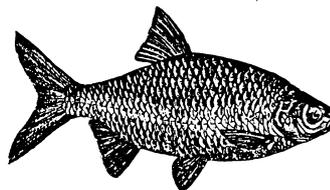
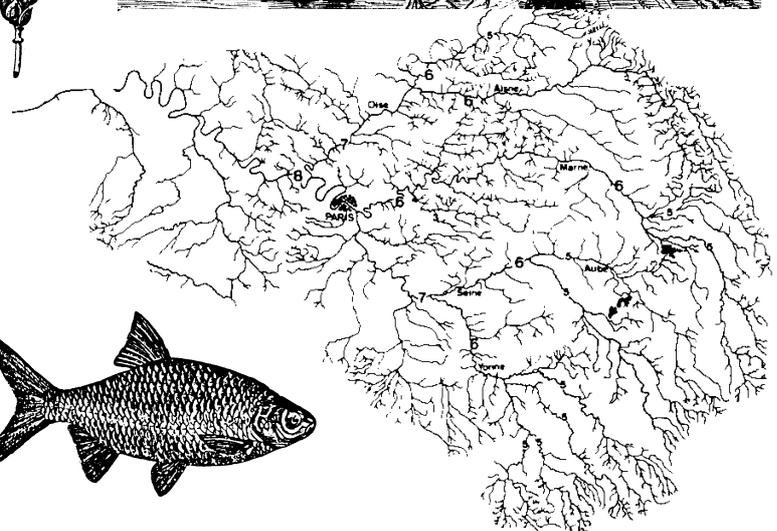
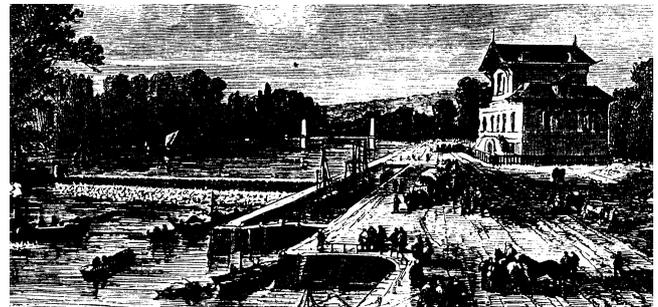
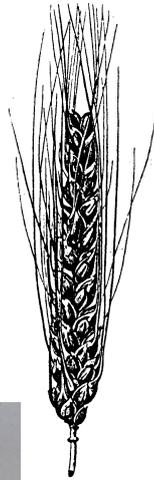


**Programme PIREN-Seine 1998-2001**

# **Modélisation du compartiment piscicole**

**Direction et secrétariat:**  
UMR CNRS 7619 Sisyphe  
Université P&M Curie  
Tour 26, 5e étage, boîte 123  
4, place Jussieu  
75005 Paris



**Rapport de Synthèse, Février 2002**

# Sommaire

<b>1</b>	<b>Test d'un modèle bioénergétique pour mieux cerner la biomasse piscicole d'un bief de la Seine.....</b>	<b>2</b>
1.1	ECOPATH II (Christensen et Pauly, 1992).....	2
1.1.1	Fondements du modèle .....	2
1.1.2	Les paramètres de base requis .....	3
1.1.3	Les paramètres estimés par le modèle .....	3
	<i>Flux vers les détritux</i> .....	3
	<i>Consommation de nourriture</i> .....	4
	<i>Quantités exportées ou ingérées</i> .....	4
	<i>Rendement net</i> .....	4
	<i>Niveaux trophiques (TL)</i> .....	4
	<i>Indice d'omnivorie (OI)</i> .....	4
	<i>Indice de sélection (Electivity)</i> .....	4
1.2	Application au bief Andrésy-Méricourt.....	5
1.2.1	Le site choisi.....	5
1.2.2	Les compartiments de l'écosystème.....	5
1.2.3	Caractéristiques bioénergétiques des compartiments .....	6
1.2.4	Résultats .....	7
1.2.5	Discussion .....	7
1.3	Conclusion .....	9
<b>2</b>	<b>Modélisation des relations milieux - peuplements.....</b>	<b>10</b>
2.1	Les réseaux de neurones.....	10
2.2	De la boîte noire à la boîte de verre. . . . .	11
2.3	Les arbres de décision .....	13
2.3.1	Principe.....	13
2.3.2	Validation .....	13
2.3.3	Construction .....	13
2.3.4	Avantages .....	15
2.4	Vers des modèles prédictifs géo-référencés .....	15
2.4.1	Données géoréférencées.....	16
2.4.2	Constitution d'une base de données Oracle™ .....	16
2.4.3	Mise en œuvre .....	16
2.4.4	Perspectives .....	19
2.5	Conclusions.....	19
	<b>Références .....</b>	<b>20</b>

# Modélisation du compartiment piscicole

La modélisation du compartiment piscicole a constitué un axe privilégié des activités récentes relatives aux poissons, dans le cadre du PIREN-Seine. Ceci, d'une part, afin de tenter d'intégrer ce maillon de l'édifice biologique dans les flux de matières et d'énergie des modèles de fonctionnement de l'écosystème Seine ; et d'autre part, pour mieux formaliser les relations entre qualité écologique des milieux aquatiques et nature du peuplement piscicole en place.

Deux approches ont ainsi été développées, l'une quantitative, visant à évaluer la possibilité d'adapter un modèle bioénergétique au cas de la Seine ; l'autre statistique, conduisant à l'élaboration de modèles prédictifs géospatialisés des peuplements piscicoles, en fonction des caractéristiques des cours d'eau, à l'échelle globale du réseau hydrographique.

## 1. Test d'un modèle bioénergétique pour mieux cerner la biomasse piscicole d'un bief de la Seine

Évaluer la biomasse totale en poisson est un point fondamental dans la connaissance d'un écosystème comme la Seine. Mais les méthodes classiques d'estimation ne sont pas adaptées à des évaluations quantitatives ou représentent un investissement matériel important. La technique de pêche électrique présente en effet des résultats d'ordre essentiellement qualitatif, comme la richesse spécifique, tandis que la mise en œuvre de méthode statistique, telle que la capture-recapture, nécessite un effort de pêche intense et le marquage d'une grande quantité de poissons, et que l'échosondage n'est pas encore au point pour des milieux aquatiques de faible profondeur.

L'approche bioénergétique peut apparaître ainsi une alternative intéressante. Depuis quelques années, l'ICLARM développe en effet de nombreux outils d'aide à la gestion des milieux aquatiques, notamment lacustres. Parmi eux, le modèle bioénergétique ECOPATH II (Christensen & Pauly 1992) qui permet de modéliser les flux de matière au sein d'un réseau trophique est appliqué au cas du bief d'Andrésey-Méricourt situé sur la Seine à l'aval de Paris.

### 1.1. ECOPATH II (Christensen et Pauly, 1992)

Conçu à l'origine pour les milieux marins et lacustres, le modèle ECOPATH est utilisé pour des rivières comme la Tamise (Mathews 1993) et la Garonne (Palomares *et al.* 1993). Il permet de modéliser les flux de matière au sein d'un réseau trophique et d'évaluer quantitativement la biomasse, la production et la consommation de chaque compartiment de l'édifice biologique. Le programme fournit des valeurs d'indices qui caractérisent l'écosystème dans son ensemble ainsi que les compartiments.

#### 1.1.1 Fondements du modèle

ECOPATH intègre les travaux de Polovina (1985) concernant les estimations de biomasses et de consommations des différents groupes définis, ainsi que l'approche de Ulanovicz (1986) pour l'analyse des flux de matière entre ces mêmes groupes (Christensen & Pauly 1992). Le cœur du programme est conçu par Polovina et Ow (1983) et Polovina (1985). Il s'agit d'un ensemble d'équations linéaires qui décrivent chacune un groupe du système écologique et sont de la forme :

$$P_i - M_{2i} - M_{0i} - C_i = 0 \quad (1)$$

Avec,

- $P_i$  la production de l'espèce  $i$ ,
- $M_{2i}$  sa mortalité par la prédation,
- $M_{0i}$  la mortalité non due à la prédation,
- $C_i$  l'exploitation par la pêche.

Cette équation peut être successivement réécrite de deux manières différentes :

$$P_i - B_i.M2_i - P_i.(1 - EE_i) - EX_i = 0$$

Avec,

$B_i$  la biomasse de  $i$ ,  
 $EE_i$  l'EfficiencE Ecotrophique de  $i$ . ( $1 - EE_i$ ) étant la mortalité non due à la prédation,  
 $EX_i$  l'exportation de  $i$  (hors du système).

$$B_i * PB_i * EE_i - \sum_j (B_j * QB_j * DC_{ji}) - EX_i = 0 \quad (2)$$

Avec,

$PB_i$  le rapport production/biomasse,  
 $QB_j$  le rapport consommation/biomasse,  
 $DC_{ji}$  la fraction de la proie  $i$  dans le régime alimentaire du prédateur  $j$ .

Les équations de la forme (2) peuvent s'écrire sous la forme d'une matrice à partir de laquelle sont estimés les différents paramètres.

### 1.1.2 Les paramètres de base requis

L'écosystème se représente sous forme de compartiments : espèce, taxon, etc., où se regroupent des organismes possédant des caractéristiques énergétiques homogènes. En effet, chacune de ces unités fonctionnelles se caractérise par une biomasse ( $B$ ), une production<sup>1</sup> ( $P/B$ ), une consommation<sup>2</sup> ( $Q/B$ ) et une efficacité écotrophique<sup>3</sup> ( $EE$ ). Il s'agit donc souvent d'espèces liées taxonomiquement ou écologiquement.

Pour chaque compartiment choisi pour décrire l'écosystème, le modèle requiert la connaissance des paramètres suivants :

- Trois valeurs parmi  $B$ ,  $P/B$ ,  $Q/B$  et  $EE$  ; la quatrième étant estimée.
- Les régimes alimentaires.
- La proportion de la consommation non assimilée.
- La biomasse accumulée ou perdue entre deux années consécutives.
- Les gains et pertes de biomasse dus aux migrations.
- La proportion du surplus de détritus (pelotes fécales, morts, etc.) qui est stockée dans le compartiment « détritus » ; le reste étant exporté hors du système.

