



**HAL**  
open science

# Valeurs-seuils pour les paramètres physico-chimiques soutenant la biologie : analyse des réponses des compartiments biologiques phytoplancton, macrophytes et ichtyofaune. Rapport d'avancement.

V. Roubex, P.A. Danis

## ► To cite this version:

V. Roubex, P.A. Danis. Valeurs-seuils pour les paramètres physico-chimiques soutenant la biologie : analyse des réponses des compartiments biologiques phytoplancton, macrophytes et ichtyofaune. Rapport d'avancement.. irstea. 2014, pp.44. hal-02601324

**HAL Id: hal-02601324**

**<https://hal.inrae.fr/hal-02601324>**


Submitted on 16 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Partenariat 2014

*Paramètres physicochimiques généraux soutenant la biologie en plans d'eau Action 15*



# **Valeurs-seuils pour les paramètres physico-chimiques soutenant la biologie : analyse des réponses des compartiments biologiques phytoplancton, macrophyte & ichtyofaune.**

***Rapport d'avancement***

**Vincent Roubex (Irstea)  
&  
Pierre-Alain Danis (Onema)**

**Relecteurs :  
Jean-Marc Baudoin (Onema)  
Christine Argillier (Irstea)**

***Décembre 2014***

## Contexte de programmation et de réalisation

En accord avec les termes de la DCE, la définition des règles d'évaluation de l'état écologique des plans d'eau implique l'établissement des valeurs-seuils correspondant aux limites de classe d'état pour les 6 paramètres physico-chimiques généraux soutenant la biologie : la concentration en nutriments, la transparence, le bilan d'oxygène, la température de l'eau, la salinité et la capacité de neutralisation des acides.

Le présent rapport présente les avancées méthodologiques réalisées en 2014 sur l'analyse des réponses des compartiments biologiques aux éléments de qualité physico-chimiques et aux caractéristiques environnementales pour, à terme, aboutir à une définition pleinement DCE compatible des valeurs-seuils des paramètres physico-chimiques soutenant la biologie. Cela implique notamment la définition de seuils permettant d'assurer le bon fonctionnement des écosystèmes et la préservation ou l'atteinte du bon état des communautés biologiques.

## Les auteurs

### **Vincent Roubéix**

*Ingénieur de recherche (Vincent.Roubéix@irstea.fr)*

*Irstea, UR HYAX, Pôle études et recherches Onema-Irstea Hydroécologie des Plans d'Eau  
13275 route de Cézanne – CS 40061, 13182 Aix-en-Provence Cedex 5*

### **Pierre-Alain Danis**

*Ingénieur Onema (Pierre-Alain.Danis@onema.fr)*

*Onema, Pôle études et recherches Onema-Irstea Hydroécologie des Plans d'Eau  
13275 route de Cézanne – CS 40061, 13182 Aix-en-Provence Cedex 5*

## Les correspondants

### **Onema :**

**Yorick Reyjol**, DAST, direction générale de l'Onema, 94300 Vincennes, *Yorick.Reyjol@onema.fr*

### **Irstea :**

**Christine Argillier**, Irstea, UR HYAX, Pôle études et recherches Onema-Irstea Hydroécologie des Plans d'Eau, 13182 Aix-en-Provence, *Christine.Argillier@irstea.fr*

**Référence du document** : Roubéix Vincent & Pierre-Alain Danis (2014) : Valeurs-seuils pour les paramètres physico-chimiques soutenant la biologie : analyse des réponses des compartiments biologiques phytoplancton, macrophyte & ichtyofaune. Rapport d'avancement. Convention Onema/Irstea 2014. Pp 44.

<b>Droits d'usage :</b>	<i>Accès libre</i>
<b>Couverture géographique :</b>	<i>France</i>
<b>Niveau géographique :</b>	<i>National</i>
<b>Niveau de lecture :</b>	<i>Professionnels, Experts</i>
<b>Nature de la ressource :</b>	<i>Document PDF</i>

**Valeurs-seuils pour les paramètres physico-chimiques soutenant la biologie**  
*Vincent Roubeix & Pierre-Alain Danis*

1. Introduction.....	4
2. Méthodologie d'analyse des réponses des compartiments biologiques .....	6
2.1. La notion de seuil écologique .....	6
2.2. Les méthodes de détection de seuils écologiques .....	9
2.3. La méthode de Gradient Forest .....	13
2.4. Les données.....	17
2.4.1. Les paramètres physico-chimiques .....	17
2.4.2. Les caractéristiques des compartiments biologiques .....	18
2.4.3. Les caractéristiques environnementales.....	19
2.5. Mise en œuvre du Gradient Forest .....	19
3. Résultats et discussion .....	20
3.1. Performances des forêts aléatoires .....	20
3.2. Importances des variables pour les communautés.....	20
3.3. Seuils identifiés.....	22
3.4. Comparaison des seuils entre compartiments.....	26
3.5. Réponses d'espèces individuelles .....	34
4. Conclusions et perspectives .....	41
4.1. Conclusions.....	41
4.2. Perspectives.....	41
5. Références.....	42

**Valeurs-seuils pour les paramètres physico-chimiques soutenant la biologie**  
Vincent Roubeix & Pierre-Alain Danis

## 1. Introduction

La définition des règles d'évaluation de l'état écologique des plans d'eau implique l'établissement des valeurs-seuils correspondant aux limites de classe d'état pour les 6 paramètres physico-chimiques généraux soutenant la biologie que sont la transparence, la température de l'eau, le bilan d'oxygène, la salinité, la capacité à neutraliser les acides et la concentration en nutriments.

Originellement, la définition de valeurs-seuils était basée sur la construction d'indicateurs physico-chimiques grâce à une modélisation statistique des variabilités inter-plans d'eau. Sur la base de variables environnementales (Danis et Argillier 2011) et d'indicateurs des niveaux de pressions anthropiques (forces motrices) (Danis *et al.* 2012), cette modélisation n'a pas abouti à des résultats robustes, en particulier pour les nutriments, du fait d'un petit nombre de données supérieures aux limites de quantification.

En 2012, l'augmentation du jeu de données disponibles a permis des analyses plus poussées de modélisation de type "Hindcasting" (Danis et Argillier 2013). Malheureusement, malgré le fait que cette méthode soit couramment utilisée dans le développement des indicateurs biologiques, elle n'a pas permis, dans le cas de la physico-chimie, d'améliorer les capacités prédictives des modèles.

En outre, ce type d'approches ne répondait pas pleinement aux exigences de la DCE qui impose la définition de seuils basés sur la réponse des compartiments biologiques aux altérations physico-chimiques, et en particulier des seuils de bon état physico-chimique permettant d'assurer le bon fonctionnement des écosystèmes et la préservation/atteinte de communautés biologiques (poissons, invertébrés, flore aquatique) en bon état écologique.

En 2013, les recherches des valeurs-seuils des paramètres physico-chimiques se sont donc orientées vers un calage vis-à-vis des valeurs-seuils des éléments de qualité biologiques (EQB). Cette méthodologie est largement utilisée par d'autres états membres et ceci principalement sur la base d'indicateurs phytoplancton fondés sur les variations de la concentration en chlorophylle  $\alpha$ . Les premiers travaux ont débuté mi-2013 pour répondre aux exigences réglementaires de la mise à jour programmée de l'arrêté évaluation et ont abouti à des valeurs-seuils pertinentes pour les métriques de l'indicateur phytoplancton français (Feret et Laplace-Treytore 2013). De tels résultats n'ont pas pu être atteints pour les compartiments macrophytes et ichtyofaune car aucune réponse significative n'a pu être mise en évidence pour ce calage des valeurs-seuils physico-chimiques (voir détails dans Danis et Roubeix 2014). Par ailleurs, cette stratégie de calage des valeurs-seuils physico-chimiques sur les valeurs-seuils des EQB est critiquable en plusieurs points. Tout d'abord, les valeurs-seuils des EQB sont pour la plupart basées sur un ou plusieurs critères statistiques (e.g. percentile des EQB

parmi les plans d'eau de référence) et/ou géométriques (e.g. divisions de la gamme des valeurs en 5 classes de tailles égales), et non pas sur l'identification de distorsions/perturbations écologiques effectives particulières des caractéristiques (abondances/composition) des espèces en réponse à des variations des concentrations des différents paramètres physico-chimiques. La limite TRES BON / BON (respectivement BON / MOYEN) devant pourtant être basée sur une distinction légère (respectivement modérée) des communautés entre les deux classes (cf Tableau 1.2 dans la DCE).

Pour ces raisons, en 2014, la stratégie suivie a été celle de l'analyse des effets écologiques des paramètres physico-chimiques sur les caractéristiques élémentaires (e.g. abondance ou présence/absence par espèce) des communautés biologiques sans passer par les informations trop synthétiques des indicateurs biologiques utilisés aujourd'hui pour l'évaluation de l'état écologique des plans d'eau, i.e. l'indice phytoplancton, IPLAC (Feret et Laplace-Treytore 2013), l'indice ichtyofaune, IIL (Argillier *et al.* 2013), l'indice macrophytes, IBML (Boutry *et al.* 2013). L'objectif est d'étudier dans leur globalité la réponse des communautés aux paramètres physico-chimiques soutenant la biologie, en tenant compte simultanément de toutes les espèces et des interactions entre les variables abiotiques. Ce travail se focalise sur la recherche de seuils physico-chimiques de part et d'autre desquels les abondances de certaines espèces sont considérablement différentes, marquant ainsi des modifications potentielles dans le fonctionnement des écosystèmes<sup>1</sup> et une déviation importante de la structure de référence des communautés biologiques. Ce travail a été réalisé par compartiment biologique (§2.4.2) et les résultats seront comparés par paramètre physico-chimique entre les compartiments biologiques (§3.3 et 3.4) et discutés en fonction des réponses de certaines espèces (§0). Cette approche permet de fournir des valeurs-seuils physico-chimiques indépendantes du processus de construction des indicateurs biologiques (et du risque de circularité des informations apportées) et de répondre aux exigences de la DCE vis à vis de l'identification de modifications dans le fonctionnement des écosystèmes et dans les caractéristiques des communautés biologiques. Par ailleurs, ce travail pourrait permettre d'identifier des pistes d'amélioration des indicateurs biologiques et/ou de leurs valeurs-seuils.

Le présent rapport détaille principalement :

- les méthodologies statistiques d'identification de seuils qui permettent d'identifier des ruptures inter-plans d'eau dans les abondances d'espèces biologiques expliquées par des gradients de paramètres physico-chimiques ou de caractéristiques environnementales (§2.2) ;
- la méthodologie statistique utilisée pour cette étude, méthode dite du "*Gradient Forest*" (§2.3) ;
- les données physico-chimiques, biologiques et environnementales utilisées pour ces analyses (§2.4) ;

---

<sup>1</sup> Remarque : sous réserve que les différentes espèces ne soient pas fonctionnellement très proches. Des travaux sont en cours sur ce sujet (cf. thèse d'Alban Sagouis à Irstea Aix-en-Provence en partenariat avec le LISC Irstea Clermont-Ferrand).

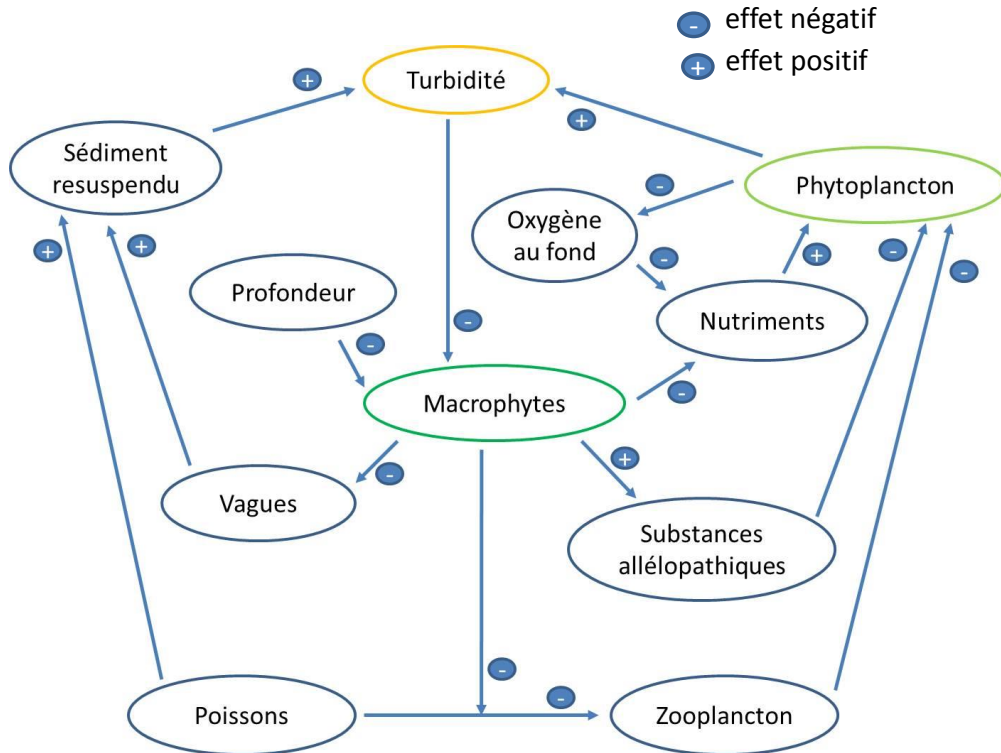
- les valeurs-seuils physico-chimiques ou environnementales identifiées par l'analyse et comparées selon les compartiments biologiques (§3.4) et ceci plus particulièrement pour la transparence, les nitrates, le phosphore total, l'ammonium, et des critères géographiques;
- les conclusions en terme de proposition de valeurs-seuils (§4.1) ainsi que les perspectives de recherche prévues pour 2015 sur la précision des valeurs-seuils selon les conditions de référence et les niveaux de pressions anthropiques (§4.2).

Une réflexion sur les incertitudes associées aux indicateurs physico-chimiques actuels et à la méthodologie présentée ici a été rapportée dans le document suivant : Pierre-Alain Danis & Roubeix Vincent (2015) : Incertitudes associées aux indicateurs physico-chimiques en plans d'eau. Rapport d'avancement. Convention Onema/Irstea 2014. Pp 17.

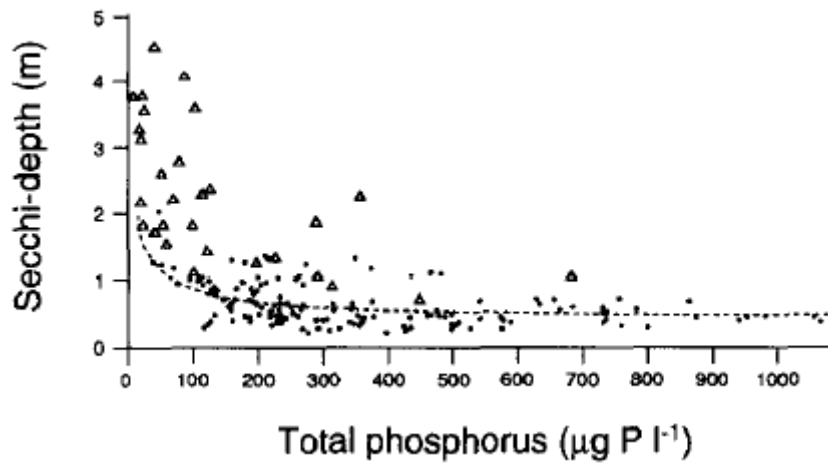
## **2. Méthodologie d'analyse des réponses des compartiments biologiques**

### **2.1. La notion de seuil écologique**

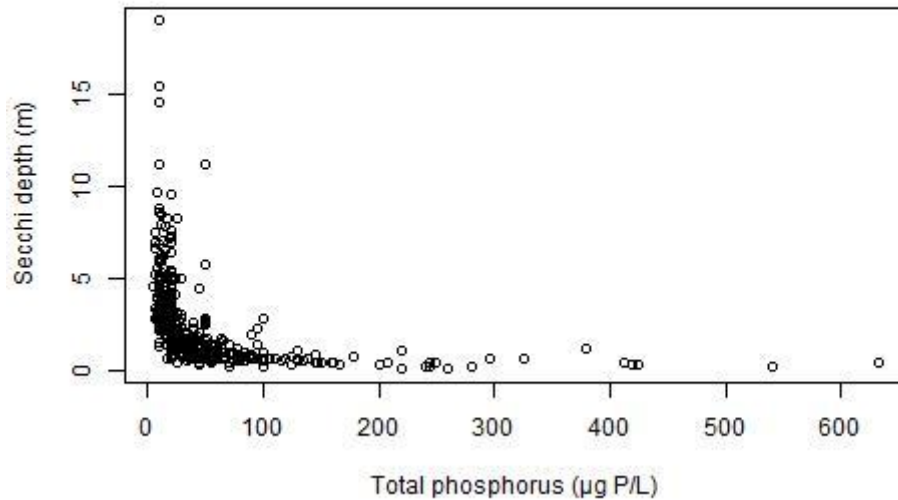
Un seuil écologique est une zone dans un gradient où de petits changements de la variable environnementale considérée entraînent des réponses importantes des caractéristiques (e.g. abondance de certaines espèces) de l'écosystème (Groffman *et al.* 2006). La notion de seuil écologique est associée à la théorie des états stables alternatifs (May 1977) qui dit qu'un système peut basculer d'un état à un autre lorsque une variable environnementale franchit un seuil écologique. Un exemple bien connu est celui de lacs peu profonds passant d'un état d'eau claire, avec beaucoup de macrophytes submergés, à un état d'eau turbide où le phytoplancton est abondant et les macrophytes sont rares (Scheffer *et al.* 1993). La stabilité de ces deux états extrêmes s'explique par des phénomènes d'amplification dus à des boucles de rétroactions positives, impliquant la compétition entre macrophytes et phytoplancton pour les nutriments, la remise en suspension de sédiments, la sécrétion par les macrophytes de substances allélopathiques et la prédation du zooplancton par les poissons (Fig1). Le changement rapide d'état peut s'expliquer par des relations non-linéaires entre certaines variables du système. Ainsi dans le cas de l'enrichissement progressif d'un lac peu profond oligotrophe, à partir d'un certain seuil de phosphore, la transparence de l'eau chute (Fig2), les macrophytes submergés sont alors inhibés par le manque de lumière et le lac passe dans un état turbide.



