



HAL
open science

**Bioindication d'état et de fonctionnement en plan d'eau :
Utilisation des traits morpho-fonctionnels du
phytoplancton pour définir des métriques de diagnostic
Rapport final**

Christophe Laplace-Treuture, J. Derot, Soizic Morin, Juliette Tison-Rosebery

► **To cite this version:**

Christophe Laplace-Treuture, J. Derot, Soizic Morin, Juliette Tison-Rosebery. Bioindication d'état et de fonctionnement en plan d'eau : Utilisation des traits morpho-fonctionnels du phytoplancton pour définir des métriques de diagnostic Rapport final. [Rapport de recherche] Irstea. 2019, pp.79. hal-02608848

HAL Id: hal-02608848

<https://hal.inrae.fr/hal-02608848v1>

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.



**AGENCE FRANÇAISE
POUR LA BIODIVERSITÉ**

Établissement public du ministère de l'Environnement

Programme 2016/2018 - Action n° 07

Bioindication d'état et de fonctionnement plan d'eau

**Utilisation des traits morpho-
fonctionnels du phytoplancton pour
définir des métriques de diagnostic**

Rapport final 2018

**Christophe Laplace-Treyture, Jonathan Derot,
Soizic Morin, Juliette Rosebery**

Février 2019



- **AUTEURS**

Christophe LAPLACE-TREYTURE, ingénieur hydrobiologiste, IRSTEA Bordeaux (Unité EABX, équipe CARMA),

christophe.laplace-treytur@irstea.fr

Jonathan DEROT, Ingénieur de recherche, IRSTEA Bordeaux (Unité EABX, équipe CARMA),

jonathan.derot@irstea.fr

Soizic MORIN, chargée de recherche en écotoxicologie, IRSTEA Bordeaux (Unité EABX, équipe CARMA),

soizic.morin@irstea.fr

Juliette ROSEBERY, chargée de recherche en hydrobiologie, IRSTEA Bordeaux (Unité EABX, équipe CARMA),

juliette.rosebery@irstea.fr

- **CORRESPONDANTS**

AFB : Jean-Claude RAYMOND, Délégation régionale Auvergne Rhône-Alpes, jean-claude.raymond@afbiodiversite.fr

Jean-Marc BAUDOIN, pôle AFB-Irstea Aix-en-Provence, jean-marc.baudoin@afbiodiversite.fr

Droits d'usage : accès libre

Niveau géographique : national

Couverture géographique : France

Niveau de lecture : experts



Bioindication d'état et de fonctionnement en plan d'eau – Utilisation des traits morpho-fonctionnels du phytoplancton pour définir des métriques de diagnostic

Rapport final

Christophe Laplace-Treytore,
Jonathan Derot, Soizic Morin,
Juliette Rosebery

**AGENCE FRANÇAISE
POUR LA BIODIVERSITÉ**

Établissement public du ministère de l'Environnement

- **RESUME**

Le présent travail s'insère dans la continuité des développements d'un indice multi-métrique pour l'évaluation de l'état écologique des plans d'eau en ce qui concerne le compartiment phytoplancton en plan d'eau. Il est ainsi en lien avec l'indice IPLAC. Il vise à terme à compléter son évaluation par des métriques dites de diagnostic permettant de rendre compte d'autres pressions et des éléments de biodiversité. Pour cela nous avons employé une approche morpho-fonctionnelle à large échelle (territoire national) des communautés phytoplanctoniques, basée sur la définition et l'utilisation de 84 traits morpho-fonctionnels du phytoplancton, couvrant l'ensemble des groupes taxonomiques, comme descripteur des communautés.

Dans les écosystèmes aquatiques, les relations et interactions entre les communautés phytoplanctoniques et les variables environnementales étant souvent assez complexes, différents outils statistiques et numériques tel que le Random Forest (RF) en mode régression et classification, les partial Dependent Plots (PDP) ont été testés pour appréhender cette complexité.

Les méthodes de RF ont montré que la profondeur du plan d'eau et l'altitude étaient des facteurs importants. Cela a conduit à la création, par la méthode des K-means, de quatre groupes typologiques, fonction de ces critères, pour en soustraire l'effet. Par la suite, divers tests de corrélations directes, sur chacun des groupes typologiques, ont été exécutés mais il en ressort des résultats plutôt décevants. Dans un second temps, des modèles RF en mode régression puis classification ont été utilisés. Cependant leurs utilisations nous éloignent de la DCE compatibilité car elles n'ont pas permis de dériver des relations et des métriques portant sur un trait ou combinaison de traits exploitables. Dans un dernier temps, une approche exploitant les PDP a été mise en œuvre. Sur les 84 traits construits, 21 ressortent des analyses avec des liens directs ou indirects avec des paramètres physico-chimiques des eaux. Il s'agit notamment des traits : forme de vie (solitaire, filamenteuse), motilité, présence de protubérances, de pseudo-vacuoles ou encore de la présence de Chlorophyceae ou d'Euglenophyceae. Le carbone organique dissous, les matières en suspensions, les nitrates ou encore la transparence mesurée au disque de Secchi sont les paramètres physico-chimiques qui présentent des relations avec des traits morpho-fonctionnels. L'utilisation des PDP permet de connaître le sens de variation de la relation existante entre le trait et le paramètres physico-chimiques. En revanche, les valeurs de seuils ne dépendent pas des groupes typologiques créés. Ce travail devrait permettre la définition de nouvelles métriques de diagnostic basées sur certains traits pour les gradients de COD et de nitrates. Cela permettra d'affiner la compréhension de l'impact des pressions anthropiques sur les écosystèmes lacustres.

- **MOTS CLES : BIO-INDICATION LACUSTRE, TERRITOIRE FRANÇAIS, TRAIT MORPHO-FONCTIONNEL DU PHYTOPLANCTON, MODELE D'AUTO-APPRENTISSAGE, RANDOM FOREST, CARBONE ORGANIQUE DISSOUS, DIRECTIVE CADRE SUR L'EAU.**



**Bioindication d'état et de
fonctionnement en plan d'eau –
Utilisation des traits morpho-
fonctionnels du phytoplancton
pour définir des métriques de
diagnostic**

Rapport final

Christophe Laplace-Treytore,
Jonathan Derot, Soizic Morin,
Juliette Rosebery

**AGENCE FRANÇAISE
POUR LA BIODIVERSITÉ**

Établissement public du ministère de l'Environnement

- **ABSTRACT**

This work is in line with the development of a multi-metric index for the assessment of the ecological status of water bodies with respect to the phytoplankton compartment in water. Thus, it is linked to the IPLAC index. In the end, it aims to complete its evaluation with so-called diagnostic metrics to account for other pressures and elements of biodiversity. For this we used a large-scale morpho-functional approach (national territory) of phytoplankton communities, based on the definition and use of 84 morpho-functional traits of phytoplankton, covering all taxonomic groups, as a descriptor of communities.

In aquatic ecosystems, relationships and interactions between phytoplankton communities and environmental variables are often quite complex. So, different statistical and numerical tools such as Random Forest (RF) in regression and classification mode, Partial Dependent Plots (PDPs) have been tested to take into account this complexity.