### 1.1.3 Les paramètres estimés par le modèle

#### *Flux vers les détritus*

Pour chaque organisme, ce flux représente la quantité de nourriture non assimilée et la mortalité naturelle non due à la prédation (vieillesse, maladies, famine). Ce flux est positif.

---

<sup>1</sup> La production est l'élaboration de tissus pendant une période donnée, elle correspond à la quantité de matière consommée, diminuée des pertes liées à l'excrétion et la respiration. Si l'espèce considérée a une mortalité totale  $Z$  constante, et que la croissance obéit au modèle de von Bertalanffy, alors le rapport  $P/B$  est égal à la mortalité totale  $Z$ .

<sup>2</sup> Le modèle ECOPATH II quantifie des flux de matière, d'énergie et de nutriments. Cela implique la nécessité de définir la « consommation » des producteurs primaires.

<sup>3</sup> Fraction de la production d'un compartiment qui est consommée par les niveaux trophiques supérieurs et/ou exportée hors du système (comprise entre 0 et 1).

### Consommation de nourriture

Elle représente l'ingestion totale de nourriture pendant un pas de temps défini. Elle est égale au produit de la biomasse par le taux d'ingestion (Q/B).

### Quantités exportées ou ingérées

La quantité totale exportée hors du système et/ou ingérée par des prédateurs est le produit du rapport P/B par la biomasse et par l'efficacité écotrophique (EE).

### Rendement net

Le rendement net correspond au rapport Production / Assimilation. L'assimilation est la quantité de nourriture consommée qui n'est pas excrétée. Elle est donc égale à la somme de la production et de la respiration. Le rendement net est positif et compris entre 0 et 1. Les organismes hébergeant des algues symbiotiques peuvent toutefois posséder un rendement net supérieur à 1.

### Niveaux trophiques (TL)

Le modèle ECOPATH II utilise la notion de niveau trophique proposée par Odum et Heald (1975). Les organismes considérés peuvent se nourrir sur plusieurs niveaux trophiques et sont donc placés à des niveaux trophiques fractionnés. Dans le cas des détritiques et des producteurs primaires, cette valeur est, par définition, égale à 1.

Par exemple, un organisme consommant 40% de plantes (TL = 1) et 60% d'herbivores (TL = 2) aura un niveau trophique de :  $TL = 1 + [(1 \times 0,4) + (2 \times 0,6)] = 2,6$

### Indice d'omnivorie (OI)

L'indice d'omnivorie a été introduit en 1987 (Pauly *et al.* 1992). Il correspond à la variance entre les niveaux trophiques des différentes proies d'un prédateur donné, et est sans unité. Lorsqu'un prédateur est très spécialisé, se nourrissant sur un unique niveau trophique, l'indice d'omnivorie est nul.

$$OI_i = \sum_j [TL_j - (TL_i - 1)]^2 * DC_{ij}$$

Avec,

TL<sub>j</sub> le niveau trophique de la proie j

TL<sub>i</sub> le niveau trophique du prédateur i

DC<sub>ij</sub> la proportion de la proie j dans la composition alimentaire du prédateur i

### Indice de sélection (Electivity)

Il s'agit de l'indice de Chesson (1983) défini par la relation suivante :

$$S_i = \frac{\frac{r_i}{P_i}}{\sum \frac{r_i}{P_i}}$$

Avec,

r<sub>i</sub> l'abondance relative d'une proie dans le régime alimentaire d'un prédateur

P<sub>i</sub> l'abondance relative de la proie dans l'écosystème

S<sub>i</sub> varie de -1 à 1 et est interprété selon le schéma suivant :

E<sub>i</sub> = -1 → Evitement de la proie

E<sub>i</sub> = 0 → Pas de sélection

E<sub>i</sub> = 1 → Choix exclusif pour la proie

## 1.2. Application au bief Andrésy-Méricourt

### 1.2.1 Le site choisi

Une rivière possède une dynamique hydrologique qui modifie en permanence le bilan de matière du site étudié. Pour se rapprocher d'un système fermé, nécessaire au modèle ECOPATH II, nous avons choisi le bief Andrésy-Méricourt (Figure 1), situé entre deux barrages, en considérant que les entrées de matière compensent les sorties. Ce bief fortement anthropisé constitue un écosystème caractérisé par une richesse spécifique faible. Le gardon y représente 80 % de la biomasse et son écologie a fait l'objet d'études pendant les années 1990-91 (Boughida, L. 1992). De plus, les travaux de recherche effectués par le PIREN-Seine, depuis 1989, offrent de nombreuses données sur les différents taxons présents dans ce secteur de la Seine.

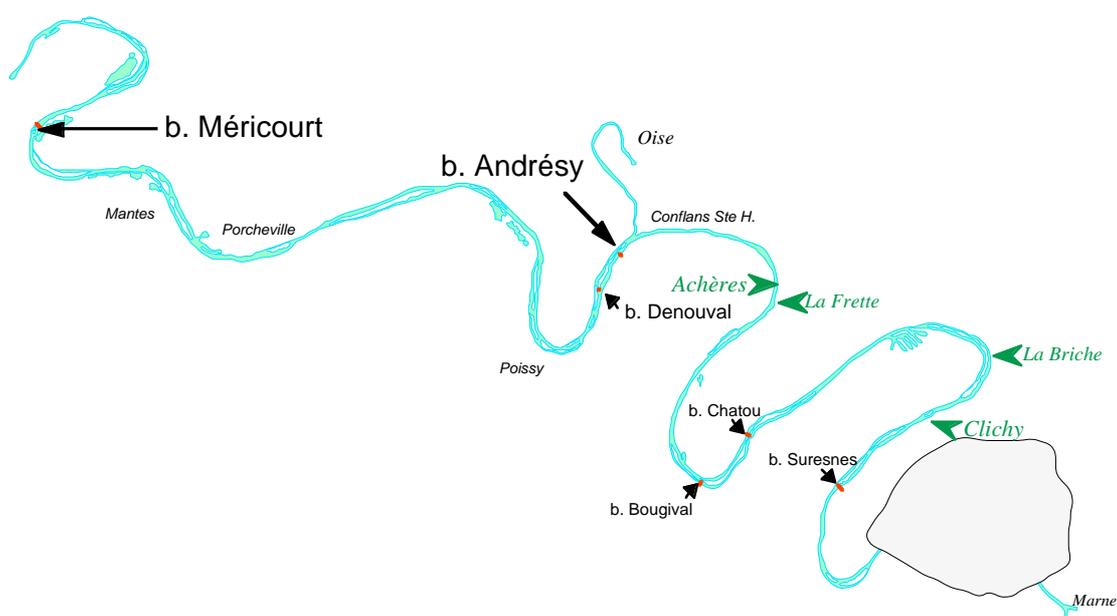


Figure 1. Localisation du site d'étude, le bief Andrésy-Méricourt

### 1.2.2 Les compartiments de l'écosystème

Sept compartiments ont été choisis pour décrire l'écosystème du bief étudié : Gardons adultes et gardons alevins, protozoaires, zooplancton, phytoplancton, bactéries et détritus.

Le modèle utilise un compartiment « Détritus » représentant la nécromasse du système. Ce dernier est alimenté par l'égestion (nourriture non assimilée) et la mortalité non due à la prédation.

Les régimes alimentaires sont décrits dans le Tableau I. Une fraction du régime alimentaire des bactéries est attribuée au phytoplancton afin de considérer l'excrétion de ce dernier, qui peut assurer jusqu'à 17% de la consommation de carbone des bactéries (Garnier et al. 1993).

#### Hypothèses

Trois hypothèses sont faites pour construire le modèle :

- Le bief est un système clos dont les entrées et sorties se compensent
- Pas d'accumulation de biomasse entre 2 années (i.e système à l'équilibre)

- Pas d'exportation de Gardons par pêche (mortalité naturelle uniquement)

Tableau I. Régimes alimentaires des compartiments trophiques choisis. Les valeurs sont exprimées en fraction du régime alimentaire du prédateur.

Prédateur					
Proie	Gardon adulte	Gardon alevin	Protozoaires	Zooplancton	Bactéries
 Protozoaires	0,01	0,01			
 Zooplancton	0,10	0,3			
 Bactéries	0,01	0,01	0,75		
 Phytoplancton	0,44	0,48	0,02	0,80	0,15
Détritus	0,44	0,20	0,23	0,20	0,85

Les données utilisées proviennent de la bibliographie et des travaux du Piren-Seine.

### 1.2.3 Caractéristiques bioénergétiques des compartiments

Les paramètres d'entrée du modèle pour chaque compartiment sont présentés dans le Tableau II. Comme il a été précisé précédemment (§ 1-2), on renseigne les champs Biomasse, Production/Biomasse, Consommation/Biomasse (ou Production/Consommation) et Efficacité écotrophique ainsi que non-assimilé/Consommation. D'après les hypothèses posées, la colonne Biomasse accumulée vaut 0.

La variabilité saisonnière est importante (température, éclairage etc.) et les flux de matières le sont donc aussi (blooms phytoplanctoniques etc.). Le pas de temps choisi est l'année et les données récoltées sont donc moyennées sur cette période.

Tableau II. Tableau récapitulatif des valeurs d'entrée du modèle.

Group name	Trophic mode	Biomass (mgC/m <sup>2</sup> )	Prod./ Biom. (year)	Cons./ Biom. (year)	Ecotrophic efficiency	Production / Consumption	± Biom.acc. (mgC/m <sup>2</sup> /year)	Unassimil./ Consumption	Detr.import (mgC/m <sup>2</sup> /year)
Gardons adultes	0,0	800,000	0,440	23,400			0,000	0,450	
Gardons alevins	0,0	80,000	7,680	213,500			0,000	0,400	
Zooplancton	0,0	110,000	65,000			0,150	0,000	0,500	
Protozoaires	0,0	330,000	120,000	400,000			0,000	0,500	
Bacteries	0,0	550,000	250,000			0,330	0,000	0,200	
Phytoplancton	1,0	5500,000	100,000				0,000		
Detritus	2,0	5500,000							0,000

Pour effectuer les conversions d'unités concernant les poissons, nous considérons que le poids sec représente 30% du poids frais et le carbone 35% du poids sec. L'estimation de la biomasse en gardons ne se fait pas directement mais par ajustements progressifs par rapport aux autres paramètres. Les valeurs de P/B pour les gardons sont calculées à partir des travaux de Boughida (1992). Les valeurs de Q/B pour les gardons adultes sont issues des travaux de Palomares (1991), pour les alevins c'est la valeur trouvée pour le Lac d'Aydat qui est retenue (Reyes-Marchant, P. 1993). La fraction non assimilée est plus élevée pour les adultes compte tenu du régime plus détritivore des adultes.

