



HAL
open science

Nouveaux indicateurs physico-chimiques soutenant la biologie en plans d'eau : Principes de construction et perspectives

V. Roubéix, P.A. Danis

► **To cite this version:**

V. Roubéix, P.A. Danis. Nouveaux indicateurs physico-chimiques soutenant la biologie en plans d'eau : Principes de construction et perspectives. [Rapport de recherche] irstea. 2015, pp.41. hal-02603633

HAL Id: hal-02603633

<https://hal.inrae.fr/hal-02603633>


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 2015

Etat écologique, biodiversité et fonctionnement des systèmes lacustres
Action 15



***Nouveaux indicateurs
physico-chimiques
soutenant la biologie
en plans d'eau :
Principes de construction
et perspectives***

Rapport final

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

Décembre 2015

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.

Ce rapport présente les avancées réalisées en 2014 et 2015 sur l'analyse des réponses des des compartiments biologiques phytoplancton, macrophytes et ichtyofaune aux éléments de qualité physico-chimiques et aux caractéristiques environnementales. Les résultats de ces travaux mettent en lumière des zones de rupture (appelée zone seuil dans ce rapport) dans les gradients inter-plans d'eau qui correspondent à des changements de composition des différentes communautés biologiques. Ce travail se focalise sur les relations écologiques entre les compartiments biologiques et l'ensemble des paramètres physico-chimiques généraux soutenant la biologie, et a comme particularité d'être pleinement indépendant des indicateurs biologiques ; à la différence des travaux présentés dans Danis et Roubeix (2014) et dont les résultats sont aujourd'hui utilisés pour l'évaluation de l'état écologique des plans d'eau DCE (Ministère de l'Ecologie et du Développement Durable 2015). Afin d'aboutir à des outils opérationnels, ces seuils doivent être (travaux prévus en 2016) confrontés aux valeurs de référence par plans d'eau ou type de plans d'eau pour, à terme, aboutir à une quantification pleinement DCE compatible des valeurs-seuils des paramètres physico-chimiques soutenant la biologie, *i.e.* des valeurs-seuils permettant d'assurer le bon fonctionnement et la préservation des écosystèmes, et l'atteinte du bon état des communautés biologiques.

Les auteurs

Vincent Roubex

Ingénieur de recherche (Vincent.Roubex@irstea.fr)

Iristea, UR RECOVER, équipe FRESHCO, Pôle études et recherches Onema-Iristea Hydroécologie des Plans d'Eau, 3275 route de Cézanne – CS 40061, 13182 Aix-en-Provence Cedex 5

Pierre-Alain Danis

Chargé de mission Onema (Pierre-Alain.Danis@onema.fr)

Onema, Pôle études et recherches Onema-Iristea Hydroécologie des Plans d'Eau, 3275 route de Cézanne – CS 40061, 13182 Aix-en-Provence Cedex 5

Les correspondants

Onema

Jean-Marc Baudoin, Pôle Onema-Iristea, Aix-en-Provence, Jean-Marc.Baudoin@onema.fr

Nolwenn Bougon, DAST Onema, Vincennes, Nolwenn.Bougon@onema.fr

Iristea

Christine Argillier, Iristea, Aix-en-Provence, Christine.Argillier@irstea.fr

Martin Daufresne, Iristea, Aix-en-Provence, Martin.Daufresne@irstea.fr

Référence du document : Roubex Vincent & Pierre-Alain Danis (2015) : *Nouveaux indicateurs physico-chimiques soutenant la biologie en plans d'eau : Principes de construction et perspectives*. Rapport final. Convention Onema/Iristea 2015. Pp 41.

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

NOUVEAUX INDICATEURS PHYSICO-CHIMIQUES SOUTENANT LA BIOLOGIE EN PLANS D'EAU :

PRINCIPES DE CONSTRUCTION ET PERSPECTIVES

Vincent Roubeix & Pierre-Alain Danis

1.	Introduction.....	5
2.	Méthodologie d'analyse des réponses des compartiments biologiques	7
2.1.	La notion de seuil écologique	7
2.2.	Les méthodes de détection de seuils écologiques	9
2.3.	La méthode de Gradient Forest	12
2.4.	Les données.....	16
2.4.1.	Les paramètres physico-chimiques	16
2.4.2.	Les descripteurs des compartiments biologiques	18
2.4.3.	Les caractéristiques environnementales.....	19
2.5.	Mise en œuvre du Gradient Forest	19
3.	Résultats du <i>Gradient Forest</i> pour les trois compartiments	20
3.1.	Phytoplancton	20
3.2.	Macrophytes.....	24
3.3.	Poissons	27
4.	Bilan et discussion	31
5.	Perspectives & Conclusions : Utilisation des seuils.....	36
6.	Références.....	39

Nouveaux indicateurs physico-chimiques soutenant la biologie en plans d'eau :

Principes de construction et perspectives

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 *a*. 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 écosystèmes¹ 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 (§**Erreur ! Source du renvoi introuvable.**) et discutés en fonction des réponses de certaines espèces (§4). 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 :

- la définition d'un seuil écologique (§2.1) ;
- 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) ;

¹ 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) et ceci plus particulièrement pour la transparence, le phosphore total, les nitrates, l'ammonium, et le carbone organique dissous.
- les conclusions en terme de proposition de valeurs-seuils (§4) ainsi que les perspectives de recherche prévues pour 2016 sur la précision des valeurs-seuils selon les conditions de référence et les niveaux de pressions anthropiques (§5).

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 (Figure 1). 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 (Figure 2), les macrophytes submergés sont alors inhibés par le manque de lumière et le lac passe dans un état turbide.

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). De manière plus qualitative, les changements rapides dans l'écosystème peuvent aussi s'accompagner de remaniements dans la composition des communautés. Par exemple, une étude paléolimnologique de 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 (Figure 3). Il est donc possible de repérer un seuil écologique en observant les variations d'abondances des espèces.

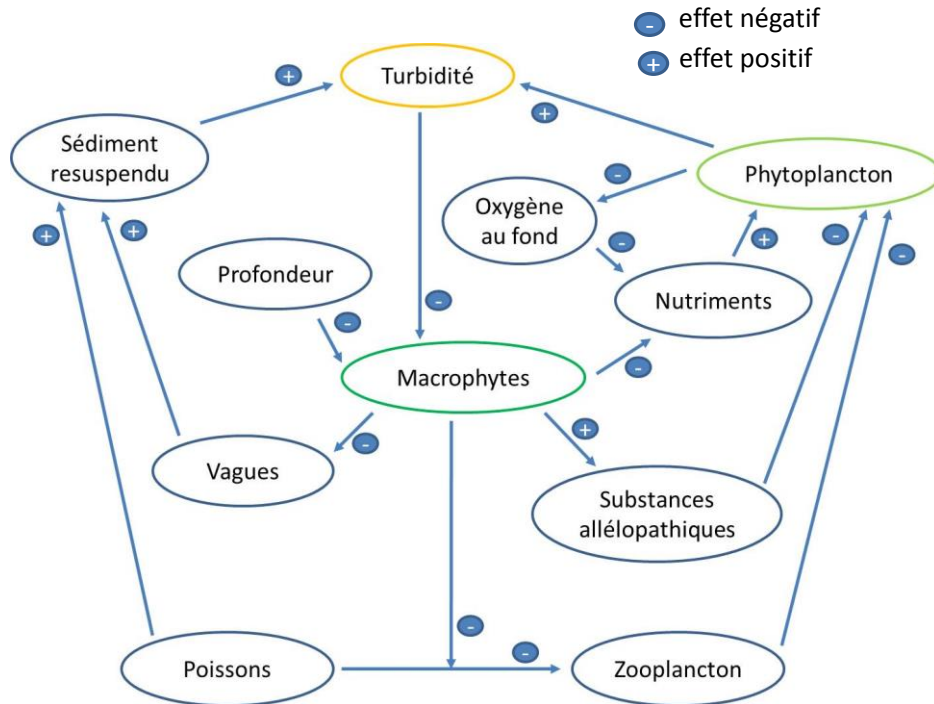


Figure 1. Interactions dans un écosystème de lac peu profond expliquant la stabilité des états 'turbides' et 'riches en macrophytes' (d'après Scheffer et al 1993).

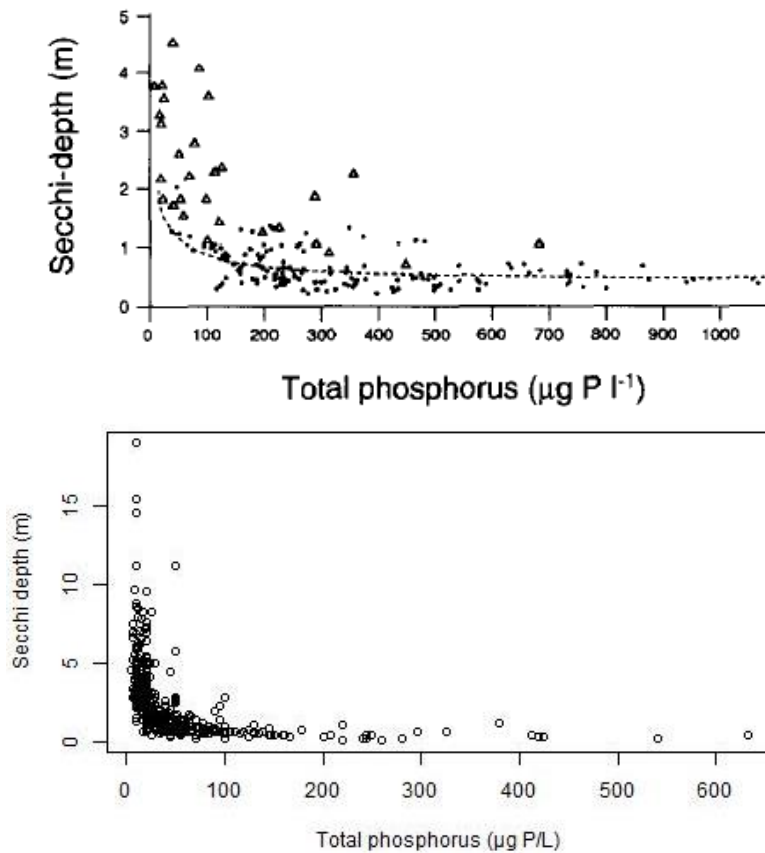


Figure 2. Relation entre la concentration en phosphore dans des lacs et la transparence de l'eau d'après Scheffer et al 1993 (en haut) et d'après les données en zone euphotique de suivi DCE des plans d'eau nationaux (en bas).

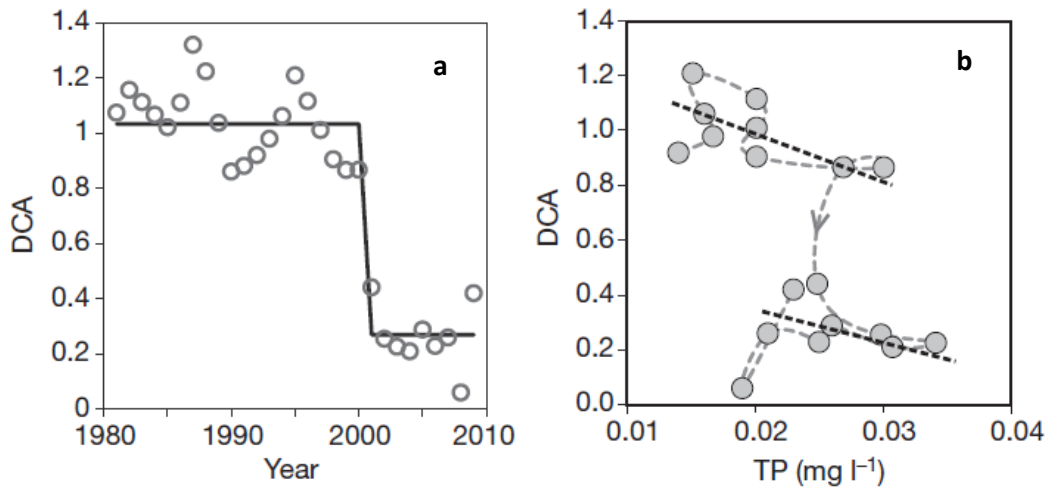


Figure 3. Evolution de la composition de la communauté de diatomées dans un lac de Chine, et lien avec la concentration en phosphore total (d'après Wang et al 2012). Les variations de composition de la communauté sont représentées par l'axe d'une analyse multivariée (DCA). La trajectoire des variations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

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

Parmi ces méthodes, l'analyse de la variance non paramétrique est à la fois une méthode en soit et aussi la base d'autres méthodes plus complexes dont celle utilisée pour cette étude (Gradient Forest, cf §2.3). 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 (Figure 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.