RF methods have shown that depth of waterbody and altitude are important factors. This led to the creation, by the K-means method, of four typological groups, based on these criteria, to subtract the effect. Subsequently, various tests of direct correlations, on each of the typological groups, were carried out but results are rather disappointing. In a second step, RF models in regression mode then classification mode were used. However, their uses drive us away from the DCE compatibility because they have not allowed us to derive relationships and metrics relating to a trait or combination of exploitable traits. Lastly, an approach using PDPs has been implemented. Of the 84 constructed traits, 21 are highlighted by the result analyses with direct or indirect links to physico-chemical water parameters. These include traits: life form (solitary, filamentous), motility, presence of protuberances, pseudo-vacuoles or the presence of Chlorophyceae or Euglenophyceae. The dissolved organic carbon, the total suspended solids, the nitrates or the transparency measured by Secchi disk are the physicochemical parameters that have relations with morpho-functional traits. The use of PDPs allows knowing the variation direction of the existing relationship between the trait and the physicochemical parameters. On the other hand, threshold values do not depend on the created typological groups. This work should allow the definition of new trait-based diagnostic metrics for DOC and nitrate gradients. This will refine the understanding of the impact of anthropogenic pressures on lake ecosystems.

- **KEY WORDS:** BIOLOGICAL INDICATOR FOR LAKES, FRENCH TERRITORY, MORPHO-FUNCTIONAL TRAIT OF PHYTOPLANKTON, MACHINE LEARNING, RANDOM FOREST, DISSOLVED ORGANIC CARBON, WATER FRAMEWORK DIRECTIVE.



**Bioindication d'état et de
fonctionnement en plan d'eau –
Utilisation des traits morpho-
fonctionnels du phytoplancton pour
définir des métriques de diagnostic**
Rapport final
Christophe Laplace-Treytore,
Jonathan Derot, Soizic Morin, Juliette
Rosebery

**AGENCE FRANÇAISE
POUR LA BIODIVERSITÉ**

Établissement public du ministère de l'Environnement

- [Cette synthèse pourra être utilisée indépendamment du rapport final, elle doit donc être utilisable telle quelle]
- [Cette synthèse est référencée en tant que telle sur le site de l'AFB et sur le portail documentaire national « Les documents techniques sur l'eau ».]
- [Cette synthèse de vulgarisation montre les résultats et leurs utilités (potentialités) pour la gestion ou la décision ; des figures, tableaux peuvent y être insérés. Les objectifs rappelés dans les fiches action de la convention peuvent servir de base pour cette rédaction.]

- **SYNTHÈSE POUR L'ACTION OPERATIONNELLE** [4 pages maximum]

*[Il est très important que le rapport final des résultats des recherches soit accompagné d'une synthèse présentant **les principaux résultats transférables**.*

Cette synthèse proposée par les scientifiques producteurs de ces connaissances (résultats, méthodes, recommandations...) doit faire l'objet d'un travail spécifique entre les auteurs du rapport et les chargés de mission de l'AFB.]

[La synthèse comporte par exemple :

° Contexte général (quelle situation, quels enjeux au niveau réglementaire, au niveau de la gestion des milieux aquatiques, ou au niveau scientifique), objectifs généraux du projet (à quelles questions veut-on répondre ? quand, avec quelles échéances et avec qui le projet est-il mené ?)

° Brève information sur les méthodes utilisées et sur les données d'entrée (indications brèves sur leurs nature, source, quantité, extensions spatiales et temporelles)

° Principaux acquis transférables obtenus (le cas échéant sous forme de tableau de synthèse) et implications pratiques et/ou recommandations et/ou réalisations pratiques. Limites de l'utilisation, et perspectives d'améliorations opérationnelles

° Pour en savoir plus : quelques références (dont les livrables, articles prévus...) ; adresse électronique et/ou site internet de(s) auteur(s)]

SOMMAIRE

Introduction.....	6
1 Base de donnée utilisée	7
1.1 Données disponibles sur la base PLAN_DEAU d'Aix-en-Provence	7
1.2 Constitution de la base de traits morpho-fonctionnels du phytoplancton.....	9
1.2.1 Sélection des traits morpho-fonctionnels	9
1.2.2 Transformation binaire de l'information des traits.....	10
2 Méthodologie mise en œuvre	12
2.1 Présentation du modèle Random Forest.....	13
2.2 Biotypologie et variabilité naturelle des communautés.....	15
2.2.1 Partitionnement via la méthode des K-moyennes	15
2.2.2 Corrélations directes entre traits morpho-fonctionnels et paramètres physicochimiques.....	15
2.3 Méthodologie d'extraction de seuils physico-chimiques	16
2.4 Utilisation des PDP plots	18
3 Résultats, réorientations méthodologiques et discussion.....	21
3.1 Groupes typologiques établis par la méthode des K-means	21
3.2 Corrélations directes	23
3.3 Limites de classes et transformation Boxcox	25
3.4 Modèle RF en mode régression	28
3.5 Modèle RF en mode classification.....	31
3.6 Résultats de l'utilisation des PDP Plots.....	34
3.6.1 Traits associés avec un paramètre physico-chimique	34
3.6.2 Paramètres physico-chimiques expliquant les traits phytoplanctoniques	36
3.6.3 Impact de la typologie sur la diversité fonctionnelle du phytoplancton.....	38
Conclusions et perspectives	39
Sigles & Abréviations.....	41
Bibliographie.....	42
Annexes.....	45

Introduction

Depuis 1945, l'anthropocène, défini par Crutzen (2006) comme une nouvelle ère géologique, serait rentré dans une phase de « grande accélération » (Steffen et al., 2007) où les activités humaines provoqueraient des modifications de plus en plus importantes sur la globalité des écosystèmes à l'échelle de notre planète.

Dans ce cadre d'anthropisation généralisée, les écosystèmes aquatiques côtiers et lacustres représentent un enjeu majeur, car ils ont un impact plus ou moins direct sur la santé humaine et l'économie (pisciculture, barrages, eaux de consommation, activités de loisirs nautiques...). Les multiples travaux de vulgarisation scientifique effectués au cours des dernières décennies par l'ensemble de la communauté scientifique, ont permis une prise de conscience de la part des organisations publiques et privées, de l'impact négatif que pouvait induire l'homme sur ces écosystèmes aquatiques.

Dans ce contexte, l'union européenne a décidé en 2000 de mettre en place des outils de mesure (bio-indication), afin d'évaluer l'état écologique des masses d'eau, via la Directive Cadre Européenne sur l'Eau (European Parliament, 2000). Dans les écosystèmes aquatiques lacustres, les perturbations anthropiques impactent directement le premier échelon du réseau trophique, qui est principalement constitué par le phytoplancton. Le phytoplancton est l'un des Eléments de Qualité Biologique (EQB), qui d'après la définition de la DCE permettent l'évaluation de l'état écologique des différentes catégories de masses d'eau.

Le présent travail s'insère pleinement dans la continuité des développements d'indice multi-métrique pour l'évaluation de l'état écologique des plans d'eau (Feret & Laplace-Treytore, 2013) en ce qui concerne le compartiment phytoplancton. Il vise à terme à compléter cette évaluation par d'autres indicateurs de fonctionnement des systèmes lacustres permettant de rendre compte d'autres pressions et des éléments de biodiversité. Pour cela nous avons employé une approche morpho-fonctionnelle à large échelle (territoire national) des communautés algales, basée sur la définition et l'utilisation de traits morpho-fonctionnels du phytoplancton comme descripteur des communautés. En effet, l'utilisation des traits et de leurs relations avec les paramètres environnementaux constituent une pièce maîtresse dans la compréhension des assemblages d'espèces (Wieher et al., 2011). De plus les traits sont assez largement utilisés pour évaluer les conditions environnementales dans plusieurs indices biotiques (invertébrés, poissons). Les communautés par le biais de leurs traits seront mises en relation avec les pressions (qualité des eaux, occupation du BV,...) afin de mettre en évidence des relations de pressions impacts.

Dans les systèmes aquatiques les relations et interactions entre les traits phytoplanctoniques et les variables environnementales sont souvent non linéaires et complexes (Edwards et al., 2016). Afin de tenir compte de cette complexité, des outils numériques tels que le machine learning et les Random Forest seront utilisés.

