

Arbre de régression multivariable : application à une communauté de poissons littoraux d'un lac du Bouclier canadien

Anik Brind'Amour

*Département de sciences biologiques,
Université de Montréal, Pavillon Marie-Victorin,
Case postale 6128, succursale Centre-ville,
Montréal (Québec) Canada, H3C 3J7*

RÉSUMÉ. La combinaison de deux approches statistiques, une analyse de partitionnement et la technique de l'arbre de régression multivariable, a été utilisée afin de développer un modèle d'utilisation d'habitats pour une communauté de poissons d'un lac du Bouclier canadien. Les résultats soulignent l'importance des variables dites associées dans le modèle et l'importance de la taille de l'habitat dans la succession d'autres variables environnementales structurantes et dans la détermination de la composition spécifique de la communauté de poissons du lac à l'étude.

MOTS-CLÉS : Arbre de régression multivariable, communauté piscicole, habitat, lac, partitionnement par la méthode des K centroïdes (K-means).

1 Introduction

La zone littorale représente l'environnement physique le plus hétérogène, diversifié et productif des lacs. Avec sa variété de structures physiques (p. ex. débris de bois, substrats, macrophytes émergentes et submergées) et de ressources alimentaires, la zone littorale se caractérise par une mosaïque de microhabitats [BOI 01]. Leur configuration spatiale et leur diversité confèrent à la zone littorale une place de choix car il s'y produit de nombreuses interactions intra et interspécifiques [WER 77]. Les communautés de poissons sont donc exposées à un environnement complexe structuré à plusieurs échelles spatiales variant du millimètre (p. ex. interstices dans les substrats rocheux) à une centaine de mètres (p. ex. distance entre deux tributaires ou entre lits de macrophytes). Par conséquent, les interactions entre les espèces de poissons et l'environnement ont le potentiel d'être fortement spatialisées. Toutefois, les modèles d'habitat développés en zone littorale incorporent rarement des données se référant à la dimension spatiale de cette dernière. De plus, ces modèles portent généralement sur des descripteurs généraux de la communauté (p. ex. abondance totale) qui ne se soucient guère de l'identité (espèce) des membres de la communauté. Dans ce contexte, le présent travail propose de développer un modèle d'utilisation d'habitats portant sur l'ensemble des espèces d'une communauté de poissons littoraux d'un lac du Bouclier canadien. Ce modèle est développé à l'aide de l'arbre de régression multivariable [DEA 00] à l'échelle des habitats dans le lac à l'étude.

2 Méthodes

2.1 Lac à l'étude

Les données ayant servi aux analyses de ce travail ont été recueillies au lac Paré (46°08' N; 73°54'W), un lac mésotrophe de 31 ha situé dans la région de Lanaudière (Québec). Le lac Paré est caractérisé par

une zone littorale diversifiée (roche, plantes aquatiques, troncs d'arbres morts, etc.) et a une profondeur maximale de 9 m. La zone littorale est peuplée par une communauté de poissons comportant huit espèces représentant cinq familles. La zone littorale du lac Paré a été divisée en 60 sites d'échantillonnage (~ 200 m²) dans lesquels des données biologiques (abondance de poissons) et des variables environnementales ont été estimées.

2.2 Analyse de partitionnement

Un partitionnement des sites d'échantillonnage par la méthode des K centroïdes (*K-means*) a été réalisé afin d'identifier les différents habitats présents dans le lac à l'étude. Le nombre de groupes optimal fut déterminé à l'aide de la statistique de Calinski-Harabasz utilisant une solution minimisant la variation intragroupe, pour $K = 2$ à $K = 10$ groupes, après 100 attributions aléatoires des objets aux groupes initiaux. La confirmation visuelle des groupes de sites similaires a été obtenue à l'aide d'une analyse en composantes principales (ACP) calculée sur la matrice des corrélations entre les variables environnementales. À la suite de cette procédure, les sites contigus appartenant à un même habitat ont été groupés. L'abondance de chaque espèce de poissons dans les sites regroupés a été sommée, puis divisée par la superficie du groupe de sites afin de contrer l'effet additif engendré par la procédure de groupement des sites. La moyenne simple des valeurs des variables environnementales a également été utilisée dans les analyses statistiques. L'analyse de partitionnement et l'ACP ont été réalisées à l'aide du Progiciel R [CAS 05].