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

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

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

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

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

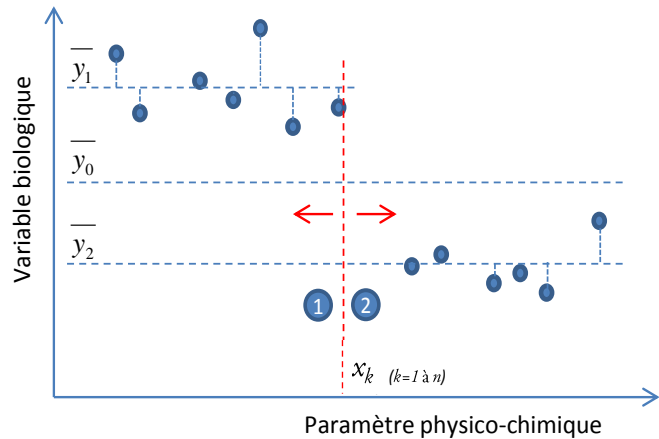


Figure 4. Mode de détermination d'un seuil par la méthode d'analyse de la variance non paramétrique. La méthode consiste à rechercher la position du seuil dans le gradient (x_k) qui sépare les groupes 1 et 2, et pour laquelle la somme des sommes des carrés intragroupes (SC_{intra}) est minimale, ce qui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

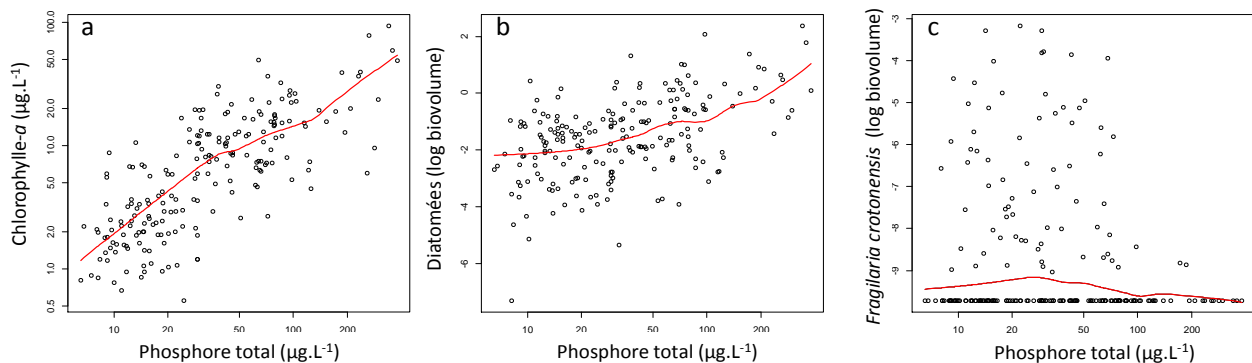


Figure 5. Effet de l'agrégation de variables biologiques sur la détection de seuil écologiques. Lorsque la biomasse totale du phytoplancton est considérée (a), la relation au phosphore apparaît linéaire. Si l'on ne considère que les diatomées du phytoplancton (b), la relation est moins linéaire avec 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 $\mu\text{g.L}^{-1}$.

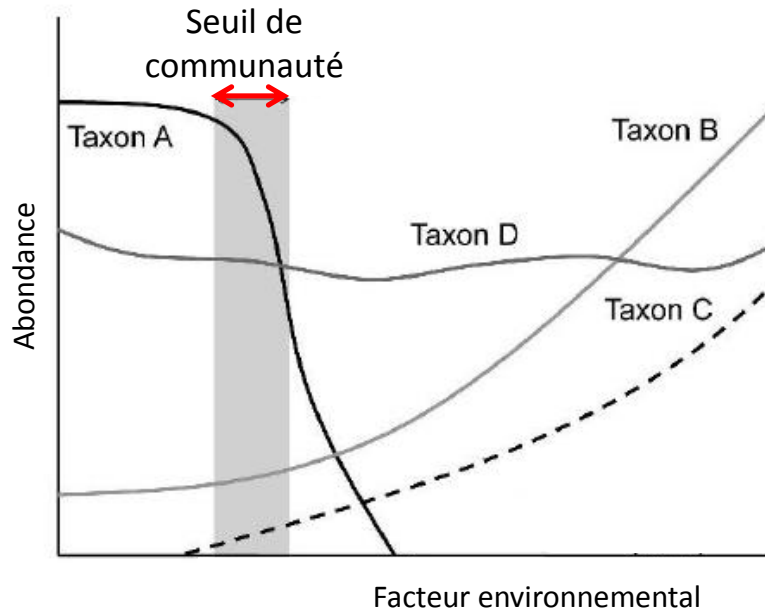


Figure 6. *Graphique conceptuel illustrant la notion de seuil de communauté avec quatre taxons (d'après King et Baker 2010). Au niveau du seuil de communauté, l'abondance du taxon A baisse brutalement, le taxon B commence à augmenter en abondance, le taxon C apparaît dans le milieu et se développe alors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

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) (Figure 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...).

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) (Figure 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. Il est possible que l'erreur mesurée $\sum_i (Y_i - \hat{Y}_i)^2$ dépasse la variance observée ce qui entraîne une valeur de R^2 négative. Les espèces concernées ne contribuent ni à l'estimation de l'importance des variables ni à l'analyse des coupures dans les gradients.

Pour des données binaires (présence/absence) :

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

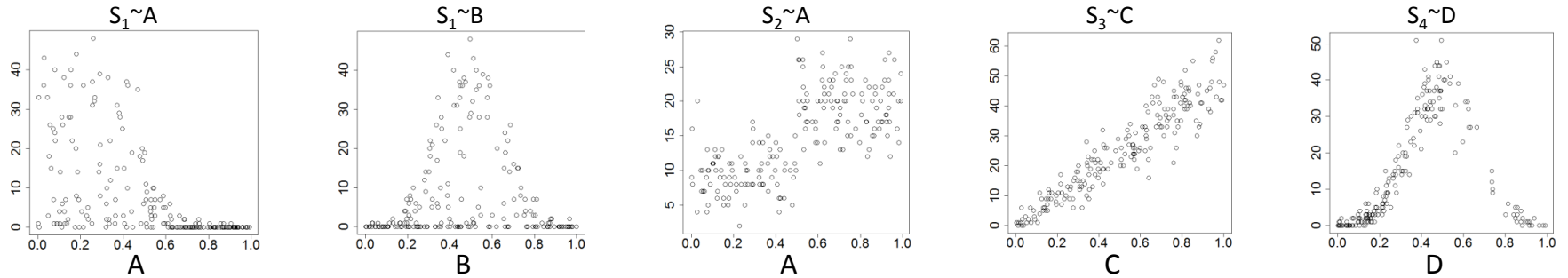
où *ErrMod* correspond au taux moyen d'erreur de classification des arbres dans les échantillons de validation, et *ErrBase* est le taux d'erreur de base : $ErrBase = 2p(1 - p)$ où p est la proportion des sites où l'espèce considérée est présente.

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) 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) indiquent la localisation probable de seuils de communauté (Figure 7). Une autre représentation consiste à sommer les importances des coupures dans le sens du gradient et à représenter cette somme sous forme de courbe d'importance cumulée. Un seuil apparaît alors comme une marche dans la courbe ou un tronçon de grande pente. Cette représentation a l'avantage de se dispenser de la calibration d'une bande passante pour les courbes de densité et elle permet aussi de superposer sur un même graphe les courbes pour chaque espèce (Figure 8). Les courbes superposées des espèces mettent en évidence les seuils de communautés où s'accumulent les composantes verticales des courbes et permettent aussi d'identifier le nombre et éventuellement l'identité des espèces qui contribuent à un seuil.

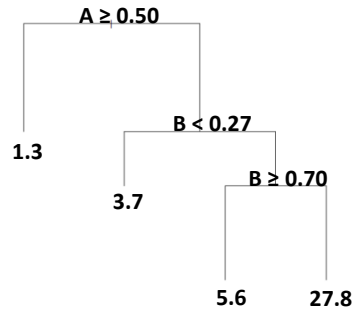
L'application de la méthode de Gradient Forest appliquée au compartiment phytoplancton a fait l'objet d'une publication publiée en 2015 (Roubeix *et al.* Sous presse).

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 : $S_1 = f(A,B,C,D,E)$



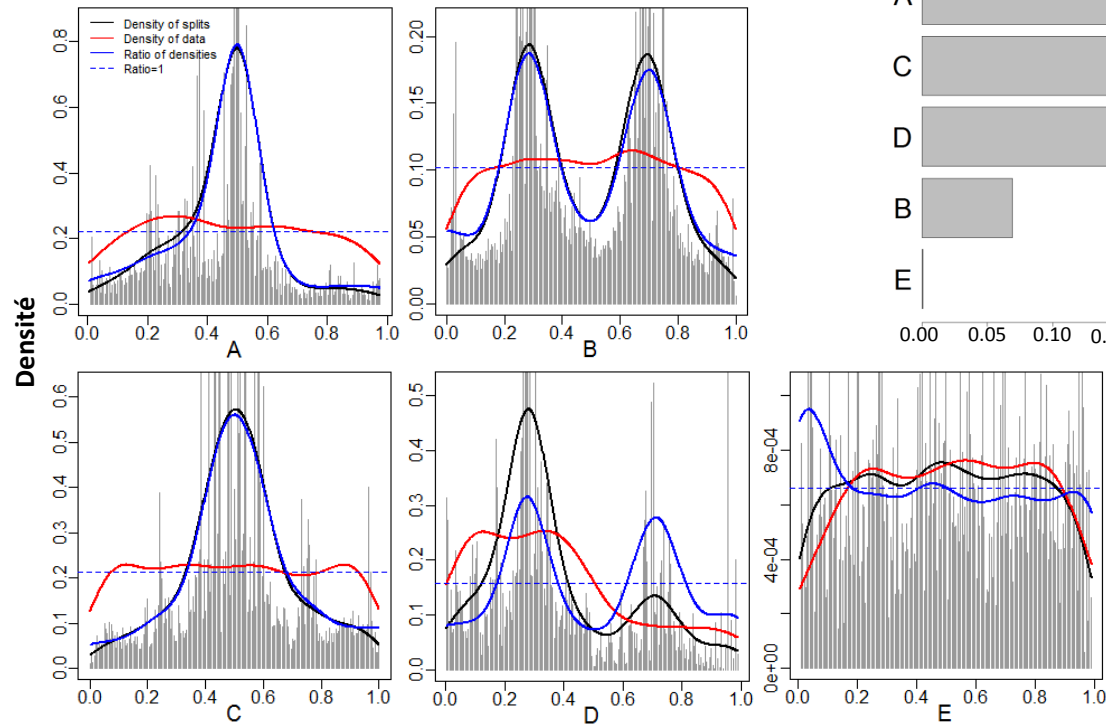
3) × n arbres => forêt aléatoire

- 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

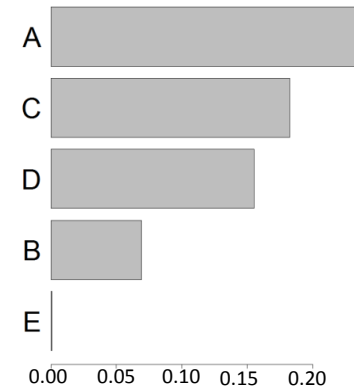


Figure 7. (page précédente) Exemple de *Gradient Forest* appliqué à des données simulées de cinq variables environnementales et d'abondances de quatre espèces. (1) Relations simulées pour cinq couples espèce-variable, suivant des fonctions logistiques, gaussiennes, en marche d'escalier ou linéaires. 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.

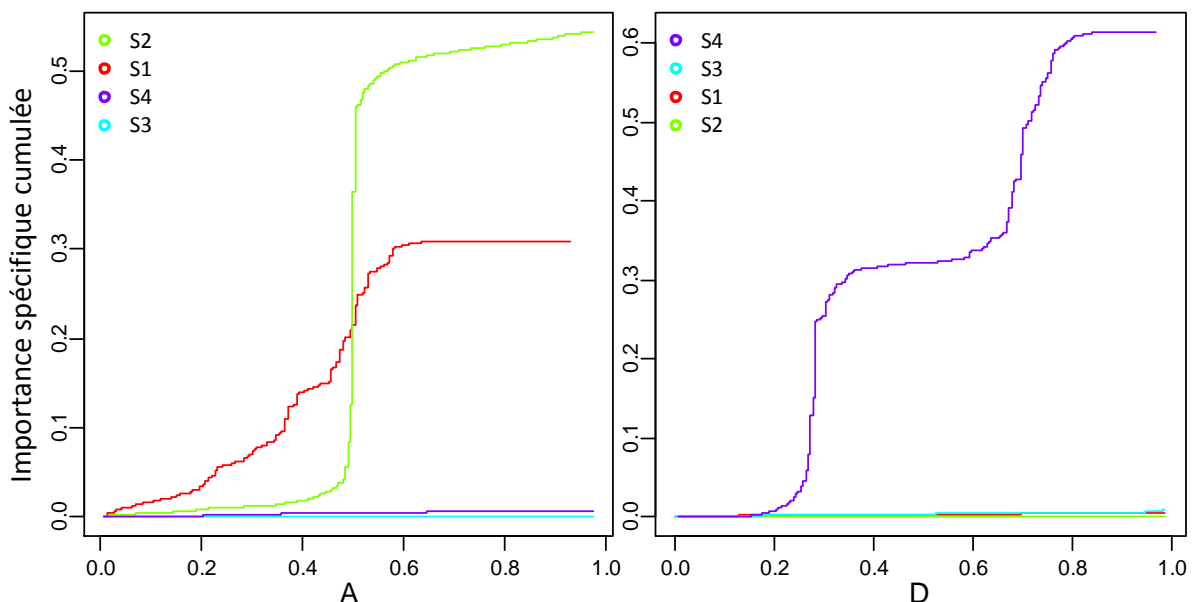


Figure 8. Représentation des résultats du Gradient Forest sous forme de courbes d'importance spécifique cumulée pour les facteurs A et D.

2.4. Les données

2.4.1. Les paramètres physico-chimiques

En 2015, 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 2).

Les valeurs utilisées sont des concentrations médianes, minimales ou maximales annuelles en zone euphotique ou au fond pour l'oxygène et la température. 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 2. Bilan du nombre de plans d'eau suivis par agences de bassin dont les données ont pu être analysées en 2015.