Les différents paramètres B et P/B retenus pour le phytoplancton proviennent des travaux du PIREN-Seine (Garnier *et al.*, 1995). Les données initiales étant exprimées par litre, une profondeur moyenne de 5,5 mètres est utilisée pour les exprimer en m<sup>2</sup>.

Les Rotifères ont un cycle de quelques jours, les Cladocères d'environ une semaine, et les Copépodes de quelques semaines. Ainsi, lors d'apparitions de blooms phytoplanctoniques, les Rotifères ont un temps de réaction bien meilleur. Il en résulte que 97 % de la biomasse zooplanctonique est attribuée aux Rotifères. Le modèle assimile le zooplancton aux Rotifères. Les valeurs utilisées sont issues des travaux de Garnier *et al.* (1995) pour B et P/B et de ceux de Mathews (1993) pour P/C et non assimilé.

Concernant les protozoaires et les bactéries, les valeurs sont issues des travaux du PIREN-Seine (Garnier et Billen, comm. pers.).

#### 1.2.4 Résultats

Les paramètres de sortie du logiciel sont réunis dans un tableau. Une partie d'entre eux permet de réaliser une sortie graphique (Figure 2) indiquant le niveau trophique des compartiments, leur importance relative (les cadres sont proportionnels au logarithme de la biomasse) et les flux de matière entre compartiments.

#### 1.2.5 Discussion

Selon les données récoltées dans la littérature par Randall *et al.* (1995), la biomasse moyenne en poisson pour 58 rivières est de 146,1 kg.ha<sup>-1</sup>. En utilisant les facteurs de conversion définis précédemment (§ 2-4), la biomasse obtenue par le modèle ECOPATH II est inférieure : environ 84 kg.ha<sup>-1</sup>. Comme il l'a été précisé au préalable, pour construire le modèle, les biomasses de gardons adultes et alevins n'ont pas été directement déduites. L'estimation s'est faite par itération et la proportion d'alevins en biomasse a été fixée arbitrairement à 10%, ce qui est légèrement inférieur à la valeur obtenue par Reyes-Marchant pour le lac d'Aydat (13,3%). Les résultats retenus pour les biomasses en poisson sont des *maxima*. En effet, si l'on essaie des valeurs de biomasse plus importante, l'efficacité écotrophique du zooplancton devient supérieure à un, ce qui est impossible. Cela est dû au fait que la totalité de la biomasse du zooplancton consommée est attribuée au gardon : 30% du régime alimentaire des alevins et 10% de celui des adultes. Le zooplancton représente le facteur biotique limitant pour le gardon.

On peut penser que la biomasse en zooplancton est sous-estimée et/ou que la composition des régimes alimentaires des poissons est à revoir.

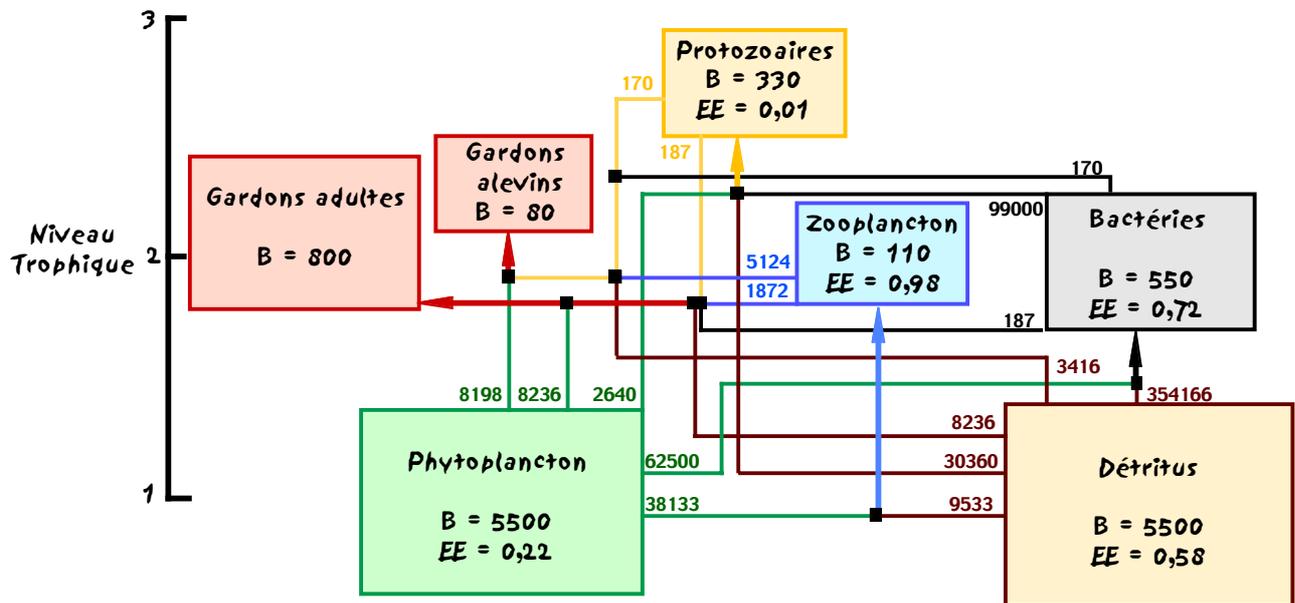


Figure 2. Représentation des flux du modèle ECOPATH II

Les valeurs de flux sont exprimées en  $\text{mg C/m}^2/\text{an}$   
 B indique la biomasse en  $\text{mg C/m}^2$   
 EE est l'efficacité écotrophique;  $EE = 0$  pour les gardons  
 représente les connexions aboutissant à un compartiment

Les Copépodes et les Nauplii n'apparaissent que très faiblement dans la biomasse zooplanctonique des prélèvements d'eau alors que des résultats d'analyses de contenus stomacaux d'alevins de gardon révèlent des importances numériques pouvant atteindre 86% pour les Nauplii (Boughida 1992). Les Copépodes sont représentés par des importances numériques faibles mais des fréquences d'occurrences élevées (jusqu'à 90%). Une hypothèse sous-jacente apparaît alors : il existe une production non négligeable de ces organismes mais insuffisante pour voir apparaître une forte biomasse dans la mesure où ils seraient consommés au fur et à mesure.

Le modèle ne comprend pas de compartiment benthique. En effet, aucune donnée précise n'est actuellement disponible. Néanmoins, il est intéressant de remarquer que Boughida (1992) a relevé des importances numériques non négligeables de Chironomes (jusqu'à 37%) et Chydorides (jusqu'à 60%) dans les contenus stomacaux d'alevins de gardons.

La proportion d'alevins, fixée dans le modèle à 10% de la biomasse des adultes, est un paramètre pouvant varier sensiblement. Cette valeur est à préciser.

Différentes simulations avec des proportions différentes offrent des résultats en biomasse très variés. Par exemple, en conservant une efficacité écotrophique maximale pour le zooplancton, pour une proportion en alevin de 5%, la biomasse en gardon est alors  $114,3 \text{ kg} \cdot \text{ha}^{-1}$  pour les adultes et  $5,7 \text{ kg} \cdot \text{ha}^{-1}$  pour les alevins. Une proportion de 5% d'alevins n'est pas nécessairement trop faible dans le cas de la Seine à l'aval de Paris. En effet, le développement des alevins se fait, dans un premier temps, en période estivale. La biomasse de zooplancton y est alors très forte et il est concevable que cette dernière puisse suppléer les besoins nutritionnels des alevins. De plus, de nombreux alevins semblent disparaître après l'été en raison des faibles concentrations en oxygène. Ainsi, à l'échelle de l'année, la

proportion d'alevins dans la population pourrait donc être très faible. Des proportions plus importantes pouvant être observées pendant la saison estivale.

L'utilisation du modèle ECOPATH II fait apparaître des incertitudes sur le régime alimentaire des poissons et de leur consommation (notamment Q/B des alevins). Il semble donc nécessaire, dans un premier temps, d'obtenir des données propres à la Seine concernant la consommation du gardon, sur les plans quantitatif et qualitatif.

### **1.3. Conclusion**

La biomasse de poisson, approchée par cette méthode, est réaliste compte tenu de la connaissance préalable du milieu et notamment des faibles densités observées par échosondeur. Cette tentative montre néanmoins la nécessité de disposer de données complémentaires pour affiner notre approche.

Il s'agirait d'abord de préciser, au plan quantitatif, le régime alimentaire du gardon dans le bief Andrézy-Méricourt, ce qui nécessiterait au moins deux cycles nyctéméraux de prélèvements réguliers de poissons, suivis des analyses de leurs contenus stomacaux au laboratoire. La consommation Q/B des alevins pourrait alors être estimée à l'aide du logiciel MAXIMS utilisant les contenus stomacaux (Jarre *et al.*, 1990).

Se pose ensuite la question d'étendre l'étude à l'ensemble du bassin, laquelle renvoie à des investigations systématiques, à considérer au minimum par rang fluvial. Mais dans ce cas, si pour tester la méthode, le cours aval de Paris a été volontairement choisi en raison de la relative simplicité de son peuplement piscicole, le problème serait fort différent dans les secteurs abritant des communautés diversifiées. Cet aspect pourrait néanmoins être résolu par une approche par guildes (groupes d'espèces exploitant des ressources alimentaires sensiblement de même type).