2.3 Arbre de régression multivariable

Un modèle d'utilisation d'habitats de poissons a été développé à l'aide de l'approche des arbres de régression multivariable. Cette méthode permet notamment de modéliser les données écologiques caractérisées par i) des relations non linéaires entre les variables réponses (abondance d'espèces) et les variables explicatives (caractéristiques environnementales), ii) une distribution non unimodale des variables réponses et iii) des devis expérimentaux non balancés (c.-à-d., groupes de différentes tailles). Les analyses ont été effectuées à l'aide de la fonction fournie par Ouellette et coll. [OUE en prép.], une version modifiée de la fonction de De'Ath [DEA 00], exécutée dans le langage R [R 04]. La composition relative des espèces à chaque partition de l'arbre a été estimée à l'aide d'une méthode de recherche des espèces indicatrices [DUF 97] qui s'est avérée complémentaire aux arbres de régression. La recherche des espèces indicatrices a été réalisée à chaque noeud de l'arbre de régression élaboré précédemment.

3 Résultats

3.1 Habitats littoraux du lac Paré

Le partitionnement a permis de grouper en quatre types d'habitats (groupes non contigus sur le pourtour du lac) les 60 sites d'échantillonnage du Lac Paré (Fig. 1). Les groupes de sites se distinguent principalement par le degré d'amplitude de la pente, le degré d'exposition au vent dominant (fetch), la densité de macrophytes submergées et émergentes, le type d'utilisation riveraine (chalet ou forêt) et la présence de sable comme substrat. Les groupes de sites sont distribués en taches de différentes tailles variant entre 120 m² et 1755 m², représentant ainsi une mosaïque d'habitats le long du littoral du lac Paré. La superficie totale des quatre habitats se répartit comme suit : Habitat O (56,9 %), Habitat Δ (26,8 %), Habitat \square (10,4 %), Habitat \diamond (5,9 %). Le partitionnement des sites fut confirmé par l'ACP et les trois premières dimensions de l'ordination représentaient 48 % de la variation totale des variables environnementales centrées réduites aux 60 sites.

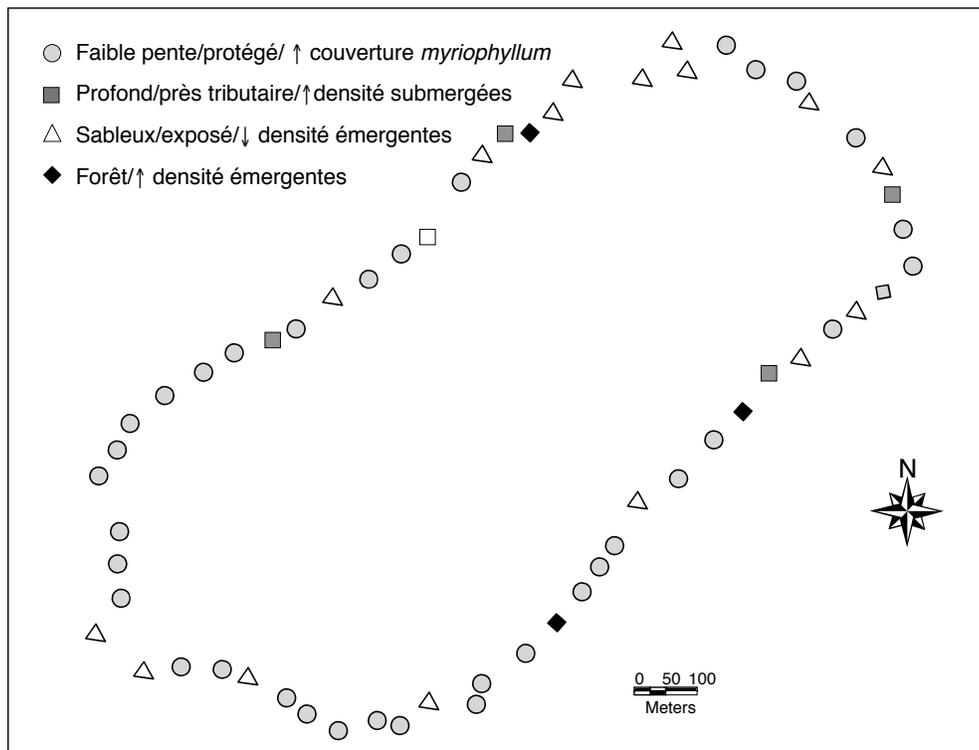


Figure 1: Carte du Lac Paré présentant le partitionnement des sites d'échantillonnage en quatre types d'habitat. La légende fournit les caractéristiques environnementales dominantes pour chaque type d'habitat.