Agences	Nombres de plans d'eau DCE	Années de suivi disponibles	Nombres de plans d'eau utilisés en 2014
AG	105	2005-2013	84
AP	5	2007-2012	5
LB	166	2005-2013	133
RM	29	2007-2012	18
RMC	115	2004-2012	107
SN	58	2008-2012	44

Tableau 3. 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=225)			Macrophytes (n=132)			Poissons (n=196)		
			min	max	med	min	max	med	min	max	med
Secchi	m	Oui	0,1	18,6	1,5	0,3	7,6	1,5	0,2	7,5	1,8
Secchim	m	Oui	0,1	12	0,9	0,1	6,7	0,8	0,1	6,7	1,0
Temp	°C	Non	7,1	22,8	16,9	9,9	21,1	17,3	9,1	22,2	16,4
Tempmin	°C	Non	0,30	14,5	6,9	1,5	13,2	7,1	1,1	18,6	6,0
Tempmax	°C	Non	8,9	29,2	21,5	12,3	33,1	22,6	12,3	27,1	21,8
TempB	°C	Non							4,1	20,0	11,3
TempBmin	°C	Non							0,3	14,2	5,4
TempBmax	°C	Non							4,2	27,0	16,6
SatO2	%	Non	53,9	126,9	94,9	63,2	140,5	95,5	60,6	158,2	95,0
SatO2min	%	Non	15,2	114,3	78,2	13,8	111,0	73,9	15,2	107,1	74,5
pH		Non	5,7	9,9	8,1	6,5	9,4	8,1	6,2	8,9	8,1
pHmin		Non	5,1	9,1	7,5	4,6	8,6	7,4	5,1	8,4	7,4
pHmax		Non	6,9	12,5	8,6	7,6	10,9	8,8	6,5	10,7	8,6
Cond	µs.cm ⁻¹	Oui	14	1307	241	16	1280	277	16	791	245
Alk	meq.L ⁻¹	Oui	0,04	4,6	1,0	0,04	4,5	1,40	0,1	4,6	1,0
DOC	mg.L ⁻¹	Oui	0,35	36,7	4,9	1,0	31,0	5,2	0,4	22,3	4,5
DOCmax	mg.L ⁻¹	Oui	0,4	56,0	6,0	1,6	39,0	7,0	0,6	38,0	5,7
NH4	µg.L ⁻¹	Oui	9	389	50	6	461	52	12	851	65
NH4max	µg.L ⁻¹	Oui	10	3300	130	26	2450	190	21	2450	150
NO3	mg.L ⁻¹	Oui	0,1	39,3	1,2	0,1	35,4	1,1	0,1	42,2	1,9
NO3max	mg.L ⁻¹	Oui	0,25	72,0	4	0,13	54	3,8	0,4	72,0	5,2
NO2	µg.L ⁻¹	Oui	4	229	18	6	340	17	3,5	260	24
NO2max	µg.L ⁻¹	Oui	6	690	50	6	940	56	6	940	60
TP	µg.L ⁻¹	Oui	6	599	30	6	425	29	6	425	27
TPmax	µg.L ⁻¹	Oui	7	1490	46	11	1490	75	10	1410	45
PO4	µg.L ⁻¹	Oui	6	972	16	6	755	16	6	515	16
PO4max	µg.L ⁻¹	Oui	6	2116	40	6	2530	42	6	1594	40
SiO2	mg.L ⁻¹	Oui	0,2	20,8	3,3	0,2	23,4	3,1	0,2	20,8	3,4

2.4.2. Les descripteurs 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 variance 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 apparaissant dans chacun des plans d'eau ont été utilisées. Les taxons non déterminés à l'espèce et ceux apparaissant dans moins de 5% des plans d'eau n'ont pas été pris en compte. Ainsi, 193 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. Un total de 225 plans d'eau tous types confondus a été considéré dans un premier temps.

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. 118 taxons ont répondu à ce critère. L'analyse a porté sur 132 plans d'eau, tous types confondus, mais caractérisés par un marnage inférieur à 2 m (contrainte pour l'échantillonnage des macrophytes (Boutry *et al.* 2013)).

c) L'ichtyofaune

Les données de biomasse par unité d'effort (BPUE) 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 3,5% a également été utilisée pour les espèces rares, ce qui a conduit à ne garder que 25 espèces pour le *Gradient Forest* (Tableau 4). Les plans d'eau artificiels (gravières, étangs de pisciculture) n'ont pas été pris en compte parce que les stocks de poissons présents dans ce type de plans d'eau sont généralement trop fortement manipulés (empoissonnements, pêches) pour pouvoir refléter correctement les conditions environnementales. Le nombre de plans d'eau considéré s'élève toutefois à 196.

Tableau 4. Codes, biomasse maximale et pourcentage d'occurrences des principales espèces de poissons dans les plans d'eau échantillonnés.

Code espèce	Nom latin	BPUE max	Occurrences (%)
ABRABR	Abramis brama	3890	75
ALBUAL	Alburnus alburnus	210	39
AMEIME	Ameiurus melas	1856	25
BARBBR	Barbatula barbatula	9	5

BLICBJ	Blicca bjoerkna	5215	53
CARACA	Carassius carassius	1167	12
CORESP	Coregonus sp	491	7
CYPRCA	Cyprinus carpio	2342	36
ESOXLU	Esox lucius	749	62
GOBIGO	Gobio gobio	1475	15
GYMNCE	Gymnocephalus cernuus	477	56
LEPOGI	Lepomis gibbosus	45	32
LEUCDE	Leucaspis delineatus	232	5
LEUCLE	Leuciscus leuciscus	57	4
ONCOMY	Oncorhynchus mykiss	853	9
PERCFL	Perca fluviatilis	3605	93
PHOXPH	Phoxinus phoxinus	49	4
RUTIRU	Rutilus rutilus	4633	95
SALMTR	Salmo trutta	1455	12
SALVUM	Salvelinus umbla	135	4
SANDLU	Sander lucioperca	1907	75
SCARER	Scardinius erythrophthalmus	2120	75
SILUGL	Silurus glanis	1500	18
SQUACE	Squalius cephalus	1452	28
TINCTI	Tinca tinca	1296	38

2.4.3. Les caractéristiques environnementales

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 :

- Zmax, la profondeur maximale du plan d'eau (en m),
- 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 du contour du plan d'eau.
- 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

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. Pour plus de souplesse dans la présentation des résultats, les figures produites par le package n'ont pas été utilisées. D'autres figures ont été créées sous R à partir des résultats numériques de l'analyse.

3. Résultats du *Gradient Forest* pour les trois compartiments

3.1. *Phytoplancton*

L'analyse a d'abord été menée sur le phytoplancton provenant de 225 lacs répartis sur l'ensemble du territoire français. Un peu plus de 140 espèces ont présenté un R^2 positif avec une moyenne de 0,2 (Figure 9). Les abondances de sept espèces étaient particulièrement bien expliquées ($R^2 > 0.5$) par les 29 variables environnementales considérées. L'estimation de l'importance des variables a révélé une très large prépondérance des variables géographiques (latitude et longitude), ce qui indique une forte structuration des communautés de phytoplancton dans l'espace (influence des zones montagneuses). L'examen des coupures dans les gradients géographiques a montré un seuil très net à environ 5°E de longitude et un seuil bien marqué à environ 44°N de latitude (Figure 10). De manière à ne pas mélanger dans une même analyse des seuils correspondants à des flores trop différentes, il a été décidé de limiter l'analyse à la région nord-ouest de la France située au nord de la latitude 44,5°N et à l'ouest de la longitude 4,5°E (Figure 11). La recherche de seuil écologique a porté ainsi sur un nombre toutefois conséquent de plans d'eau (130) et sur 147 espèces réparties de façon plus homogène que lorsque l'ensemble du territoire était considéré.

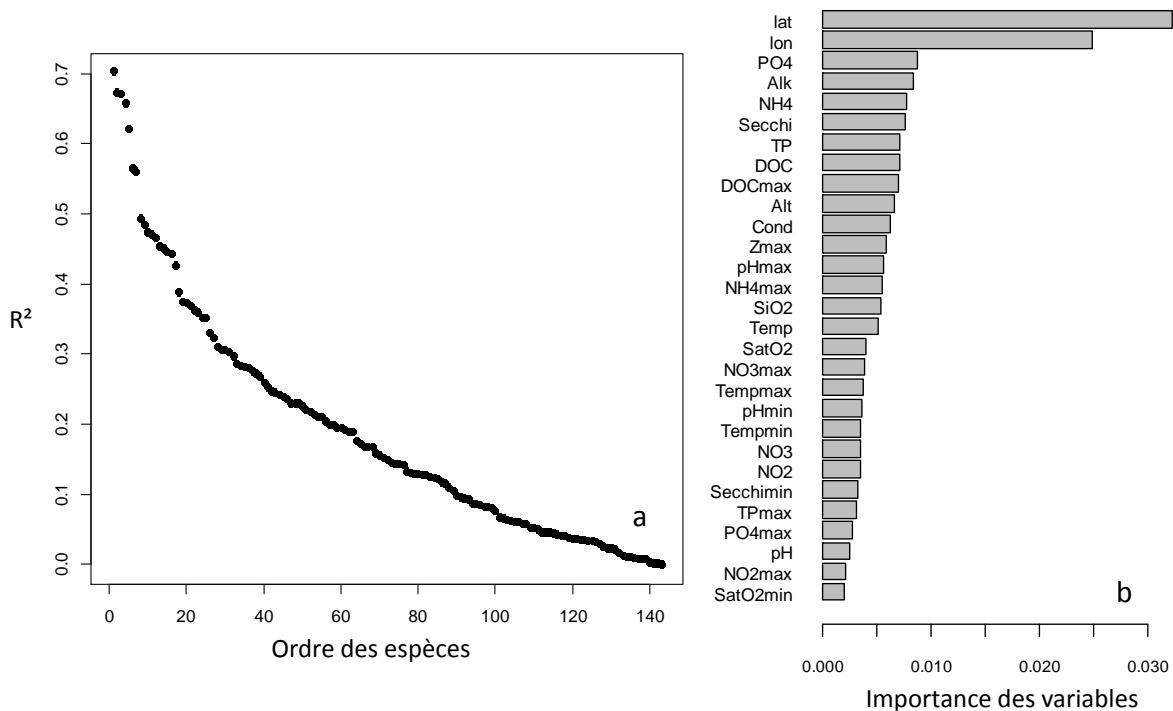


Figure 9. Performances des forêts aléatoires (R^2) générées pour chaque espèce de phytoplancton (a) et importances globales des variables explicatives pour l'ensemble du compartiment phytoplancton, exprimées en composantes du R^2 moyen des espèces (b).

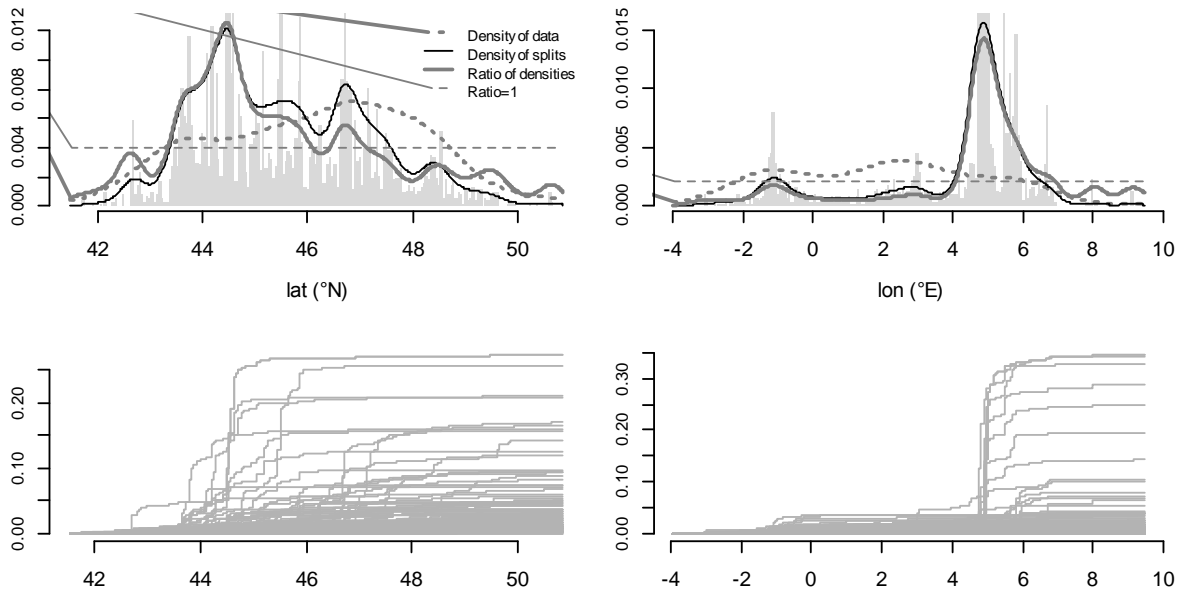


Figure 10. Résultats du *Gradient Forest* pour le phytoplancton concernant les variables géographiques latitude et longitude. En haut, densités des coupures (courbe noire fine), densité des observations (courbe discontinue) et ratio des densités (courbe grise épaisse). En bas, importances cumulées spécifiques pour chaque espèce de phytoplancton.

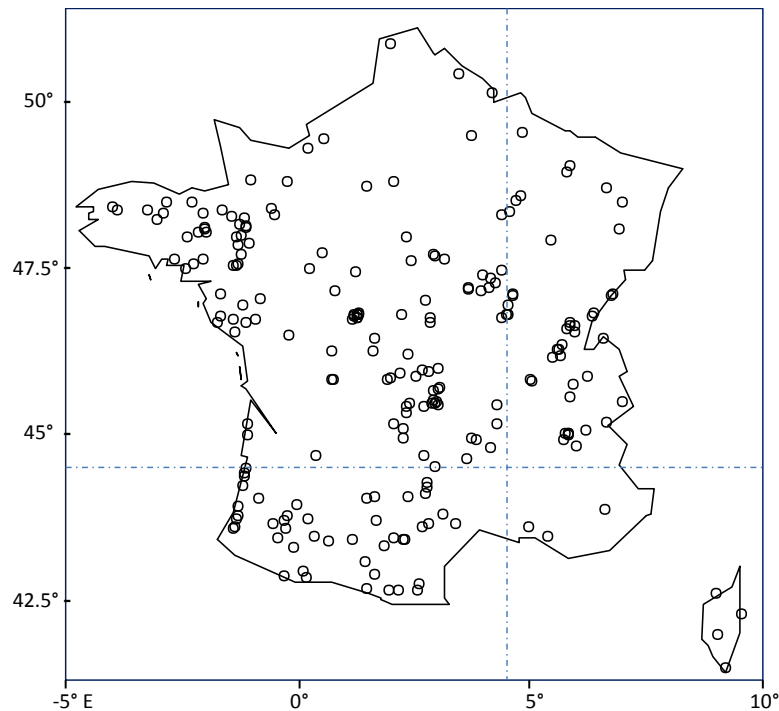


Figure 11. Carte des plans d'eau considérés pour l'analyse du phytoplancton. Les lignes discontinues indiquent la latitude 44,5° N et la longitude 4,5° E.