Dans tous les cas cependant, il convient de souligner que ces efforts n'auraient pas de sens sans une meilleure connaissance du compartiment benthique, lequel constitue un maillon essentiel, exploité par de nombreuses espèces de poissons.

## 2. Modélisation des relations milieux - peuplements

Pour approfondir la connaissance du peuplement piscicole et préciser davantage l'importance relative des variables de milieu dans les mécanismes de structuration des communautés, la modélisation est une étape indispensable. Une telle modélisation permettrait aussi de simuler ensuite l'impact de différents aménagements.

Ces travaux s'appuient sur l'exploitation d'une importante base de données qui couvre l'ensemble du bassin de la Seine. Ces données sont issues d'échantillonnages réalisés au moyen de pêches électriques. Elles 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 elles ont l'avantage d'être comparables et de couvrir un vaste espace. Il faut donc trouver des méthodes d'analyse adaptées.

Une première synthèse a été réalisée au moyen d'analyses multivariées. 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 (Belliard *et al.* 1997).

Pour aller plus loin dans l'analyse, il s'agit de modéliser les relations entre les poissons et les descripteurs physiques de l'habitat. Une telle modélisation se heurte toutefois à la complexité des systèmes étudiés. Ces derniers sont en effet constitués de nombreuses composantes dont l'ensemble des interactions est encore mal connu et les relations entre les poissons et les descripteurs physiques de l'habitat, par exemple, sont *a priori* non-linéaires. Il convient donc d'utiliser des méthodes adaptées : non-linéaires et robustes, c'est-à-dire peu sensibles au bruit des données, convenant mieux à nos données. Les réseaux de neurones fournissent un exemple de tels modèles.

Leur application à la prédiction de la présence ou de l'absence des espèces de poissons à l'échelle du bassin de la Seine s'avère en effet pertinente (Boët & Fuhs 2000). La mise en œuvre effective de cette technique apparemment facile est toutefois délicate. C'est pourquoi, avec l'aide de collègues du LISC, je me suis orienté aujourd'hui vers une autre approche : les arbres de décision.

### 2.1. Les réseaux de neurones

Constitués de composants simples, les neurones formels, interconnectés en grand nombre, les réseaux de neurones permettent de représenter toute fonction calculable (McCulloch & Pitts 1943). Ce qui fait le succès de ces modèles réside dans leur capacité à modéliser des phénomènes non-linéaires. Certes, les statisticiens classiques ne sont pas dépourvus devant ces non-linéarités (modèles linéaires généralisés, estimations non-paramétriques, etc.) mais, par leur simplicité d'utilisation, les réseaux de neurones constituent depuis leur avènement des compétiteurs crédibles. D'ailleurs, les ponts entre les deux communautés sont de plus en plus actifs (Cheng & Titterton 1994).

Nous avons testé l'efficacité de ces modèles pour prédire la présence ou l'absence de poissons à partir des caractéristiques du milieu. Le problème posé est un problème de discrimination, pour lequel l'utilisation de réseaux connexionnistes multicouches entraînés par l'algorithme de rétropropagation du gradient a montré son intérêt (Rumelhart *et al.* 1986). Cette démarche a donc été privilégiée au cours de cette approche. Nous avons choisi d'utiliser l'implémentation MASS décrite par Ripley (1996) et Venables & Ripley (1997).

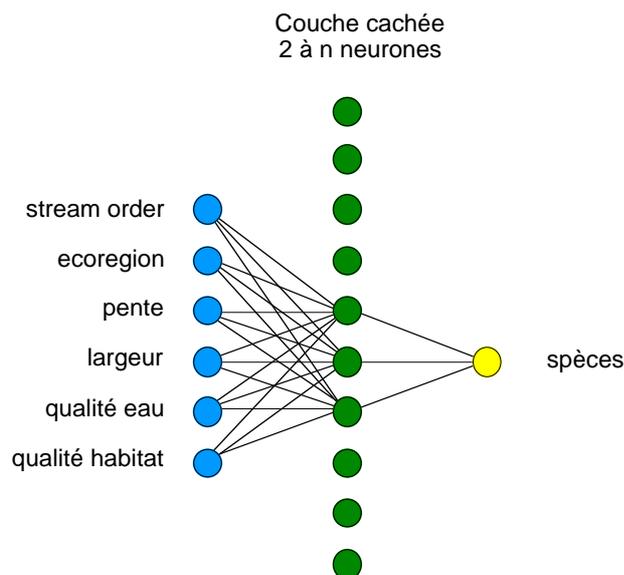


Figure 3. Schéma des réseaux mis en œuvre (d'après Boët & Fuhs 2000).

Éprouvée à l'échelle du bassin de la Seine et en fonction de descripteurs très globaux de la qualité du milieu aquatique (six variables synthétiques d'entrée, cf. Figure 3), la prédiction en termes de présence ou d'absence d'une espèce par des réseaux connexionnistes multicouches s'avère pertinente (Boët & Fuhs 2000).

Vingt-six espèces sont testées, choisies parmi les plus représentatives présentes dans le bassin (ie occurrence  $\geq 9\%$ ). Et alors que les données d'entrée sont assez fortement bruitées, les taux de réussite en généralisation varient de 67,6 à plus de 87 % selon les espèces, ce qui représente des performances très appréciables car une erreur de mesure de l'ordre de 10 à 20 % sur ce type de données est très probable.

Les meilleurs résultats sont obtenus pour la truite, le chabot, la loche franche et le vairon. Ces espèces sont aussi parmi celles dont les profils écologiques sont les plus nets sur le bassin de la Seine. Elles font partie du cortège faunistique classique rencontré dans les zones amont, où les habitats restent parmi les moins perturbés.

À l'inverse, les moindres performances s'expliquent par l'écologie particulière de certaines espèces dans le bassin de la Seine (cas du goujon ou de l'anguille).

Par ailleurs, pour certaines espèces, l'examen détaillé des résultats révèle des différences sensibles entre prédiction de présence ou prédiction d'absence. Celles-ci paraissent liées à des difficultés d'échantillonnage (brème, tanche) ou des manipulations humaines des populations (brochet). Néanmoins, il est bien sûr possible également que les variables prédictives d'entrée du modèle ne sont pas suffisantes pour prédire efficacement ces différentes espèces et que d'autres caractéristiques sont nécessaire pour affiner leur prédiction.

## 2.2. De la boîte noire à la boîte de verre...

Les prédictions obtenues à l'aide des réseaux de neurones sont donc de bonne qualité. Les taux d'erreur s'échelonnent en effet de 13 à 32 %, ce qui est tout à fait intéressant lorsqu'on les compare aux erreurs de mesure qui peuvent fréquemment atteindre 20 %. En ce sens, les réseaux de neurones se révèlent une technique non-linéaire tout à fait pertinente dans la prédiction de la faune ichtyologique à l'échelle du bassin versant.

Compte tenu de la nature des données traitées et du caractère très synthétique des variables d'entrée, ces modèles s'avèrent très satisfaisants. Ils sont déjà très proches de ceux obtenus à l'aide de

méthodes classiques, comme les analyses discriminantes et les régressions multiples utilisées par exemple par Pouilly (1994) ou Capra (1995). Mais ces derniers, 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.

Ces essais sont donc encourageants si l'on considère qu'à l'heure actuelle il n'existe guère de modèles prédictifs de poissons à l'échelle d'un bassin fluvial. Oberdorff *et al.* (1999, 2001) viennent d'appliquer avec succès des procédures de régressions logistiques pour élaborer des modèles qui décrivent la présence des espèces piscicoles et la richesse des peuplements dans les grands cours d'eau français. La variabilité expliquée est dans ce cas autour de 73 %. Seuls, Mastroiello *et al.* (1997, 1998) ou Guégan *et al.* (1998) ont déjà démontré les capacités des réseaux de neurones artificiels pour prédire la richesse spécifique des communautés de poissons à large échelle.

Néanmoins, la mise en œuvre des réseaux de neurones n'est pas toujours aisée (Boët & Fuhs 2000). En effet, 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.