**Figure 1.** Interactions dans un écosystème de lac peu profond expliquant la stabilité des états 'turbides' et 'riches en macrophytes' (d'après Scheffer et al 1993).

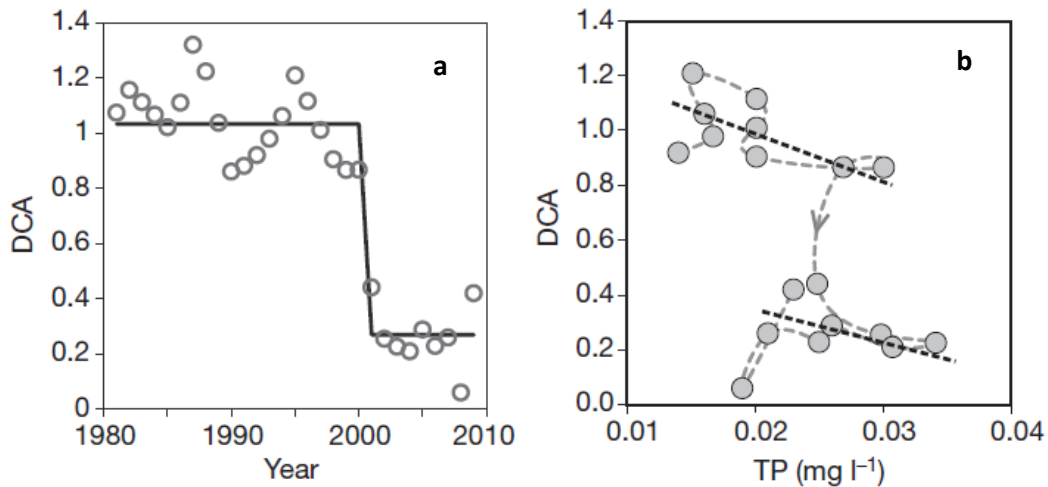






**Figure 2.** Relation entre la concentration en phosphore dans des lacs et la transparence de l'eau d'après Scheffer et al 1993 (en haut) et d'après les données en zone euphotique de suivi DCE des plans d'eau nationaux (en bas).

Cet exemple montre que le franchissement d'un seuil écologique a des effets importants sur la biomasse des communautés aquatiques (notamment macrophytes, phytoplancton et poissons). Mais de manière plus qualitative, les changements rapides dans l'écosystème peuvent aussi s'accompagner de remaniements dans la composition des communautés. Une étude paléolimnologique récente Wang *et al.* (2012) a montré que les communautés de diatomées d'un lac chinois subissant une eutrophisation progressive, ont changé brutalement alors que la concentration en phosphore s'accroissait de façon graduelle (Fig3). Il est donc possible de repérer un seuil écologique en observant les variations d'abondances des espèces.



**Figure 3.** Evolution de la composition de la communauté de diatomées dans un lac de Chine, et lien avec la concentration en phosphore total (d'après Wang et al 2012). Les variations de composition de la communauté sont représentées par l'axe d'une analyse multivariée (DCA). La trajectoire des variations de la concentration en phosphore (ligne courbe discontinue, b) indique un phénomène d'hystérésis impliquant l'existence de deux seuils.

Note : Le phénomène d'hystérésis implique qu'il y ait un seuil lorsque le système se dégrade et qu'il existe un seuil distinct lorsque le système est en voie de restauration (Andersen *et al.* 2009). Si l'on suppose que les plans d'eau étudiés sont dans une même trajectoire de dégradation, alors les seuils écologiques identifiés correspondent à des seuils de protection des systèmes. Ces seuils ne sont pas applicables pour fixer des objectifs de restauration si on suspecte un phénomène d'hystérésis. De plus, si l'on considère ensemble des plans d'eau en voie de dégradation et d'autres en voie de restauration, l'hypothèse d'hystérésis doit mener à la détection d'au moins deux seuils ou de seuils diffus.

## 2.2. Les méthodes de détection de seuils écologiques

La détection de seuils écologiques revient à rechercher des sauts ou ruptures dans des relations non-linéaires entre des variables biologiques et des variables environnementales. L'analyse peut se faire dans le temps ou dans l'espace. Un seuil écologique peut être identifié à partir de données temporelles concernant un site particulier qui connaît une dégradation progressive de son état écologique (par ex. Fig 3). Néanmoins, lorsqu'un changement soudain de l'écosystème est observé dans le temps, il convient de vérifier qu'il est apparu suite à des variations graduelles de facteurs environnementaux. En effet, un changement brutal dans le temps d'un facteur environnemental peut générer une réponse importante de l'écosystème sans qu'il existe nécessairement un seuil écologique (Andersen *et al.* 2009).

Plus communément dans la littérature, des seuils écologiques sont recherchés dans des gradients spatiaux à partir de données concernant un grand nombre de sites : le temps est remplacé par l'espace. Les données biologiques et environnementales des différents sites sont mises en relations

pour mettre en évidence d'éventuels points de changement importants pouvant correspondre à des seuils écologiques. Diverses méthodes existent pour la détection de seuils dans des gradients spatiaux (Brenden *et al.* 2008, Dodds *et al.* 2010, Qian 2014). Elles peuvent être classées par leur caractère plus ou moins exploratoire selon la connaissance acquise préalablement sur les relations biote-pressions. Lorsque les relations sont bien connues, l'analyse peut se limiter à l'examen de la réponse d'une variable biologique à un facteur environnemental. L'analyse peut se complexifier lorsque différentes variables environnementales et/ou biologiques sont prises en compte simultanément, afin de rechercher les meilleurs bioindicateurs de seuils ainsi que les variables environnementales les plus susceptibles de déclencher un changement d'état de l'écosystème alors même que ces variables peuvent agir en interaction. Le tableau 1 présente les principales méthodes utilisées pour la détection de seuils écologiques dans des gradients environnementaux.

**Tableau 1.** Principales méthodes de détection de seuils classées selon le nombre de variables considérées sur les plans de la biologie et de l'environnement.

		Description environnementale	
		1 paramètre	Multi-paramètres
Réponse biologique	Univariée	<ul style="list-style-type: none"> <li>- Régression linéaire par morceaux</li> <li>- Analyse de courbes de fréquences cumulées (Utz <i>et al.</i> 2009)</li> <li>- Analyse de variance non paramétrique (Qian <i>et al.</i> 2003)</li> <li>- SiZer : Significant Zero crossings (Sonderregger <i>et al.</i> 2009)</li> </ul>	<ul style="list-style-type: none"> <li>- Arbres de régression ou de classification (De'ath et Fabricius 2000)</li> <li>- Arbres de régression par quantiles (Kail <i>et al.</i> 2012)</li> </ul>
	Multivariée (nombreux taxons)	<ul style="list-style-type: none"> <li>- Analyse bayésienne sur données multinomiales (Qian <i>et al.</i> 2004)</li> <li>- TITAN (Threshold Indicator Taxa Analysis) (Baker et King 2010)</li> </ul>	<ul style="list-style-type: none"> <li>- Arbres de régression multivariés (De'Ath 2002)</li> <li>- Classification des communautés, analyse discriminante et comparaison des moyennes par cluster (Catalan <i>et al.</i> 2009, Smith et Tran 2010)</li> <li>- Gradient Forest (Ellis <i>et al.</i> 2012)</li> </ul>

Parmi ces méthodes, l'analyse de la variance non paramétrique est à la fois une méthode en soi et aussi la base d'autres méthodes plus complexes dont celle utilisée pour cette étude (cf plus bas). Elle permet de localiser un seuil dans la relation entre une variable biologique et une variable environnementale. Elle est basée sur la minimisation de la somme des carrés des écarts de la variable de réponse dans les deux groupes formés par le seuil (Fig 4). De nombreuses études de seuils écologiques l'ont utilisée dans des contextes variés (Chambers *et al.* 2012, Evans-White *et al.* 2009, Holt *et al.* 2003, Smith et Tran 2010, Soranno *et al.* 2008). Cette méthode constitue aussi le critère de construction des arbres de régression lorsque plusieurs facteurs sont considérés en même temps. Les arbres de régression permettent d'évaluer simultanément les effets de nombreuses variables environnementales en prenant en compte leurs interactions.

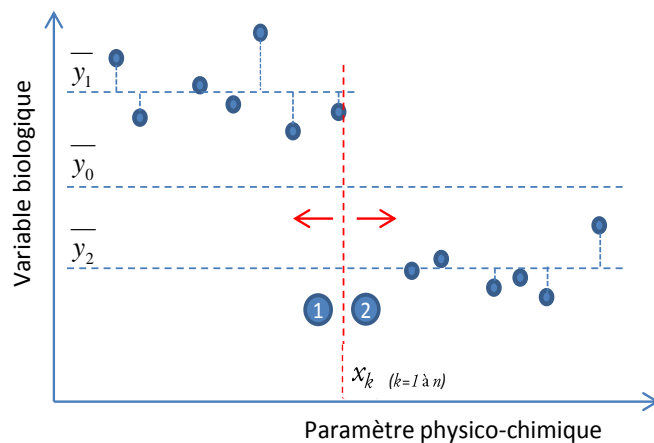
$$SC_{intra} = SC_1 + SC_2$$

$$SC_{intra} = \sum_{i=1}^k (y_i - \bar{y}_1)^2 + \sum_{i=k+1}^n (y_i - \bar{y}_2)^2$$

$$SC_{inter} = SC_{tot} - SC_{intra}$$

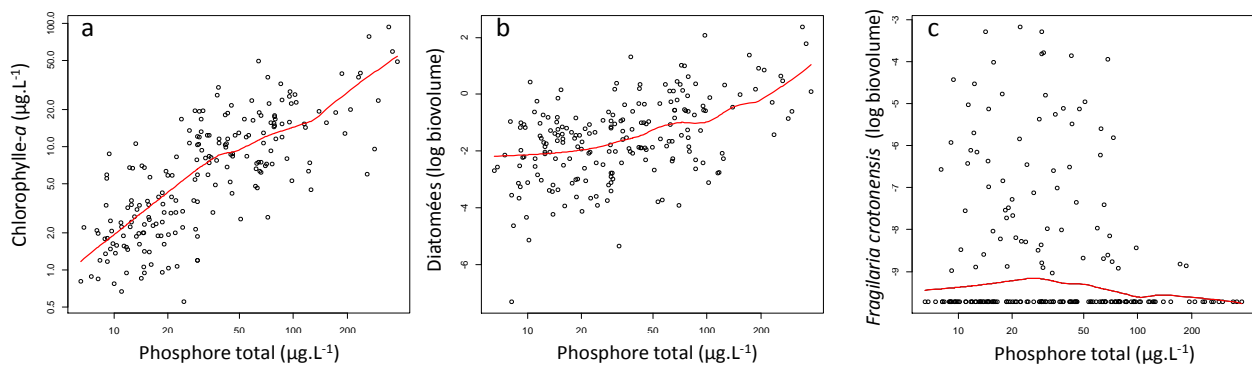
$$SC_{tot} = \sum_{i=1}^n (y_i - \bar{y}_0)^2$$

$$SC_{inter} = k \cdot (\bar{y}_1 - \bar{y}_0)^2 + (n-k) \cdot (\bar{y}_2 - \bar{y}_0)^2$$



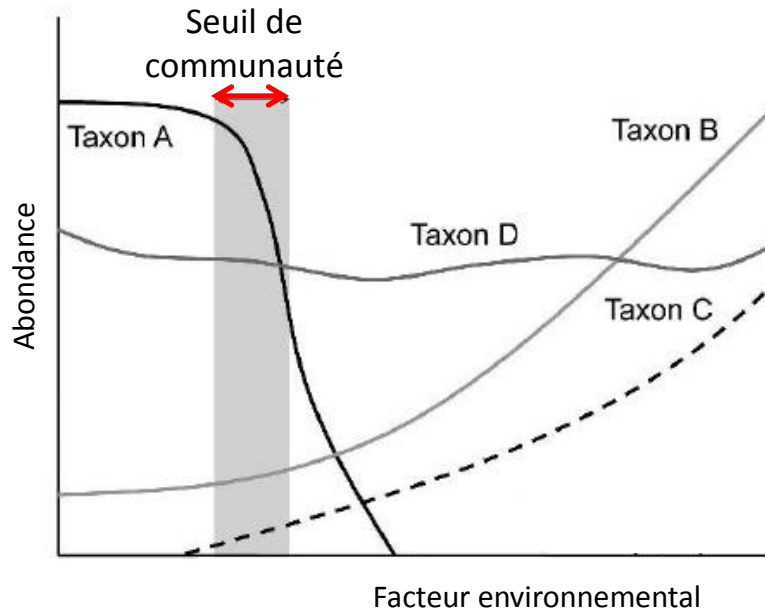
**Figure 4.** Mode de détermination d'un seuil par la méthode d'analyse de la variance non paramétrique. La méthode consiste à rechercher la position du seuil dans le gradient ( $x_k$ ) qui sépare les groupes 1 et 2, et pour laquelle la somme des sommes des carrés intragroupes ( $SC_{intra}$ ) est minimale, ce qui équivaut à la maximisation de la somme des carrés intergroupes ( $SC_{inter}$ ).

La réponse de la biologie aux gradients peut être appréhendée par une seule variable, qui peut être l'abondance d'une espèce particulière ou d'un groupe fonctionnel, ou n'importe quelle métrique jugée pertinente par rapport au facteur considéré. Il faut noter que l'agrégation de données de communautés dans des variables intégratives tend à linéariser les relations aux variables environnementales ce qui rend la détection d'éventuels seuils plus difficile (King et Baker 2010, Utz *et al.* 2009) (Fig 5). Pour ne pas perdre d'information, il est préférable d'analyser les données biologiques à l'espèce lorsque la résolution taxonomique des données le permet. Lorsque l'étude porte sur un vaste territoire, il est alors nécessaire d'inclure des variables spatiales (par ex. longitude et latitude) dans l'analyse afin de tenir compte des limites des aires de distributions géographiques des espèces.



**Figure 5.** Effet de l'agrégation de variables biologiques sur la détection de seuil écologiques. Lorsque la biomasse totale du phytoplancton est considérée (a), la relation au phosphore apparaît linéaire. Si l'on ne considère que les diatomées du phytoplancton (b), la relation est moins linéaire avec une phase de relative indépendance en début de gradient. En ne regardant l'abondance que d'une seule espèce de diatomée (c), la relation n'est pas linéaire avec un seuil apparent entre 50 et 100  $\mu\text{g.L}^{-1}$ .

Une autre approche consiste à analyser les réponses aux gradients environnementaux de toutes les espèces inventoriées dans une étude. Elle permet de capter un maximum d'information sur les communautés et n'impose aucune sélection *a priori* des variables biologiques. Les seuils écologiques correspondent alors aux zones dans le gradient où coïncident les variations importantes des abondances de plusieurs espèces : on parle de seuil de communauté (Baker et King 2010) (Fig 6). Les réponses similaires de différentes espèces peuvent s'expliquer par leur coévolution dans une gamme de variations naturelles d'un facteur environnemental. Les espèces sont alors affectées simultanément dès lors que le facteur environnemental considéré sort de cet intervalle suite à des pressions anthropiques sur l'écosystème. Le remaniement des communautés peut aussi être amplifié au franchissement d'un seuil par les interactions biotiques entre espèces (associations, antagonismes, relations à des prédateurs ou des parasites...).



**Figure 6.** Graphe conceptuel illustrant la notion de seuil de communauté avec quatre taxons (d'après King et Baker 2010). Au niveau du seuil de communauté, l'abondance du taxon A baisse brutalement, le taxon B commence à augmenter en abondance, le taxon C apparaît dans le milieu et se développe alors que le taxon D garde la même abondance tout au long du gradient (le taxon est dit "indifférent" au facteur considéré).