La deuxième analyse de *Gradient Forest* a pris en compte 75 espèces ayant un R^2 positif, avec un R^2 moyen de 0,16 (Figure 12). Dans ce jeu de données restreint, les variables géographiques ont eu une importance plus modérée. La variable la plus importante était la transparence (Secchi), suivie par, à

des niveaux similaires, l'alcalinité, la latitude, le phosfore total. Puis on trouve notamment dans l'ordre d'importance décroissante, l'ammonium, les nitrates (concentration maximale) et le carbone organique dissous.

Les densités de coupures et les seuils identifiés sont ensuite présentés pour les variables les plus importantes d'un point de vue de l'analyse et d'un point de vue écologique (Figure 13 à Figure 15). Pour la variable Secchi, la zone seuil s'est révélée large dans le gradient. Le graphe d'importances spécifiques cumulées permet de délimiter cette zone entre 0,2 et 1 m (Figure 13). Le graphe de densité indique toutefois un pic d'effet plus net de la transparence autour de 0,4 m. En ce qui concerne le phosphore total, un seuil assez net est apparu autour de $50 \mu\text{g.L}^{-1}$, marqué par la réponse d'un certain nombre d'espèces comme en témoignent les importances spécifiques cumulées. On peut aussi noter la présence d'un autre seuil plus diffus se situant entre 100 et $200 \mu\text{g.L}^{-1}$. Pour les nutriments azotés, des seuils ont pu également être repérés (Figure 14). Un seuil d'ammonium a été détecté autour de $120 \mu\text{g.L}^{-1}$. Un autre seuil plus léger est décelable à environ $60 \mu\text{g.L}^{-1}$. Les pics du ratio de densité aux extrémités du gradient ne sont pas suffisamment fiables pour être pris en compte (cf discussion §4). Quant aux nitrates, les résultats sont plus clairs. Un seuil unique en fin de gradient est mis en évidence par l'analyse en considérant la valeur maximale annuelle (NO_3max). Le graphe de densité suggère un début d'effet à partir de 10mg.L^{-1} mais les importances spécifiques cumulées montrent que les effets les plus importants se situent plutôt à partir de 20mg.L^{-1} . Enfin, un effet significatif du carbone organique dissous sur les espèces de phytoplancton n'a été décelé qu'à partir de 10mg.L^{-1} (Figure 15). Deux seuils ont pu être identifiés, un premier à 10mg.L^{-1} et, de façon moins précise, un second autour de 20mg.L^{-1} .

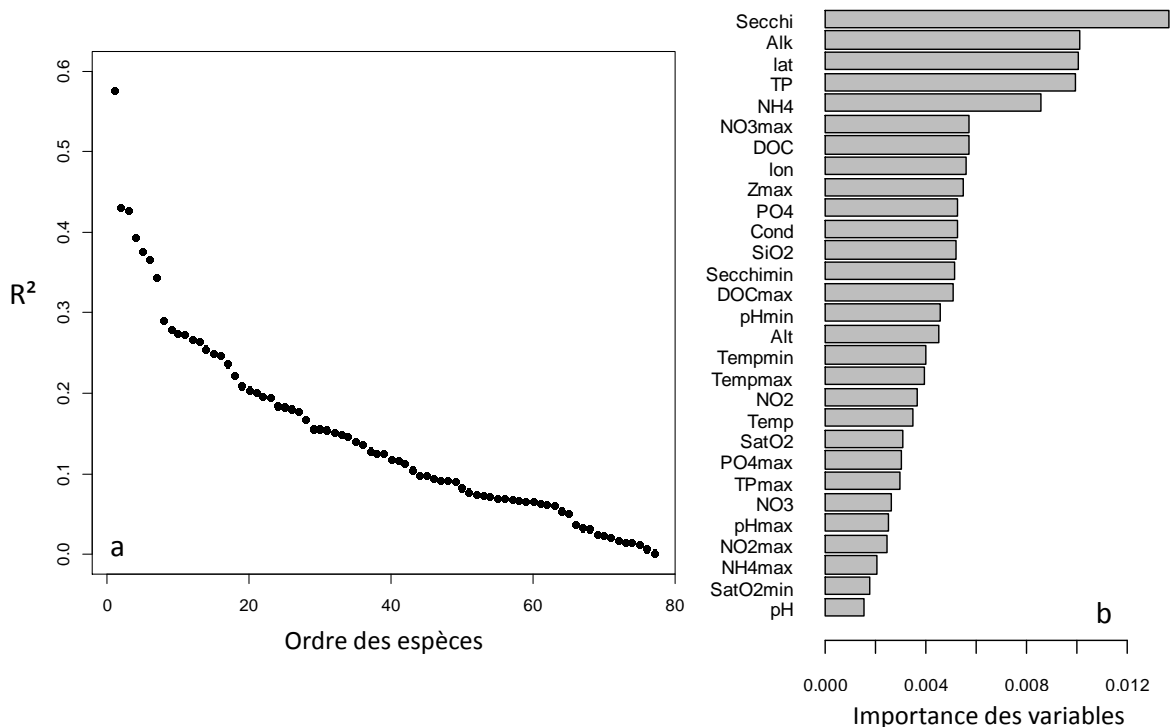


Figure 12. Résultats de l'analyse restreinte à la partie nord-ouest de la France. Performances des forêts aléatoires (R^2) générées pour chaque espèce de phytoplancton (a) et importances globales des variables explicatives pour l'ensemble du compartiment phytoplancton, exprimées en composantes du R^2 moyen des espèces (b).

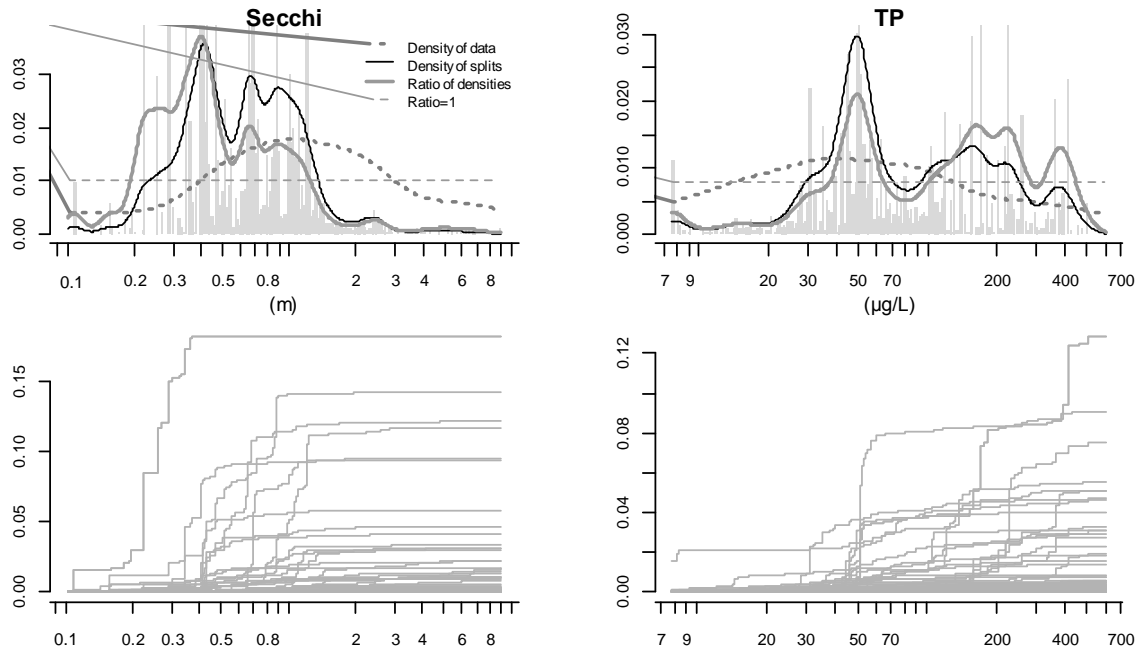


Figure 13. Résultats de l'analyse restreinte à la partie nord-ouest de la France pour les variables Secchi et TP. En haut, densités des coupures (courbe noire fine), densité des observations (courbe discontinue) et ratio des densités (courbe grise épaisse). En bas, importances cumulées spécifiques pour chaque espèce de phytoplancton.

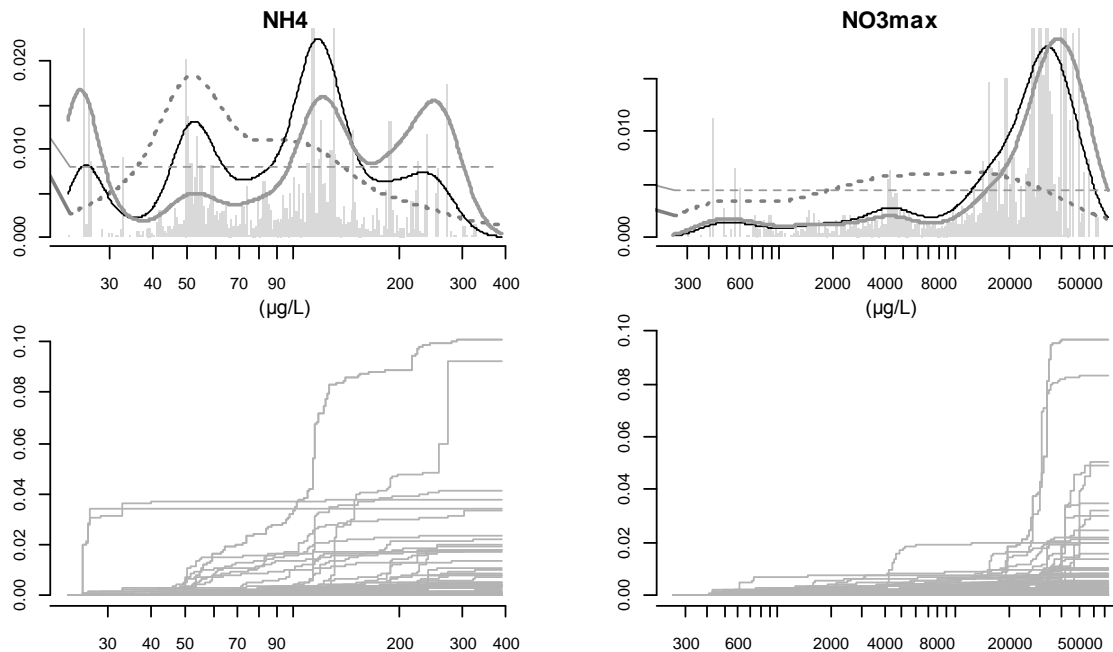


Figure 14. Résultats de l'analyse restreinte à la partie nord-ouest de la France pour les variables NH4 et NO3max. En haut, densités des coupures (courbe noire fine), densité des observations (courbe discontinue) et ratio des densités (courbe grise épaisse). En bas, importances cumulées spécifiques pour chaque espèce de phytoplancton.

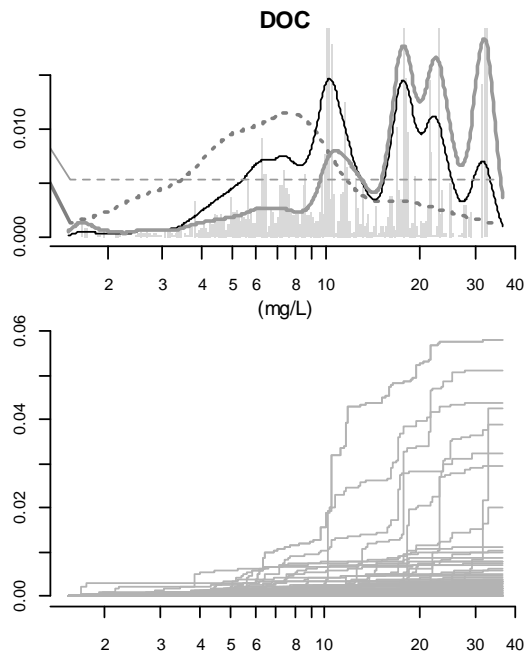


Figure 15. Résultats de l'analyse restreinte à la partie nord-ouest de la France pour la variable DOC. En haut, densités des coupures (courbe noire fine), densité des observations (courbe discontinue) et ratio des densités (courbe grise épaisse). En bas, importances cumulées spécifiques pour chaque espèce de phytoplancton.

3.2. Macrophytes

Les valeurs de R^2 obtenues avec les macrophytes ne sont pas directement comparables aux valeurs des autres compartiments biologiques parce qu'elles sont issues de données en présence/absence et non en abondances. Dans notre cas, l'ensemble des taxons ont eu un R^2 positif sauf un (117 taxons) et les valeurs étaient assez peu variables autour d'une moyenne de 0,36 (Figure 16). Les deux variables les plus importantes étaient la longitude et l'altitude, reflétant comme pour le phytoplancton, une hétérogénéité du territoire français et de la répartition des espèces. Même si ces deux variables se distinguaient nettement des autres, le faible nombre de plans d'eau avec des données macrophytes et le manque de seuils bien définis pour ces variables n'a pas conduit à restreindre l'analyse à une zone géographique comme pour le phytoplancton. L'importance des autres variables décroît ensuite doucement à partir de celle de l'ammonium. On note l'importance pour les macrophytes du paramètre morphologique DLIac, car certains taxons semblent avoir une préférence pour les plans d'eau de contour plutôt rond, qui sont principalement des lacs naturels en France .