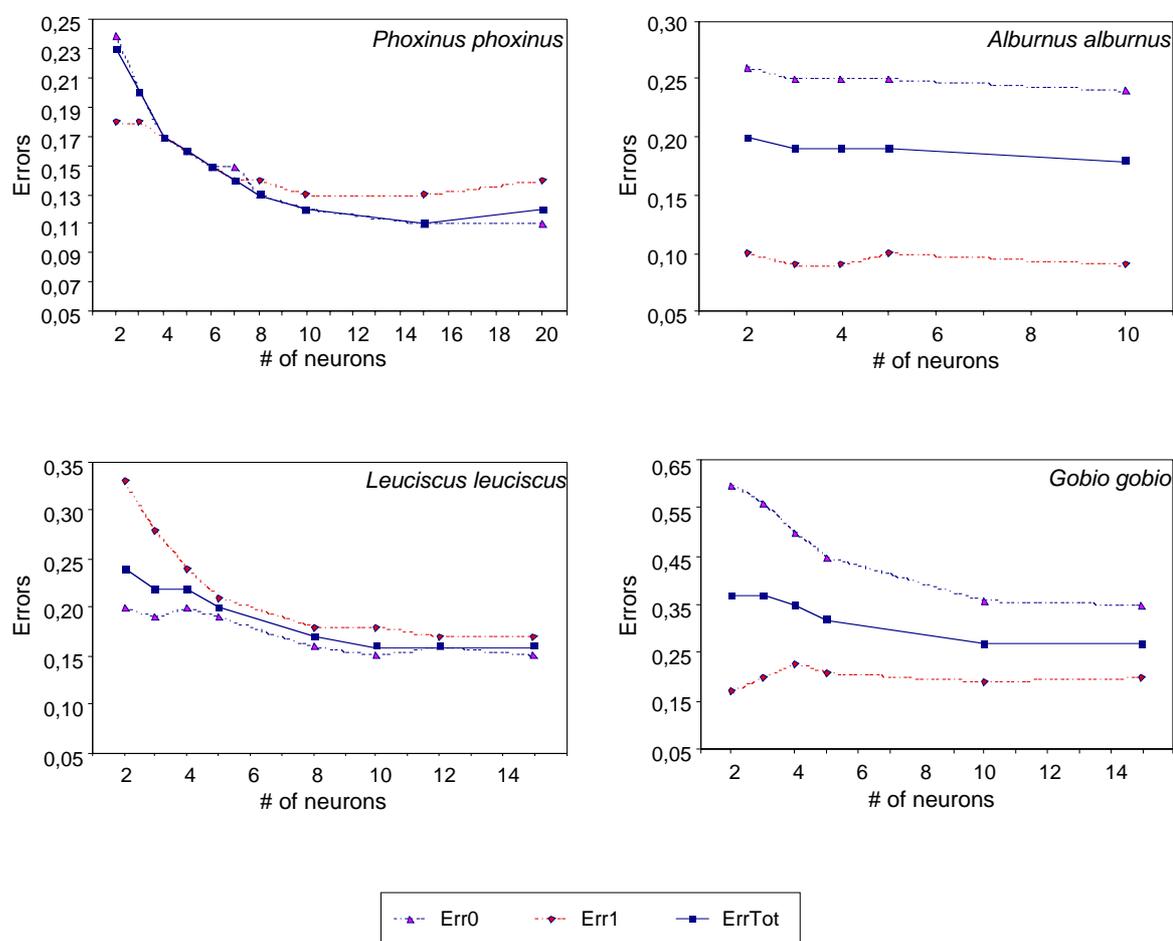


Figure 4. Comportement de l'erreur de prédiction en fonction du nombre d'unités dans la couche cachée des réseaux. Err0 : erreur de prédiction sur l'absence, Err1 : erreur de prédiction sur la présence, ErrTot : erreur de prédiction totale (d'après Boët & Fuhs 2000).

Le nombre d'unités de la couche d'entrée du réseau est lié aux variables prédictives choisies. Mais pour chacune des espèces, nous avons cherché à choisir le meilleur nombre d'unités en couche cachée et ceci s'est avéré plus délicat que prévu. La Figure 4 montre en effet le taux d'erreur de prédiction en fonction du nombre d'unités dans la couche cachée pour 4 espèces. Dans la plupart des cas, l'erreur décroît assez rapidement avec l'augmentation du nombre de neurones, mais on observe

ensuite un plateau qui ne permet pas de sélectionner de façon rigoureuse la taille optimale des réseaux construits.

La gestion des variables qualitatives est également délicate. Dans le cas présent, l'écorage est une variable purement qualitative, qu'il faut découper en autant d'unités d'entrée que de modalités moins un (ie 6 modalités - 1). Les variables quantitatives étant au nombre de 5, la couche d'entrée du réseau comprend donc en réalité 10 unités.

Enfin, les réseaux de neurones apparaissent comme une boîte noire au biologiste : les coefficients du réseau ne peuvent en effet avoir une interprétation biologique. Ce dernier point semble le plus préjudiciable dans la perspective d'une utilisation de ces modèles prédictifs pour l'aide à la gestion au niveau du bassin versant. Il est par exemple délicat de prédire l'impact d'un aménagement sur la présence d'une espèce particulière sans proposer « d'explication » à cet impact, et ce qu'il soit positif ou négatif.

C'est pourquoi, j'ai aussi testé une autre méthode de discrimination appelée « arbres de décision ».

### **2.3. Les arbres de décision**

Les arbres de décision constituent une méthode statistique applicable aux domaines de la discrimination et de la régression. Le problème de la prédiction de la présence d'une espèce en fonction des caractéristiques du tronçon de rivière considérée étant un problème de discrimination, seul ce cas est présenté dans ce qui suit.

À partir d'un échantillon de données représentatives du phénomène considéré, la discrimination statistique consiste à construire un modèle permettant de prédire la classe d'une nouvelle observation. Les pêches électriques qui décrivent les caractéristiques physico-chimiques et morphologiques de la station ainsi que les espèces détectées sont un tel échantillon. La classe d'une observation (une variable qualitative) est ainsi la variable binaire indiquant la présence ou l'absence d'une espèce donnée.

#### **2.3.1 Principe**

L'objectif de cette méthode est le partitionnement récursif de l'espace des observations en sous-domaines les plus homogènes possibles quant à la classe de leurs éléments. La construction se fait en partant de la partition triviale, contenant toutes les observations, à laquelle on attribue la classe majoritaire. On tente alors de séparer cet ensemble suivant une des variables de l'échantillon. Cette variable et la valeur de séparation sont déterminées de manière à engendrer des sous-ensembles plus homogènes quant à la classe de leurs éléments. Un nœud est alors construit qui contient le test sur la variable. Conventionnellement, si le test réussit, l'observation est affectée à la feuille gauche de l'arbre. Ce découpage est poursuivi récursivement sur chaque branche de l'arbre jusqu'à obtention de nœuds totalement homogènes ou d'effectifs trop faibles pour rester représentatifs.

#### **2.3.2 Validation**

Comme toute méthode statistique inférentielle basée sur un échantillon de taille finie, les arbres de décision présentent une erreur sous-évaluée. En particulier, un arbre développé au maximum, c'est-à-dire dont toutes les feuilles sont homogènes, a une erreur nulle. Mais appliqué à un autre échantillon, il présentera une erreur certainement plus importante. C'est pourquoi il est nécessaire de choisir un arbre ayant de bonnes capacités de *généralisation*, c'est-à-dire un arbre qui reste pertinent vis-à-vis d'un nouvel échantillon.

Pour cela, les arbres de décision utilisent la validation croisée et un critère d'homogénéité pénalisant la taille de l'arbre (son nombre de nœuds). L'arbre retenu est alors celui qui réalise le meilleur compromis entre erreur et taille.

#### **2.3.3 Construction**

L'algorithme de construction des arbres de décision se déroule en deux phases successives.

La **première phase** est un algorithme glouton de partitionnement récursif binaire. À chaque pas, est déterminée la meilleure séparation linéaire des observations de l'échantillon d'apprentissage, puis récursivement sur chaque sous-ensemble de l'échantillon ainsi formé. La récursion stoppe dès que le sous-ensemble est homogène (toutes les observations sont de même classe), ou si son effectif est trop faible, ce seuil étant décidé par l'utilisateur.

En d'autres termes cela consiste à effectuer une partition de l'espace X des observations vers l'ensemble C des différentes classes du problème. Cependant, la distribution des données dans l'espace des X induit fréquemment des chevauchements. Il n'existe donc pas de partition décrivant complètement les classes. Ainsi, pour chaque feuille de l'arbre, il y aura une probabilité de distribution selon les classes et la règle de décision de Bayes choisira la classe avec la plus forte probabilité (Fuhs 1998). Il s'agit donc de faire des partitions en utilisant les probabilités de plus grande non-classification. La méthode utilisée pour effectuer la construction de l'arbre fait appel à une procédure « pas à pas ». Cela signifie qu'elle effectue la séparation suivante de manière optimale sans tenter d'optimiser les performances à l'échelle de l'arbre entier.

La **seconde phase** est « l'élagage » de l'arbre. Il s'agit de nettoyer l'arbre de ses branches les moins significatives statistiquement. En effet, tout modèle statistique construit à partir d'un échantillon fini est sujet à ce qu'on nomme « le dilemme de l'apprentissage ». Cela signifie que si le modèle est très complexe (dans notre cas, un arbre avec de nombreuses feuilles et branches), il sera trop « proche » de l'échantillon ayant servi à le construire et généralisera mal à des observations étrangères. À l'inverse, un modèle trop fruste (peu de branches et de feuilles) ne distinguera pas suffisamment les contours de la surface de décision (frontière entre les observations des différentes classes).

Breiman *et al.* (1984) ont montré qu'il existait un élagage optimal de tout arbre de décision. Cette propriété vient du fait qu'il est possible de classer les nœuds de tout arbre de décision suivant sa résistance à la pénalisation du critère d'erreur. Notons  $R(T)$  ce critère non pénalisé, et  $R\alpha(T)$  le critère pénalisé obtenu par  $R\alpha(T) = R(T) + \alpha \text{NombreNœuds}(T)$ . Il est immédiat que si nous faisons augmenter  $\alpha$ , la taille de l'arbre le pénalisera de plus en plus. Breiman *et al.* ont alors montré qu'il existait une suite optimale d'arbres emboîtés issus de l'arbre construit dans la première phase et dont les éléments (des sous-arbres de cet arbre initial) correspondent à un  $\alpha$  donné.

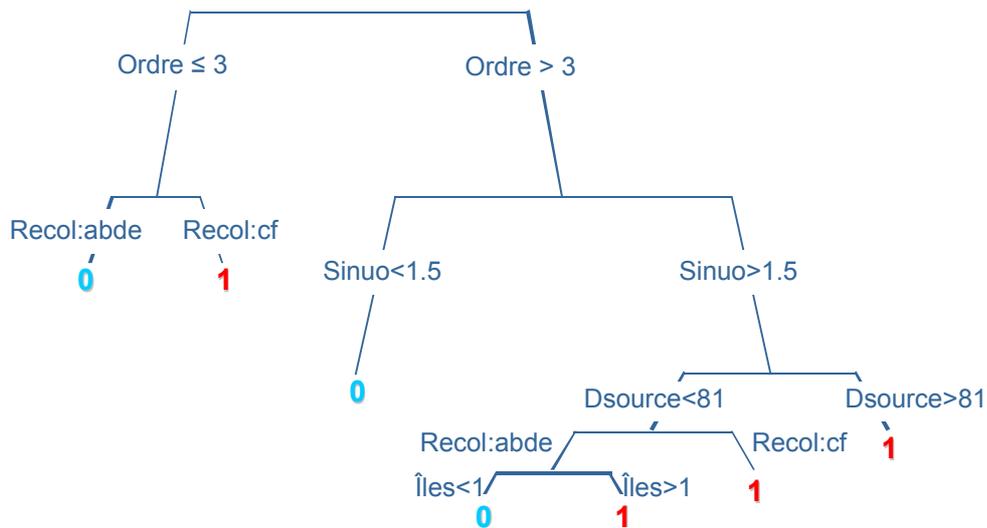


Figure 5. Modélisation de la présence-absence du brochet au moyen d'arbres de décision. Dans cet exemple, seules 5 variables sont retenues par le modèle parmi 19 variables possibles (d'après Boët *et al.* 2000).

