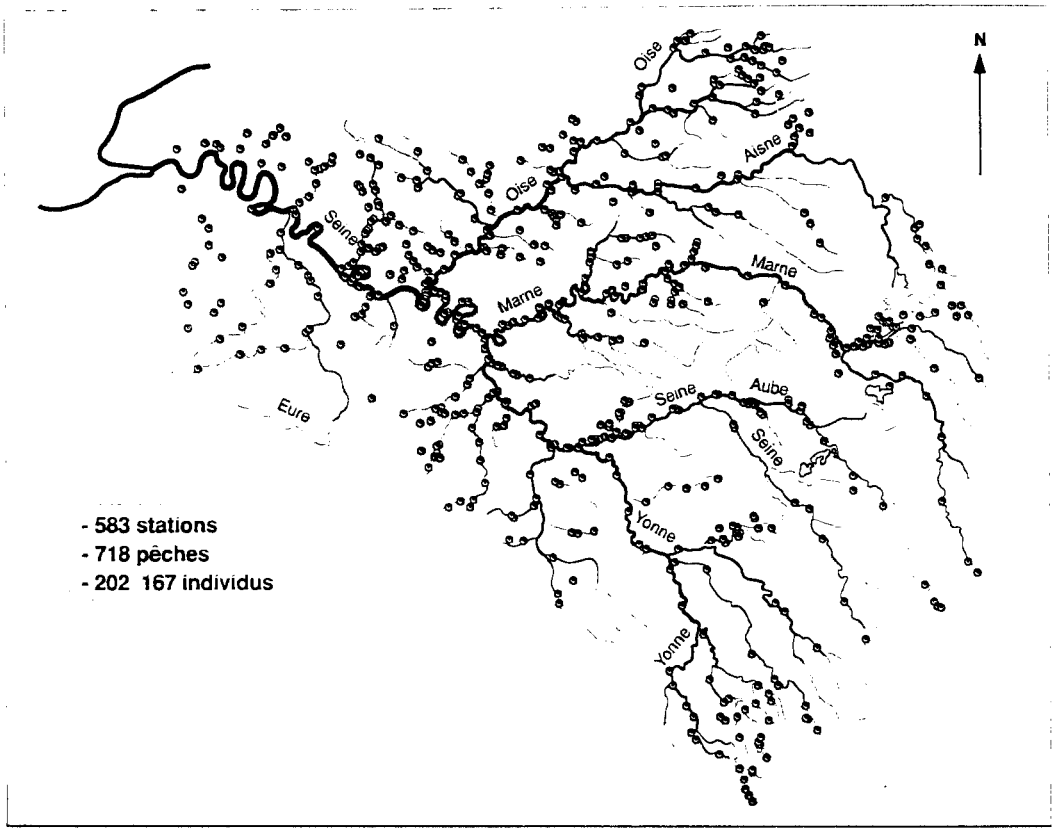
Comment prévoir la diversité piscicole à partir des caractéristiques des cours d'eau ? Cette question complexe préoccupe les chercheurs qui travaillent sur la Seine, fleuve de dimensions modestes, fortement soumis aux multiples pressions des activités humaines. Il s'agit de mieux connaître, comprendre et prévoir les conséquences des perturbations d'origine anthropique ou naturelle sur son fonctionnement. Identifier, hiérarchiser et évaluer les différents facteurs responsables de la composition des communautés aujourd'hui est indispensable pour conserver ou restaurer la faune et les milieux.

Les travaux consistent à exploiter une importante base de données qui couvre l'ensemble du bassin de la Seine et est issue d'échantillonnages réalisés au moyen de pêches électriques. Plus de 700 pêches, réalisés dans 583 stations, ont permis jusqu'à aujourd'hui de capturer plus de 200 000 poissons, représentant 39 espèces (figure 1 et 2). Ces données sont hétérogènes et bruitées, d'une part parce qu'elles résultent d'échantillonnages répondant à des objectifs différents et, d'autre part en raison de biais liés à la pêche électrique, dont l'efficacité est limitée dans les grands cours d'eau. Elles manquent de précision mais ont l'avantage d'être comparables et de couvrir un vaste espace. Il fallait donc trouver des méthodes d'analyse adaptées.

Une première synthèse de ces données a déjà été réalisée après utilisation d'analyses multidimensionnelles. Elle a permis de dégager les principaux facteurs qui conditionnent l'organisation actuelle du peuplement piscicole à l'échelle de l'ensemble du réseau hydrographique de la Seine (Belliard, 1994). Les caractéristiques du milieu, liées à l'organisation longitudinale et aux spécificités régionales du bassin, se sont avérées déterminantes. Les communautés s'enrichissent progressivement de l'amont vers l'aval, ce qui confirme les schémas théoriques de la zonation piscicole. Des espèces s'ajoutent, d'autres sont remplacées. En dépit de la relative homogénéité du bassin de la Seine, des facteurs locaux influencent cette évolution. Les successions sont plus ou moins rapides selon les écorégions drainées par les cours d'eau, et la richesse en espèces diffère pour des tronçons de rivière comparables.