### 2.3. La méthode de Gradient Forest

Pour analyser les réponses des compartiments biologiques (i.e. phytoplancton, macrophytes et ichtyofaune) des plans d'eau aux gradients physico-chimiques, et identifier des seuils écologiques potentiels, une méthode statistique récente, appelée 'Gradient Forest', a été utilisée (Ellis *et al.* 2012). C'est une méthode exploratoire qui prend en compte simultanément l'ensemble des espèces inventoriées d'un compartiment et un grand nombre de paramètres environnementaux, naturels ou sous influence anthropique, afin de déceler des seuils de communauté dans les gradients. Elle peut s'appliquer successivement aux trois communautés considérées dans cette étude (i.e. poissons, macrophytes, phytoplancton) et nécessite très peu de connaissances *a priori* sur les relations espèces/environnement.

Son principe de base est celui des arbres de régression ou des arbres de classification, c'est-à-dire le partitionnement récursif des valeurs d'une variable de réponse en fonction des variables explicatives incluses dans l'analyse. Les valeurs de la variable de réponse sont successivement divisées en deux groupes selon la valeur de coupure de la variable explicative qui génère les groupes les plus homogènes. Le critère d'homogénéité est celui de la somme des carrés des écarts dans le cas d'une variable de réponse quantitative (régression) (Fig.4) et celui de l'indice de Gini ou de Shannon pour une variable qualitative (présence/absence par ex.)(De'ath et Fabricius 2000). Les groupes formés (les nœuds) sont reliés entre eux par des branches, chaque bipartition rajoutant deux branches, on

obtient au final un arbre. Les valeurs de coupures peuvent être interprétées comme des seuils dont l'importance peut être mesurée par le gain d'homogénéité (ou baisse d'impureté) issu de la coupure.

La distribution de chaque espèce d'un type de communauté pourrait ainsi être modélisée par un arbre. Or la méthode de *Gradient Forest* ne se limite pas à la construction d'un seul arbre par espèce, mais génère une forêt aléatoire pour chaque espèce ('randomforest') (Breiman 2001). Une forêt aléatoire est composée d'un grand nombre d'arbres construits chacun à partir d'une sélection aléatoire des valeurs de la variable de réponse et en examinant pour chaque partitionnement, les effets d'une sélection aléatoire de variables explicatives. Les relations qui lient la variable de réponse aux variables explicatives sont ainsi explorées plus en profondeur et les prédictions d'une forêt, qui correspond à la moyenne des prédictions de tous les arbres qui la composent, sont généralement plus performantes que celles d'un seul arbre. La performance d'une forêt aléatoire ( $R^2$ ) est évaluée en considérant les échantillons de validation qui constituent les observations ayant été exclues aléatoirement pour la construction de chacun des arbres.

Pour des données quantitatives (abondance, biomasse) :

$$R^2 = 1 - \frac{\sum_i (Y_i - \hat{Y}_i)^2}{\sum_i (Y_i - \bar{Y})^2}$$

où  $Y_i$  correspond à l'observation  $i$  de la variable expliquée,  $\hat{Y}_i$  est la moyenne des prédictions sur tous les arbres pour lesquels l'observation  $i$  se trouve dans l'échantillon de validation et  $\bar{Y}$  est la moyenne générale des observations.

Pour des données binaires (présence/absence) :

$$R^2 = 1 - \frac{ErrMod}{ErrBase}$$

où *ErrMod* correspond au taux moyen d'erreur de classification des arbres dans les échantillons de validation, et *ErrBase* est le taux d'erreur de base :  $ErrBase = 2p(1 - p)$  où  $p$  est la proportion des sites où l'espèce considérée est présente.

L'importance des variables dans une forêt aléatoire correspond à la baisse relative de  $R^2$  lorsque les valeurs de cette variable sont permutées au hasard entre les observations. Toutefois cette méthode d'estimation de l'importance des variables présente un biais en faveur des variables corrélées. Une méthode de permutation conditionnelle proposée par Strobl *et al.* (2008) a été appliquée pour mieux appréhender le véritable impact de chaque variable. Cette approche consiste à permuter les valeurs de la variable  $X_i$  uniquement à l'intérieur de partitions des données réalisées sur la base de variables  $X_j$  corrélées à  $X_i$ .

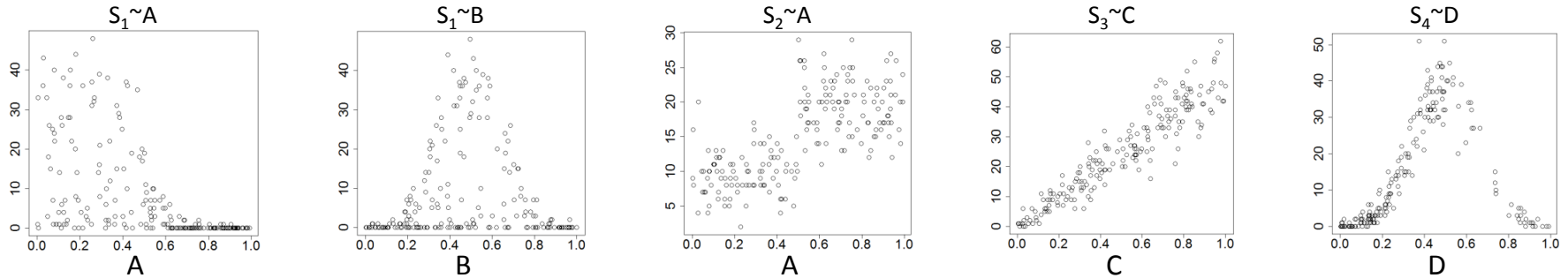
Le *Gradient Forest* réalise des forêts aléatoires pour chaque espèce appartenant à un même compartiment et agrège les résultats de toutes les forêts pour en tirer des informations sur les communautés. Ainsi les importances des variables pour les espèces sont moyennées pour obtenir une importance globale pour les communautés. Aussi, les coupures et leurs importances provenant de chaque arbre de chaque forêt sont synthétisées le long des gradients environnementaux sous

forme de courbes de densités. La densité des coupures (courbe noire, partie 4-a de la Figure 7) est rapportée à celle des données (courbe rouge, partie 4-a de la Figure 7) pour corriger le biais créé par la distribution non uniforme des valeurs dans le gradient environnemental. Les pics de la courbe de ratio des densités (courbe bleue, partie 4-a de la Figure 7) indiquent la localisation probable de seuils de communauté (Fig. 7).

L'application de la méthode de *Gradient Forest* appliquée au compartiment phytoplancton a fait l'objet d'une publication soumise en 2014 (Roubeix *et al.* Soumis).

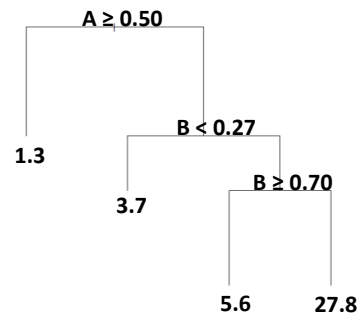


# 1) Relations entre les abondances d'espèces (S1, S2, S3, S4) et des variables environnementales (A,B,C,D,E)



## 2) Arbres de régression par espèce

➤ Ex :  $S_1 = f(A, B, C, D, E)$



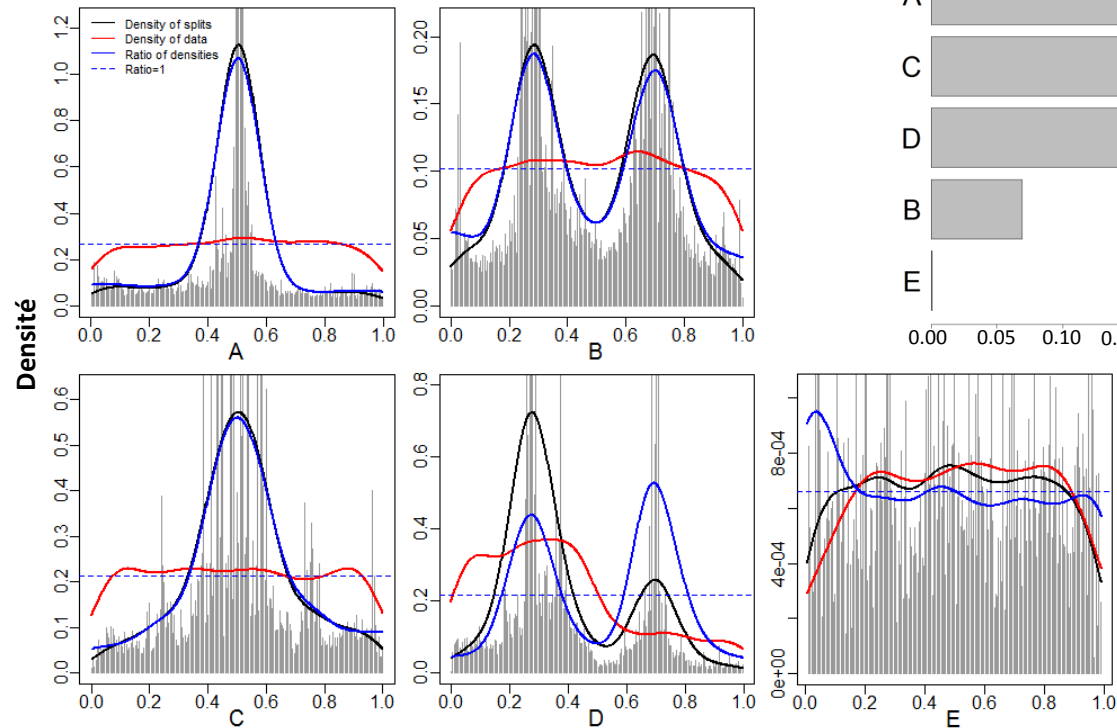
## 3) × n arbres => forêt aléatoire

- Performance de prédiction de chaque espèce  $R^2$
- Importance des variables pour chaque espèce

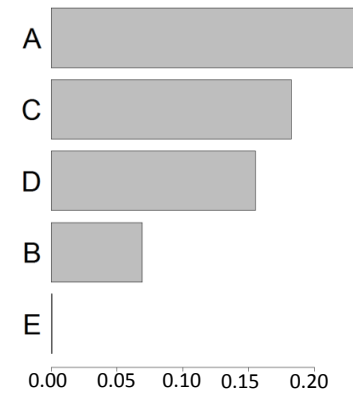
## 4) × p espèces => gradient forest

➤ Agrégation des résultats des  $p$  forêts aléatoires

### 4-a) Densité des coupures dans les gradients (tirées de $n \times p$ arbres)



### 4-b) Importances globales des variables



**Figure 7. (page précédente)** Exemple de 'gradient forest' appliqué à des données simulées de cinq variables environnementales et d'abondances de quatre espèces. (1) Relations simulées pour cinq couples espèce-variable, suivant des fonctions logistiques, gaussiennes, en marche d'escalier ou linéaires. Les autres couples non représentés sont caractérisés par une indépendance espèce/variable. (2) Exemple d'arbre de régression pour une espèce. Les valeurs de coupure et les variables correspondantes sont indiquées au niveau des nœuds. Les feuilles indiquent les valeurs moyennes d'abondance dans les groupes. (3) Pour chaque espèce,  $n$  arbres sont construits, constituant une forêt à partir de laquelle un  $R^2$  et les importances des variables sont déterminés pour l'espèce. (4) Résultats du gradient forest : (4-a) densité corrigée des coupures dans les gradients (courbe bleue) et (4-b) importances des variables pour les communautés. On peut identifier un seuil dans le gradient A et deux dans le gradient B. La variable E ne montre aucun seuil (indépendance). Il y a deux seuils d'importances similaires dans le gradient D après correction par la distribution des données et un seuil non pertinent dans le gradient C. Remarque : la méthode identifie artificiellement des seuils même lorsque les relations sont linéaires. En effet, pour la variable C, le seuil bien marqué à la valeur de 0,5 (cf 4-a) est un artefact car la dépendance de S3 à C est linéaire (cf 1) et les autres espèces sont indépendantes de C.

## 2.4. Les données

### 2.4.1. Les paramètres physico-chimiques

En 2014, les données physico-chimiques récentes et transmises par les agences ont été contrôlées et analysées. Les analyses statistiques présentées dans ce rapport ont porté sur les données issues de 389 plans d'eau, tous types confondus (Tableau 1).

Les valeurs utilisées sont des concentrations moyennes annuelles en zone euphotique. Seules les données des lacs pour lesquels des données sur les compartiments biologiques étaient disponibles, ont été utilisées. Les distributions des valeurs des paramètres physico-chimiques utilisées dans les analyses de *Gradient Forest* sont synthétisées dans le Tableau 2.

**Tableau 1 : Bilan du nombre de plans d'eau suivis par agences de bassin dont les données ont pu être analysées en 2014.**

Agences	Nombres de plans d'eau DCE	Années de suivi disponibles	Nombres de plans d'eau utilisés en 2014
AG	105	2005-2012	82
AP	5	2007-2012	5
LB	166	2005-2012	134
RM	29	2007-2012	18
RMC	115	2004-2012	107
SN	58	2008-2012	41

**Tableau 2 : Caractéristiques (valeurs minimales, maximales et médianes) des paramètres physico-chimiques utilisés pour l'analyse des différents compartiments biologiques. Ces valeurs diffèrent entre les compartiments biologiques car les données correspondent à des listes de plans d'eau différentes pour lesquelles les données physico-chimiques et biologiques sont simultanément disponibles.**

	Unités	Transf. log	Phytoplancton (n=220)			Macrophytes (n=122)			Poissons (n=197)		
			min	max	med	min	max	med	min	max	med
Secchi	M	Oui	0,1	18,9	1,6	0,2	8,4	1,6	0,3	8,4	1,9
Temp	°C	Non	6,8	21,0	15,6	9,4	19,5	15,9	9,4	21,9	15,4
SatO2	%	Non	53,3	133,0	95,8	69,8	123,5	95,0	60,6	158,2	95,0
pH		Non	6,0	10,2	8,0	6,5	9,4	8,1	6,2	8,9	8,1
Cond25	µs.cm <sup>-1</sup>	Oui	14	1280	237	29	1279	273	24	835	261
Alc	meq.L <sup>-1</sup>	Oui	0,1	4,6	0,9	0,1	4,5	1,3	0,1	4,6	1,0
COD	mg.L <sup>-1</sup>	Oui	0,4	35,5	4,9	1,0	28,3	5,4	0,4	23,6	4,8
NH4	µg.L <sup>-1</sup>	Oui	8	324	52	12	851	79	12	851	65
NO3	mg.L <sup>-1</sup>	Oui	0,1	32,1	13,1	0,1	34,0	1,7	0,1	41,5	2,4
NO2	µg.L <sup>-1</sup>	Oui	6	248	19	6	361	22	6	361	30
TP	µg.L <sup>-1</sup>	Oui	6,5	380,3	29,2	8	458	34	8	458	29
PO4	µg.L <sup>-1</sup>	Oui	6	602	17	6	820	19	6	603	20
SiO2	mg.L <sup>-1</sup>	Oui	0,3	18,5	3,2	0,3	21,9	3,3	0,3	21,9	3,6

## 2.4.2. Les caractéristiques des compartiments biologiques

### a) Le phytoplancton

Pour les arbres de régression, la recherche de coupures étant basée sur la somme des carrés, il est préférable de transformer la variable de réponse (Y) afin de supprimer la dépendance entre variations de Y et moyenne de Y dans les groupes formés récursivement (De'ath et Fabricius 2000). Ainsi, les données brutes de biovolume des espèces de phytoplancton ont subi une transformation logarithme de type  $\log(Y_i + \min\{Y, Y > 0\})$  (Pitcher *et al.* 2012).

Les moyennes arithmétiques des biovolumes des espèces de phytoplancton ont été utilisées. Les taxons non déterminés à l'espèce et ceux apparaissant dans moins de 10% des plans d'eau n'ont pas été pris en compte. Ainsi, 120 espèces ont été considérées dans l'analyse. Seules les années comprenant au moins 4 campagnes avec une détermination du phytoplancton et des mesures physico-chimiques de l'eau disponibles, ont été retenues pour l'analyse.

### b) Les macrophytes

Pour les macrophytes, il n'y a pas de données de biomasse des taxons aux stations échantillonnées. L'analyse en *Gradient Forest* a donc été faite sur des données de présence/absence en tenant compte de tous les relevés faits sur chaque plan d'eau, i.e. au niveau de la rive et des profils

perpendiculaires. La limite d'occurrence des espèces a été fixée à 5% des plans d'eau échantillonnés. 45 espèces ont répondu à ce critère.

### c) L'ichtyofaune

Les données de biomasse par unité d'effort de pêche provenant de l'ensemble des filets benthiques ont été considérées. La même transformation logarithme que pour les données de phytoplancton a été appliquée. La limite de 5% a également été utilisée pour les espèces rares, ce qui a conduit à ne garder que 20 espèces pour le *Gradient Forest*.