Le présent rapport présente les recherches méthodologiques qui ont été effectuées afin de préfigurer de futures métriques d'anthropisation basées sur les traits morpho-fonctionnels du phytoplancton. Il présente différentes méthodes permettant d'établir des relations traits-paramètres, discute des résultats obtenus et propose finalement la méthode retenue, les traits et les

paramètres environnementaux pour lesquels des relations ont pu être mises en évidence et à même de servir de futures métriques de diagnostic complémentaire à l'indice IPLAC.

1 Base de donnée utilisée

Dans le cadre de cette étude, nous avons utilisé la base de données DCE PLAN_DEAU d'Aix-en-Provence comme source des données de relevés de phytoplancton et de physico-chimie associée. Ces données couvrent ainsi l'ensemble du territoire métropolitain et tous les types de plans d'eau existant en France.

Afin de disposer d'informations concernant les traits morpho-fonctionnels du phytoplancton, une nouvelle table a été créée avec des données quantitatives et qualitatives caractérisant ces traits pour un certain nombre de taxons. Nous allons par conséquent dans ce chapitre développer les différentes étapes qui nous ont conduit à cette « nouvelle » base de données enrichie.

1.1 Données disponibles sur la base PLAN_DEAU d'Aix-en-Provence

La base DCE PLAN_DEAU, administrée au centre Irstea d'Aix-en-Provence, comporte 469 plans d'eau (PE) soumis à la DCE dont 28 sont classés dans la catégorie « Plan d'eau de référence ».

Un certain nombre de ces plans d'eau ont fait l'objet de campagnes de suivis du phytoplancton. Ainsi la base de données comporte 943 « année-lacs¹ » ayant été échantillonnées entre 2005 et 2016 sur le compartiment phytoplancton dont 28 venant de plans d'eau de référence (voir la carte de la Figure 1 ci-dessous). Cela représente alors jusqu'à 1740 données (ammonium par exemple) au maximum par paramètres avec une moyenne de 1308 données et un minimum de 244 (pour l'oxygène dissous). Ces données ont été relevées selon le protocole d'acquisition des données en plan d'eau (Laplace-Treyture et al., 2009) afin d'être comparables et de couvrir la période de végétation. Il est important de rappeler qu'un plan d'eau est classé comme référence sur des critères abiotiques par chaque agence de bassin. Ainsi, un plan d'eau est dit de référence si aucune ou très peu de pressions anthropiques s'y exercent.

Dans la base PLAN_DEAU, les caractéristiques génériques de chaque plan d'eau telles que l'altitude, la profondeur moyenne, la surface du bassin versant,... sont disponibles. Les données relatives à l'occupation des bassins versants sont aussi renseignées.

¹ On entend par ce terme, une année de relevés (plusieurs campagnes, habituellement 4 par an) de données phytoplanctoniques et physicochimiques sur un plan d'eau donné.

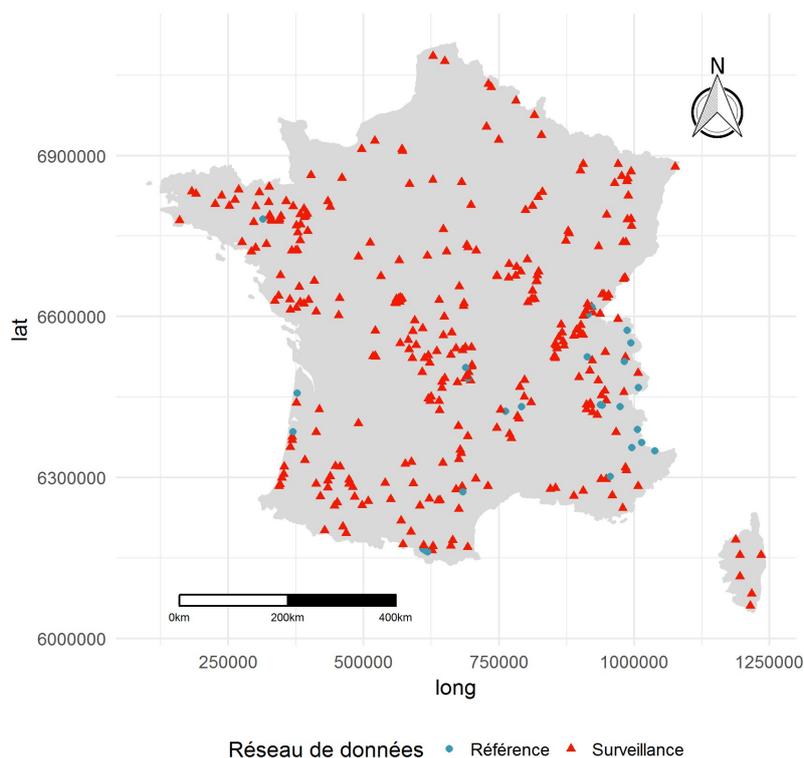


Figure 1. Carte de répartition des plans d'eau utilisés pour cette étude. En bleu les plans d'eau de référence et en rouge les autres plans d'eau (réseau de surveillance). © Jonathan Derot

Nous disposons dans la base de données des relevés de 1409 taxons du phytoplancton différents. Afin de pouvoir s'en servir dans des traitements statistiques, nous avons uniquement conservé ceux qui sont présents dans au moins 15 échantillons. Par conséquent, nous avons utilisé un total de 582 taxons dont l'abondance est exprimée en biovolume (comme pour l'indice IPLAC) dans les analyses présentées dans ce rapport. En ce qui concerne la physico-chimie, le nombre de données disponibles pour chacun des paramètres est récapitulé dans le tableau ci-dessous.

Tableau 1. Disponibilité des données de physico-chimie des eaux des plans d'eau, par paramètre sur la période 2005 à 2016

Paramètres physico-chimiques	Nombre de données présentes	Pourcentage de données disponibles
Chlorophylle a	1724	83,17 %
Conductivité	728	35,1 %
COD	1735	83,70 %
Ammonium	1740	83,98 %
Nitrite	1740	83,98 %
Nitrate	1738	83,84 %
Oxygène dissous	244	11,77 %
pH	924	44,55 %
Phosphate	1739	83,94 %
Phosphore total	1739	83,84 %
Silicium	1233	59,48 %
Température	814	39,27 %
MES	1512	72,94 %
Transparence au disque de Secchi	1081	52,15 %
TAC	933	45,01 %

1.2 Constitution de la base de traits morpho-fonctionnels du phytoplancton

1.2.1 Sélection des traits morpho-fonctionnels

Comme indiqué ci-dessus, notre travail s'est concentré sur les taxons rencontrés dans au moins 15 échantillons (occurrence supérieure à 15) de la base de données PLAN_DEAU, soit 582 taxons. Pour chacun de ces taxons, considéré comme « pertinent » à l'échelle nationale, nous avons décidé, à dire d'expert et en se basant sur un travail préalable (Prévost, 2014), de retenir une liste de 43 traits morpho-fonctionnels suivants :

(1) groupe algal, (2) taille minimale des individus, (3) taille maximale des individus, (4) forme de vie standard de l'individu (solitaire, filamenteuse ou coloniale), (5) autre forme de vie moins fréquente de l'individu (solitaire essentiellement), (6) forme de la cellule (classification PHYTOBS), (7) forme de l'individu (classification PHYTOBS), (8) motilité, (9) présence de flagelle(s), (10) nombre de flagelles, (11) présence d'aérotopes, (12) présence de vacuoles contractiles, (13) présence de mucilage, (14) présence d'akinètes, (15) présence d'hétérocytes, (16) présence de kystes, (17) présence de chloroplastes, (18) nombre de chloroplastes, (19) présence de gaine, (20) présence de tractus, (21) présence de squelette siliceux, (22) présence de lorica, (23) présence de plaques, (24) présence d'écailles, (25) présence d'ornementations, (26) présence de protubérances, (27) type de protubérance, (28) taille de protubérance, (29) nombre de protubérances, (30) biovolume cellulaire, (31) biovolume individuel, (32) largeur minimale, (33) largeur maximale, (34) longueur minimale, (35) longueur maximale, (36) présence de chlorophylle b, (37) présence de chlorophylle c, (38) présence de xanthophylles, (39) présence de phycobilines, (40) présence de toxines, (41) type de trophie, (42) type de reproduction, (43) type d'eau (douce et/ou saumâtre).