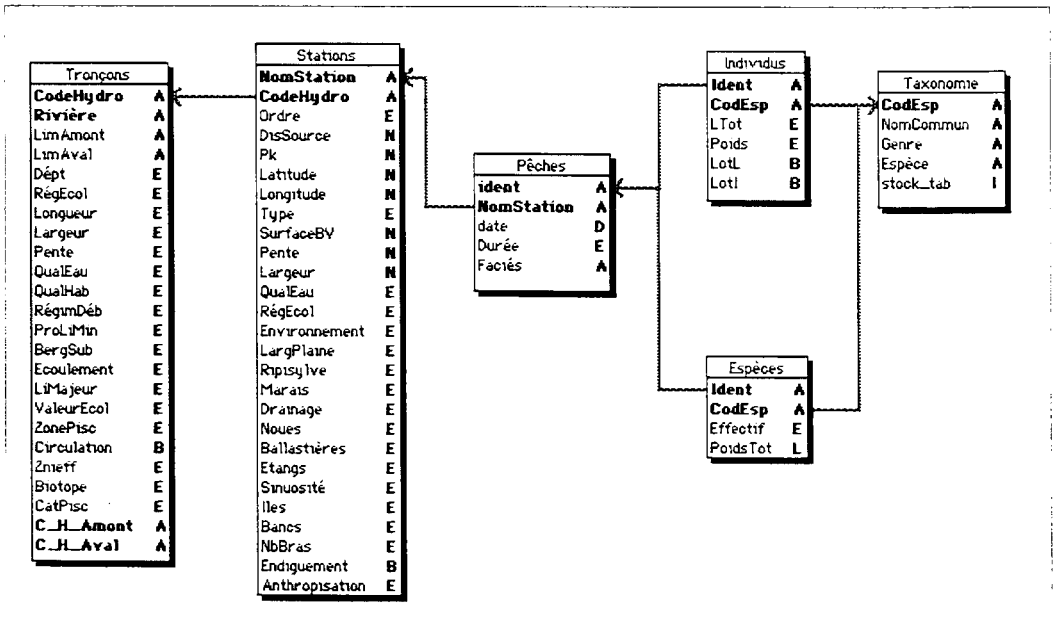
Il devenait nécessaire d'aller plus loin dans l'analyse. Pour approfondir notre connaissance et préciser davantage l'importance relative des variables de milieu dans les mécanismes de structuration des communautés, la modélisation est l'étape suivante. Elle permettrait notamment de simuler l'impact de différents aménagements. Une telle modélisation se heurte toutefois à la complexité des systèmes étudiés. Ces derniers sont constitués

Philippe Boët
Cemagref
14, av. de Saint-Mandé
75012 Paris
Thierry Fuhs
ENSAI
Timbre J220
3, av. Pierre Larousse
92245 Malakoff Cedex



▲ Figure 1. – Carte du bassin de la Seine et localisation des stations de pêche échantillonnées.

▼ Figure 2. – Structure de la base de données des pêches et des caractéristiques des sites échantillonnés.



de nombreuses composantes dont l'ensemble des interactions est encore mal connu. Les relations entre les poissons et les descripteurs physiques de l'habitat ne sont, par exemple, pas *a priori* linéaires. Parmi les différentes possibilités qui s'offraient, nous avons donc recherché des modèles non-linéaires et robustes, c'est-à-dire peu sensibles au bruit des données, mieux adaptés à nos données.

Les réseaux de neurones fournissent un exemple de tels modèles. Nous présentons ici les résultats d'une première utilisation de cet outil séduisant par sa simplicité d'application. Nous détaillons certaines difficultés qui ont dû être surmontées pour sa mise en œuvre effective. Nos premiers résultats sont satisfaisants. Cette approche s'avère pertinente pour la prédiction des poissons à l'échelle d'un bassin comme la Seine et des possibilités d'améliorations sont à confirmer. Néanmoins, d'intéressantes perspectives se dessinent déjà qui mériteraient d'être dévelopées.

Méthododologie et architecture du réseau

L'objectif de notre modèle est donc de prédire la présence ou l'absence de poissons à partir des caractéristiques du milieu. Le problème posé est analogue, au fond, à la discrimination, pour lequel l'utilisation de réseaux connexionnistes multicouches entraînés par l'algorithme de rétropropagation du gradient a montré son intérêt (encadré 1 et 2). Cette démarche a donc été privilégiée au cours de cette approche initiale.

La mise en œuvre effective de cette technique apparemment facile s'avère toutefois délicate.

▬ *Une espèce prédite par huit variables*

Ces modèles d'apprentissage offrent une très grande richesse de structure mais n'apportent à l'utilisateur aucune aide méthodologique, même empirique, pour dimensionner correctement un réseau en fonction du problème à résoudre.