## 2.4.3. Les caractéristiques environnementales

### a) Morphologie de la cuvette

Les variables environnementales décrivant la morphologie générale des cuvettes lacustres sont décrites dans les précédents rapports (Danis et Argillier 2011). Elles ont toutes été log-transformées. Il s'agit de :

- $Z_{max}$ , la profondeur maximale du plan d'eau (en m),
- $SurfLAC$ , la surface du plan d'eau (en  $km^2$ )
- $DLLac$ , critère de forme du plan d'eau (« Shorelinedevelopment Index » selon Hutchinson 1957),  $DLLac = \frac{Périm_{lac}}{(2 \times \sqrt{3,1415 \times SurfLAC})}$  -> indice de développement des berges qui exprime leur caractère plus ou moins découpé/sinueux
- $Hd$ , paramètre décrivant le rapport entre le profil bathymétrique et un cône -> indice décrivant la forme de la cuvette (plus ou moins convexe ou concave)
- $IBP = \frac{Volume_{lac}}{Périm_{lac}}$  -> (Index of Basin Permanence) indice décrivant l'importance des zones littorales

### b) Facteurs externes

Des facteurs hydromorphologiques à plus grande échelle ont également été considérés. Il s'agit de la surface du bassin versant ( $SurfBV$  en  $km^2$ ) et du temps de résidence de l'eau ( $TpsRes$  en années)(Danis et Argillier 2011). Ces deux paramètres ont subi une transformation logarithme avant analyse.

## 2.5. Mise en œuvre du Gradient Forest

L'analyse de *Gradient Forest* a été réalisée sous R à l'aide des packages "extended-Forest" et "gradientForest". Les paramètres par défaut ont été utilisés : 500 arbres pour chaque forêt aléatoire et un seuil de corrélation de 0,5 pour les permutations conditionnelles.

## 3. Résultats et discussion

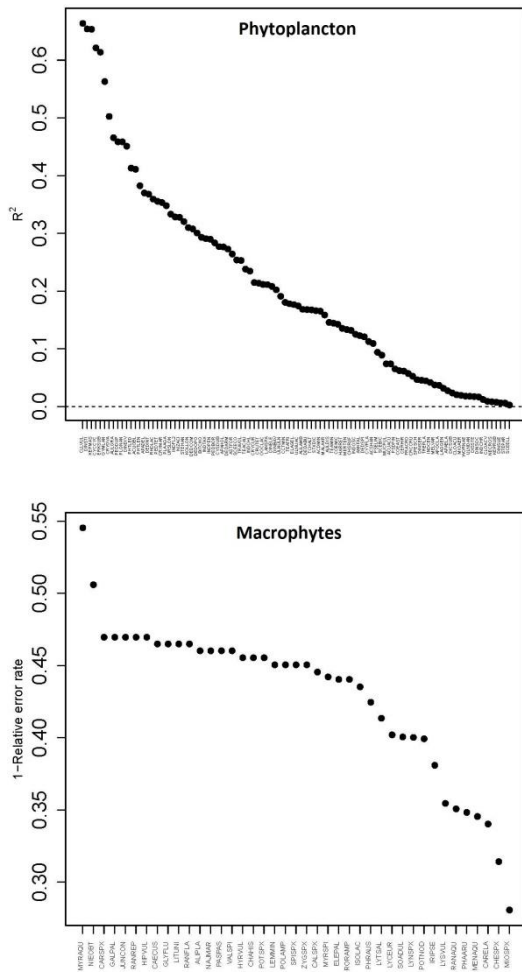
### 3.1. Performances des forêts aléatoires

Les performances des forêts aléatoires pour chaque compartiment nous informent sur la qualité de la modélisation de chaque taxon par les variables environnementales et géographiques considérées. La figure 8 présente les valeurs de  $R^2$  des taxons de chaque compartiment rangées par ordre décroissant. Pour le phytoplancton, les espèces les mieux modélisées ayant des  $R^2$  supérieurs à 0,6 sont la chlorophycée *Chlorella vulgaris*, les Cryptophycées *Dinobryon sociale var. stipitatum*, *Kephyrion mastigophorum*, *Erkenia subaequiciliata*, la diatomée *Cyclotella cyclopuncta* et la Dinophycée *Gymnodinium lantzschii*. Les deux espèces de macrophytes dont les occurrences sont particulièrement bien prédites ( $R^2 > 0.5$ ) sont l'Hélophyte *Myriophyllum aquaticum* et la Charophyte *Nitellopsis obtusa*. Quant aux poissons, les espèces qui se distinguent par leur  $R^2$  sont : *Squalius cephalus* (le Chevesne), *Coregonus sp.* (le Corégone), *Abramis brama* (la Brème commune), *Blicca bjoerkna* (la Brème bordelière) et *Sander lucioperca* (le Sandre) ( $R^2 > 0.3$ ).

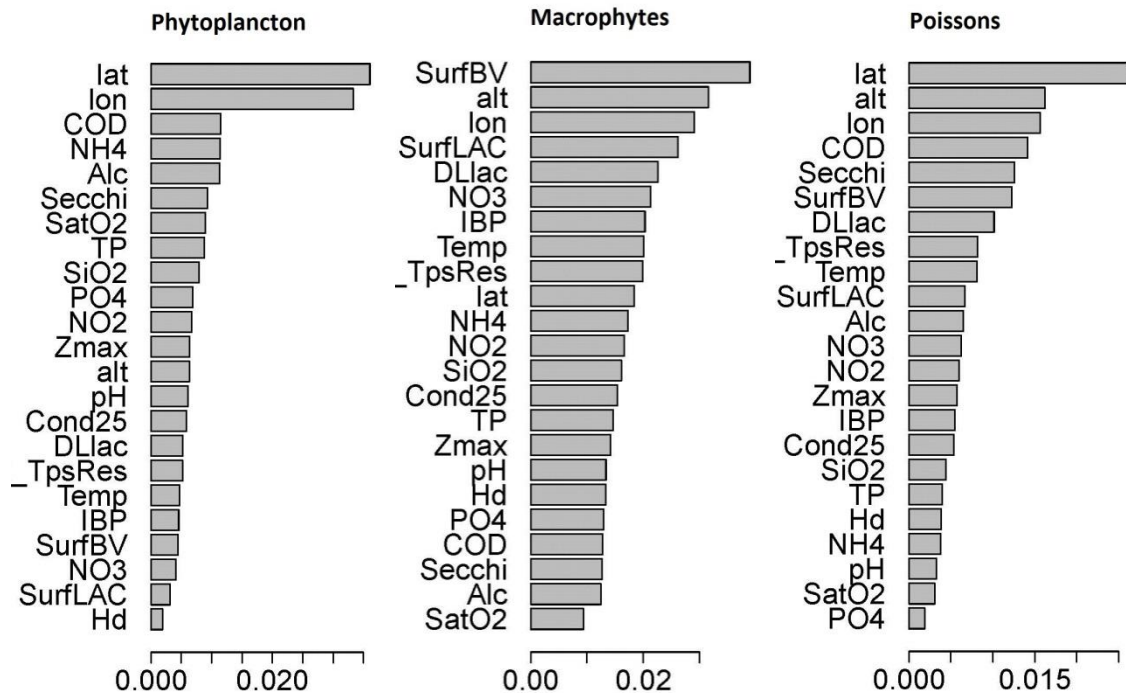
### 3.2. Importances des variables pour les communautés

Les importances relatives des variables environnementales sont différentes entre les trois compartiments (Figure 9). On remarque toutefois l'importance systématique des variables géographiques telles que longitude et latitude, ce qui indique une forte structuration spatiale des communautés à l'échelle du territoire français et quel que soit le type d'organismes. On peut noter aussi l'importance pour les compartiments phytoplancton et poissons du carbone organique dissous (COD) qui peut provenir soit du bassin versant (il donne alors une couleur à l'eau), soit de la production par les végétaux du lac (transparent et plus labile) (Bade *et al.* 2007). De même, la transparence de l'eau est un facteur déterminant pour les abondances de phytoplancton et de poissons. Les variables morphologiques telles que la surface du bassin versant ou le DLlac (cf §2.4.3) sont importantes pour les poissons et surtout les macrophytes, beaucoup moins pour le phytoplancton. Ce dernier compartiment est considérablement influencé par l'ammonium alors que les nutriments qui influencent le plus les macrophytes et les poissons sont les nitrates et le phosphore total respectivement.

De manière générale, ces résultats montrent que le compartiment phytoplancton est plus influencé par les paramètres physico-chimiques de l'eau que les deux autres compartiments. Ces derniers sont relativement plus sensibles aux paramètres morphologiques qui conditionnent leurs habitats.



**Figure 8.** Résultats de Gradient Forest appliqué aux trois compartiments biologiques : phytoplancton (haut), macrophytes (bas) et poissons (droite). Performance des forêts aléatoires ( $R^2$ ) générées pour chaque taxon. Les valeurs moyennes sont respectivement 0,21, 0,41 et 0,19.



**Figure 9.** Importances des variables estimées par Gradient Forest pour les compartiments Phytoplancton, Macrophytes et Poissons exprimées en composantes du R<sup>2</sup> moyen du compartiment (cf Fig. 8).

### 3.3. Seuils identifiés

Les figures 10 à 12 présentent les courbes de densités dans les gradients des 12 variables les plus importantes pour chaque compartiment. La courbe bleue indique l'ampleur des remaniements de la composition des communautés le long des gradients des différentes variables. Les pics de cette courbe localisent des seuils potentiels qui sont d'autant plus significatifs qu'ils sont étroits et qu'ils montent largement au-dessus de la ligne de base (trait discontinu). Les pics situés aux extrémités des gradients doivent être considérés avec prudence car, du fait qu'ils reposent sur seulement quelques valeurs, ils ne sont généralement pas très significatifs.

Ainsi pour le phytoplancton (Fig. 10), un seuil bien net apparaît à la longitude 5. Pour la latitude, on distingue plusieurs seuils moins marqués. Pour le COD, un seuil ressort à 0,5\*<sup>2</sup> (i.e. 3 mg.L<sup>-1</sup>) alors que l'autre seuil en fin de gradient n'est probablement pas significatif. Comme la latitude, l'ammonium présente plusieurs petits seuils dont le plus marqué se situe autour de 1,65\* (45 µg.L<sup>-1</sup>). L'analyse révèle un seuil diffus pour l'alcalinité à partir de 0,45\* (3 meq.L<sup>-1</sup>). On remarque un seuil assez net de transparence de l'eau à -0,35\* (0,45 m de profondeur de Secchi) et un autre moins marqué à 0,4\* (2,5 m). On note l'absence de seuil significatif pour l'oxygène, alors que pour le phosphore total il y a clairement un seuil diffus entre 1,5\* et 1,8\* (30-60 µg.L<sup>-1</sup>) et un autre seuil plus précis à 2,2\* (160 µg.L<sup>-1</sup>). Les nutriments silice et phosphates présentent des seuils à 0\* (1 mg.L<sup>-1</sup>) et 2\* (100 µg.L<sup>-1</sup>) et les nitrites, plusieurs petits seuils à 1,65\* et 1,85\* (45 et 70 µg.L<sup>-1</sup>). Enfin, le premier paramètre

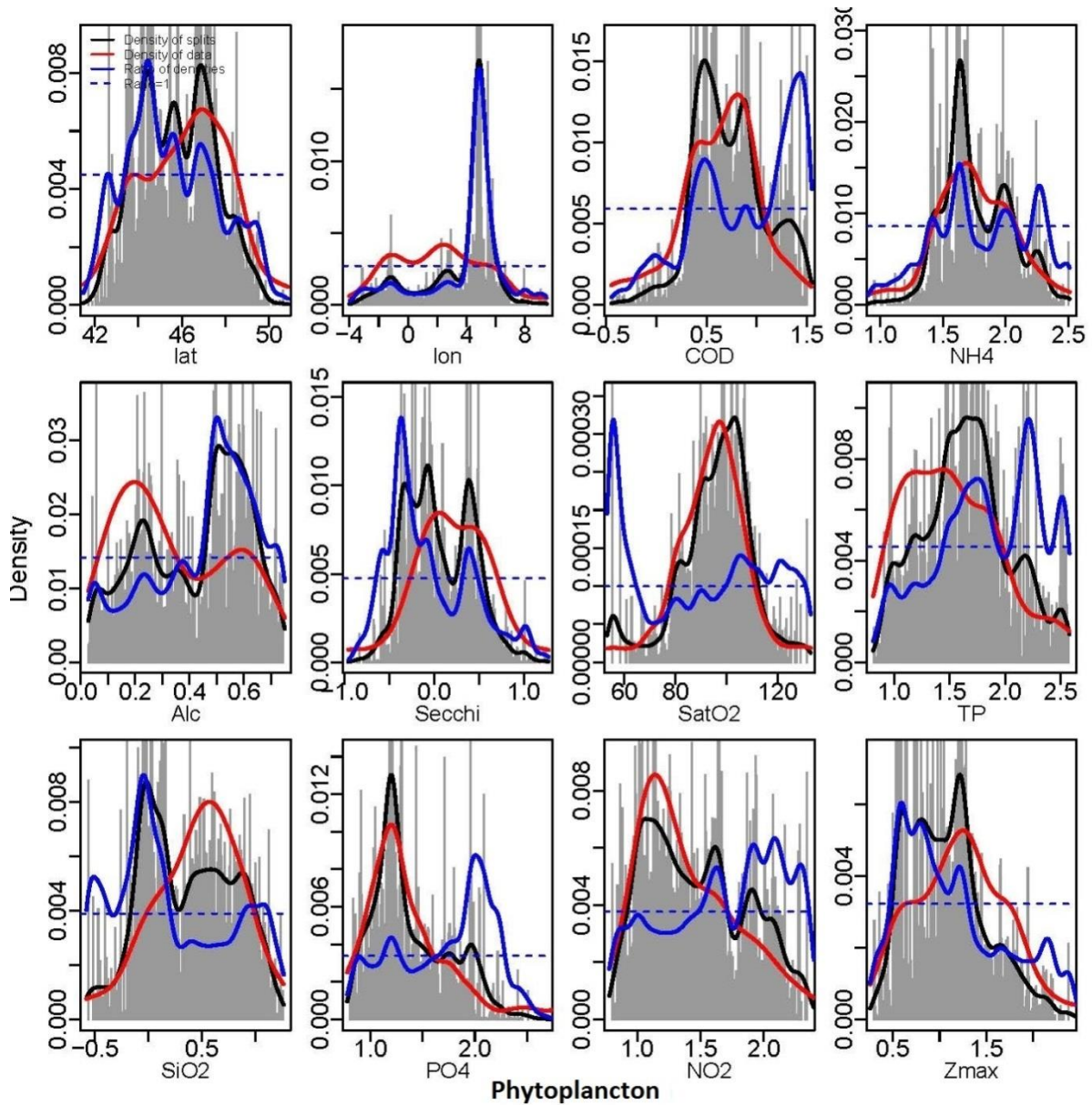
<sup>2</sup> Les valeurs en log10 sont annotées par une étoile (\*) dans la suite du paragraphe

morphologique par ordre d'importance, la profondeur maximale, se caractérise par une zone de changements importants entre 3 et 8 m (seuil diffus).

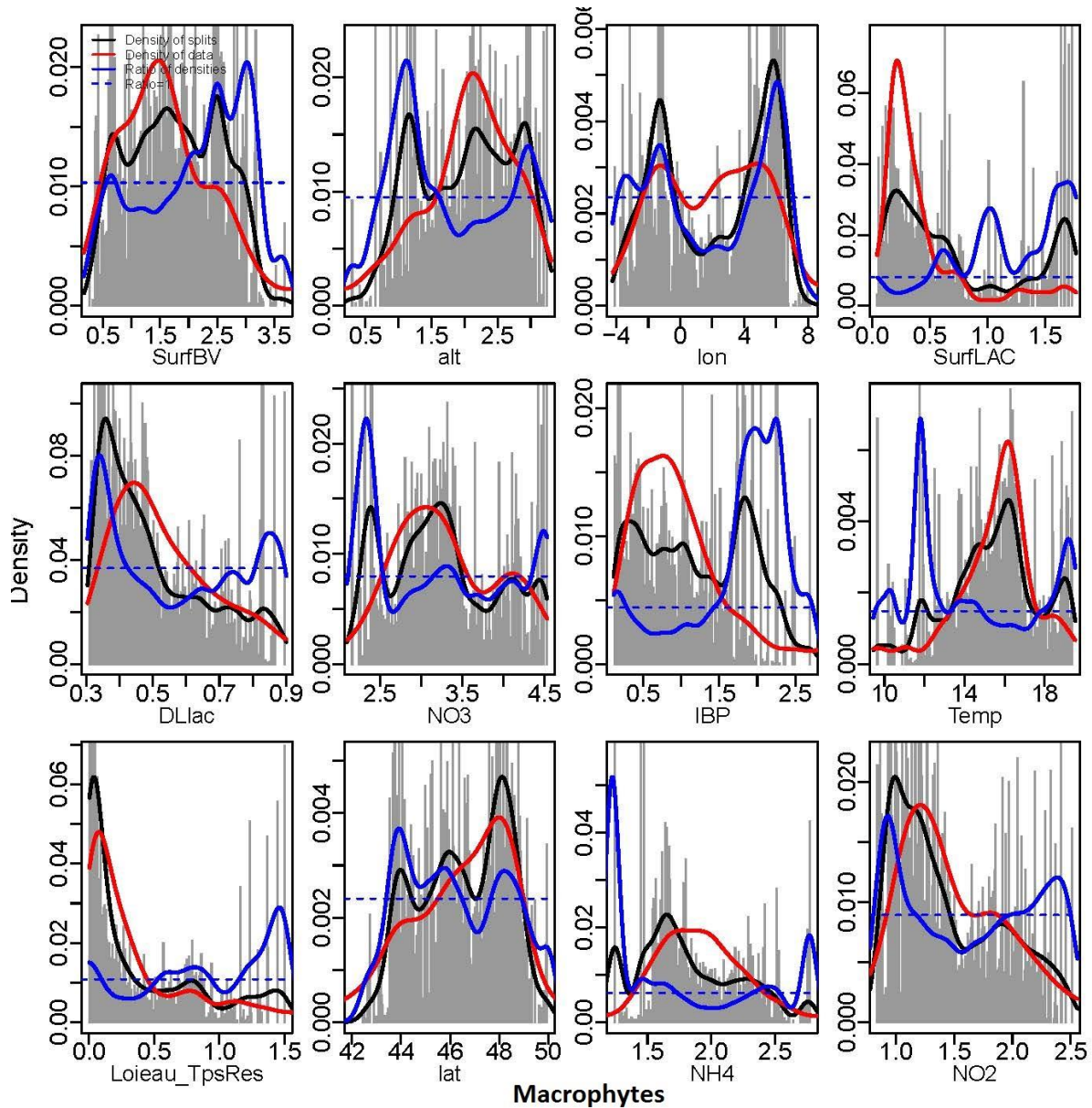
Concernant les macrophytes (Fig. 11), on observe des seuils bien clairs pour les trois variables les plus importantes : la surface du bassin versant (300 et 1000 km<sup>2</sup>), l'altitude (10 et 1000 m) et la longitude (-1,5 et 5). On trouve aussi des indications de seuils pour les trois paramètres morphologiques importants : surface du lac (4,5 et 10 km<sup>2</sup>), le DLLac (2,2) et l'IBP (60-200). Un faible seuil est détecté pour les nitrates (2 mg.L<sup>-1</sup>), la variable chimique la plus importante. On peut noter un seuil de température à 12°C et plusieurs seuils de latitude (44, 46 et 48 °N).