Au final une table de traits pour la liste des 582 taxons phytoplanctoniques a été créée. Les informations concernant ces traits morpho-fonctionnels ont été tirées de la bibliographie suivante :

- En ce qui concerne les ouvrages généraux (tous taxa confondus): John et al. (2002); Reynolds (2006); Wehr et al. (2015) ;
- Pour les chlorophycées spécifiquement : Huber Pestalozzi (1974); Komárek et al. (1983) ;
- Pour les chrysophycées spécifiquement : Starmach (1985) ;
- Pour les cyanobactéries spécifiquement : Komarek et Anagnostidis (2005); Komárek (2013); Komárek et Anagnostidis (2000)
- Pour les diatomées spécifiquement : Krammer (1988); Krammer et Lange-Bertalot (1986); Krammer et Lange-Bertalot (1991a); Krammer et Lange-Bertalot (1991b); Krammer et Lange-Bertalot (2000) ;
- Pour les dinophycées spécifiquement : Popovsky et Pfiester (1990) ;
- Pour les euglenophycées spécifiquement : Huber-Pestalozzi (1955) ;
- Pour les desmidiées spécifiquement : Coesel et Meesters (2007) ;
- Pour les xanthophycées spécifiquement : Ettl (1978); Rieth (1980).

1.2.2 Transformation binaire de l'information des traits

Dans un second temps, pour les besoins de nos analyses nous avons transformé les traits quantitatifs contenus dans cette table de traits en format « binaire » (0 ou 1) pour chacun des traits concernés de la manière suivante :

(1) présence de diatomées, (2) présence de chlorophycées, (3) présence de chrysophycées, (4) présence de cryptophycées, (5) présence de cyanobactéries, (6) présence de desmidiées, (7) présence de dinophycées, (8) présence de d'euglenophycées, (9) présence de xanthophycées, (10) catégorie de taille pico-plancton (minimale), (11) catégorie de taille nano-plancton (minimale), (12) catégorie de taille micro-plancton (minimale), (13) catégorie de taille méso-plancton (minimale), (14) catégorie de taille macro-plancton (minimale), (15) catégorie de taille pico-plancton (maximale), (16) catégorie de taille nano-plancton (maximale), (17) catégorie de taille micro-plancton (maximale), (18) catégorie de taille méso-plancton (maximale), (19) catégorie de taille macro-plancton (maximale), (20) forme de vie solitaire, (21) forme de vie coloniale, (22) forme de vie filamenteuse, (23) forme sphérique, (24) forme ellipsoïde rotationnelle, (25) forme ellipsoïde aplatie, (26) forme cylindrique, (27) forme cône avec demi-sphère, (28) forme double cône, (29) forme parallélépipède, (30) forme cylindre ovale, (31) forme cymbelloïde, (32) forme prisme sur une base parallélogramme, (33) forme prisme sur une base triangle, (34) forme *Staurastrum*, (35) forme *Ceratium*, (36) motilité, (37) absence de flagelle, (38) présence de 1 flagelle, (39) présence de 2 flagelles, (40) présence de flagelles supérieure à 2, (41) présence d'aérotopes, (42) présence de vacuoles contractiles, (43) présence de mucilage, (44) présence d'akinètes, (45) présence d'hétérocytes, (46) présence de kystes, (47) absence de chloroplaste, (48) présence de 1 chloroplaste, (49) présence de 2 chloroplastes, (50) présence de chloroplastes supérieur à 2, (51) présence de gaine, (52) présence de tractus, (53) présence de squelette siliceux, (54) présence de lorica, (55) présence de plaques, (56) présence d'ornementations, (57) présence d'écailles, (58) présence de protubérances, (59) absence de protubérance, (60) protubérance de type aiguille, (61) protubérance de type cône, (62) protubérance de type dent, (63) protubérance de type granule, (64) protubérance de type poil, (65) protubérance de type tractus, (66) protubérance de type bouton, (67) protubérance de type écaille siliceuse, (68) protubérance de type plaque, (69) protubérance de petite taille, (70) protubérance de grande taille, (71) une protubérance, (72) deux protubérances, (73) plus de deux protubérances, (74) présence de chlorophylle b, (75) présence de chlorophylle c, (76) présence de xanthophylles, (77) présence de phycobilines, (78) présence de toxines, (79) autotrophie, (80) mixotrophie, (81) hétérotrophie, (82) reproduction sexuée, (83) reproduction asexuée, (84) reproduction sexuée/asexuée.

Soit au final un pool de 84 traits codés en 0 ou 1 pour chacun des 582 taxons retenus.

Cette table de traits a été appliquée aux relevés phytoplanctoniques disponibles dans la base PLAN_DEAU. Ceci permet d'obtenir, pour chaque campagne, une occurrence de chaque trait que l'on a multiplié par le biovolume (cf. Figure 2). De cette manière, il est possible de considérer l'importance quantitative de ce trait au sein des communautés considérées. Afin d'homogénéiser ce nouveau type d'information, nous l'avons aussi exprimé sous la forme d'un pourcentage de présence/absence (cf. Figure 2).