Cela implique donc un examen très attentif des données d'exemples disponibles. Idéalement, en effet, un réseau multicouches devrait prendre en entrée les paramètres de milieu (une quinzaine de variables) et en sortie la présence-absence d'espèces (39 au total). Avec une seule couche cachée de N neurones, ceci représente 15xNx39 paramètres à calculer (les poids des connexions) ; soit plus de 2 000 paramètres à calculer pour quatre neurones en couche cachée (N=4). Avec seulement 700 exemples de pêches, c'est impossible.

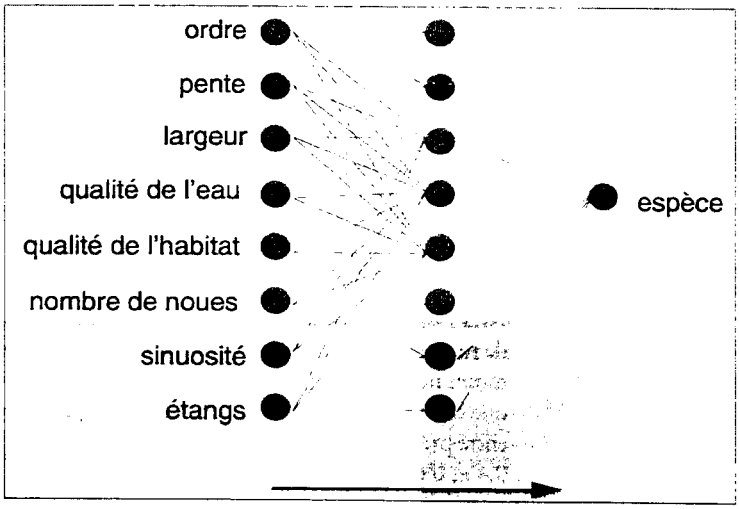
C'est pourquoi la taille des réseaux a dû être réduite de deux manières (figure 3).

En entrée, les variables les plus pertinentes sont sélectionnées (résultat des analyses multidimensionnelles ; Belliard, 1994). Huit ont finalement été retenues.

L'*ordre*, ou rang fluvial (*stream order* ; Strahler, 1957), est un paramètre de position de la station de pêche au sein du gradient amont-aval dans le réseau hydrographique ; par son caractère très synthétique, il rend compte de nombreuses variables physiques et fonctionnelles du cours d'eau. La *pente* et la *largeur* sont des descripteurs morphologiques classiques en hydrobiologie. La *qualité de l'eau* est la note des Agences de l'Eau. La *qualité de l'habitat* est un indice synthétique qui rend compte du degré d'altération de l'habitat physique : lit majeur et lit mineur, nature des berges et du substrat, degré d'artificialisation de l'écoulement. *Nombre de noues*, *sinuosité* et *étangs* sont une façon de caractériser l'écorégion où est située la station de pêche.

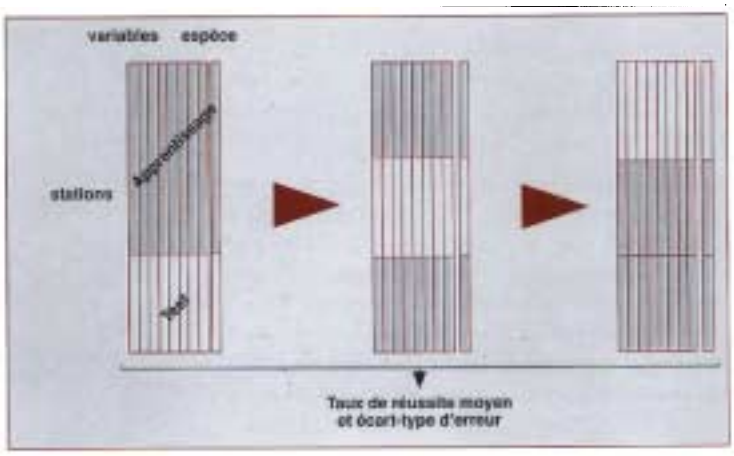
En sortie, l'étude est effectuée sur une seule espèce à la fois ce qui permet de n'avoir qu'une unité. Ceci, bien entendu, au détriment du nombre de réseaux entraînés, puisqu'il faut autant de réseaux que d'espèces, au lieu d'un seul pour toutes les espèces.

Figure 3. – Architecture des réseaux utilisés : une couche cachée de trois à huit neurones ▼



En outre, les interactions entre espèces sont négligées malgré leur éventuelle importance.

Pour la couche cachée, la meilleure valeur du nombre de cellules est recherchée par essais successifs.



▲ Figure 4. – Schéma de la validation croisée par locs.

▬ *Constitution des bases d'apprentissage et de test : une étape décisive*

Une attention particulière doit être portée à la constitution des ensembles d'apprentissage et de généralisation. Il s'agit, d'une part, d'assurer un nombre d'exemples suffisamment représentatifs pour la construction du modèle et, d'autre part, de maintenir un ensemble de généralisation de taille suffisante pour que l'estimation du taux d'erreur soit significative.