Quant aux poissons (Fig. 12), le *Gradient Forest* a identifié un seuil net de latitude à 46,5°N, un seuil d'altitude à 300 m et un seuil de longitude à 5°E. Deux seuils peuvent être retenus pour le COD (1,4 et 4 mg.L<sup>-1</sup>) et pour la transparence (0,6 et 3,2 m). On peut également distinguer un seuil pour les trois paramètres hydromorphologiques, surface du bassin versant (10 km<sup>2</sup>), DLLac (4,5) et temps de résidence (4,5 ans). Enfin, un seuil de nitrates apparaît à 3 mg.L<sup>-1</sup>.

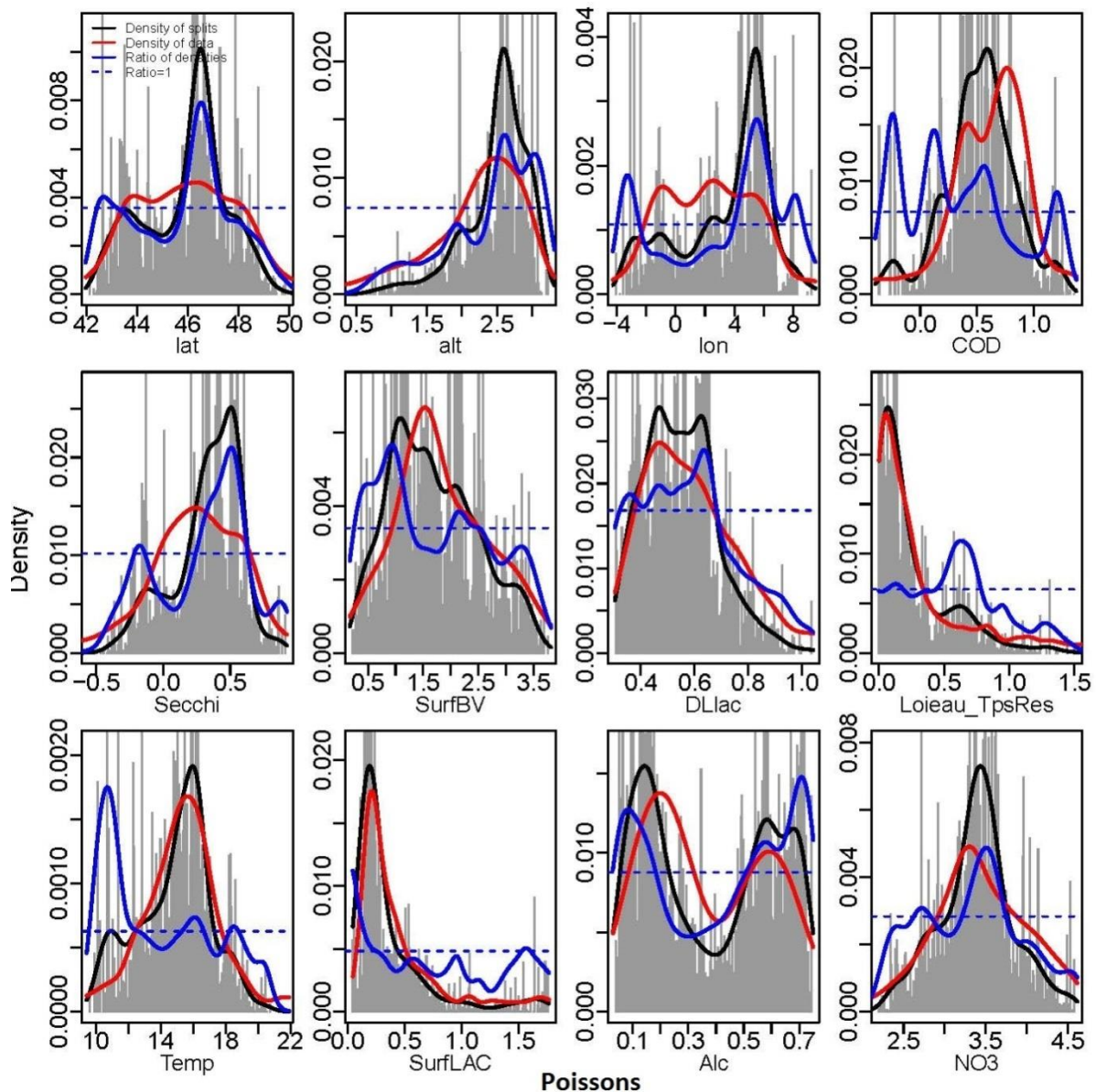




**Figure 10.** Résultats du Gradient Forest pour le compartiment phytoplancton et les 12 variables les plus importantes : densité des coupures (courbes noires), densité des observations (courbe rouge) et ratio des densités (courbe bleue). Les pics de la courbe bleue indiquent la localisation de seuils potentiels.



**Figure 11.** Densité des coupures pondérée par leur importance, pour le compartiment macrophytes et les 12 variables les plus importantes : densité des coupures (courbes noires), densité des observations (courbe rouge) et ratio des densités (courbe bleue). Les pics de la courbe bleue indiquent la localisation de seuils potentiels.



**Figure 12.** Densité des coupures pondérée par leur importance : densité des coupures (courbes noires), densité des observations (courbe rouge) et ratio des densités (courbe bleue). Les pics de la courbe bleue indiquent la localisation de seuils potentiels.

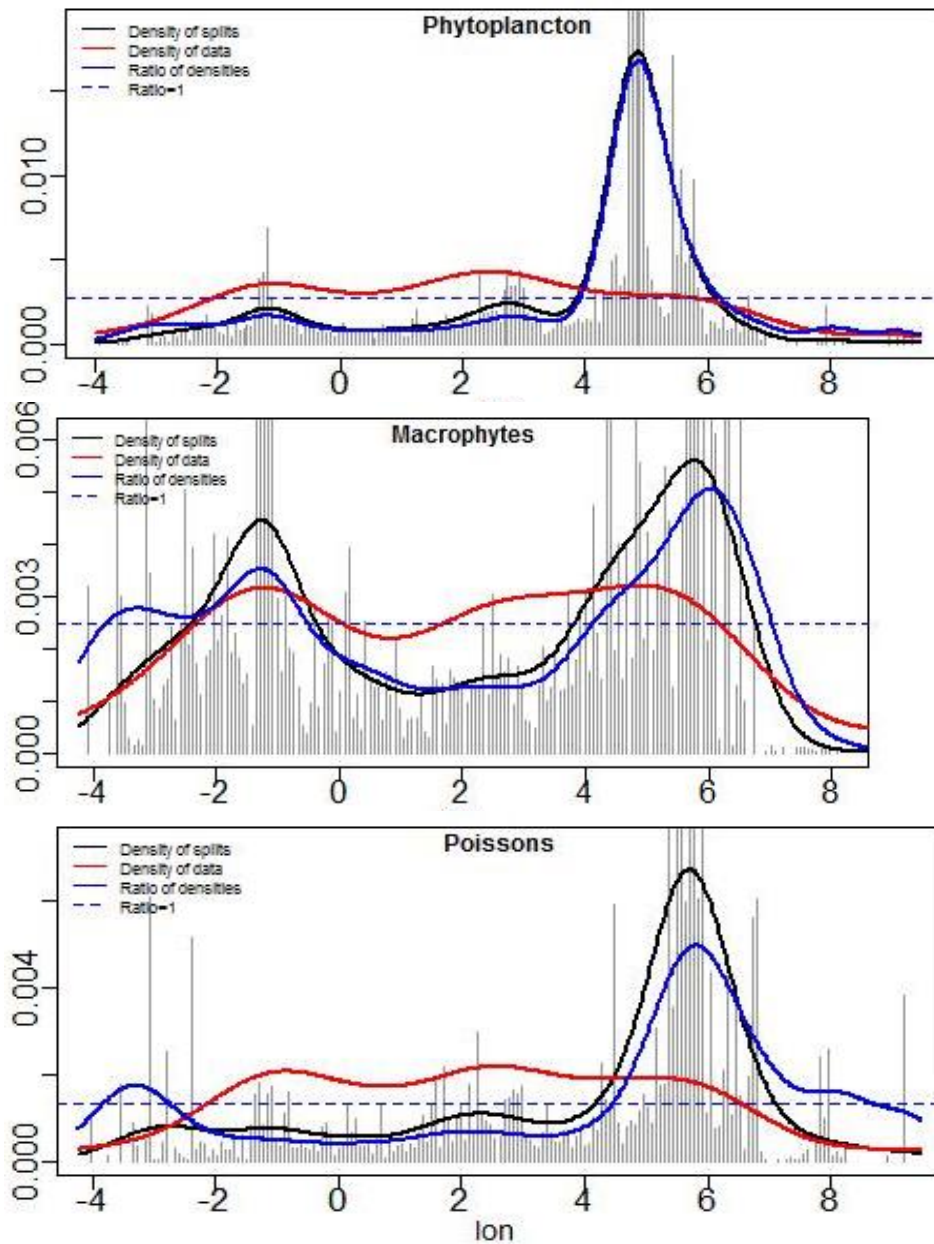
### 3.4. Comparaison des seuils entre compartiments

Il est intéressant de comparer les seuils révélés par chacun des compartiments, d'autant plus que les seuils d'état des éléments de qualité physico-chimique devraient témoigner de leur influence sur l'état de l'ensemble des éléments de qualité biologique. Pour certaines variables, on trouve des coïncidences alors que pour d'autres les seuils sont dispersés dans le gradient. Un des résultats les plus marquants est le seuil géographique constitué par la longitude 5° pour les trois compartiments (Fig. 13). Cette longitude sépare les lacs des massifs montagneux situés à l'est du pays, qui ont probablement des communautés bien distinctes de celles des autres lacs du territoire national. On

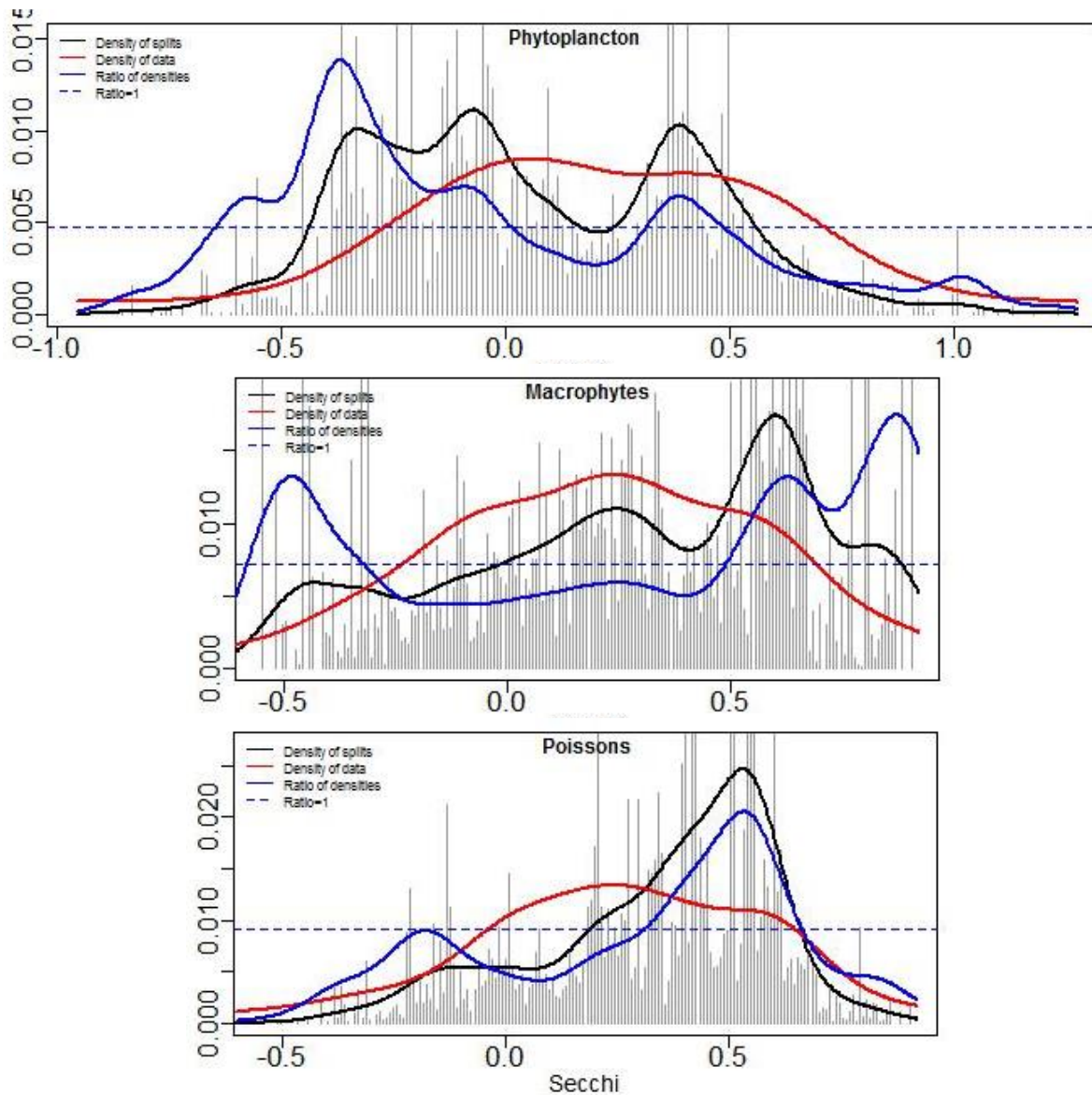
peut également rapprocher les seuils de transparence détectés pour chaque compartiment autour de la valeur 0,5 (3,2 m), même si les seuils ne concordent pas aussi bien que ceux de la longueur (Fig. 14). Cette similitude des réponses suggère que la transparence de l'eau est un paramètre pouvant bouleverser profondément la structure des communautés.

On peut relever des seuils communs entre deux compartiments pour plusieurs autres variables environnementales. Les macrophytes et les poissons font apparaître un seuil du paramètre morphologique DLlac en début de gradient, autour de 2,2 (Fig. 15). Un faible niveau de découpage des rives serait donc critique pour la richesse spécifique de ces compartiments. Le seuil de nitrates identifié avec les poissons en milieu de gradient peut être mis en relation avec les seuils très légers repérés avec le phytoplancton et les macrophytes autour de 3 mg.L<sup>-1</sup> (Fig. 16). Enfin, les seuils diffus de phosphore total qui ressortent en considérant le phytoplancton et les poissons présentent une zone de chevauchement entre 1,5 et 1,7 (30-50 µg.L<sup>-1</sup>) (Fig. 17).