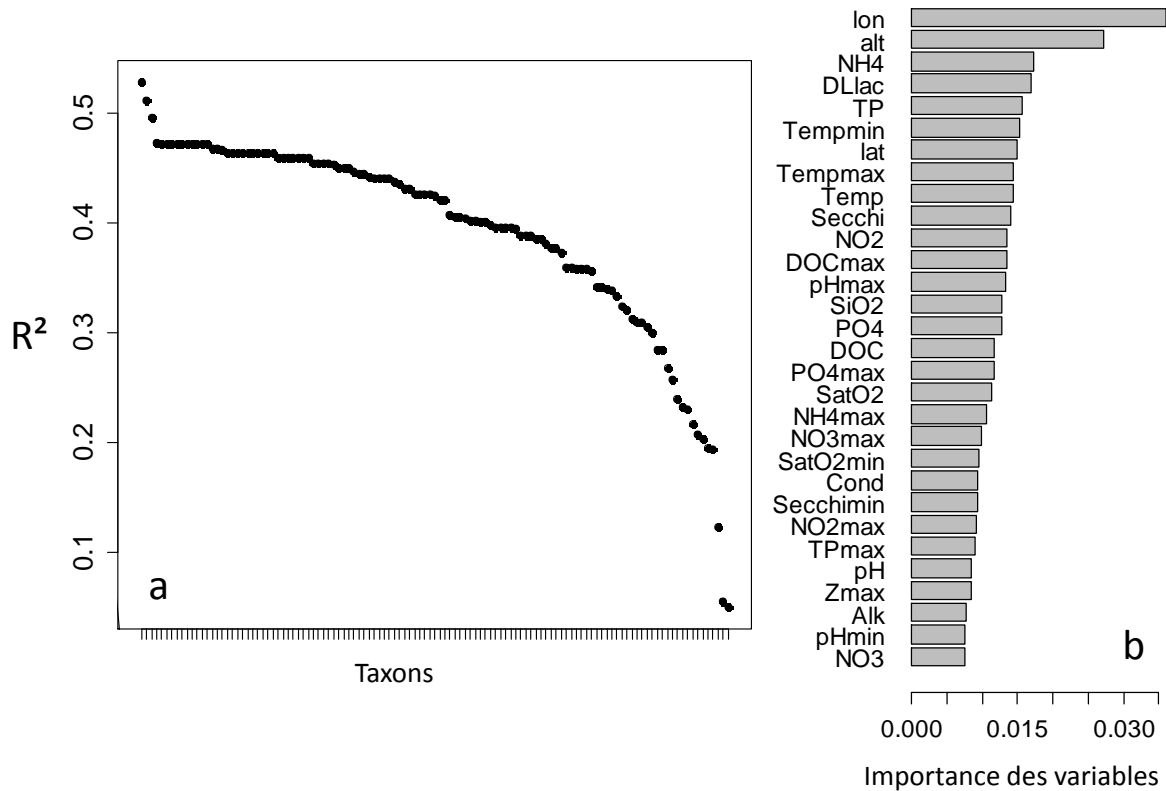


Figure 16. Performances des forêts aléatoires (R^2) générées pour chaque macrophyte (a) et importances globales des variables explicatives pour l'ensemble du compartiment macrophytes, exprimées en composantes du R^2 moyen des espèces (b).

Seulement quatre variables d'intérêt pour la gestion des plans d'eau ont présenté des seuils. Il s'agit par ordre d'importance, de la température minimale annuelle, de la concentration en phosphore total, de la transparence de l'eau et de la concentration maximale en nitrates (Figure 17 et Figure 18). Pour Tempmin, un seuil assez large se dégage autour de 5-6°C et un autre plus léger autour de 10°C. Dans le gradient de TP, un pic de densité de coupures autour de 12 $\mu\text{g.L}^{-1}$ coïncide avec des sauts de courbes d'importances spécifiques (Figure 17) et marque un seuil important à considérer. Un deuxième pic moins important apparaît un peu au-dessus de 20 $\mu\text{g.L}^{-1}$. Pour la transparence, un seuil ressort clairement entre 3 et 4,5 m de profondeur de Secchi. Enfin, on peut compter deux seuils dans le gradient de NO3max. Le premier se situe autour de 3 mg.L^{-1} et le second plus diffus s'étend entre 10 et 50 mg.L^{-1} avec toutefois un pic entre 10 et 20 mg.L^{-1} .

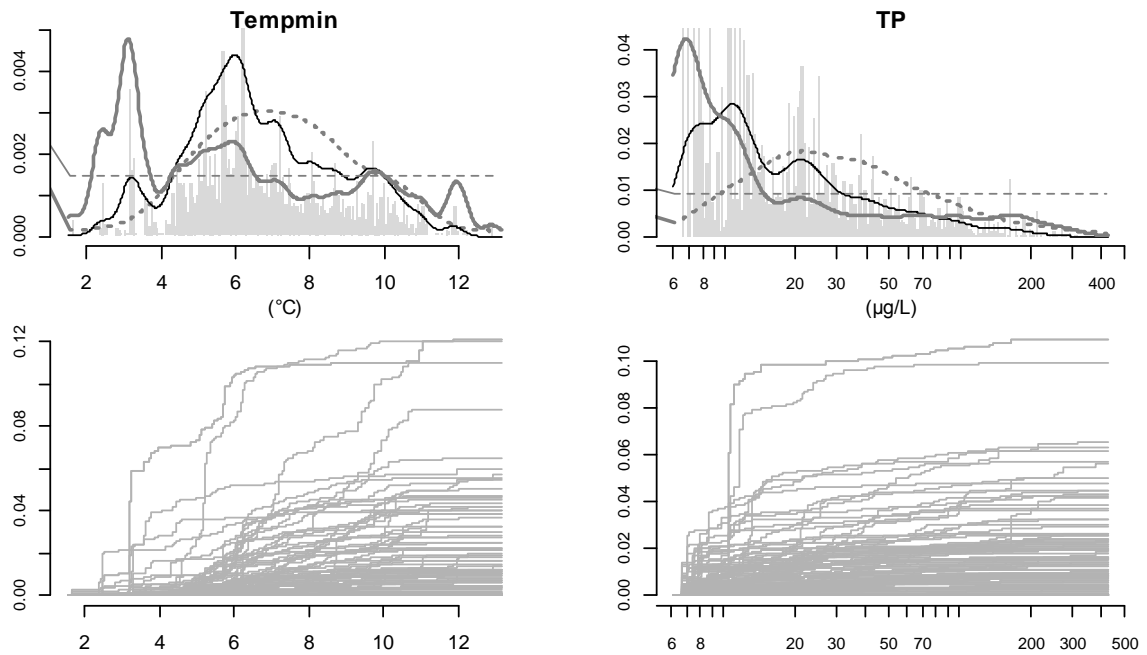


Figure 17. Résultats du *Gradient Forest* pour les macrophytes concernant les variables Tempmin et TP. En haut, densités des coupures (courbe noire fine), densité des observations (courbe discontinue) et ratio des densités (courbe grise épaisse). En bas, importances cumulées spécifiques pour chaque taxon de macrophytes.

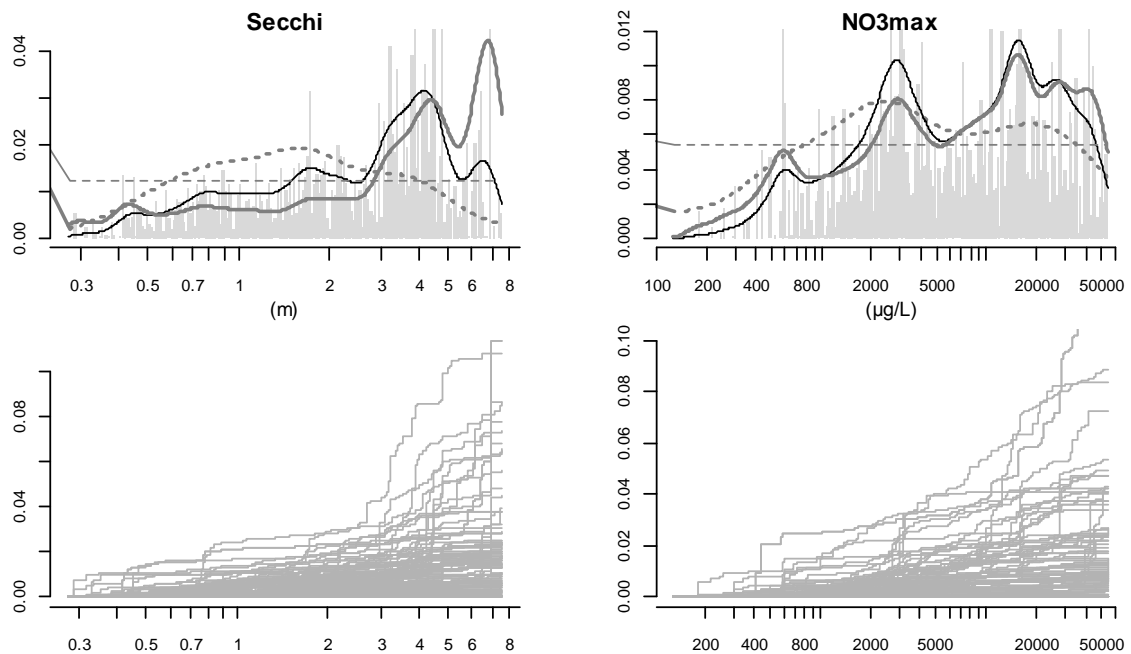


Figure 18. Résultats du *Gradient Forest* pour les macrophytes concernant les variables Secchi et NO3max. En haut, densités des coupures (courbe noire fine), densité des observations (courbe discontinue) et ratio des densités (courbe grise épaisse). En bas, importances cumulées spécifiques pour chaque taxon de macrophytes.

3.3. Poissons

Sur les 25 espèces de poissons prises en compte dans l'analyse, 21 ont présenté un R^2 positif avec un étalement des valeurs assez régulier entre 0 et 0,5 (Figure 19). La variable qui s'est avérée la plus importante est la température maximale annuelle (Tempmax). Puis dans l'ordre décroissant, on retrouve les variables géographiques et d'autres variables d'intérêt pour la gestion des plans d'eau, telles que le carbone organique dissous et la transparence. Les nutriments sont apparus relativement peu importants comme en témoigne les positions de NO3 et TP dans l'histogramme.

Etant donné le faible nombre d'espèces de poissons, on peut examiner facilement l'importance des variables pour certaines espèces individuellement. La Figure 20 montre ainsi pour les douze espèces ayant les R^2 les plus élevés, les importances spécifiques des variables les plus importantes au niveau du compartiment ainsi que les importances des nitrates et du phosphore total. On peut ainsi repérer aisément les espèces dont l'abondance est déterminée principalement par des variables géographiques (CORESP, BLICBJ ou SQUACE), par la température (SALMTR, AMEIME) ou par la transparence (ABRABR, TINCTI).

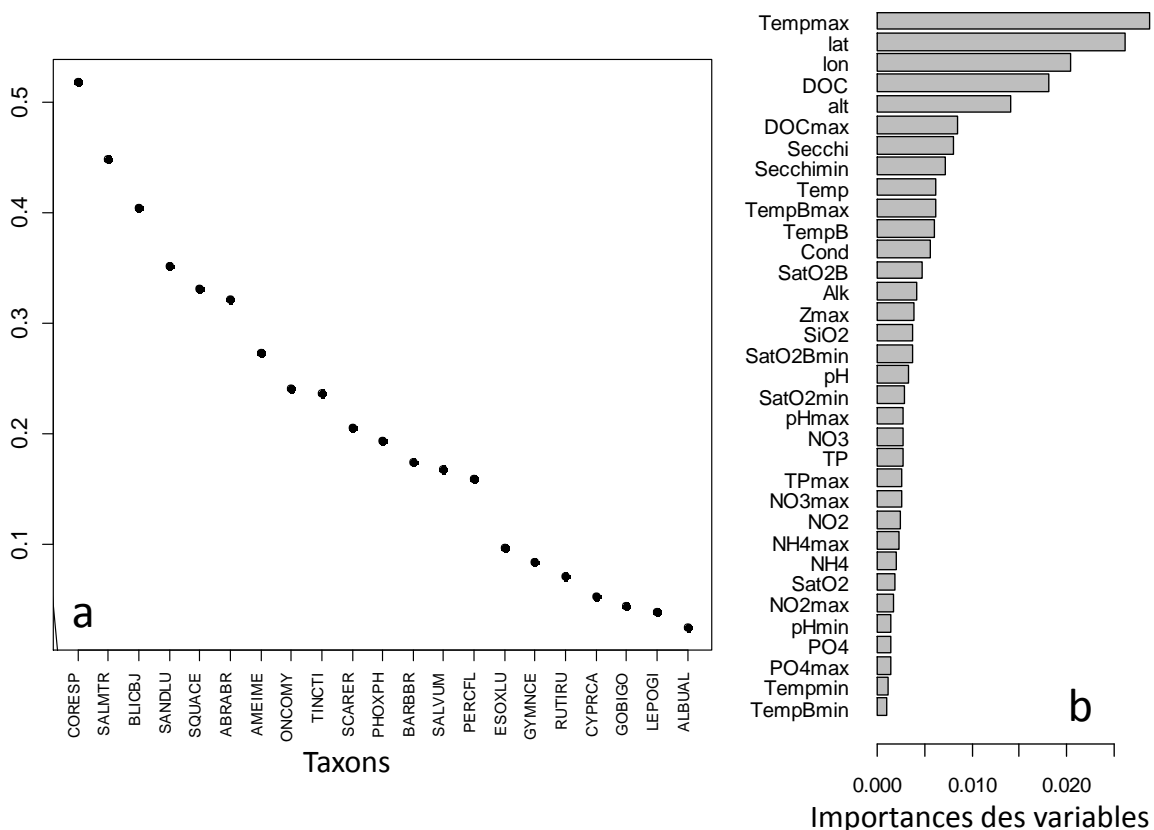


Figure 19. Performances des forêts aléatoires (R^2) générées pour chaque espèce de poisson (a) et importances globales des variables explicatives pour l'ensemble du compartiment ichthyofaune, exprimées en composantes du R^2 moyen des espèces (b).