Pour y parvenir, deux types de problèmes liés à l'échantillonnage ont dû être surmontés.

• Les données étaient, en général, inégalement réparties entre présence et absence. La classe la plus nombreuse aurait alors eu une influence excessive dans le calage des poids du réseau.

Dans le cas extrême de la Carpe commune, par exemple, présente dans moins de 10 % des pêches, le réseau prédisait systématiquement son absence... et, bien sûr, se trompait rarement !

Pour éviter ce biais, nous avons choisi de dupliquer de manière aléatoire certains exemples de la classe la moins nombreuse. Ceci permet d'obtenir, pour chaque espèce, une répartition des exemples équilibrée entre présence et absence – en fait, un rapport entre 0,4 et 0,6 était considéré comme satisfaisant.

• Afin de constituer les deux bases d'apprentissage et de test, la méthode la plus classique consiste à diviser de manière arbitraire l'ensemble de tous les exemples disponibles. Selon la taille de l'ensemble de départ, la proportion d'exemples placés dans la base de test varie de 10 à 50 % ; dans la pratique, on utilise souvent une répartition de 2/3, 1/3.

Une telle répartition présente cependant deux inconvénients. D'une part, le découpage de la base des pêches en base de test et base d'apprentissage ne peut être *a priori* homogène à cause du nombre limité de pêches. Or, le choix de la base d'apprentissage influe sur le résultat et ceci peut entraîner des taux de généralisation inutilisables car sans rapport avec la distribution réelle étudiée. D'autre part, une partie importante de l'information disponible, contenue dans la base de test est ignorée au stade de l'élaboration du modèle, ce qui est d'autant plus pénalisant que la base de données totale est réduite.

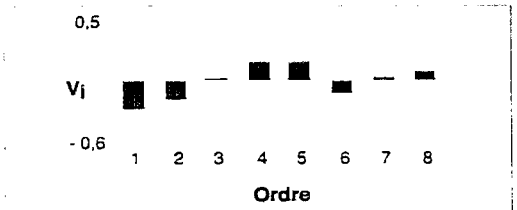
Une validation croisée permet de remédier à ces deux inconvénients. Elle a été effectuée « par blocs ». L'ensemble de la base était d'abord découpé en trois parties d'effectif égal. Puis, chaque réseau était entraîné sur deux tiers de la base et ensuite testé sur le troisième tiers pour estimer sa performance en généralisation. Ceci était répété trois fois, en permutant les trois parties (figure 4). Trois modèles sont donc obtenus dont le taux moyen de généralisation ainsi que l'écart-type associé sont enfin calculés.

Des résultats interprétables par l'écologie

Éprouvée à l'échelle du bassin de la Seine et en fonction de descripteurs très globaux de la qualité du milieu aquatique (huit variables synthétiques d'entrée), la prédiction en termes de présence ou d'absence d'une espèce par des réseaux connexionnistes multicouches s'avère pertinente.

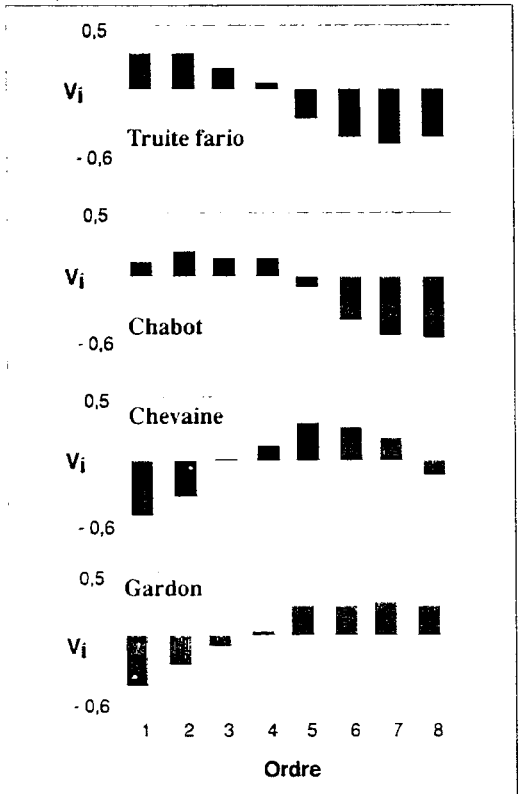
Dix-huit espèces ont été testées, choisies parmi les plus représentatives présentes dans le bassin.

Le nombre de neurones requis dans la couche cachée varie selon les espèces de trois à huit. Pour quatorze espèces, un réseau avec cinq neurones en couche cachée donne de bons résultats (tableau 1). Pour les autres, davantage de neurones sont nécessaires : six pour la brème bordelière, le vairon et la vandoise, huit dans le cas de l'ablette.