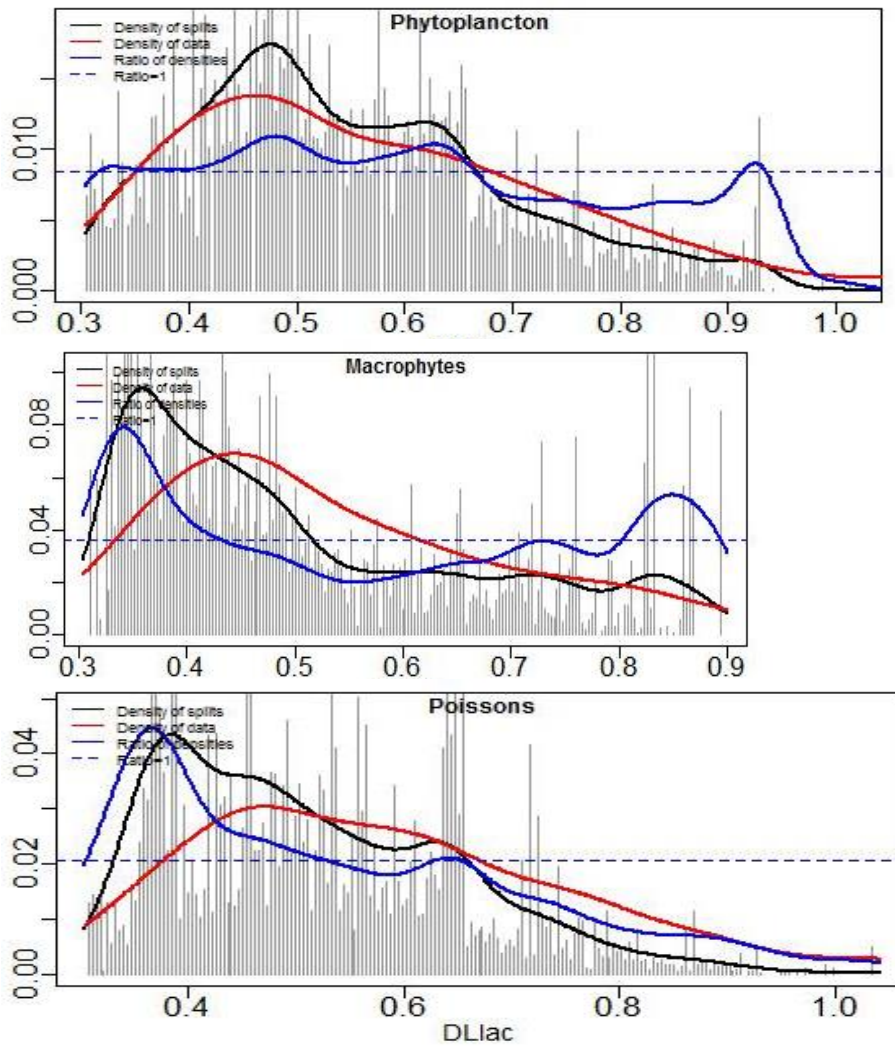
Pour certaines variables, il n'y a pas de seuil commun entre les compartiments qui semblent répondre de façons différentes aux gradients. C'est le cas par exemple de l'ammonium qui ne présente pas de seuil significatif si l'on considère les macrophytes, et les seuils du phytoplancton et des poissons associés à ce paramètre ne sont pas en phase dans le gradient. On peut toutefois noter des pics de densité des coupures de chaque compartiment autour de 1,6-1,7 (40-50 µg.L<sup>-1</sup>).



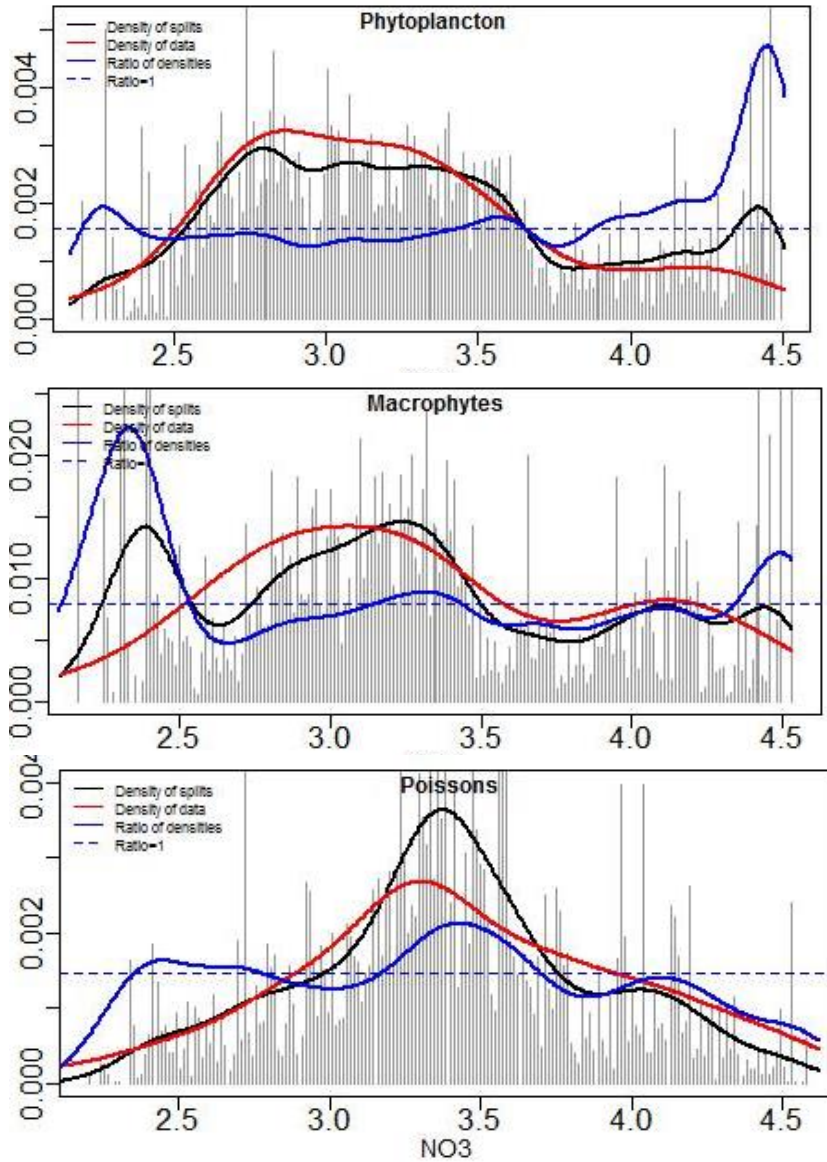
**Figure 13.** Comparaison des seuils de longitude obtenus pour chaque compartiment biologique. Un seuil net apparaît dans les trois cas autour de 5°E.



**Figure 14.** Comparaison des seuils de transparence obtenus pour chaque compartiment biologique. Une zone critique commune apparaît autour de 0,5 (3,2 m).

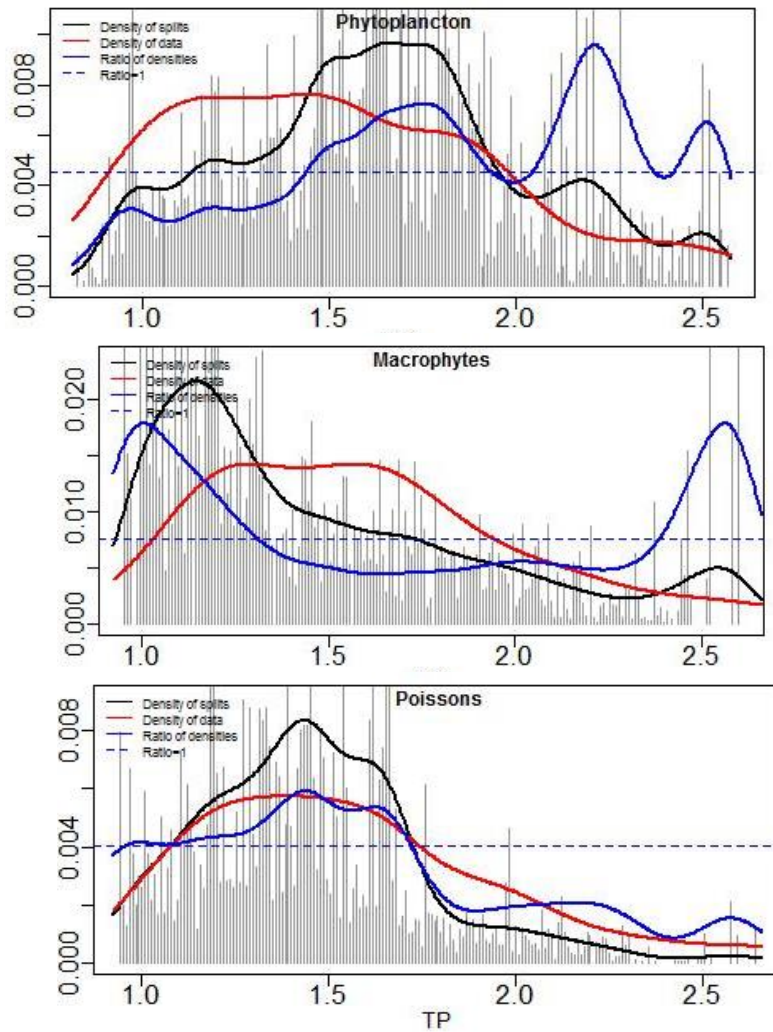


**Figure 15.** Comparaison des seuils obtenus avec le paramètre DLlac pour les trois compartiments biologiques. On remarque un seuil commun pour les macrophytes et les poissons en début de gradient, autour de 0,35 (2,2 en valeur réelle).

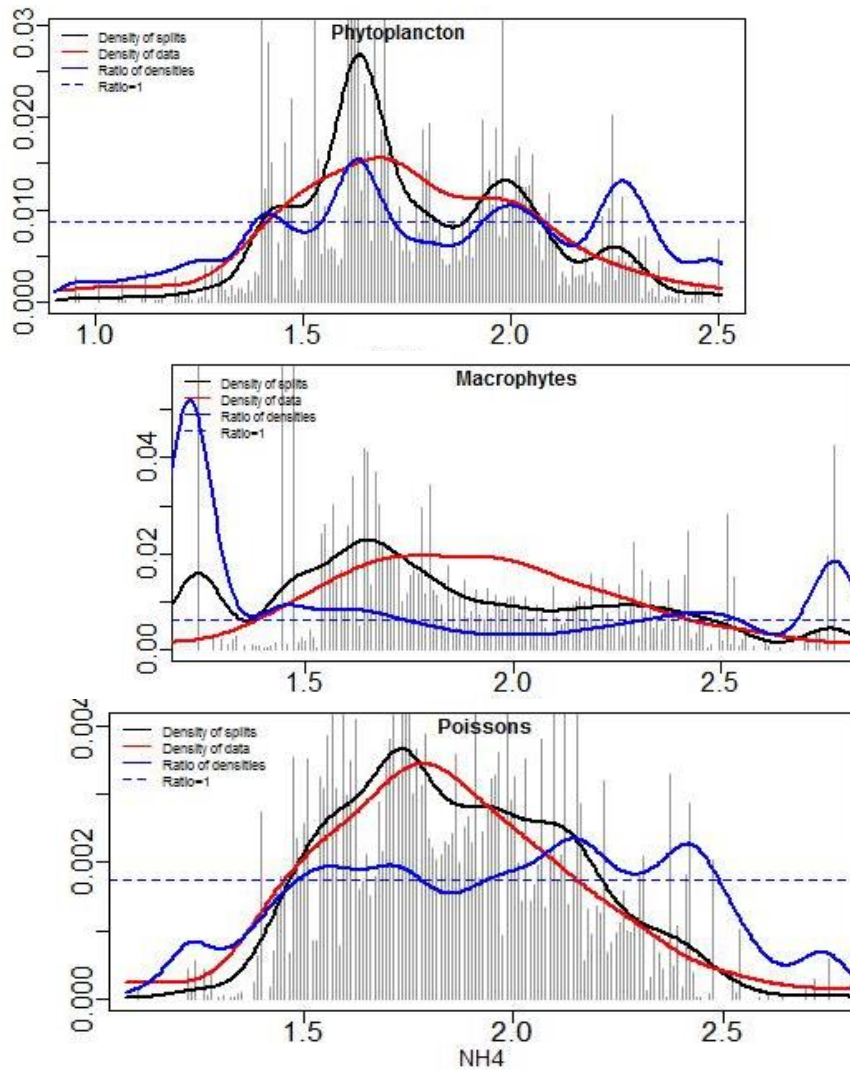


**Figure 16.** Comparaison des seuils de nitrates entre les trois compartiments. On peut distinguer une certaine concordance autour de 3,5 (i.e. 3,2 mg.L<sup>-1</sup>).





**Figure 17.** Comparaison des seuils de phosphore total obtenus pour chaque compartiment biologique. On peut noter des similitudes dans les réponses du phytoplancton et des poissons.

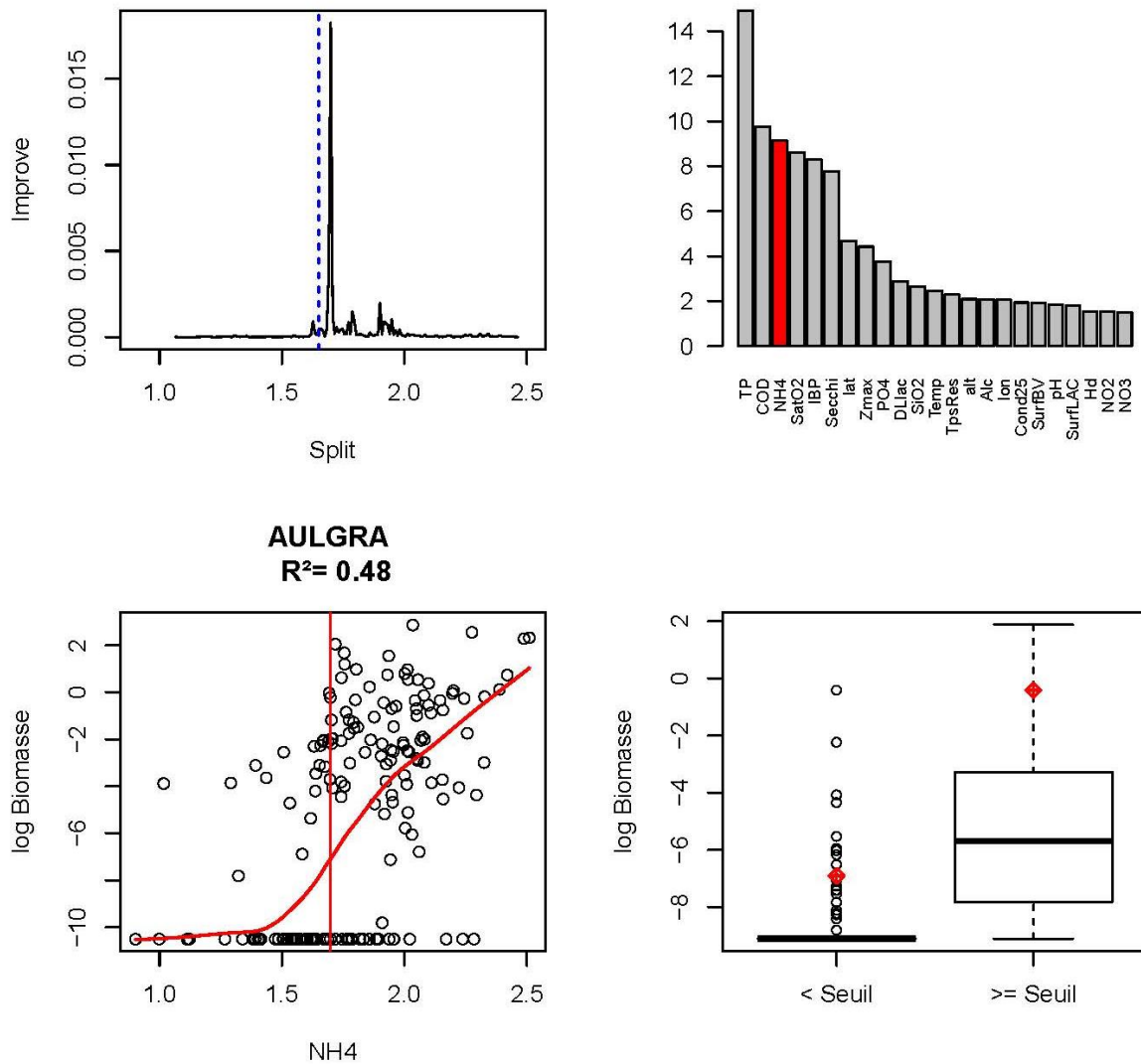


**Figure 18.** Comparaison des seuils d'ammonium obtenus pour chaque compartiment biologique. Il ne ressort pas de seuil commun pour ce paramètre.