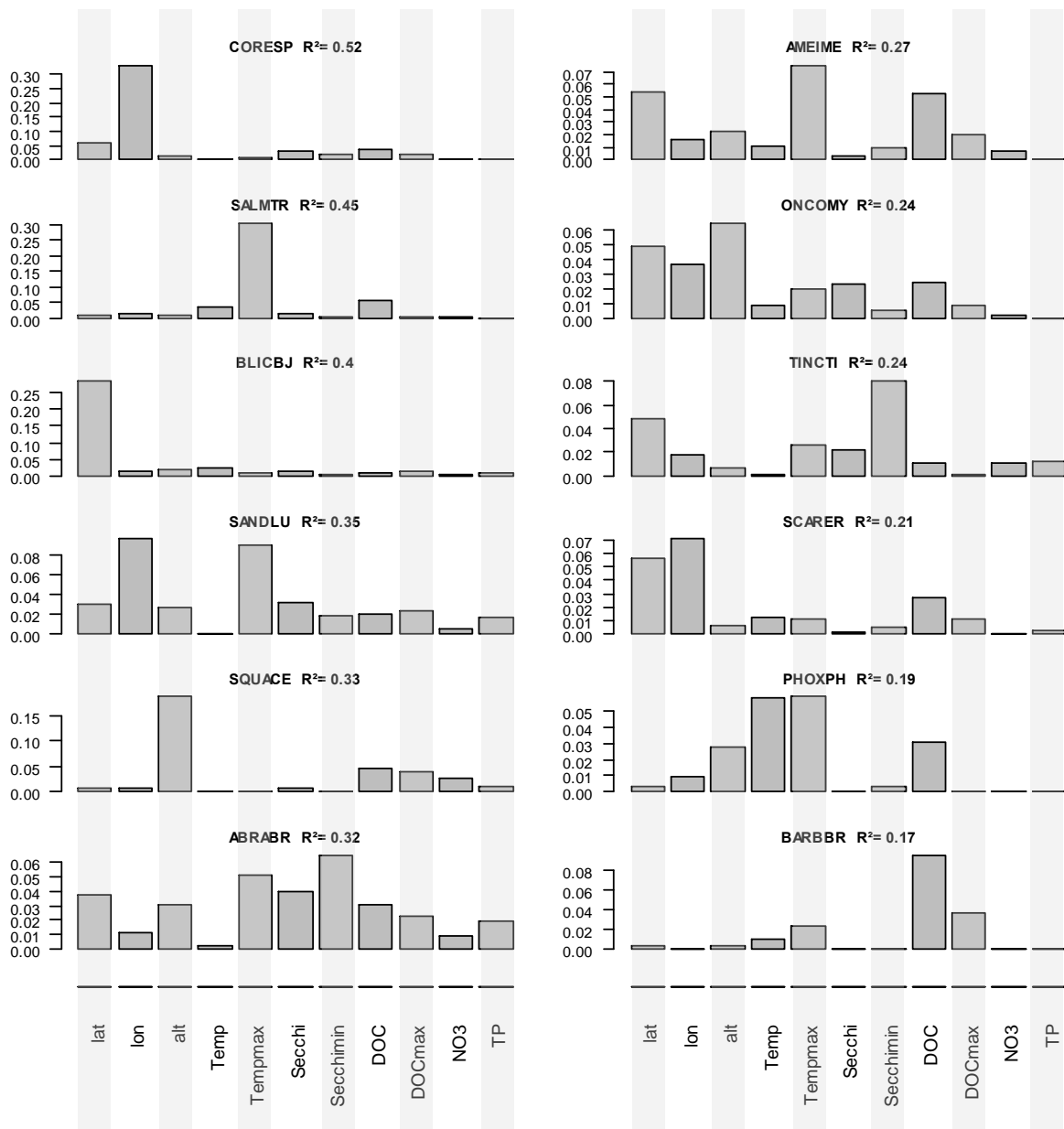


Figure 20. Importances des variables pour les 12 espèces de poissons ayant les R² les plus élevés. Les variables sont limitées dans les figures aux plus importantes globalement pour le compartiment poissons et aux nutriments les plus importants (NO₃ et TP).

La recherche de seuils écologiques s'est faite sur les paramètres physico-chimiques les plus importants pour les poissons ainsi que dans les gradients des nitrates et du phosphore total (Figure 21 à Figure 23) compte tenu de leur implication dans les processus d'eutrophisation. Pour Tempmax, un seuil très net est apparu autour de 18°C qui correspond à la réponse très marquée de SALMTR et à celles plus modérées d'autres espèces telles que PHOXPH et RUTIRU. D'autres espèces ont présenté des réponses à des niveaux plus élevés de température (20° et 22°C) mais les seuils associés sont à peine visibles sur le graphe de densités (Figure 21). De même, un seuil assez net ressort pour le DOC entre 1,2 et 2 mg.L⁻¹ du fait des réponses de plusieurs espèces à ce niveau du gradient. Un petit seuil autour de 4 mg.L⁻¹ est tout juste décelable. En ce qui concerne la transparence, un seuil

bien marqué autour de 3 m est mis en évidence aussi bien par les courbes de densités que par les courbes d'importance cumulée de plusieurs espèces (Figure 22). Pour les nitrates, on constate deux seuils dans le gradient correspondant chacun à la réponse de deux espèces. Les seuils sont localisés autour de 0,7 et 2,5 mg.L⁻¹. Enfin dans le gradient de TP, le seuil le plus important auquel trois espèces semblent répondre, se situe autour de 30 µg.L⁻¹ (Figure 22). Deux autres seuils plus légers sont visibles aux environs de 13 µg.L⁻¹ et 90 µg.L⁻¹.

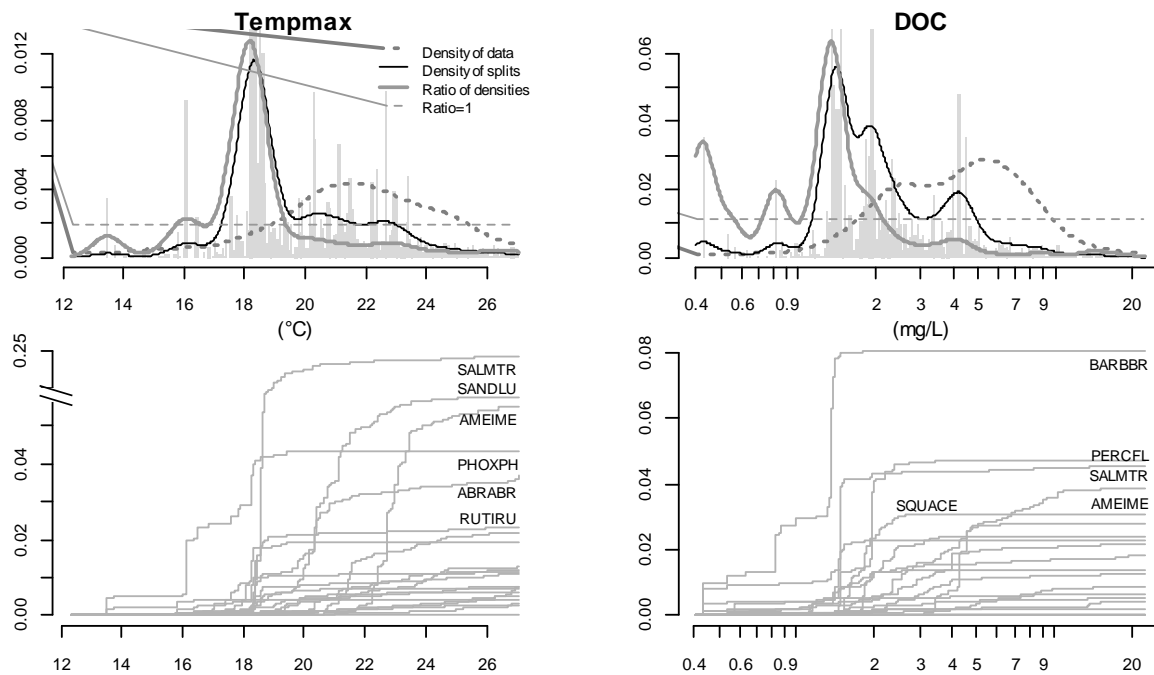


Figure 21. Résultats du *Gradient Forest* pour les poissons concernant les variables Tempmax et DOC. En haut, densités des coupures (courbe noire fine), densité des observations (courbe discontinue) et ratio des densités (courbe grise épaisse). En bas, importances cumulées spécifiques pour chaque espèce de poissons. Note : Pour Tempmax, un segment de l'axe des ordonnées a été extrait de manière à pouvoir représenter entièrement la courbe de SALMTR.

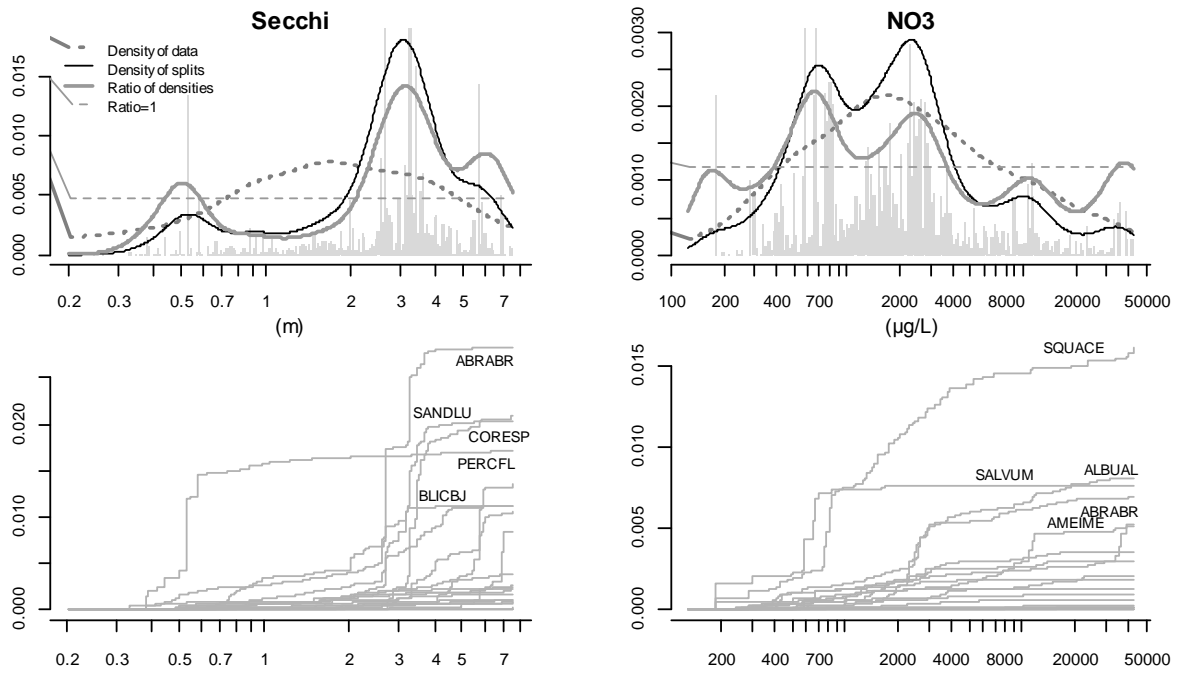


Figure 22. Résultats du *Gradient Forest* pour les poissons concernant les variables Secchi et NO3. En haut, densités des coupures (courbe noire fine), densité des observations (courbe discontinue) et ratio des densités (courbe grise épaisse). En bas, importances cumulées spécifiques pour chaque espèce de poissons.

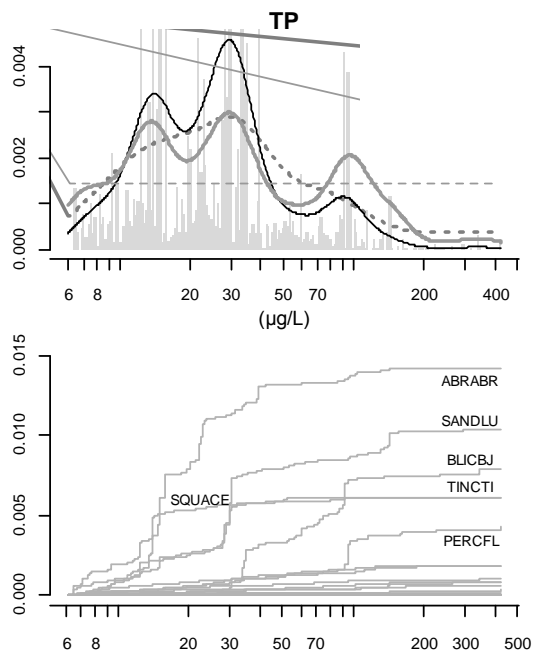


Figure 23. Résultats du *Gradient Forest* pour les poissons concernant la variable TP. En haut, densités des coupures (courbe noire fine), densité des observations (courbe discontinue) et ratio des densités (courbe grise épaisse). En bas, importances cumulées spécifiques pour chaque espèce de poissons.

4. Bilan et discussion

Le *Gradient Forest* est une méthode d'analyse exploratoire qui s'applique à de grands jeux de données biologiques avec de nombreuses variables explicatives potentielles. L'interprétation des résultats en termes de seuils écologiques nécessite une bonne compréhension des incertitudes associées à l'analyse. En effet, lorsqu'un pic apparaît dans une courbe de densité associée au gradient d'une variable, plusieurs aspects sont à considérer pour évaluer la pertinence du pic sur le plan écologique.

Tout d'abord, des pics importants peuvent se former aux extrémités des gradients. Or, du fait de la faible densité des données dans ces zones, leur importance est souvent amplifiée par le ratio des densités. Il ne faut donc pas les prendre en compte comme des seuils importants car, aux extrémités des gradients, les coupures effectuées par l'analyse ne reposent que sur une ou quelques données biologiques et ne peuvent refléter qu'une tendance sans être vraiment significatives. De manière générale, l'incertitude sur les seuils identifiés augmente lorsque leur localisation se rapproche d'un bout ou de l'autre du gradient (Andersen *et al.* 2009). Pour limiter ce problème, la méthode TITAN de recherche de seuils écologiques basée également sur l'analyse de la variance, impose un nombre minimal de 5 observations en-dessous ou au-dessus de chaque seuil (Baker et King 2010). Ce type de précaution n'est pas implémenté dans le package *Gradient Forest* et doit donc être considéré *a posteriori*.