Pour une classe i, la valeur du profil est égale à Vi = Fi - Frot où, Fi est la fréquence relative de l'espèce dans les relevés de la classe i et Frot la fréquence relative de l'espèce sur l'ensemble des relevés. Cette valeur est donc d'autant plus fortement positive (ou négative) que l'espèce est sur-représentée (ou sous-représentée). Le profil permet donc de déterminer les exigences biologiques d'une espèce pour un facteur donné, ici l'ordre. Il est ici très significatif au risque P 0,01 ; test du χ_2.

▲ Figure 5. – Profil écologique du Goujon en fonction de l'ordre, ou rang fluvial, des cours d'eau (d'après Belliard, 1994)



La truite et le chabot sont des espèces caractéristiques des têtes de bassin, le chevaine occupe une position intermédiaire, tandis que le gardon est typique des parties aval. Voir à la figure précédente les explications du calcul des profils. Tous ces profils sont très significatifs au risque

▲ Figure 6. – Profils écologiques de quatre espèces en fonction de l'ordre, ou rang fluvial, des cours d'eau (d'après Belliard, 1994)

Espèce	Apprentissage		Test	
	Taux de réussite (%)	Ecart-type d'erreur	Taux de réussite (%)	Ecart-type d'erreur
Ablette	-	-	-	-
Barbeau fluviatile	81,0	0,5	75,9	3,1
Brochet	83,7	0,5	83,4	15,4
Brème bordelière	-	-	-	-
Carpe	82,1	0,9	79,0	2,6
Chabot	87,4	1,9	85,4	4,8
Chevaine	84,3	0,4	80,1	2,5
Gardon	83,7	2,8	82,3	5,1
Goujon	77,8	0,4	72,3	2,5
Hotu	85,8	5,3	77,6	9,2
Loche franche	84,9	3,4	79,5	10,0
Lote de rivière	80,3	0,8	72,3	2,3
Perche	83,2	1,3	79,2	3,4
Rotengle	82,9	1,9	81,8	1,9
Truite arc-en-ciel	78,4	1,1	72,6	2,6
Truite fario	87,0	1,9	86,3	2,5
Vairon	-	-	-	-
Vandoise	-	-	-	-

▲ Tableau 1. – Taux de réussite moyens et écarts types d'erreur avec cinq neurones en couche cachée.

Alors que les données d'entrée sont assez fortement bruitées, les taux de réussite en généralisation varient de 70 à plus de 85 % selon les espèces. Ceci représente des performances très appréciables, car une erreur de mesure de l'ordre de 10 à 20 % sur ce type de données est en effet tout à fait possible. Les plus faibles résultats sont observés pour la truite arc-en-ciel et le goujon, en raison de leurs particularités écologiques.

La truite arc-en-ciel est en effet une espèce qui fait l'objet de nombreux réempoissonnements ; sa présence est donc très indépendante des caractéristiques du milieu aquatique.

Dans le cas du goujon, un phénomène plus subtil peut expliquer les mauvaises performances du réseau. Dans le bassin de la Seine, cette espèce se distribue en effet en deux groupes distincts (figure 5). Dans les parties amont, des populations composées de petits individus sont classiquement inféodées à des eaux vives de bonne qualité s'écoulant sur des substrats de sable ou de gravier. A l'aval, dans des zones plus riches en matière organique, se trouvent au contraire de gros goujons pouvant parfois mesurer plus de 25 cm de longueur. Il conviendra certainement de bien séparer ces deux sous-ensembles pour mieux entraîner les réseaux et améliorer leur qualité de prédiction.

En revanche, les meilleurs taux de réussite sont obtenus pour les espèces dont les profils écologiques sont nets comme le montre par exemple la figure 6.

L'examen approfondi des résultats est également intéressant. Ainsi, dans le cas du barbeau fluviatile (tableau 2), l'erreur commise pourrait sembler importante mais elle s'avère essentiellement due à la prédiction de la présence de ce poisson par le réseau alors qu'il ne figure en fait pas dans le relevé des pêches. Or, même présent dans un milieu, le barbeau est une espèce difficile à capturer par la pêche à l'électricité. Ceci illustre l'importance de la représentativité de l'échantillonnage qui pose le problème, déjà largement débattu, de la signification écologique à donner à l'absence d'une espèce dans des relevés faunistiques.

Tableau 2 - Erreur de prédiction du Barbeau fluviatile. ▼

	Apprentissage	Total	Test	
			Présence prédite / absence	Absence prédite / présence
Taux moyen (%)	15,3	21,4	16,3	5,1
Ecart-type	1,6	11,6	13,4	2,7

Un outil de prédiction à développer ?