3.2 Modèle multivariable d'utilisation des habitats par les poissons

L'arbre de régression multivariable développé à l'aide des densités de poissons de la communauté littorale du lac Paré est composé de six nœuds et explique 48.3 % de la variation totale de la composition en espèces de la communauté du lac (Fig. 2). Dans cet arbre, les deux premiers nœuds sont associés à des variables environnementales dites « associées », c'est-à-dire estimées à l'extérieur de l'unité d'échantillonnage et qui se réfèrent à la spatialité des habitats. La distance à un tributaire et l'exposition au vent dominant (fetch) sont les deux variables qui contrôlent les premiers nœuds de l'arbre. La taille d'un habitat semble être également un facteur déterminant du modèle puisque cette variable contrôle le troisième nœud qui permet de séparer en deux groupes 85 % (29/34) des taches d'habitat de la zone littorale du lac. L'arbre indique que lorsque la taille d'un habitat est supérieure à 196 m², la présence d'une forêt riveraine et la distance qui sépare cet habitat d'un habitat similaire deviennent alors des facteurs déterminants dans la structure de la communauté autour du lac. La composition relative (%) en espèces des trois partitions subséquentes au nœud associé à la taille de l'habitat est similaire ; ces trois nœuds ne sont pas conservés par validation croisée. La communauté de ces 5 groupes est représentée majoritairement par les cyprinidés (Seat_S, Copl_S, Copl_L) et les perchaudes de grande taille (Pefl_L).

Lorsque la taille de l'habitat est inférieure à 196 m², la complexité de l'habitat, ici associée à la densité de macrophytes émergentes, devient un facteur déterminant de la structure de la communauté de poissons. La communauté est alors majoritairement composée de cyprinidés et des poissons de petite taille. Toutefois, les poissons de petite taille des espèces benthiques adultes (Amne_S et Caco_S) montrent des abondances relatives de 10 % supérieures dans les habitats arborant des densités de macrophytes émergentes inférieures à 6.5 tiges·m⁻² par rapport aux habitats exposant des densités de macrophytes émergentes supérieures à 6.5 tiges·m⁻².

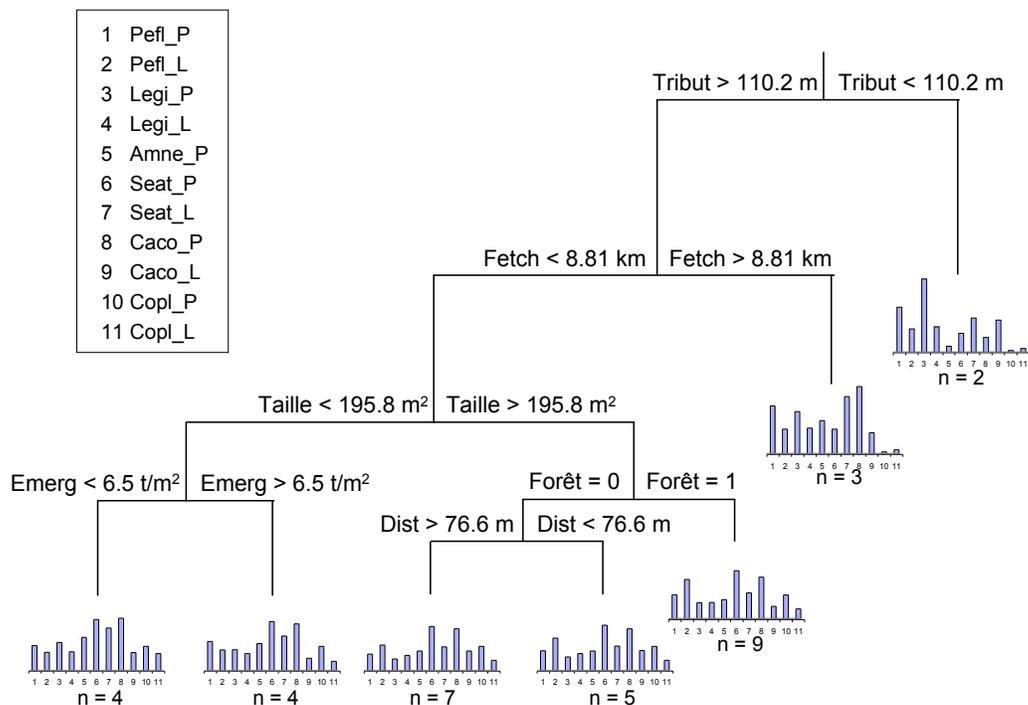


Figure 2 : Arbre de régression multivariable développé pour la communauté de poissons littoraux du lac Paré. P: petite taille; L: grande taille; Pefl: *Perca flavescens*; Legi: *Lepomis gibbosus*; Amne: *Amiurius nebulosus*; Seat: *Semotilus atromaculatus*; Caco: *Catostomus commersoni*; Copl: *Couesius plumbeus*.