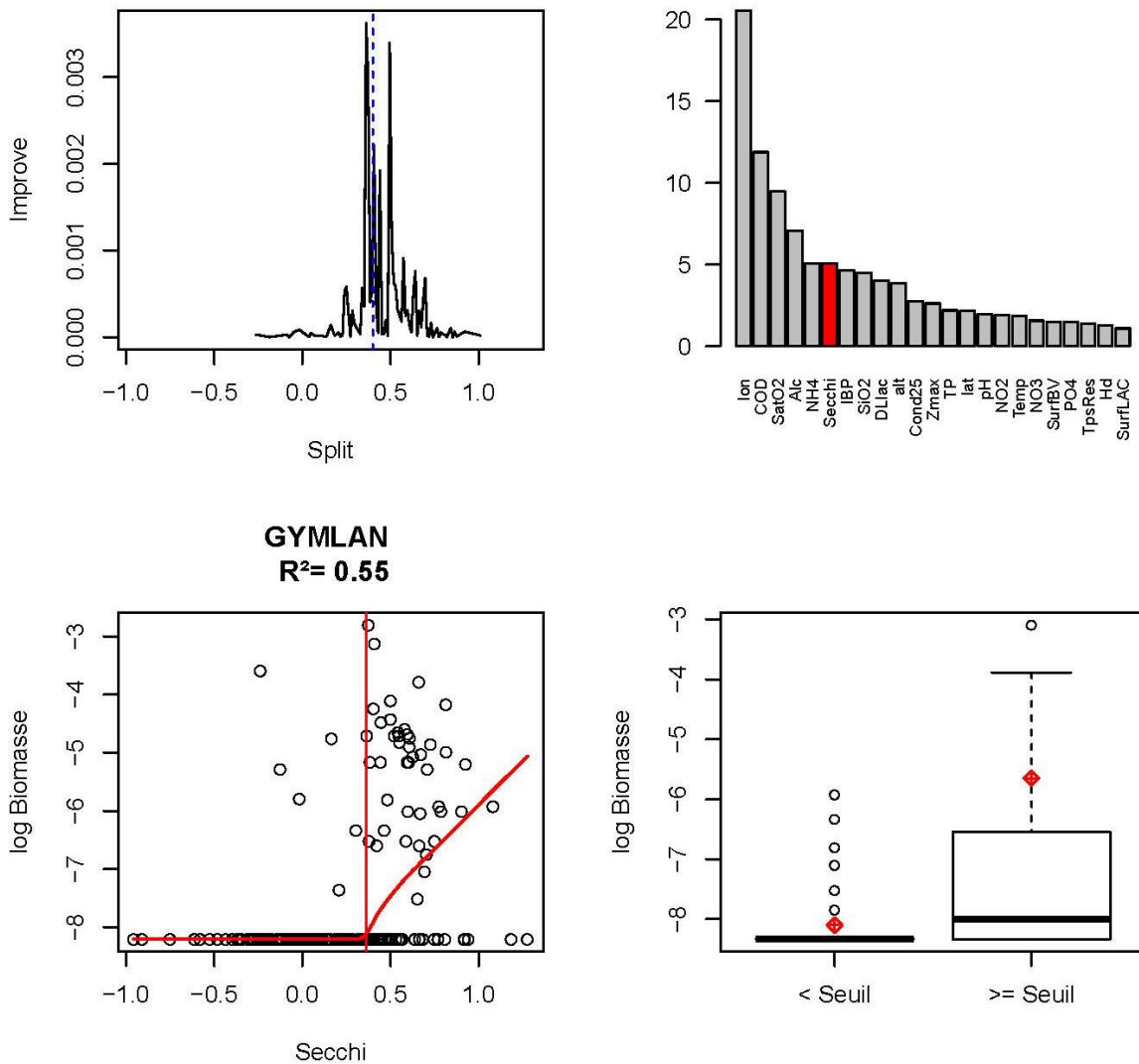
### **3.5. Réponses d'espèces individuelles**

Le *Gradient Forest* combine les réponses des différentes espèces d'un compartiment biologique pour extraire des signaux à l'échelle des communautés. Il peut être intéressant de regarder ensuite les réponses des espèces individuellement ainsi que les forêts aléatoires associées. Ceci permet d'identifier les espèces expliquant un seuil de communauté et de vérifier la vraisemblance des résultats compte tenu des connaissances sur l'écologie de ces espèces. Cet examen détaillé est d'autant plus justifié que le nombre d'espèces est faible et que les liens entre ces espèces et les caractéristiques de leur environnement sont bien connues (pour les poissons par ex.).

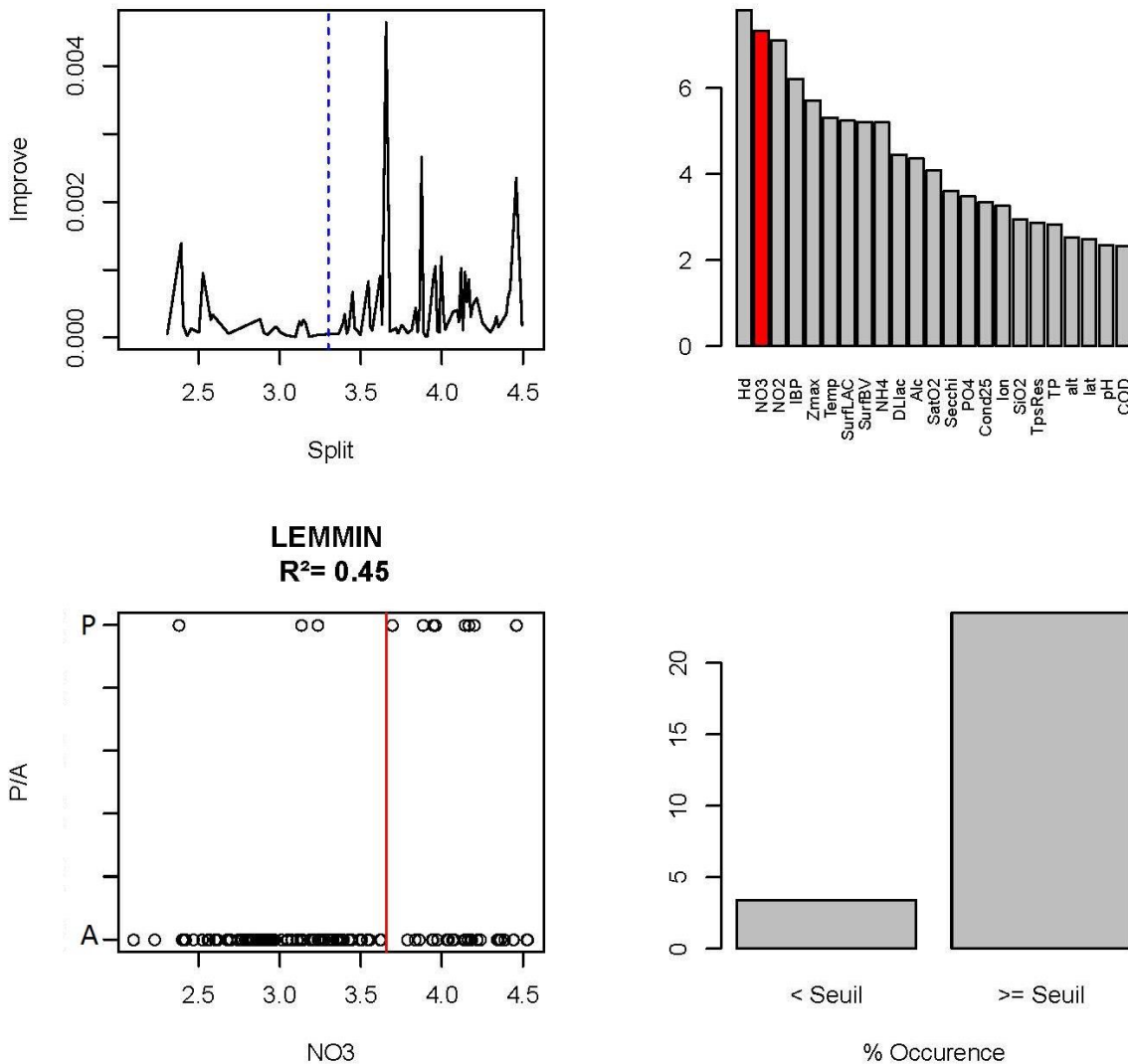
Deux exemples sont fournis ici pour chaque compartiment avec des variables influencées par des effets anthropiques (Fig. 19 à 24). La diatomée *Aulacoseira granulata* présente un seuil de réponse à l'ammonium qui est très proche du seuil de communauté identifié par *Gradient Forest*. L'abondance de l'espèce au-delà de 50  $\mu\text{g.L}^{-1}$  est cohérente avec sa préférence pour les milieux eutrophisés (Reynolds *et al.* 2002). L'espèce *Gymnodinium lantzschii* illustre bien le seuil de transparence détecté par *Gradient Forest* à 2,5 m de profondeur de Secchi. Elle est plus abondante dans les eaux claires. L'occurrence de l'hydrophyte *Lemna minor* est influencée par la concentration en nitrates de l'eau et un seuil peut être identifié à 4,5  $\text{mg.L}^{-1}$  au-delà duquel l'espèce est plus fréquemment recensée. Un autre hydrophyte, *Potamogeton nodosus*, se développe préférentiellement dans les eaux plus transparentes et la forêt aléatoire a décelé un seuil autour de 3,5 m de profondeur de Secchi. Parmi les poissons, la Brème commune (*Abramis brama*) a une biomasse plus élevée au-dessus d'un seuil de phosphore total de 45  $\mu\text{g.L}^{-1}$ , légèrement supérieur au seuil de communauté identifié par *Gradient Forest*. Enfin, la biomasse du Poisson-chat (*Ameiurus melas*) est relativement plus importante dès que la concentration en carbone organique dissous dépasse 4,5  $\text{mg.L}^{-1}$ . La répartition de ces espèces dans les gradients chimiques sont en accord avec la littérature, puisque *A. brama* est typique des milieux eutrophes et *A. melas* est caractéristique des eaux turbides, fortement eutrophisées et riches en matière organique (Schlumberger et Elie 2008).



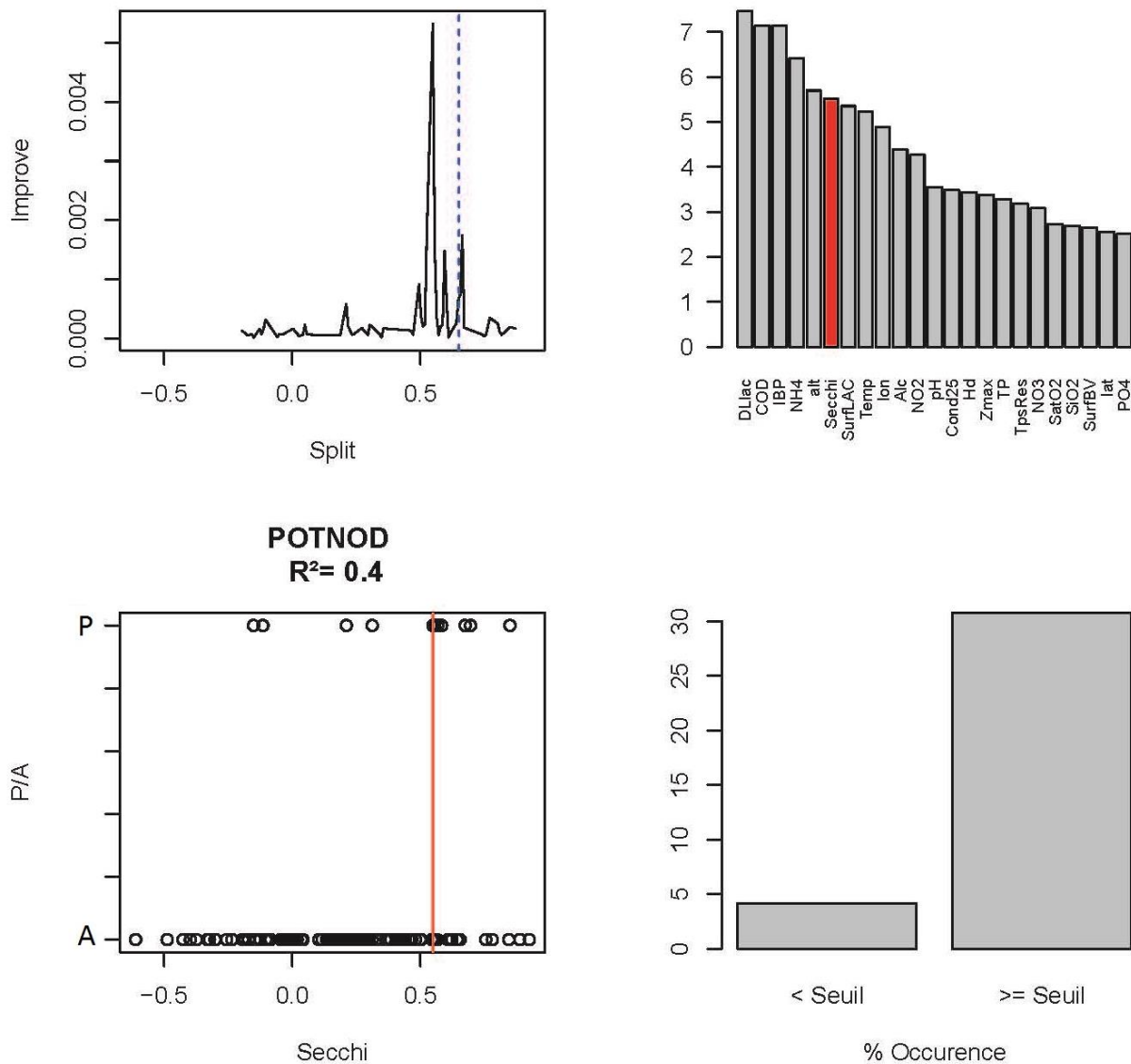
**Figure 19.** Résultat de la forêt aléatoire pour l'espèce de phytoplancton *Aulacoseira granulata*. Densité des coupures dans le gradient (en haut à gauche, la ligne pointillée verticale indique la localisation du seuil de communauté), importance des variables (en haut à droite), biovolume de l'espèce avec une courbe de lissage et une droite verticale indiquant la localisation du maximum de densité des coupures (=seuil) (en bas à gauche) et distribution des valeurs de biomasse de part et d'autre du seuil de l'espèce (en bas à droite), les losanges rouges représentent le pourcentage d'occurrences mis à l'échelle des variations de biovolume.



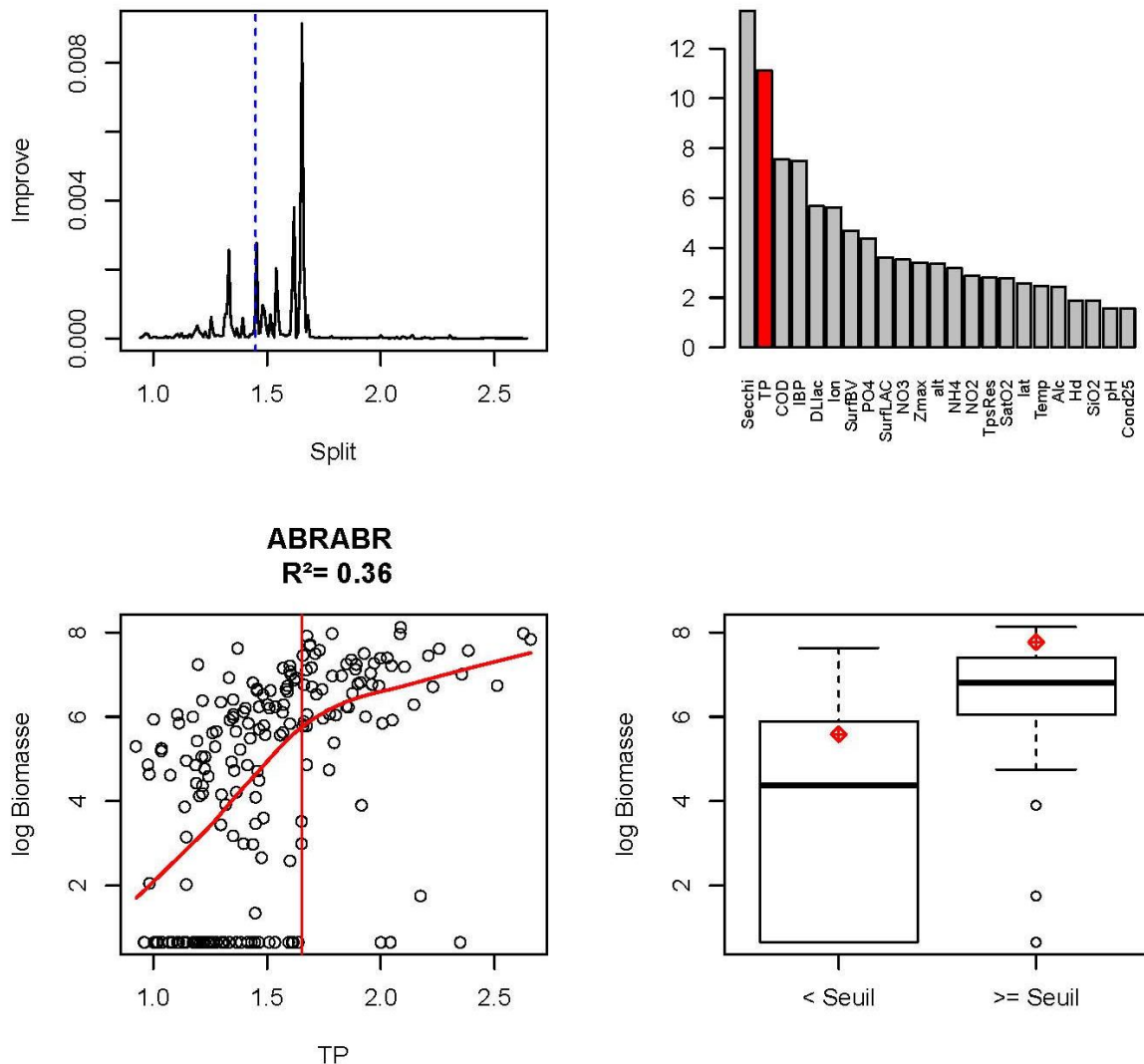
**Figure 20.** Résultat de la forêt aléatoire pour l'espèce de phytoplancton *Gymnodinium lantzschii*. Densité des coupures dans le gradient (en haut à gauche, la ligne pointillée verticale indique la localisation du seuil de communauté), importance des variables (en haut à droite), biovolume de l'espèce avec une courbe de lissage et une droite verticale indiquant la localisation du maximum de densité des coupures (=seuil) (en bas à gauche) et distribution des valeurs de biomasse de part et d'autre du seuil de l'espèce (en bas à droite), les losanges rouges représentent les pourcentages d'occurrences mis à l'échelle des variations de biovolume.