Muni de ce résultat théorique essentiel, il faut maintenant choisir le meilleur arbre de cette suite qui maximise la capacité de généralisation, c'est-à-dire la qualité du modèle sur de nouvelles données. Bien entendu, si nous estimons l'erreur à l'aide de l'échantillon initial, il est évident que c'est l'arbre initial qui est optimal par sa construction même. Mais l'erreur ainsi calculée appelée erreur empirique n'est qu'une approximation qui sous-évalue la véritable erreur de discrimination donnée par l'arbre.

Une meilleure estimation de cette véritable erreur consiste à utiliser la « validation croisée ». Cette méthode consiste à découper l'échantillon initial en N parties égales (N=2, 3, 5, 10 sont des valeurs fréquemment choisies), à construire un arbre et la suite de ces arbres élagués sur chacun des N échantillons constitués de N-1 des parties ci-dessus. Alors, au lieu de calculer l'erreur empirique sur ces échantillons, nous la déterminons sur la N<sup>ième</sup> partie non utilisée dans le développement de l'arbre. En moyennant ensuite sur les N suites d'arbres, nous obtenons une estimation meilleure du critère pénalisé  $R\alpha(T)$  pour tout  $\alpha$ . Breiman *et al.* ont montré que celui-ci est une fonction linéaire par morceaux de  $\alpha$ , laquelle, hormis les cas pathologiques, est convexe : elle diminue tout d'abord lorsque  $\alpha$  augmente pour atteindre un minimum, puis augmente à nouveau. Le paramètre  $\alpha$  correspondant au minimum est le paramètre déterminant finalement le meilleur arbre élagué.

Un exemple d'un tel arbre est montré par la Figure 5, lequel modélise la présence-absence du brochet.

### 2.3.4 Avantages

Les avantages de la technique des arbres de décision sont multiples.

En premier lieu, les paramètres sont tous explicites : les tests aux nœuds de l'arbre portent chacun sur une seule variable. Si celle-ci est quantitative, le test indique un seuil ; si elle est qualitative, le test indique l'appartenance à un sous-ensemble des modalités possibles de la variable. Dans les deux cas, ceci est immédiatement interprétable par le biologiste. Ainsi, le modèle s'apparente beaucoup plus à une boîte de verre !

En deuxième lieu, les arbres de décision font la sélection des variables les plus pertinentes. Si une variable n'apparaît en aucun nœud de l'arbre, c'est qu'elle est peu pertinente pour la discrimination recherchée. Cette sélection se fait automatiquement lors de la construction de l'arbre, puisque seules les variables discriminantes y sont retenues.

Enfin, la qualité de la prédiction est généralement bonne, voire meilleure que celle des réseaux de neurones. Ceci apparaît sur le Tableau III, lequel recense les résultats obtenus pour la présence du brochet. L'erreur de prédiction y est très comparable à celles des réseaux de neurones, mais l'arbre est plus compact que le réseau correspondant qui contient 8 unités sur la couche cachée.

Tableau III. Prédiction de la présence-absence du brochet : comparaison des méthodes réseaux de neurones et arbres de décision (d'après Boët *et al.* 2000).

Méthode	Variabes	Err. moy	Err. min	Err. max	Taille
R.N.	6	20,8	13,1 (0)	32 (1)	8 cachés
Arbres	6 sur 6	20,3	13,4	40	15 feuilles
Arbres	5 sur 19	20,1	10	28,5	7 feuilles

### 2.4. Vers des modèles prédictifs géo-référencés

Il s'agit de bâtir un modèle d'organisation et de structuration des peuplements piscicoles à l'échelle globale du bassin. Ce dernier est intégré à un système d'information géographique, afin

d'expliciter les contraintes spatiales par la prise en compte de la topologie arborescente du chevelu hydrographique.

Des développements récents de modèles d'arbres de décision, couplés à une base de données spatialisées au moyen d'un SIG, ont conduit aujourd'hui à un outil expérimental permettant de prédire les espèces de poissons en fonction des caractéristiques des habitats aquatiques (Boët *et al.* 2001).

#### **2.4.1 Données géoréférencées**

La base géographique sur laquelle reposent les modèles est actuellement constituée de 6 couches d'informations :

- le réseau hydrographique (échelle 1 : 50 000), fourni par l'Agence de l'Eau Seine-Normandie ;
- les stations de pêches, localisées et spatialisées ;
- les bassins versants, définis et numérisés sur la base de la couverture IGN du bassin (1 : 100 000) ;
- l'occupation du sol, définie par la base de données Corine Land Cover (1 : 100 000) ;
- un modèle numérique de terrain (Gtopo 30) de résolution 650 m ;  
les écorégions de Dupias et Rey (1985) à l'échelle 1 : 100 000.

Grâce à cet outil, il est ainsi possible de calculer des variables dérivées caractérisant les cours d'eau à l'échelle de l'ensemble du bassin (rang fluvial, pente des tronçons de cours d'eau, surface des bassins versants, distance entre les stations de pêche, distance de chaque station de pêche à l'exutoire final, etc.). Peuvent également être pris en compte des descripteurs extérieurs au cours d'eau lui-même caractérisant le bassin versant. Corine Land Cover est par exemple une façon d'intégrer le degré d'anthropisation et une nouvelle couverture plus détaillée, d'une meilleure résolution, doit bientôt être disponible.

#### **2.4.2 Constitution d'une base de données Oracle™**

L'ampleur des données ainsi rassemblées a conduit à la nécessité d'optimiser leur gestion au moyen d'une base de données relationnelle. Toutes sont maintenant rassemblées dans une base Oracle™ : variables dérivées du système d'information géographique, données physiques décrivant localement les caractéristiques des stations de pêche et données biologiques afférentes. L'intérêt de la constitution d'un tel outil est de disposer maintenant d'un référentiel validé et sécurisé ; celui-ci est en outre conçu pour être consulté et interrogé à distance via un navigateur Internet.

Une centaine de variables, réparties en 7 tables, sont aujourd'hui renseignées. Les données biologiques sont constituées des résultats de plus de 1000 pêches électriques, réalisées sur 578 stations.

#### **2.4.3 Mise en œuvre**

Les routines mises en œuvre utilisent la bibliothèque RPART de S-Plus (Therneau & Atkinson 1997). Le meilleur arbre élagué est choisi par validation croisée en découpant le tableau initial des observations en 20 parties égales.

Sur cette base, des modèles préliminaires de prédiction de présence-absence des espèces piscicoles sont construits à l'échelle de l'ensemble du bassin de la Seine. À titre illustratif, une généralisation de prédiction du brochet est présentée (Figure 6).

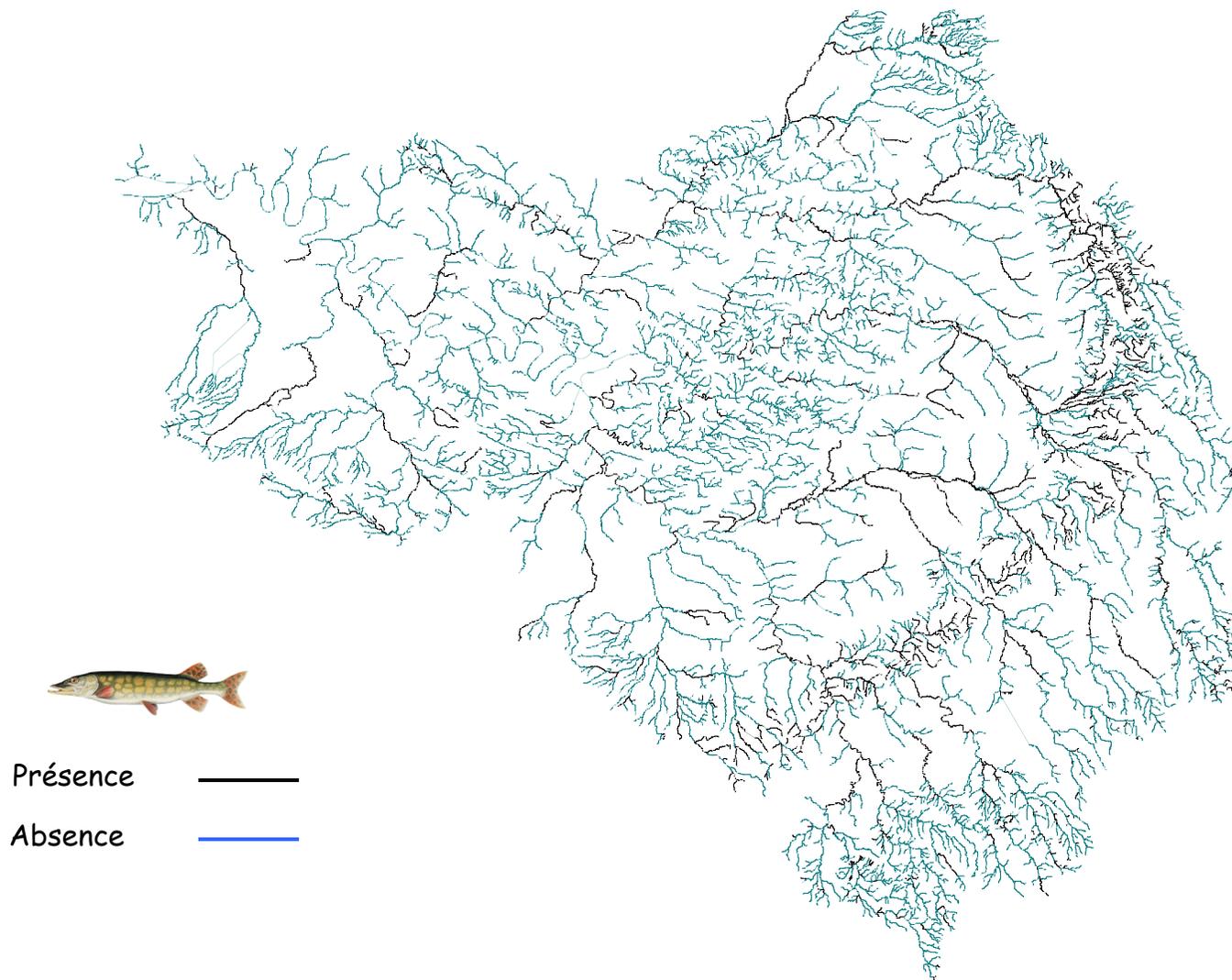


Figure 6. Exemple de généralisation de la prédiction du brochet à l'ensemble du réseau hydrographique de la Seine (d'après Boët et al. 2001).