Ensuite, les résultats de cette étude montrent que les seuils peuvent être plus ou moins bien marqués dans les gradients, *i.e.* que les pics de densité apparaissent tantôt nets et étroits, tantôt larges et diffus. Un seuil plus large peut s'expliquer par le fait que les espèces n'ont pas exactement la même sensibilité au paramètre considéré, induisant des réponses plus étalées dans le gradient, ou que les réponses de certaines espèces sont moins franches et plus graduelles à ce seuil par rapport à d'autres espèces. Une autre cause potentielle d'élargissement des seuils peut être un phénomène d'hystérésis qui implique l'existence de deux seuils proches mais différents selon le sens d'évolution des écosystèmes dans le gradient, *i.e.* écosystèmes en cours de dégradation ou de restauration (Andersen *et al.* 2009). Une forte hystérésis pourrait même conduire à la détection de deux seuils voisins dans un même gradient.

Hormis les cas de forte hystérésis, la détection de plusieurs seuils dans un même gradient révèle différents niveaux d'effet d'un paramètre sur les différentes espèces impliquées. Afin de hiérarchiser l'importance relative des seuils obtenus par le *Gradient Forest*, plusieurs critères doivent être considérés. Un seuil est d'autant plus important sur la plan écologique que (1) la variable considérée est importante à l'échelle du compartiment biologique, (2) le seuil est bien marqué avec un pic de densité étroit et au-dessus des autres pics, (3) le nombre d'espèces répondant au seuil est grand et (4) plusieurs compartiments biologiques sont sensibles au même seuil.

Dans les figures 24 à 26, les seuils identifiés pour les principales variables à partir des trois compartiments ont été replacés dans les gradients. Un constat général est qu'il n'y a pas de seuil fort qui se dégage concernant tous les compartiments à la fois. Les trois compartiments indiquent plutôt des seuils à différents endroits dans les gradients. Pour une variable donnée, la même importance ne doit pas être accordée à tous les seuils. En appliquant les règles citées ci-dessus, les importances

relatives des seuils ont été approximativement retranscrites par l'épaisseur des traits de couleur dans la partie inférieure des figures. Il faut noter que les gradients n'ont pas la même longueur pour les différents compartiments. Ceci peut être dû au fait que les lacs échantillonnés ne sont pas les mêmes et qu'une partie seulement du territoire national a été considérée pour le phytoplancton (*i.e.* latitude > 44,5 °N et longitude < 4,5°E, Figure 11). Seuls les couples compartiment/variable pertinents par rapport aux résultats du *Gradient Forest* (*i.e.* importance de la variable et présence d'un seuil) ont été représentés dans les figures.

On peut ainsi constater que dans le gradient de transparence (Secchi), les seuils obtenus par l'ichtyofaune et par les macrophytes sont quasiment les mêmes alors que ceux du phytoplancton sont bien marqués à des niveaux nettement plus bas (Figure 24). Pour le phosphore total, les seuils identifiés sont assez dispersés dans le gradient, les plus hauts et les plus importants étant fournis par les données de phytoplancton. Pour les nitrates, on constate une certaine concordance des seuils obtenus avec les données phytoplancton et macrophytes en bout de gradient (Figure 25, valeurs maximales annuelles). Les seuils indiqués par les données poissons ne sont pas directement comparables car ils correspondent à des concentrations médianes annuelles, mais leurs positions semble correspondre aux seuils de maxima des macrophytes. Pour l'ammonium, seul le phytoplancton a permis de détecter des seuils pertinents. Pour le carbone organique dissous, on peut voir que le gradient pour le phytoplancton est décalé vers les valeurs les plus élevées et ceci s'explique probablement par le fait que les lacs de montagnes les plus pauvres en COD ne sont pas pris en compte par le découpage géographique (Figure 26). Le seuil indiqué par les poissons est nettement plus bas que ceux associés au phytoplancton. Pour la température, deux indications de seuils sont données par les macrophytes et l'ichtyofaune, mais les statistiques annuelles ne sont pas les mêmes (Figure 27).

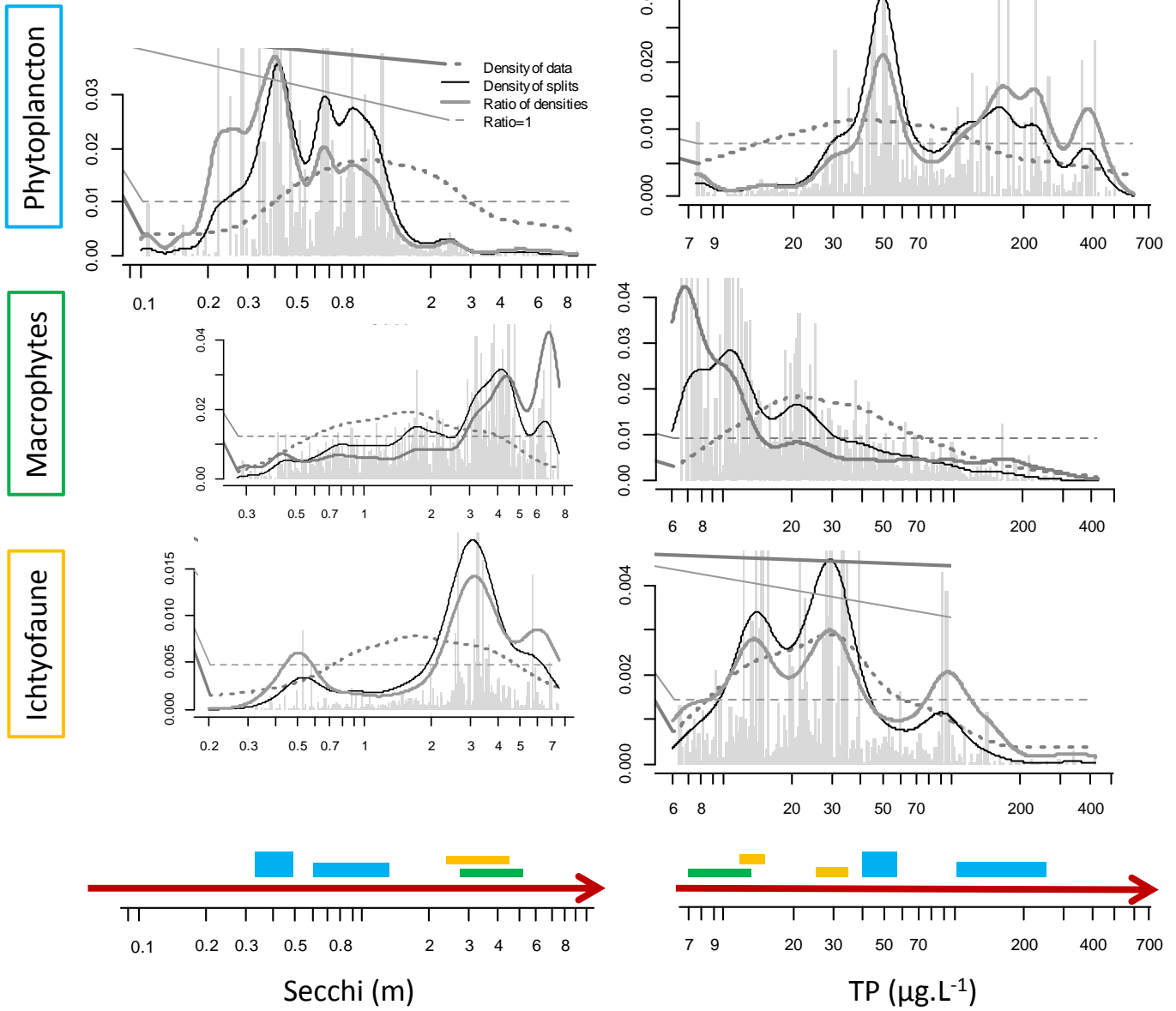


Figure 24. Bilan des seuils identifiés avec les différents compartiments biologiques pour les variables Secchi et TP (phosphore total).

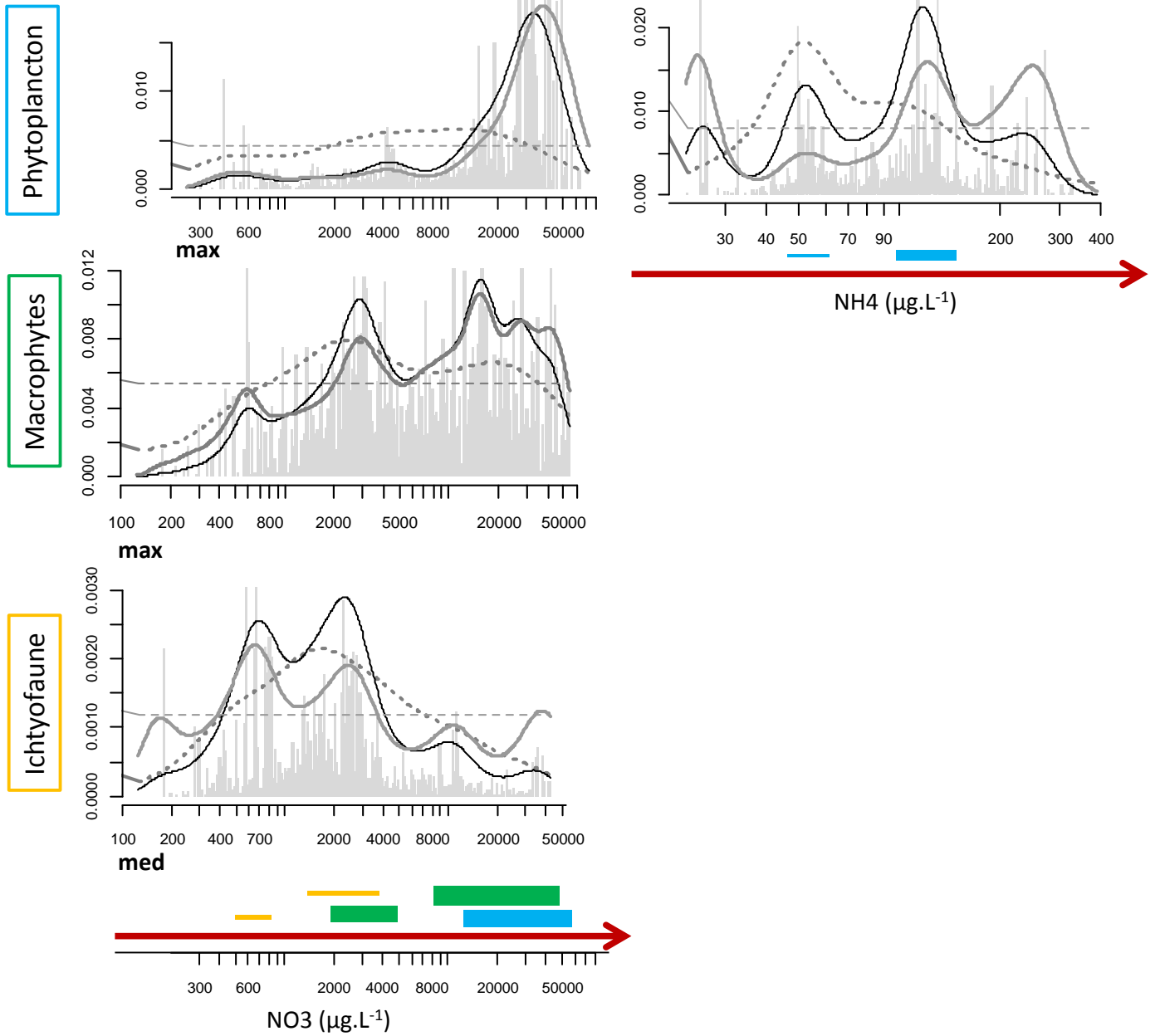


Figure 25. Bilan des seuils identifiés avec les différents compartiments biologiques pour les variables NO3 et NH4. Les seuils de nitrate relatifs à l'ichtyofaune correspondent à des valeurs médianes annuelles, à la différence des autres compartiments pour lesquels il s'agit de valeurs maximales annuelles.

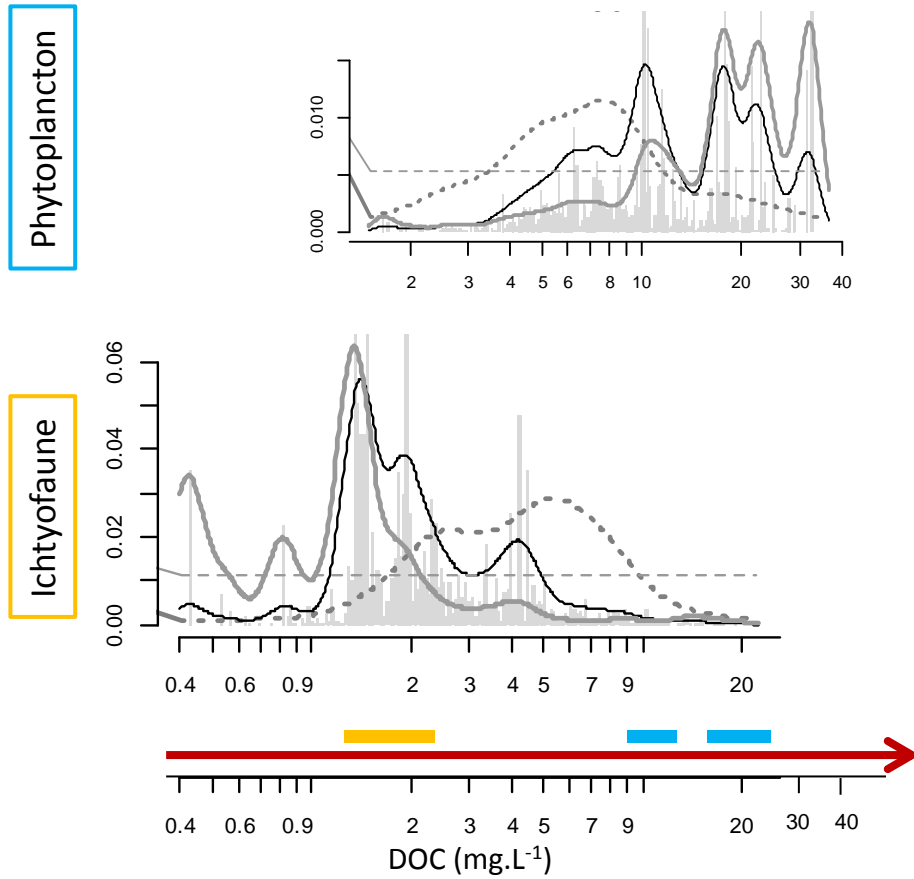


Figure 26. Bilan des seuils de DOC identifiés avec les compartiments phytoplancton et ichtyofaune.