Compte tenu de la nature des données traitées et du caractère très synthétique des variables d'entrée, ces résultats exploratoires s'avèrent satisfaisants. Certes, il faut maintenant les comparer avec des méthodes multivariées classiques. Toutefois, ils sont déjà très proches de ceux obtenus à l'aide d'analyses discriminantes et de régressions multiples par Pouilly (1994) ou Capra (1995), lesquels, travaillant à l'échelle du micro-habitat, disposent de données très fiables de description de l'habitat et d'échantillonnage de la faune en place.

Néanmoins, ces résultats semblent encore susceptibles d'améliorations. En particulier, malgré l'influence connue des facteurs régionaux sur la composition du peuplement, le paramètre « région écologique » n'a pas été pris en compte en raison des problèmes de codage liés à son caractère purement qualitatif. Il conviendra cependant de vérifier les performances des réseaux en présence de ce paramètre. La considération des classes d'abondance relative des différentes espèces devrait aussi conduire à de meilleures prédictions.

Ces résultats se concentrent sur la prédiction de présence ou d'absence d'une espèce donnée, alors que l'ambition initiale était de considérer directement tout le peuplement. Ceci ne pouvant être fait avec l'ensemble des espèces en même temps, il faudrait retraiter les données des pêches afin d'identifier les différents types de peuplement en place et entraîner ensuite des réseaux multicouches où la sortie ne serait plus une espèce particulière mais un type de peuplement donné. Les problèmes sont donc avant tout de définir écologiquement ces peuplements.

Ces premiers essais sont donc encourageants si l'on considère qu'à l'heure actuelle il n'existe guère de modèles prédictifs « poissons » à l'échelle de l'ensemble d'un bassin fluvial. La zonation piscicole de Thienneman (1925), reformulée par Huet (1949, 1959) et connue comme « *règle des pentes* », permet seulement de distinguer quatre zones (à truite, ombre, barbeau, ou brème) le long du gradient amont-aval d'un cours d'eau. Plus récemment, Verneaux (1977, 1981) a proposé un calcul permettant d'approcher le groupement d'espèces caractéristiques, ou biocœnotype théorique, d'un secteur de cours d'eau, en fonction de différents paramètres, malheureusement pas toujours faciles à renseigner, comme par exemple « *la température maximale moyenne du mois le plus chaud* ».

Ainsi, il serait d'ores et déjà intéressant d'étudier, avec un tel réseau connexionniste, les conséquences des changements de milieu d'origine naturelle ou anthropique sur la composition des peuplements de poissons à l'échelle du bassin hydrographique. Parmi les variables d'entrée, certaines décrivent en effet la morphologie du milieu ou sa position dans le gradient amont-aval et ont un caractère figé. D'autres, au contraire, peuvent traduire une perturbation (physique ou chimique) et sont susceptibles de constituer un premier élément du diagnostic d'un éventuel facteur de déséquilibre du peuplement piscicole en place ; encore très synthétiques actuellement, comme par exemple la note de qualité de l'eau, ces variables pourraient être décomposées afin d'affiner un tel diagnostic.

A terme, s'entrevoient des applications concrètes comme, par exemple, l'établissement du peuplement théorique de référence permettant de quantifier plus aisément l'impact éventuel d'un aménagement en un lieu donné. □

Encadré 1

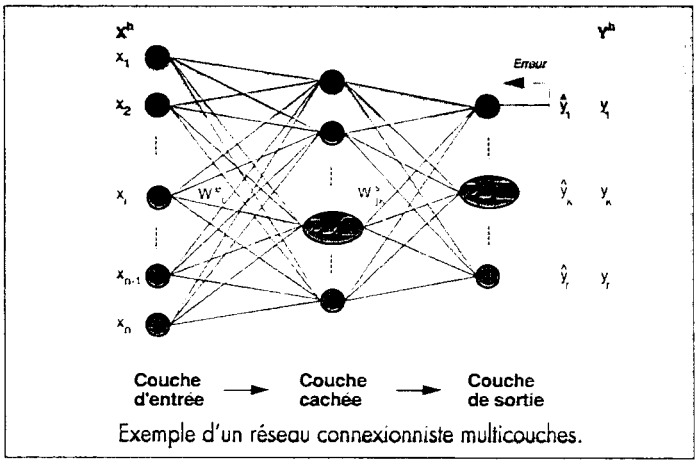
Présentation des réseaux de neurones

Comme leur nom l'indique, les réseaux de neurones ont une origine marquée par la biologie du cerveau. L'objectif de McCulloch et Pitts (1943) était de simuler le fonctionnement du cerveau humain à l'aide de composants simples, les neurones formels, interconnectés en grand nombre (figure). L'idée sous-jacente était que *le tout est plus puissant que la somme des parties.*

Une liste des modèles formels les plus importants issus de cette idée est dressée par M. Steinmetz (1995 ; p. 25).

McCulloch et Pitts ont en quelque sorte réussi puisqu'ils ont montré que ces réseaux, à l'instar de la machine de Turing, permettaient de calculer toute fonction calculable.