**Figure 21.** Résultat de la forêt aléatoire pour l'espèce de macrophyte *Lemna minor*. Densité des coupures dans le gradient (en haut à gauche, la ligne pointillée verticale indique la localisation du seuil de communauté), importance des variables (en haut à droite), occurrences de l'espèce dans le gradient avec une droite verticale indiquant la localisation du maximum de densité des coupures (=seuil) (en bas à gauche) et pourcentages d'occurrences de part et d'autre du seuil de l'espèce (en bas à droite).

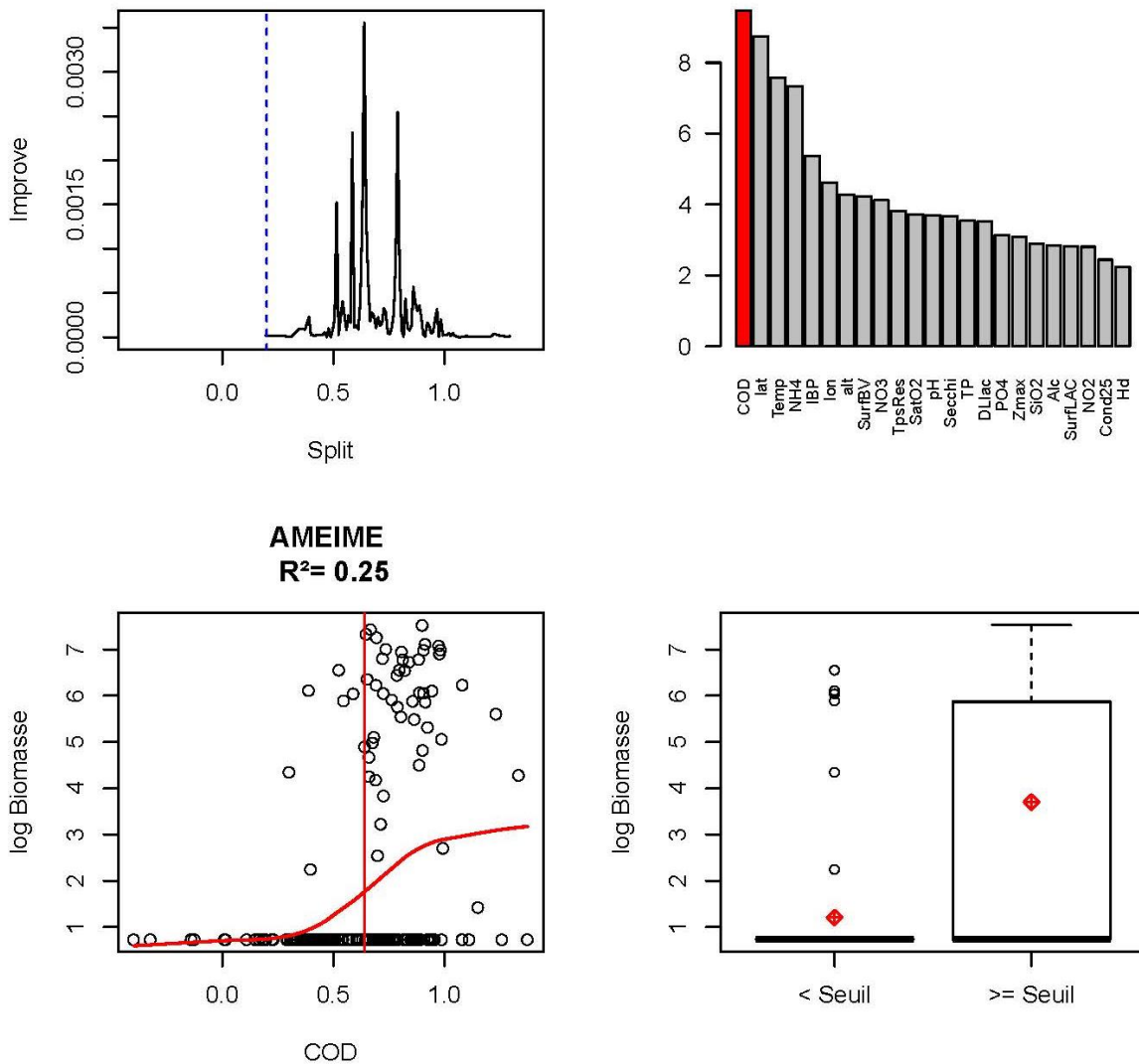


**Figure 22.** Résultat de la forêt aléatoire pour l'espèce de macrophyte *Potamogeton nodosus*. Densité des coupures dans le gradient (en haut à gauche, la ligne pointillée verticale indique la localisation du seuil de communauté), importance des variables (en haut à droite), occurrences de l'espèce dans le gradient avec une droite verticale indiquant la localisation du maximum de densité des coupures (=seuil) (en bas à gauche) et pourcentages d'occurrences de part et d'autre du seuil de l'espèce (en bas à droite).



**Figure 23.** Résultat de la forêt aléatoire pour l'espèce de poisson *Abramis brama*. Densité des coupures dans le gradient (en haut à gauche, la ligne pointillée verticale indique la localisation du seuil de communauté), importance des variables (en haut à droite), biovolume de l'espèce avec une courbe de lissage et une droite verticale indiquant la localisation du maximum de densité des coupures (=seuil) (en bas à gauche) et distribution des valeurs de biomasse de part et d'autre du seuil de l'espèce (en bas à droite), les losanges rouges représentent les pourcentages d'occurrences mis à l'échelle des variations de biomasse.





**Figure 24.** Résultat de la forêt aléatoire pour l'espèce de poisson *Ameiurus melas*. Densité des coupures dans le gradient (en haut à gauche, la ligne pointillée verticale indique la localisation du seuil de communauté), importance des variables (en haut à droite), biovolume de l'espèce avec une courbe de lissage et une droite verticale indiquant la localisation du maximum de densité des coupures (=seuil) (en bas à gauche) et distribution des valeurs de biomasse de part et d'autre du seuil de l'espèce (en bas à droite), les losanges rouges représentent les pourcentages d'occurrences mis à l'échelle des variations de biomasse.

## 4. Conclusions et perspectives

### 4.1. Conclusions

La méthode de *Gradient Forest* a permis de traiter la question de l'existence de seuils écologiques à travers l'analyse d'une large base de données biologiques et physico-chimiques. Cette approche exploratoire a révélé, dans les gradients de différents paramètres influencés par les activités anthropiques, des zones sensibles pour les communautés pouvant être interprétées comme des seuils écologiques. Ces valeurs critiques pourront servir de base à l'établissement de seuils opérationnels pour la gestion des plans d'eau. Avant cela, un travail de synthèse est nécessaire entre les trois compartiments, incluant la définition des seuils communs, la hiérarchisation des seuils lorsque plusieurs sont identifiés dans un même gradient et la manière de traiter les seuils diffus. Il sera aussi important de prendre en compte la variabilité naturelle des valeurs des paramètres physico-chimiques afin de déterminer un niveau de référence pour chaque plan d'eau et de considérer uniquement les seuils supérieurs à ce niveau. Etant donné l'importance des variables géographiques dans les analyses effectuées, la définition des seuils pourrait être affinée en considérant séparément différentes parties du territoire, comme par exemple les régions situées de part et d'autre de la longitude 5.

### 4.2. Perspectives

La suite du travail consiste à relier les variations des paramètres physico-chimiques des plans d'eau aux caractéristiques environnementales à différentes échelles spatiales, i.e. de la cuvette du plan d'eau au bassin-versant. Les relations seront établies avec des variables naturelles et des forces motrices. Ces relations permettront de cibler les éléments qui contrôlent principalement les variations des paramètres physico-chimiques, et d'identifier possiblement les points d'action pour envisager une restauration. De plus, elles serviront à estimer des valeurs de référence pour les plans d'eau en fonction de leur type.

## 5. Références

- Andersen, T., J. Carstensen, E. Hernández-García et C.M. Duarte. 2009. Ecological thresholds and regime shifts: approaches to identification. *Trends in Ecology & Evolution* 24: 49-57.
- Argillier, C., S. Caussé, M. Gevrey, S. Pédron, J. De Bortoli, S. Brucet, et al. 2013. Development of a fish-based index to assess the eutrophication status of European lakes. *Hydrobiologia* 704: 193-211.
- Bade, D.L., S.R. Carpenter, J.J. Cole, M.L. Pace, E. Kritzberg, M.C. Van de Bogert, et al. 2007. Sources and fates of dissolved organic carbon in lakes as determined by whole-lake carbon isotope additions. *Biogeochemistry* 84: 115-129.
- Baker, M.E. et R.S. King. 2010. A new method for detecting and interpreting biodiversity and ecological community thresholds. *Methods in Ecology and Evolution* 1: 25-37.
- Boutry, S., V. Bertrin et A. Dutartre. 2013. Méthode d'évaluation de la qualité écologique des plans d'eau basée sur les communautés de macrophytes. *Indice Biologique Macrophytique en Lac (IBML) - Rapport d'avancement. Rapport convention Onema/Irstea 2012. Irstea, REBX, Bordeaux.* p. 26+annexes.
- Breiman, L. 2001. Random forests. *Mach Learn* 45: 5-32.
- Brenden, T.O., L. Wang et Z. Su. 2008. Quantitative Identification of Disturbance Thresholds in Support of Aquatic Resource Management. *Environmental Management* 42: 821-832.
- Catalan, J., M.G. Barbieri, F. Bartumeus, P. Bitusik, I. Botev, A. Brancelj, et al. 2009. Ecological thresholds in European alpine lakes. *Freshwater Biology* 54: 2494-2517.
- Chambers, P.A., D.J. McGoldrick, R.B. Brua, C. Vis, J.M. Culp et G.A. Benoy. 2012. Development of Environmental Thresholds for Nitrogen and Phosphorus in Streams. *Journal of Environmental Quality* 41: 7-20.
- Danis, P.-A. et C. Argillier. 2011. Seuils des paramètres physico-chimiques soutenant la biologie : Plans d'eau. *Rapport convention Onema/Irstea 2010. Irstea, UR HYAX, Pôle Onema/Irstea, Aix-en-Provence.* p. 49.
- Danis, P.-A. et C. Argillier. 2013. Seuils des paramètres physico-chimiques soutenant la biologie en plans d'eau. *Rapport convention Onema/Irstea 2012. Irstea, UR HYAX, Pôle Onema/Irstea, Aix-en-Provence.* p. 33.
- Danis, P.-A., R. Ferrer, M. Gevrey et C. Argillier. 2012. Seuils des paramètres physico-chimiques soutenant la biologie : Plans d'eau naturels. *Rapport convention Onema/Irstea 2011. Irstea, UR HYAX, Pôle Onema/Irstea, Aix-en-Provence.* p. 60.
- Danis, P.-A. et V. Roubéix. 2014. Physico-chimie soutenant la biologie des plans d'eau nationaux : Principes et méthodes de définition des valeurs-seuils & Amélioration des connaissances par la télédétection. *Rapport convention Onema/Irstea 2013. Irstea, UR HYAX, Pôle Onema/Irstea, Aix-en-Provence.* p. 81.
- De'Ath, G. 2002. Multivariate regression trees: a new technique for modeling species-environment relationships. *Ecology* 83: 1105-1117.
- De'ath, G. et K.E. Fabricius. 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology* 81: 3178-3192.
- Dodds, W.K., W.H. Clements, K. Gido, R.H. Hilderbrand et R.S. King. 2010. Thresholds, breakpoints, and nonlinearity in freshwaters as related to management. *J N Am Benthol Soc* 29: 988-997.
- Ellis, N., S.J. Smith et C.R. Pitcher. 2012. Gradient forests: calculating importance gradients on physical predictors. *Ecology* 93: 156-168.
- Evans-White, M.A., W.K. Dodds, D.G. Huggins et D.S. Baker. 2009. Thresholds in macroinvertebrate biodiversity and stoichiometry across water-quality gradients in Central Plains (USA) streams. *J N Am Benthol Soc* 28: 855-868.

- Feret, T. et C. Laplace-Treytoure. 2013. IPLAC : l'indice Phytoplancton Lacustre : Méthode de développement, description et application nationale 2012. Rapport convention Onema/Irstea 2012. Irstea, UR REBX, Bordeaux. p. 69.
- Groffman, P., J. Baron, T. Blett, A. Gold, I. Goodman, L. Gunderson, et al. 2006. Ecological thresholds: The key to successful environmental management or an important concept with no practical application? *Ecosystems* 9: 1-13.
- Holt, C.A., N.D. Yan et K.M. Somers. 2003. pH 6 as the threshold to use in critical load modeling for zooplankton community change with acidification in lakes of south-central Ontario: accounting for morphometry and geography. *Can J Fish Aquat Sci* 60: 151-158.
- Kail, J., J. Arle et S.C. Jaehnig. 2012. Limiting factors and thresholds for macroinvertebrate assemblages in European rivers: Empirical evidence from three datasets on water quality, catchment urbanization, and river restoration. *Ecol Indic* 18: 63-72.
- King, R.S. et M.E. Baker. 2010. Considerations for analyzing ecological community thresholds in response to anthropogenic environmental gradients. *J N Am Benthol Soc* 29: 998-1008.
- May, R.M. 1977. Thresholds and Breakpoints in Ecosystems with a Multiplicity of Stable States. *Nature* 269: 471-477.
- Pitcher, C.R., P. Lawton, N. Ellis, S.J. Smith, L.S. Incze, C.L. Wei, et al. 2012. Exploring the role of environmental variables in shaping patterns of seabed biodiversity composition in regional-scale ecosystems. *J Appl Ecol* 49: 670-679.
- Qian, S.S. 2014. Ecological threshold and environmental management: A note on statistical methods for detecting thresholds. *Ecol Indic* 38: 192-197.
- Qian, S.S., R.S. King et C.J. Richardson. 2003. Two statistical methods for the detection of environmental thresholds. *Ecological Modelling* 166: 87-97.
- Qian, S.S., Y.D. Pan et R.S. King. 2004. Soil total phosphorus threshold in the Everglades: a Bayesian changepoint analysis for multinomial response data. *Ecol Indic* 4: 29-37.
- Reynolds, C.S., V. Huszar, C. Kruk, L. Naselli-Flores et S. Melo. 2002. Towards a functional classification of the freshwater phytoplankton. *J Plankton Res* 24: 417-428.
- Roubeix, V., T. Feret, J.-M. Baudouin et P.-A. Danis. Soumis. A multivariate tree-based approach to identify ecological thresholds from variations in phytoplankton communities among French lakes. *Ecol Indic*.
- Scheffer, M., S.H. Hopper, M.L. Meijer, B. Moss et E. Jeppesen. 1993. Alternative Equilibria in Shallow Lakes. *Trends in Ecology & Evolution* 8: 275-279.
- Schlumberger, O. et P. Elie. 2008. Poissons des lacs naturels français - Ecologie des espèces et évolution des peuplements.
- Smith, A.J. et C.P. Tran. 2010. A weight-of-evidence approach to define nutrient criteria protective of aquatic life in large rivers. *J N Am Benthol Soc* 29: 875-891.
- Sonderegger, D.L., H. Wang, W.H. Clements et B.R. Noon. 2009. Using SiZer to detect thresholds in ecological data. *Frontiers in Ecology and the Environment* 7: 190-195.
- Soranno, P.A., K.S. Cheruvilil, R.J. Stevenson, S.L. Rollins, S.W. Holden, S. Heaton, et al. 2008. A framework for developing ecosystem-specific nutrient criteria: Integrating biological thresholds with predictive modeling. *Limnology and Oceanography* 53: 773-787.
- Strobl, C., A.L. Boulesteix, T. Kneib, T. Augustin et A. Zeileis. 2008. Conditional variable importance for random forests. *Bmc Bioinformatics* 9.
- Utz, R.M., R.H. Hilderbrand et D.M. Boward. 2009. Identifying regional differences in threshold responses of aquatic invertebrates to land cover gradients. *Ecol Indic* 9: 556-567.
- Wang, R., J.A. Dearing, P.G. Langdon, E. Zhang, X. Yang, V. Dakos, et al. 2012. Flickering gives early warning signals of a critical transition to a eutrophic lake state. *Nature* 492: 419-422.

Onema  
Hall C – Le Nadar  
5 square Félix Nadar  
94300 Vincennes  
01 45 14 36 00  
[www.onema.fr](http://www.onema.fr)

Irstea  
Parc de Tourvoie  
BP 44,  
92163 Antony cedex  
01 40 96 61 21  
[www.irstea.fr](http://www.irstea.fr)