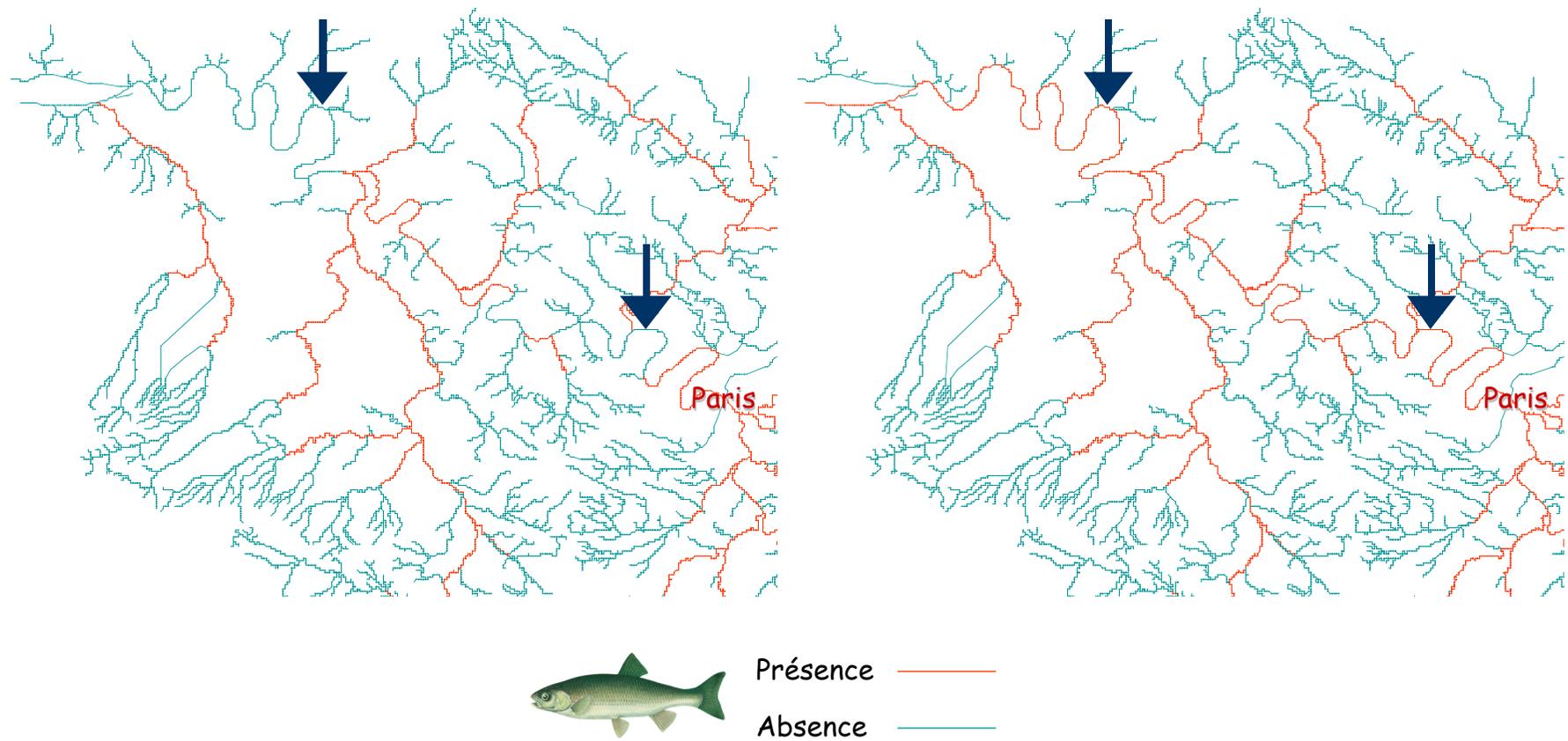


Figure 7. Exemple de prédictions du chevesne : dans les conditions de milieu actuelles (à gauche) ; en simulant des conditions idéalisées d'un traitement optimal des phosphates et des nitrates (à droite). L'espèce est alors prédite sur tout le cours aval de l'agglomération parisienne (d'après Boët et al. 2001).

Dans un même esprit, un essai de simulation de scénario est également montré : il s'agit de la prise en compte de données de qualité de l'eau fournies par le modèle SENEQUE (Poulin *et al.* 1998) considérant des conditions optimales de traitement des eaux usées sur le bassin (abattement des phosphates et des nitrates). Une telle amélioration des conditions de milieu se traduit bien par une modification sensible, prédite par le modèle, de la répartition du chevesne à l'aval de l'agglomération parisienne, comparé à la situation actuelle (Figure 7).

Cet exercice ne vise toutefois qu'à montrer les possibilités susceptibles d'être offertes par cet outil encore en cours de réalisation, de nombreuses mises au point restant en effet nécessaires.

#### **2.4.4 Perspectives**

Une « architecture » générale est donc actuellement en place, qu'il convient en partie de consolider mais dont il faut surtout maintenant aménager « l'intérieur », afin de finaliser nos modèles de prédiction des peuplements de poissons à l'échelle du bassin de la Seine.

Quelques variables explicatives doivent encore compléter cet ensemble, comme la température et l'hydrologie notamment, ou la carte des rejets de l'Agence de l'Eau Seine-Normandie et celle des barrages de la DIREN Île-de-France.

Les données récentes du suivi du Réseau Hydrobiologique et Piscicole du Conseil supérieur de la pêche sont en cours d'intégration et utilisées à des fins de validation des modèles.

Nous envisageons également de figurer l'incertitude des prédictions dans la représentation spatialisée des résultats. Doivent aussi être prises en compte, les abondances relatives des espèces afin de prédire les peuplements piscicoles.

L'essentiel du travail encore à fournir doit toutefois porter maintenant sur l'analyse systématique et détaillée des réponses de chacune des espèces piscicoles aux variables de milieu des modèles, afin de préciser l'importance relative de chacune dans les mécanismes de structuration des communautés.

Plus fondamentalement, il s'agit maintenant de prendre en compte les corrélations spatiales le long du réseau hydrographique. Cette question n'est pas classique, puisqu'il convient de mettre en œuvre des statistiques spatiales gérant les relations de voisinages dans la topologie particulière du réseau hydrographique (graphe orienté). Diverses méthodes doivent être explorées : treillis, corrélogrammes, variogrammes, etc. (cf. p. ex. Ripley 1981, Arnaud & Emery 2000). Ce travail particulier fait actuellement l'objet d'une thèse qui vient de démarrer (École doctorale E2M2 ; Dir. Alain Pavé). Il est entrepris également en collaboration avec le LIS (François Goraud).

## **2.5. Conclusions**

Ces travaux devraient conduire à une meilleure connaissance des relations milieu-peuplements, par l'identification et la hiérarchisation des variables caractérisant non seulement le cours d'eau, mais son bassin versant.

Un tel outil pourrait aussi permettre une première analyse des aires de répartition des espèces, susceptibles de fournir des éléments préliminaires concernant la fragmentation éventuelle de l'aire de certaines espèces menacées sur le bassin de la Seine, comme la lote de rivière par exemple qui semble en régression actuellement.

Au plan appliqué, il s'agit de bâtir un outil d'aide à l'évaluation des choix en matière de gestion du territoire, capable de fournir une simulation cartographique dynamique permettant de prédire les peuplements en fonctions de scénarios d'aménagements, de gestion d'ouvrages, ainsi que de pollutions.

Notre maquette logicielle est en effet conçue de manière modulaire et évolutive afin de pouvoir s'enrichir ultérieurement de nouvelles fonctionnalités, tant par la prise en compte de nouveaux compartiments biologiques (algues, invertébrés) que de descripteurs du milieu fluvial (typologie des

zones humides en cours d'élaboration par le PIREN-Seine) ou de types d'intervention humaine (rejets ponctuels, pollution diffuse, changement d'occupation du sol).

L'objectif est bien de rendre opérationnel un outil capable d'évaluer à large échelle les effets positifs ou négatifs d'éventuelles modifications de milieu, susceptible de fournir des critères d'analyses et d'anticipation à la disposition de la décision publique, permettant d'améliorer la portée et la pertinence des études d'impact de grands aménagements (barrage, canalisation, etc.), voire d'opérations de restauration.

À terme il pourrait s'appliquer à d'autres contextes. En soulignant bien que par la prise en compte explicite de variables descriptives extérieures au cours d'eau proprement dit, cet outil a bien vocation à s'adresser aux territoires.

À cet égard, il est prévu d'appliquer les modèles d'arbres de décision au bassin de la Loire. Un très important travail d'analyse morphologique et hydrologique a en effet été réalisé sur ce bassin (Cohen *et al.* 1998, Andriamahefa 1999) conduisant à caractériser et affiner le découpage hydro-écorégional initialement proposé (Wasson *et al.* 1993). Toutes ces informations, spatialisées et intégrées dans un SIG, ont encore peu été confrontées à la composante biologique (Ivol-Rigaut 1998, Cattaneo & Breil 2000). Or, la réponse des communautés de poissons au découpage *a priori* des hydro-écorégions s'est déjà avérée très forte (Ivol *et al.* 1996). Il s'agirait maintenant d'explicitier davantage ces différences. Les données disponibles permettent également de définir maintenant des « conditions de référence ». Couplées à la composante piscicole, pourraient alors être enfin établis des référentiels faunistiques régionalisés.