Cette caractéristique est toujours d'actualité, mais ce qui fait le succès grandissant de ces modèles réside dans leur capacité à modéliser des phénomènes non-linéaires. Or, les régressions statistiques habituellement utilisées sont linéaires et ne peuvent donc qu'approcher d'assez loin des phénomènes fortement non-linéaires. Bien entendu, les statisticiens classiques ne sont pas complétement dépourvus devant ces non-linéarités (modèles linéaires généralisés, estimations non-paramétriques, etc.) mais, par leur simplicité d'utilisation, les réseaux de neurones ont constitué depuis leur avènement des compétiteurs crédibles. D'ailleurs, les ponts entre les deux communautés sont de plus en plus actifs.



Exemple d'un réseau connexionniste multicouches.

Voyons d'où vient ce comportement non linéaire.

La fonction de transfert d'un neurone isolé est déjà non-linéaire. Elle s'écrit en effet

$$\hat{y} = f(\sum_i \alpha_i x_i)$$

où f est une fonction sigmoïde. On peut néanmoins montrer que ce type de fonction est équivalent à un séparateur linéaire dans le cas d'un problème de discrimination, comme en relève la détection de la présence du poisson.

Par contre, lorsque nous combinons de tels neurones formels en plusieurs couches, la fonction de transfert devient beaucoup plus puissante. Les x_i représentent les données du problème, dans notre cas les caractéristiques de l'habitat. $\hat{y}_k$ représente la prédiction du système, ici une valeur binaire de présence ou d'absence de poissons. En supposant une seule couche cachée, la valeur prédite de chaque $\hat{y}_k$ s'écrit :

$$\hat{y}_k = f\left(\sum_j w^s_{jk} t_j\right) = f\left(\sum_j w^s_{jk}\left(\sum_i w^e_{ij} x_i\right)\right)$$

où t_j sont des valeurs d'activations des neurones de la couche cachée, w^e_{ij} le poids de la connexion entre le neurone d'entrée i et le neurone caché j et w^s_{jk} le poids de la connexion entre le neurone caché j et le neurone de sortie k.

Des résultats mathématiques ont montré que si l'on considère n'importe quelle fonction continue des entrées vers la sortie, et une précision ε donnée, si petite soit elle, on peut trouver, parmi tous les réseaux à une couche cachée ayant le même nombre d'entrées et de sorties, un réseau situé à une distance inférieure à ε de cette fonction. Décrites ainsi, les choses semblent idylliques. Malheureusement, ce résultat n'est qu'un résultat d'existence : il ne donne d'aucune manière le nombre de neurones nécessaires sur la couche cachée pour approcher à moins de ε la fonction recherchée ! Et bien entendu, plus la précision demandée est faible, plus grand est le nombre de neurones cachés. Par contre, à architecture fixée, plusieurs algorithmes permettent de calculer les poids des connexions à partir de l'échantillon des (x, y) considéré. Le plus célèbre est la rétropropagation du gradient qui converge effectivement vers un minimum local de l'erreur (voir aussi l'encadré de l'article de Steinmetz, 1995). Ces algorithmes minimisent l'erreur entre la valeur prédite $\hat{y}_k$ et la valeur observée y. L'erreur quadratique $\left(\sum_k y_k - \hat{y}_k\right)^2$ est la plus utilisée.

Encadré 2

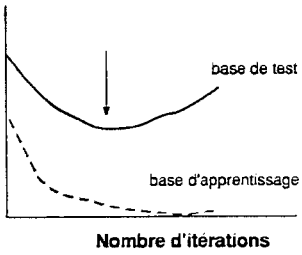
Le dilemme de l'apprentissage

Récapitulons : nous avons une fonction inconnue des entrées vers la sortie ; nous ne connaissons pas le nombre de neurones du réseau le plus proche de cette fonction ; nous avons un échantillon de taille finie (nos pêches) ; nous disposons d'un algorithme de calcul des poids d'un réseau donné.

Nous sommes ainsi typiquement dans le cadre de la statistique inférentielle classique : une fonction inconnue, un ensemble non paramétrique (l'ensemble de tous les réseaux à une couche cachée) dense dans l'ensemble des fonctions considérées et une estimation paramétrique possible dans des sous-ensembles de ce gros ensemble grâce à un échantillon de la fonction inconnue.

Comme toute inférence statistique, nous sommes soumis au dilemme biais/variance ou dilemme de l'apprentissage (Geman *et al.*, 1992) (figure). On pourrait en effet penser que l'idéal est de considérer l'espace de recherche le plus grand possible, c'est-à-dire les réseaux ayant le plus grand nombre possible de neurones cachés. Malheureusement, cela n'est pas une stratégie gagnante.

En effet, si nous recherchons un modèle complexe, donc un réseau à beaucoup de paramètres, nous aurons plus de chance de trouver un réseau proche de la fonction inconnue et ainsi le biais pourra être faible. Par contre, si cet espace de recherche est très grand, on pourra trouver de nombreux réseaux qui collent à l'échantillon d'apprentissage, réseaux pouvant être contradictoires sur de nouveaux échantillons : la variance sera alors forte.