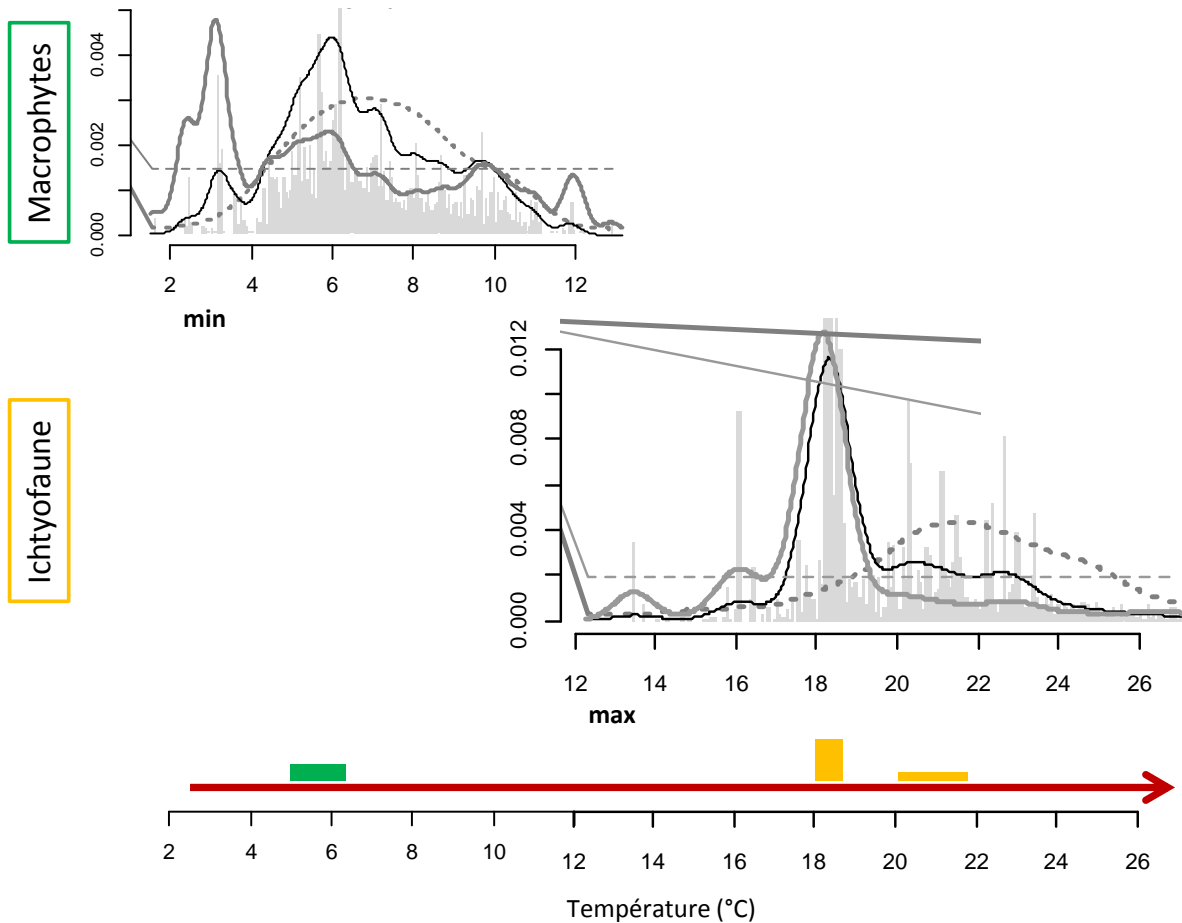


Figure 27. Bilan des seuils de température identifiés avec les compartiments macrophytes (Tempmin) et ichtyofaune (Tempmax).

5. Perspectives & Conclusions : Utilisation des seuils

Les seuils identifiés dans cette étude à travers la méthode de *Gradient Forest* pourront par la suite servir de base pour fixer des nouvelles limites de classes (valeurs-seuils) pertinentes sur le plan écologique. Lorsque plusieurs seuils sont reconnus, ils peuvent servir à définir différentes limites de classes.

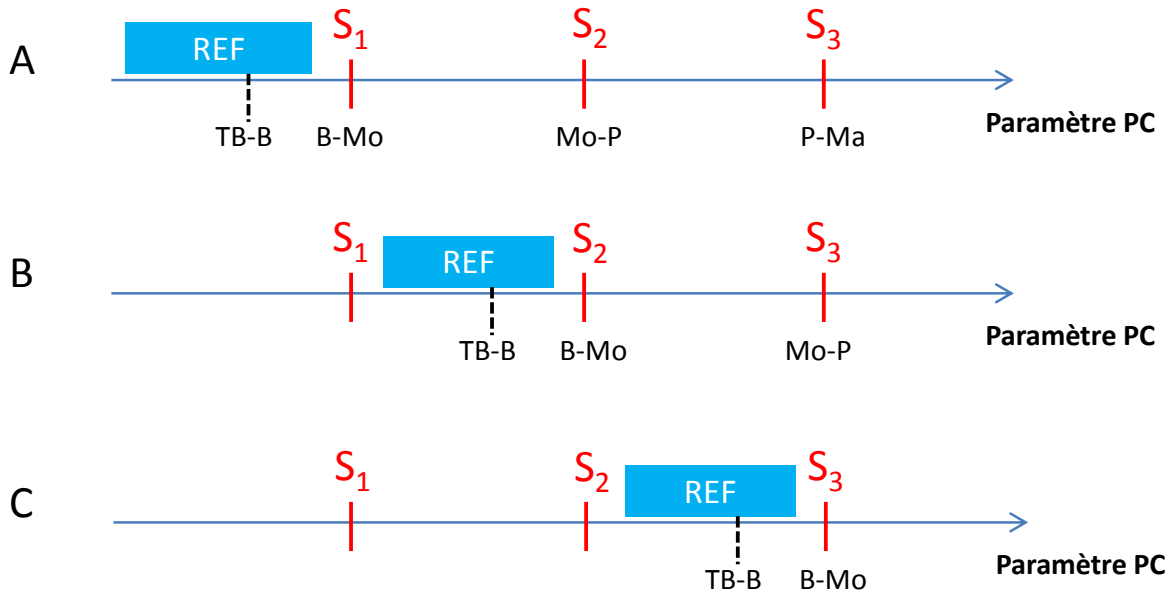


Figure 28. Schéma représentant trois positions possibles (A, B et C) du niveau de référence (REF, carrés bleus) d'un paramètre physico-chimique dans un gradient où trois valeurs-seuils ont été identifiées. Ces trois possibilités peuvent correspondre à des lacs ou des types de lacs se distinguant par des caractéristiques naturelles différentes, par exemple hydromorphologiques.

Nous proposons ici une méthode pour fixer des limites de classe pour un lac donné, inspirée de celle publiée par Soranno *et al.* (2008) pour des lacs du Michigan. La première étape consiste à déterminer le ou les seuils importants dans le gradient d'un paramètre physico-chimique à considérer, à partir d'une analyse globale des plans d'eau, comme effectuée par *Gradient Forest*. Cela nécessite pour chaque seuil retenu d'extraire une ou des valeurs à partir d'un intervalle : il peut être choisi de prendre la valeur marquant le milieu du seuil ou celle située au début (plus préventif). Dans le cas d'un seuil large, les bornes inférieure et supérieure définissant le seuil peuvent être retenues. Les choix qui sont faits aboutissent à la définition de quelques valeurs-seuils écologiques pour chaque gradient.

Ces valeurs-seuils doivent définir les classes de qualité DCE, mais elles ne peuvent pas être utilisées de la même manière pour l'ensemble des plans d'eau français. Il est nécessaire de tenir compte de la variabilité naturelle entre les plans d'eau des paramètres physico-chimiques soutenant la biologie. En effet, pour un plan d'eau donné, il faut déterminer les conditions physico-chimiques de référence. Le niveau de référence d'un paramètre peut être estimé à partir de données historiques ou paléolimnologiques, par comparaison avec un plan d'eau de référence (peu anthropisé) du même type (besoin d'une typologie), ou bien par l'utilisation de modèles statistiques établis à partir d'un ensemble de plans d'eau de référence ou ayant des niveaux d'altération variés (approche de type hindcasting). La limite Très Bon-Bon représente en générale la limite supérieure de la zone de référence dans le sens de la dégradation. Elle est souvent déterminée à partir d'un quantile ou d'un intervalle de prédiction. La première valeur-seuil à considérer ensuite devient en conséquence la limite Bon-Moyen qui est la plus critique pour la gestion des écosystèmes. La Figure 28 présente schématiquement la position théorique possible du niveau de référence dans un gradient avec trois

valeurs-seuils identifiées. Selon le niveau de référence, la limite Bon-Moyen (B-Mo) est ainsi représentée par la valeur-seuil S_1 , S_2 ou S_3 . Les valeurs-seuils situées au-delà de la limite Bon-Moyen donnent les limites Moyen-Passable (Mo-P) et Passable-Mauvais (P-Ma). Dans le cas peu probable où le niveau de référence serait au-delà de toutes les valeurs-seuils, il n'y aurait pas de repère pour fixer les limites autres que Très Bon-Bon (TB-B). Pour déterminer des limites de classes manquantes sans pouvoir s'appuyer sur un effet écologique, on pourra réaliser une simple division de l'intervalle entre la dernière limite de classes estimée et la valeur minimale ou maximale (selon le sens de la dégradation) du paramètre considéré. Dans ces cas-là, il est important de souligner que les limites de classes déduites de la sorte sont arbitraires et géométriques et n'ont pas de justification écologique. De plus, le choix des valeurs minimales ou maximales est cruciale et pourrait par exemple prendre en compte la variabilité européenne et non pas seulement la variabilité nationale.

Conclusion

L'utilisation de seuils écologiques de communauté pour caler les limites de classe des paramètres physico-chimiques soutenant la biologie permet de faire le lien avec les éléments de qualité biologiques dans la construction des indices physico-chimiques, comme préconisé par la DCE. La méthode de *Gradient Forest* est adéquate pour détecter efficacement des seuils à partir de grands jeux de données. En considérant les trois compartiments phytoplancton, macrophytes et ichtyofaune, des indications de seuils ont été obtenues pour les variables transparence, phosphore total, ammonium, nitrates, carbone organique dissous et température. Afin d'aboutir à des indicateurs opérationnels, il reste encore à déterminer des valeurs-seuils précises pour chaque variable et un niveau de référence pour chaque plan d'eau. Les limites de classes pourront ensuite être définies par plan d'eau ou type de plans d'eau. En suivant cette approche, les valeurs-seuils sont fixes mais elles peuvent constituer des limites de classes différentes selon le niveau de référence estimé. C'est une avancée fondamentale par rapport aux valeurs-seuils proposées dans Danis et Roubeix (2014) qui varient de manière continue avec la profondeur moyenne des plans d'eau, critère de prise en compte de l'état de référence dans l'indice phytoplancton lacustre (IPLAC).

Il n'a pas été possible d'identifier des seuils pour des variables relatives à certains éléments de qualité physico-chimiques tels que la salinité, le pH ou l'oxygénation. Des études plus ciblées dans le choix des sites et des descripteurs pourront éventuellement permettre à l'avenir d'en identifier.

6. 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. Pédrón, 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.
- 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. *Machine Learning* 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. Pôle Onema-Irstea "Hydroécologie des Plans d'eau", 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. *Pôle Onema-Irstea "Hydroécologie des Plans d'eau", Aix-en-Provence. p. 60.*
- Danis, P.-A. et V. Roubeix. 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. *Journal of the North American Benthological Society* 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. *Journal of the North American Benthological Society* 28: 855-868.
- Feret, T. et C. Laplace-Treyture. 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. *Canadian Journal of Fisheries and Aquatic Sciences* 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. *Ecological Indicators* 18: 63-72.
- King, R.S. et M.E. Baker. 2010. Considerations for analyzing ecological community thresholds in response to anthropogenic environmental gradients. *Journal of the North American Benthological Society* 29: 998-1008.
- May, R.M. 1977. Thresholds and breakpoints in ecosystems with a multiplicity of stable states. *Nature* 269: 471-477.
- Ministère de l'Ecologie et du Développement Durable. 2015. Arrêté du 27 juillet 2015 modifiant l'arrêté du 25 janvier 2010 relatif aux méthodes et critères d'évaluation de l'état écologique, de l'état chimique et du potentiel écologique des eaux de surface pris en application des articles R. 212-10, R. 212-11 et R. 212-18 du code de l'environnement. *Journal Officiel de la République Française*.
- 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. *Journal of Applied Ecology* 49: 670-679.
- Qian, S.S. 2014. Ecological threshold and environmental management: A note on statistical methods for detecting thresholds. *Ecological Indicators* 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. *Ecological Indicators* 4: 29-37.
- Roubeix, V., P.-A. Danis, T. Feret et J.-M. Baudouin. Sous presse. Identification of ecological thresholds from variations in phytoplankton communities among lakes: contribution to the definition of environmental standards. *Environmental Monitoring and Assessment*.
- 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.
- Smith, A.J. et C.P. Tran. 2010. A weight-of-evidence approach to define nutrient criteria protective of aquatic life in large rivers. *Journal of the North American Benthological Society* 29: 875-891.
- Sonderregger, 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. *Ecological Indicators* 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

Irstea
Parc de Tourvoie
BP 44,
92163 Antony cedex
01 40 96 61 21
www.irstea.fr