## Références

- Andriamahefa H., 1999. Les hydro-écorégions du bassin de la Loire. Morphologie, hydrologie, pressions anthropiques sur les cours d'eau et les bassins versants (vol. 1 et 2). Thèse Univ. Jean Monnet - Saint Etienne, 272 + annexes.
- Arnaud M. & X. Emery, 2000. Estimation et interpolation spatiale - méthodes déterministes et méthodes géostatistiques., Hermes Sciences Publications, Paris, 221 p.
- Belliard J., 1994. Le peuplement ichthyologique du bassin de la Seine : rôle et signification des échelles temporelles et spatiales. *Thèse Doct. Univ. Paris VI*, 197 p.
- Belliard J., P. Boët & E. Tales, 1997. Regional and longitudinal patterns of fish community structure in the Seine River basin, France. *Environ. Biol. Fish.*, 50, 133-147.
- Boët P. & T. Fuhs, 2000. Predicting presence of fish species in the Seine River basin using artificial neural networks. p. 131-142, in: S. Lek & J.-F. Guégan (Eds), *Artificial neuronal networks. Application to ecology and evolution*, Environmental Science, Springer-Verlag, Berlin/Heidelberg
- Boët P., T. Fuhs, & G. Gorges, 2000. Modélisation prédictive des peuplements piscicoles, à l'échelle du bassin. Rapport d'activité PIREN-Seine, février 2000, Paris, 10 p.
- Boët P., T. Fuhs, G. Gorges & L. Toupotte, 2001. Modélisation prédictive des peuplements piscicoles, à l'échelle du bassin de la Seine. Rapport d'activité Piren-Seine, février 2001, Paris, 9 p.
- Boughida, L., 1992. Eléments d'écologie du gardon (*Rutilus rutilus* L. Cyprinidae). Croissance, reproduction et développement de deux populations de la Seine, en amont et en aval de l'agglomération parisienne Thèse de Doctorat de L'Université Paris VI. 176p.
- Breiman L., J.H. Friedman, R.A. Olshen & C.J. Stone, 1984. *Classification and regression trees.*, Chapman & Hall, New York, 358 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*, 281 p.
- Cattaneo F. & P. Breil, 2000. *A quantification of representative water use effects on natural flow regimes in the Loire River basin, France.*, EISORS (Eight International Symposium On Regulated Streams), Toulouse, France, 17-21 Juillet 2000.
- Cheng B. & D.M. Titterton, 1994. Neural networks: a review from statistical perspective. *Statistical Science*, 9 (1), 2-54.

- Chesson, J., 1983. The estimation and analyses of preference and its relationships to foraging model. *Ecology*, 64 : 1297-1304.
- Christensen, V. & D. Pauly., 1992. A guide to the ECOPATH II program (version 2.1). ICLARM Software 6, 72 p.
- Cohen P., H. Andriamahefa & J.-G. Wasson, 1998. Towards a regionalization of aquatic habitat: distribution of mesohabitats at the scale of a large basin. *Regulated Rivers*, 14, 391-404.
- Fuhs T., 1998. Les membranes linéaires par morceaux : une approche géométrique de la boucle abduction-induction dans les arbres et listes de décision. *Thèse Université Caen*, 192 p.
- Garnier, J. & G. Billen & M. Coste, 1995. Seasonal succession of diatoms and Chlorophyceae in the drainage network of the Seine River : Observations and modeling. In *Limnol. Oceanogr.*, 40 : 750-765.
- Garnier, J. & P. Servais & A. Barillier & G. Billen, 1993. Modélisation de la dégradation de la matière organique par les bactéries hétérotrophes dans la Seine. In *Rapport de Synthèse – mai 1993, PIREN-Seine*.
- Guégan J.-F., S. Lek & T. Oberdorff, 1998. Energy availability and habitat heterogeneity predict global riverine fish diversity. *Nature*, 391 (22), 382-391.
- Ivol J.-M., J. Belliard & J.-G. Wasson, 1996. Constitution d'un référentiel régionalisé de bioindicateurs des écosystèmes d'eau courante du bassin de la Loire. Rapport intermédiaire Agence de l'Eau Loire Bretagne Cemagref Lyon BEA/LHQ et Paris QE/FEA, janvier 1996, 66 p.
- Ivol-Rigaut J.M., 1998. Hydro-écorégions et variabilité des communautés du macrobenthos sur le bassin de la Loire. Essai de typologie régionale et référentiel faunistique. *Thèse de Doctorat Université Claude Bernard Lyon I*, 271 p.
- Jarre, A. & M.L. Palomares & M.L. Soriano & V.C. Sambilay Jr. & D. Pauly, 1990. A user's manual for MAXIMS. A computer program for estimating the food consumption of fishes from diel stomach contents data and population parameters. ICLARM Software 4, International Center for Living Aquatic Resources Management, Manila, Philippines, 27 p.
- Mastrorillo S., F. Dauba, T. Oberdorff, J.-F. Guégan & S. Lek, 1998. Predicting local fish species richness in the Garonne River basin. *C.R. Acad. Sci. Paris, Sciences de la vie / Life Sciences*, 321, 423-428.
- Mastrorillo S., S. Lek, F. Dauba & A. Belaud, 1997. The use of artificial neural networks to predict the presence of small-bodied fish in a river. *Freshwater Biology*, 38, 237-246.
- Mathews, C.P., 1993. Productivity and energy flows at all trophic levels in the River Thames, England : Mark 2, p. 161-171. In V. Chistensen and D. Pauly (eds.) *Trophic models of aquatic ecosystems*. ICLARM Conf. Proc. 26, 390 p.
- McCulloch W.S. & W. Pitts, 1943. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115-133.
- Oberdorff T., Boët P., D. Chessel, B. Hugueny, D. Pont & J.-P. Porcher, 1999. Mise au point d'un indice Poisson sur l'ensemble du réseau hydrographique national. Rapport final de la phase II du programme national "Indice Poisson", 31 janvier 1999, 60 p. + annexes.
- Oberdorff T., D. Pont, B. Hugueny, P. Boët, J.-P. Porcher & D. Chessel, 2001. Adaptation à l'ensemble du réseau hydrographique d'un indice de qualité écologique fondé sur les peuplements de poissons : résultats actuels et perspectives. p. 95-124, in: J. Lemoalle, F. Bergot & M. Robert (Eds), *État de santé des écosystèmes aquatiques. De nouveaux indicateurs biologiques*, Cemagref, Paris.
- Oberdorff T., D. Pont, B. Hugueny & D. Chessel, 2001. A probabilistic model characterizing fish assemblages of French rivers : a framework for environmental assessment. *Freshwater Biology*, 46, 399-415.
- Odum, W.E. & E.J. Heald, 1975. The detritus-based food web of an estuarine mangrove community, p. 265-286. In L.E. Cronin (ed.) *Estuarine Research*. Vol. 1. Academic Press, N.Y.
- Palomares, M.L. & B. Yulianto & L. Puy & D. Bengen & A. Belaud, 1993. A preliminary model of the Garonne River (Toulouse, France) ecosystem in spring, p. 172-179. In V. Chistensen and D. Pauly (eds.) *Trophic models of aquatic ecosystems*. ICLARM Conf. Proc. 26, 390 p.
- Palomares, M.L., 1991. La consommation de nourriture chez les poissons : Etude comparative, mise au point d'un modèle prédictif et application à l'étude des réseaux trophiques. Thèse de Doctorat de L'Institut National Polytechnique de Toulouse. 211 p.

- Pauly, D. & M. Soriano & M.L. Palomares, 1992. On improving the construction, parametrization and interpretation of steady-states multispecies models. In V. Chistensen and D. Pauly (eds.) *Trophic models of aquatic ecosystems*. ICLARM Conf. Proc. 26, 390 p.
- Polovina, J.J. & M.D. Ow, 1983. *ECOPATH : a user's manual and program listings*. Nat. Mar. Fish. Serv., NOAA, Honolulu Admin. Rep. H-83-23. 46p.
- Polovina, J.J., 1985. An approach to estimating an ecosystem box model. *U.S. Fish Bull.* 83(3) : 457-460.
- 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.
- Poulin M., S. Even, G. Billen, J.-M. Mouchel, J. Garnier, A. Levassor & T. Leviandier, 1998. *Modèles : des processus au bassin versant*. p. 679-720, in: M. Meybeck, G. de Marsily & E. Fustec (Eds), *La Seine en son bassin. Fonctionnement écologique d'un système fluvial anthropisé*, Elsevier, Paris.
- Randall, R.G. & J.R.M. Kelso & C.K. Minns, 1995. Fish production in freshwaters : Are rivers more productive than lakes ? *Can. J. Fish. Aquat. Sci.* 52 : 631-643.
- Reyes-Marchant, P., 1993. Impact des alevins de gardon (*R. rutilus*) sur les communautés planctoniques littorales : modélisation des relations trophiques du lac d'Aydat. Thèse de doctorat de l'Université Blaise Pascal Clermont-Ferrand, France. 185 p.
- Ripley B.D., 1981. *Spatial statistics.*, John Wiley & Sons, New York.
- Ripley B.D., 1996. *Pattern recognition and neural networks.*, Cambridge University Press, Cambridge, 403 p.
- Rumelhart D.E., J.L. McClelland & the PDP Research Group, 1986. *Parallel distributed processing.*, MA:MIT Press/Bradford Books, Cambridge.
- Therneau T.M. & E.J. Atkinson, 1997. An introduction to recursive partitioning using the RPART routines. Technical report Mayo Foundation, September 3, 1997, [Distributed in Poscript with the rpart package], 52 p.
- Ulanovicz, R.E., 1986. *Growth and development : Ecosystem phenomenology*, Springer Verlag, New-York. 203p.
- Venables W. & B.D. Ripley, 1997. *Modern Applied statistics with S-Plus*. (2nd Ed.), Springer-Verlag, New-York, 548 p.
- Wasson J.-G., J. Bethemont, J.-N. Degorce, B. Dupuis & T. Joliveau, 1993. Approche écosystémique du bassin de la Loire : éléments pour l'élaboration des orientations fondamentales de gestion . Phase I : Etat initial - Problématique. Rapport final Cemagref Lyon BEA/LHQ & Univ. Saint Etienne CRENAM, juillet 1993, 102 p + Atlas 70 pl. et annexes p.