Comportement de l'erreur sur la base d'apprentissage et la base de test : c'est le minimum d'erreur en base de test qu'il faut trouver.

A l'inverse, si le modèle recherché est simple, le biais obtenu sera fort (par exemple, un séparateur linéaire pour un problème fortement non-linéaire) mais la variance restera bonne, les résultats n'étant que peu modifiés sur un nouvel échantillon.

La capacité de généralisation d'un réseau, c'est-à-dire son pouvoir prédictif sur de nouveaux exemples (de nouvelles pêches) est alors le résultat du compromis obtenu sur la richesse de l'espace de recherches. Un espace trop riche introduit une variance forte entre les échantillons alors qu'un espace trop pauvre nous laisse nous éloigner d'une solution satisfaisante.

Résumé

Les communautés de poissons dépendent étroitement des caractéristiques de l'environnement aquatiques où elles évoluent. Pour modéliser ces relations complexes et *a priori* non linéaires, nous testons les possibilités offertes par les réseaux de neurones. Nous disposons pour cela d'une importante base de données de pêches électriques réalisées sur l'ensemble du bassin de la Seine. Les principaux problèmes rencontrés concernent l'architecture du réseau et la constitution des bases d'apprentissage et de test. Les premiers résultats obtenus sur 18 espèces montrent que la prédiction, en termes, de présence ou d'absence, à partir de huit descripteurs synthétiques est pertinente. Des améliorations à confirmer laissent entrevoir d'intéressantes perspectives en matière de gestion.

Abstract

Fish communities are closely dependent on the characteristics of the aquatic environment in which they evolve. We have investigated the use of neural networks for modelling those complex non-linear relationships. This study is based on a large database of electric fishing carried out in the whole Seine basin. The basic problems raised by the use of neural network are twofold : on the one hand the network architecture is to be chosen by a trial-and-error procedure, and on the other hand the training and test sets are to be wisely chosen. The first experiments have been driven using eight empirical input parameters on 18 different species. The results show a reasonnably good accuracy in the prediction of presence or absence of fish. Further experiments are still requiered to confirm this results, but the latter already show interesting possibilities for environmental management.

Bibliographie

BELLIARD, J , 1994. Le peuplement ichtyologique du bassin de la Seine - rôle et signification des échelles temporelles et spatiales. *Thèse Doct Paris VI*, 197 p.

CAPRA, H., 1995. Amélioration des modèles prédictifs d'habitat de la truite fario : échelles d'échantillonnage ; intégration des chroniques hydrologiques. *Thèse Doc. Univ. Claude Bernard - Lyon I*, soutenance fin novembre.

FAUSCH, K.D., LYONS, J., KARR, J.R., ANGERMEIER, P.L., 1990. *Fish communities as indicators of environmental degradation.* p. 123-144, *In* : S. M. Adams (Ed.), Biological indicators of stress in fish, American Fishery Society Symposium n° 8.

GEMAN, S., BIENENSTOCK, E., DOURSAT, R., 1992. Neural network and the bias/variance dilemma. Neural Computation 4, p. 1-58.

HUET, M., 1949. Aperçu des relations entre la pente et les populations piscicoles des eaux courantes. *Scheiw. Z. Hydrol.*, 11 (3-4), p. 332-351.

HUET, M., 1959. Profiles and biology of western european streams as related to fish management. *Trans. Am. Fish. Soc.*, 88 (3), p. 155-163.

MCCULLOCH, W.S., PITTS, W., 1943. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, p. 115-133.

POUILLY, M., 1994. Relations entre l'habitat physique et les poissons des zones à cyprinidés rhéophiles dans trois cours d'eau du bassin rhodanien : vers une simulation de la capacité d'accueil pour les peuplements. *Thèse Doc. Univ. Claude Bernard - Lyon I*, 256 p.

RUMELHART, D.E., MCCLELLAND, J.L., the PDP Research Group, 1986. *Parallel distributed processing.* MA : MIT Press/Bradford Books, Cambridge.

STEINMETZ, V., 1995. Les neurones dans le champagne. *Ingénieries-EAT*, 3, p. 23-28.

THIENEMANN, A., 1925. Die Binnengewässer Mitteleuropas. Eine limnologische Einfuhrung. *Die Binengewasser, Stuttgart*, 1, p. 1-255.

VERNEAUX, J., 1977. Biotypologie de l'écosystème « eau courante ». Détermination approchée de l'appartenance typologique d'un peuplement ichtyologique. *C R. Acad. Sci. Paris*, t. 284 (sér. D), p. 675-678.

VERNEAUX, J., 1981. Les poissons et la qualité des cours d'eau. *Ann. Sci. Univ. Franche-Comté, Besançon, Biol. Anim.*, 4ème sér. (fasc. 2), p. 33.