4 Discussion

Les approches statistiques basées sur l'étude des communautés d'espèces (c.-à-d., approches multivariées) représentent une avancée de taille dans le développement de modèles d'utilisation des habitats par les poissons. Dans le présent travail, la technique des arbres de régression multivariable a permis de considérer l'identité spécifique des membres d'une communauté dans l'élaboration du modèle d'utilisation d'habitat d'une communauté piscicole et, par conséquent, de définir implicitement les relations fonctionnelles des espèces face aux caractéristiques environnementales des habitats [OLD 04]. Les résultats ont montré l'importance notamment de la taille d'un fragment d'habitat comme facteur déterminant dans la structure des communautés littorales de poissons lacustres.

Parallèlement à l'intérêt des études multivariées des communautés piscicoles, ce travail souligne l'importance d'incorporer des informations portant sur la spatialité des habitats de poissons dans les lacs. Quatre des six nœuds de l'arbre de régression multivariable impliquaient des variables associées. Des conclusions similaires ont été obtenues récemment par Brind'Amour [BRI soumis]. Ces derniers ont mis en évidence l'importance des variables dites associées dans la performance des modèles d'utilisation d'habitat.

Alors que la zone littorale des lacs est de plus en plus perçue comme un paysage composé de multiples habitats de taille et de qualité variables, ce travail montre que la combinaison de méthodes statistiques telles les méthodes d'analyses multivariées offre un cadre spatialisé très intéressant pour améliorer l'étude des relations poisson-habitat dans la zone littorale des lacs.

5 Remerciements

Je remercie P. Legendre pour la révision de ce travail et M.-H. Ouellet pour m'avoir permis d'utiliser sa fonction en langage R. Merci à M. Coinçon et J. Guimmond-Cataford pour leur aide lors de l'échantillonnage au lac Paré. Le support financier a été fourni par D. Boisclair et par les bourses d'études octroyées à A. Brind'Amour par le CRSNG et le FCAR.

6 Bibliographie

- [BOI 01] BOISCLAIR D., "Fish habitat modelling: from conceptual framework to functional tools". *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 58, n° 1, 2001, p. 1-9, NRC Research press, Ottawa.
- [BRI prep] BRIND'AMOUR A., BOISCLAIR D., "The effect of the spatial arrangement of habitats on the development of fish habitat models in the littoral zone of a Laurentian lake", (soumis à CJFAS).
- [CAS 05] CASGRAIN P., LEGENDRE P., *The R Package for multivariate and spatial analysis, Version 4.0 (development release 10) – User's manual*, Département de sciences biologiques, Université de Montréal, available from <<http://www.bio.umontreal.ca/legendre/>>, 2005.
- [DEA 00] DE'ATH G., FABRICIUS K.E., "Classification and regression trees: A powerful yet simple technique for ecological data analysis". *Ecology*, vol. 81, n° 11, 2000, p. 3178-3192, ESA, Ithaca.
- [DUF 97] DUFRÊNE M., LEGENDRE P., "Species assemblages and indicator species: The need for a flexible asymmetrical approach". *Ecological Monographs*, vol. 67, n° 3, 1997, p. 345-366, ESA, Ithaca.
- [OLD 03] OLDEN J.D., "A species-specific approach to modeling biological communities and its potential for conservation". *Conservation Biology*, vol. 17, n° 3, 2003, p. 854-863, Blackwell Publishing, Gainesville.
- [OUE prep] OUELLETTE M.-H., DESGRANGES J.-L., LEGENDRE P., BORCARD D. "Multivariate regression tree analysis: bird assemblages from the Saint Lawrence corridor", (en préparation).
- [R 04] R DEVELOPMENT CORE TEAM, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, 2004.
- [WER 77] WERNER E.E., HALL D. J.D., LAUGHLIN R., WAGNER D.J., WILSMANN L. A., FUNK, F.C. "Habitat partitioning in a freshwater fish community". *Journal of Fisheries and Research Board of Canada*, vol. 34, 1977, p. 360-370, NRC Research press, Ottawa.