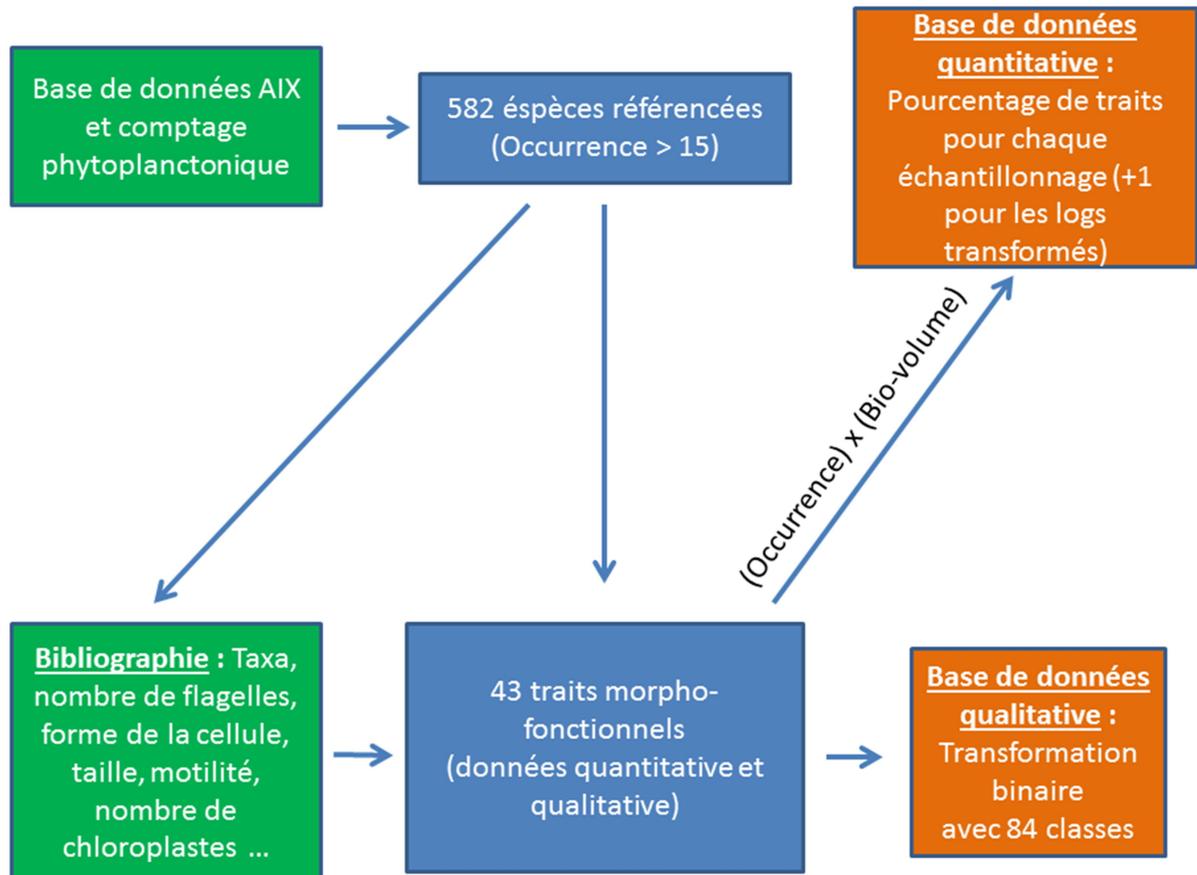


Figure 2. Schéma conceptuel de l'implémentation des valeurs de traits appliquée aux données phytoplanctoniques de la base de données PLAN_DEAU

La Figure 2 reprend le cheminement chronologique des modifications que nous avons faites subir aux données de phytoplancton contenues dans la base PLAN_DEAU étape par étape, en passant de la création de la table de traits à la transformation de certaines données quantitatives. Quant à la Figure 3, elle permet de visualiser le nombre d'occurrences de chacun des traits référencés ; pour des besoins de lisibilité nous avons affiché les traits avec leur libellé « numérique » (cf. paragraphe ci-dessus pour la correspondance). Il est important de noter dans cette Figure 3, qu'un certain nombre de traits sous la forme d'occurrence binaire comporte une majorité de zéros (cf. barres noirs, par exemple traits 19, 27, 45, ...), ce qui peut poser problème lors de l'utilisation de méthodes statistiques paramétriques simples et complexes.

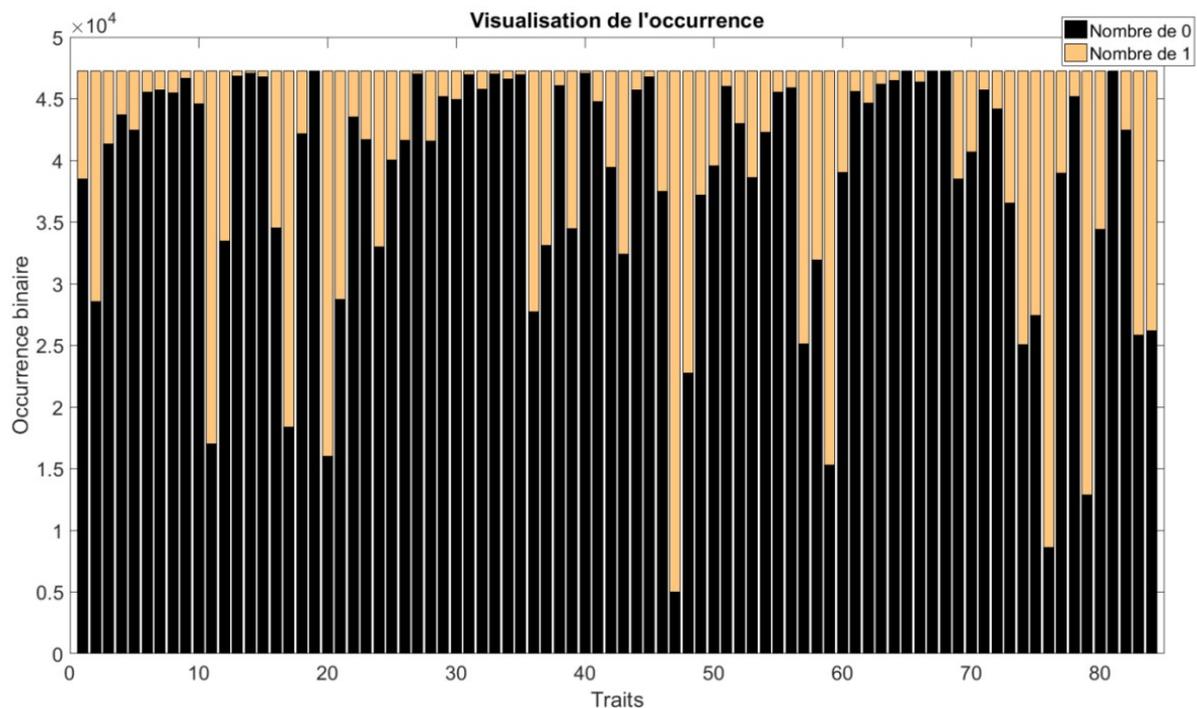


Figure 3. Visualisation de l'occurrence binaire de chacun des traits du phytoplancton après la transformation des traits. Les barres noires représentent le nombre de 0 et les barres orange le nombre de 1

2 Méthodologie mise en œuvre

L'objectif principal consiste à mettre en évidence des relations de pressions- impacts entre un trait et un ou des paramètres abiotiques disponibles, et de déterminer pour chaque trait pertinent une ou des valeurs seuils critiques pour la qualification de la pression (paramètre). Cependant comme indiqué plus haut, les relations entre les traits morpho-fonctionnels et les paramètres environnementaux sont complexes et bien souvent non linéaires.

Aussi, afin d'extraire de tels relations et seuils, nous avons testé différents outils et méthodes statistiques. Dans un premier temps nous avons utilisé un modèle à structure arborescente appelé « Gradient Forest », dérivé des modèles Random Forest (RF) afin d'explorer le jeu de données. L'importance des prédicteurs de type hydromorphologique (ou typologique) dans les modèles obtenus, nous a ensuite conduits à segmenter notre base donnée en plusieurs groupes plus homogènes de façons à créer une typologie. Sur ces groupes typologiques, nous avons testé les corrélations directes entre traits et paramètres du milieu par des méthodes classiques. Par la suite nous avons testé une méthodologie d'extraction de seuils physico-chimiques basée sur les modèles RF. Enfin, nous avons recherché les paramètres les plus influents sur les traits morpho-fonctionnels au sein des différentes groupes typologiques afin de visualiser leur relation avec les traits au travers de « Partial Dependence Plots ».

2.1 Présentation du modèle Random Forest

Nous commençons ce chapitre méthodologique par un rappel du fonctionnement théorique d'un modèle Random Forest afin de faciliter la compréhension ultérieure.

Il est possible de classer les modèles d'intelligence artificielle par type d'apprentissage, dans le cas du modèle RF, nous avons à faire à une structure de type arborescente. Il est primordial de rappeler que le modèle RF est une amélioration du modèle CART (Classification And Regression Tree) (Breiman et al., 1984). Dans un modèle CART la phase d'apprentissage est uniquement basée sur un seul arbre. Alors que dans un modèle RF la phase d'apprentissage est basée sur un nombre prédéfini d'arbres (par exemple 200). C'est aussi pour cette raison que le terme de forêt est utilisé dans le nom du modèle, car il crée une forêt de modèles CART. Quant au terme de « Random » il provient du tirage aléatoire avec remise qui est effectué parmi les prédicteurs en aval de la création des arbres. En anglais cette phase est appelée « bootstrap » et permet théoriquement d'augmenter la performance du modèle. Pour aboutir au résultat final le modèle RF va effectuer un « vote majoritaire » au sein de cette « forêt ».

L'apprentissage via une structure en arborescence pourrait être trivialement résumé de la manière suivante : ce type de modèle sépare les données dans des « sacs » à chaque nœud en utilisant un seuil provenant de prédicteurs définis en entrée du modèle. Le nœud racine est la première séparation effectuée au niveau d'un arbre lors de la phase d'apprentissage. Ce dernier aboutit sur deux autres nœuds dit « enfants », qui se subdivisent ensuite en deux autres nœuds et ainsi de suite, jusqu'à arriver aux « nœuds terminaux » qui contiennent les prédictions finales du modèle (cf. Figure 4). Une fois la phase d'apprentissage terminée, il est possible de récupérer l'importance de chacun des nœuds (IN = Importance du Nœud) qui est calculé à partir de 2 paramètres :

- La position du nœud, c'est-à-dire que le nœud racine est considéré comme un bon séparateur à large échelle, contrairement aux nœuds terminaux qui eux, sont considérés comme des séparateurs à petite échelle.
- L'impureté du nœud, qui mesure le taux d'erreur lors de la ségrégation des données au niveau du nœud considéré.

IN = Importance du nœud

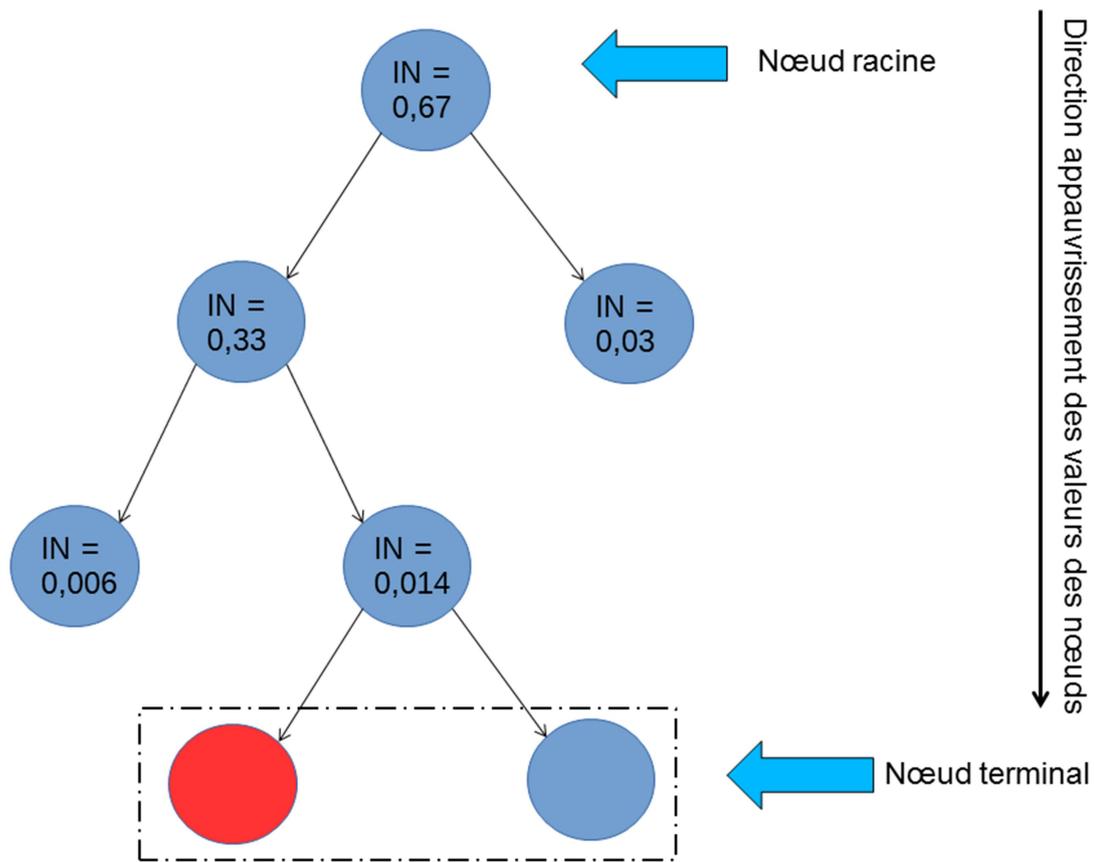


Figure 4. Schéma conceptuel de l'importance du nœud à l'échelle d'un arbre de Random Forest

Il est important de noter que chaque nœud possède aussi une valeur « seuil » de séparation. Par exemple, lors de la phase d'apprentissage le modèle va considérer que la température de l'eau de surface est le prédicteur le plus pertinent pour créer le nœud racine. Ce nœud sera donc le 1^{er} avec une faible valeur d'impureté, pour une séparation vers les deux nœuds enfants effectuée à 12°C. L'un de ces nœuds enfants sera donc en deuxième position avec une valeur d'impureté un peu plus importante et utilisera par exemple le pH comme séparateur avec un seuil de 8,1. Et ainsi de suite jusqu'aux nœuds terminaux qui auront très probablement des valeurs d'impureté plus élevées (cf. Figure 4). Le modèle RF peut être utilisé pour la prédiction de deux types de données qualitatives et quantitatives. Pour la prédiction de données discrètes, ce type de modèle devra être utilisé en mode classification. Et pour effectuer des prédictions sur des données continues, il faudra utiliser ce dernier en mode régression. Quel que soit le mode utilisé, l'architecture du modèle RF est assez souple pour supporter des prédicteurs sous la forme de données continues et/ou discrètes. Dans ce contexte, il est envisageable de prédire un signal constitué de données qualitatives avec uniquement des prédicteurs constitués de données quantitatives.

2.2 Biotypologie et variabilité naturelle des communautés

Dans le cadre législatif de la DCE, les notes indicielles doivent permettre de rendre compte d'un écart à la référence, exprimé en EQR (Ecological Quality Ratios ou en français ratios de qualité écologique). L'idée sous-jacente de cette démarche est de prendre en considération l'impact de l'hydro-morphologie entre les lacs via un système typologique. C'est-à-dire que l'on considère par exemple qu'un lac situé en altitude héberge une flore phytoplanctonique différente de celle d'un lac de plaine, en conditions non impactées. Dans l'annexe V de la DCE (cf. conditions normatives ; page 25), il y existe ainsi deux systèmes proposés pour caractériser les différents types de masse d'eau. Le premier système (« A »), est basé sur l'utilisation des HER (Hydro-Eco-Région), qui subdivise le territoire français en 22 zones. Le deuxième système (« B »), est quant à lui un peu plus flexible, à condition de tenir compte des facteurs géographiques obligatoires suivants : l'altitude, la latitude, la longitude, la profondeur, l'alcalinité.

2.2.1 Partitionnement via la méthode des K-moyennes

Dans le cas de l'IPLAC (Indice Phytoplancton lacustre), les auteurs utilisent une typologie basée sur la profondeur et l'altitude des lacs (Laplace-Treytore et Feret, 2016) car la typologie actuelle des plans d'eau n'est pas satisfaisante. Ce qui leur permet de s'affranchir de ce système de « groupe géographique » (système A), tout en conservant l'idée initiale. Nous nous sommes inspirés de cette approche afin de déterminer des groupes typologiques de lacs au sein de notre base. La méthode des K-means (ou K-moyennes en français) fait aussi partie de la famille des algorithmes de « machine learning ». Mais contrairement au modèle RF, cette méthode est utilisée afin de partitionner les données (Hartigan et Wong, 1979), ce qui correspond bien à notre problématique de zonations. En amont de la phase d'apprentissage, il faut définir le nombre de groupes que l'on veut obtenir. Dans un premier temps nous avons donc essayé de restreindre le nombre de groupes.

2.2.2 Corrélations directes entre traits morpho-fonctionnels et paramètres physicochimiques

Une fois les groupes typologiques définis, nous avons testé les potentielles corrélations directes qui pourraient exister à l'intérieur de ces derniers, entre les traits morpho-fonctionnels et les paramètres physico-chimiques. En complément d'une recherche de corrélations linéaires *stricto sensu*, nous avons aussi examiné les corrélations de type : loi de puissance (échelle log-log sur les deux axes), logarithmique (échelle log sur l'axe X) et exponentielle (échelle log sur l'axe Y). Un nombre conséquent de tests statistiques classiques sont uniquement compatibles avec des données qui suivent une distribution normale.

Conséquemment, ces transformations en échelle logarithmique permettent d'homogénéiser la forme des distributions de probabilités dans les jeux de données. Dans ce contexte, nous avons aussi utilisé la méthode des « Boxcox » qui permet de log-transformer une distribution de probabilité en base lambda (λ). Ce coefficient λ est calculé de manière automatique par cette méthode, pour que la distribution transformée soit le plus proche possible d'une loi normale. Pour chaque paire, nous calculons la corrélation de Pearson, l'erreur quadratique moyenne (MSE en anglais pour Mean Squared Error). Simultanément, les résidus ont été analysés à l'aide d'un test de Shapiro-Wilk (Shapiro

et Wilk, 1965) et d'un diagramme Q-Q (Quantile-Quantile) (Wilk et Gnanadesikan, 1968), afin de vérifier qu'ils suivent bien une distribution gaussienne. De cette manière il est possible de contrôler la « qualité » d'une régression de type linéaire.

Ainsi, nous avons testé toutes les combinaisons possibles via une structure de contrôle informatique de type boucle. Pour être plus précis nous avons imbriqué deux boucles itératives de type « for », c'est-à-dire que dans un premier temps notre script informatique va prendre un trait donné et tester les différents types de corrélation (linéaire, loi de puissance, ...) avec un paramètre physico-chimique. Puis notre script va boucler en conservant le premier trait sélectionné et passer au paramètre physico-chimique suivant et ainsi de suite. Une fois que toutes les combinaisons avec un trait ont été effectuées, la deuxième boucle va prendre le relais et changer le trait qu'il va tester de nouveau avec tous les paramètres physico-chimiques en retournant dans la première boucle. De cette manière, nous pouvons tester toutes les combinaisons possibles au sein d'un groupe, avant de réutiliser notre script pour réaliser le même travail sur le groupe suivant. Nous avons aussi implanté des structures de contrôle de type alternatives (ou test logique) avec un test « if » (si). Cela nous permet de récupérer les paires en fonction d'un seuil. Par exemple, il est possible de collecter les couples traits/physico-chimie qui ont un coefficient de corrélation supérieur au seuil préalablement fixé (exemple $R^2 > 0,3$).

2.3 Méthodologie d'extraction de seuils physico-chimiques

Des travaux précédents réalisés dans le cadre de la DCE par Irstea (Roubeix et al., 2016) utilisaient une méthode dérivée des modèles Random Forest, appelée Gradient Forest, et permettant d'établir des seuils physicochimiques en se basant sur des communautés d'espèces. Dans le cadre d'études exploratoires, nous avons tenté de reprendre et réadapter certains des concepts de ce modèle à notre problématique des traits morpho-fonctionnels. Afin de faciliter la compréhension, nous utiliserons un exemple concret dans la section suivante. Dans ce test nous essayons de prédire le pourcentage de présence des diatomées (cf. partie 1.2.1 ci-dessus). Nous allons qualifier ce pourcentage de présence comme le signal cible. En entrée du modèle RF, nous avons utilisé 21 prédicteurs géographiques et physicochimiques : la longitude, l'altitude, la profondeur maximale, le volume (des lacs), la surface (des lacs), l'alcalinité, la pénétration de la lumière, la matière organique en suspension, la température de surface, l'oxygène dissout à saturation, le pH, le carbone organique dissout, la conductivité, la concentration en silice, phosphore total, ortho-phosphate, nitrate, nitrite et ammonium. Une fois la phase d'apprentissage effectuée, on contrôle la force de chaque prédicteur et la qualité de la prédiction. Dans notre cas d'étude (cf. Figure 5), le modèle arrive à prédire correctement plus de 80% des observations et la silice est le prédicteur le plus important suivi de près par la pénétration de la lumière dans l'eau exprimée par l'intermédiaire du disque de Secchi. Quant à la concentration en nitrite, elle est considérée comme le prédicteur ayant l'impact le plus faible.

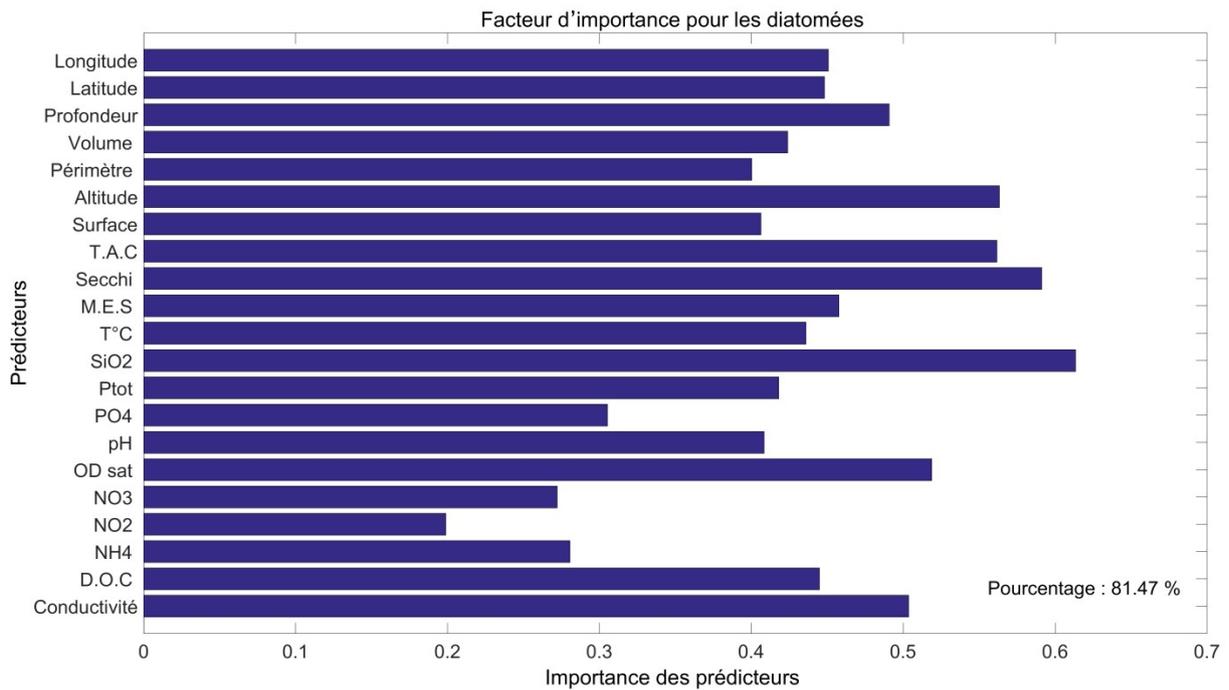


Figure 5. Visualisation de l'importance des prédicteurs en aval de l'apprentissage avec comme signal cible la présence de diatomées

Il est possible de récupérer l'importance des nœuds, ainsi que les seuils de séparation pour un prédicteur donné pour chaque arbre qui constitue la « forêt ». En regroupant ces informations sous la forme d'un histogramme (cf. Figure 6, à gauche), on peut déterminer visuellement les valeurs du prédicteur sélectionné qui sont les plus utilisées pour la séparation dans les nœuds lors de la phase d'apprentissage en fonction d'une classe d'importance moyenne des nœuds. Il est aussi possible de visualiser ces informations sous la forme d'un histogramme cumulatif (cf. Figure 6 à droite). Cette méthodologie permet via la rupture de pente de déterminer un seuil général. Dans notre exemple, on pourrait interpréter cette Figure 6 de la manière suivante : la silice est un élément important pour modéliser le signal cible (pourcentage de présence de diatomées), et globalement le modèle utilise une concentration d'environ 15 mg/L (cf. Figure 6 à gauche dans l'histogramme le pic le plus élevé) pour séparer les données lors de sa phase d'apprentissage. Ce qui potentiellement correspond à une limite biologique dans les écosystèmes lacustres qui influence la présence des diatomées.

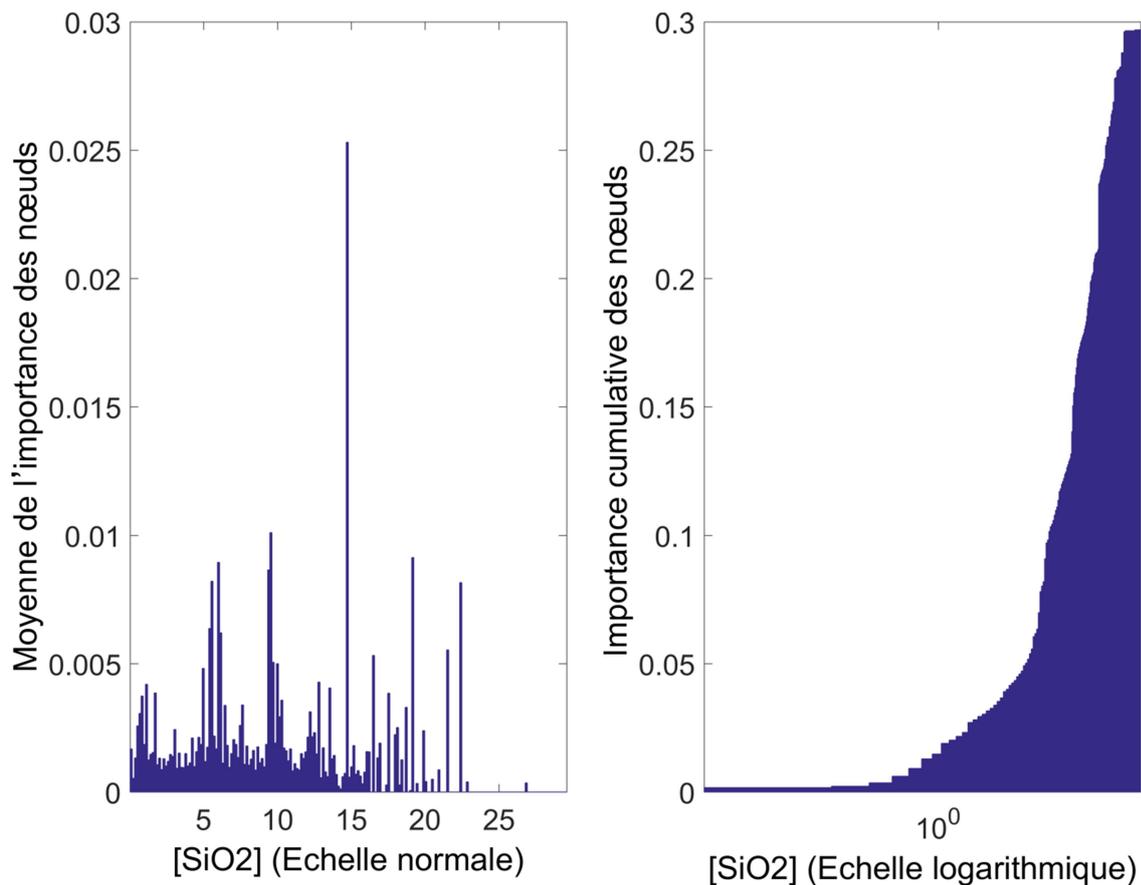


Figure 6. Exemple de seuil pour la silice. A gauche l’histogramme de l’importance des nœuds (échelle normale) moyennée sur l’ensemble de la forêt d’arbres en fonction des teneurs en silice (exprimées sous forme de classes) et à droite le même graphique sous la forme d’une courbe cumulative (échelle logarithmique)

S’il est possible de déterminer une valeur seuil pour un prédicteur donné, il reste encore difficile de savoir s’il exerce une influence positive ou négative sur l’importance du trait dans la communauté. Nous ne pouvons pas déduire directement de cette méthodologie si l’augmentation du prédicteur, pris en compte, a tendance à faire diminuer ou augmenter la réponse du signal cible. Il reste néanmoins possible de regarder la tendance de co-variation de ces deux paramètres de manière classique.

2.4 Utilisation des PDP plots

Afin d’aller plus loin vers notre objectif de déterminer des traits morpho-fonctionnels du phytoplancton, utilisables potentiellement comme métriques de diagnostics, des approches statistiques supplémentaires ont été mises en œuvre en 2018. Pour faire une sélection des traits pertinents au sein des 84 traits renseignés, nous avons testé la qualité (et valeur statistique) de la relation entre chaque trait et les 15 paramètres physico-chimique (prédicteurs). La qualité de la relation a été évaluée via le pseudo r^2 (breiman 2001, Large et al., 2015, Teicher et al., 2016) calculé selon l’Équation 1 suivante

$$\text{Équation 1 : } \text{pseudo } r^2 = 1 - \frac{\text{MSE}}{\text{var}(y)}$$

avec $\text{var}(y)$ qui représente la variance de l'occurrence du trait pondéré par son biovolume et MSE qui correspond à l'erreur quadratique moyenne (Mean Squared Error).

La deuxième étape de cette sélection a été de vérifier si le modèle créé était significativement différent d'un modèle nul. C'est-à-dire de voir si les valeurs prédites par le modèle n'étaient pas dues à l'effet du hasard (North et al., 2002, Teicher et al., 2016). Pour chaque trait nous avons permuté de manière aléatoire les valeurs des prédicteurs d'entrée (valeurs physico-chimiques) puis comparé tous les pseudo r^2 de ces permutations avec ceux issus de la première étape sans permutation en utilisant l'Équation 2

$$\text{Équation 2 : } P_{\text{value}} = \frac{(r+1)}{(n+1)}$$

dans laquelle n est le nombre de permutations (ici $n = 999$) et r le nombre de fois où le pseudo r^2 des permutations est supérieur au pseudo r^2 des modèles non permutés. Pour chaque trait nous avons réalisé 1 000 exécutions du modèle RF (999 avec les prédicteurs permutés et un sans permutation). Chaque modèle RF étant composé de 200 arbres, 16 800 000 arbres de modèles ont été créés à la fin de cette analyse.

En troisième étape, seuls les traits, parmi les 84, qui avaient une qualité de la relation (pseudo r^2) supérieure à 0,3 (Moller & Jennions, 2002, Kruk & Segura, 2012) et une p-value significative ($\alpha < 0,05$) ont été conservés.

Afin de tenir compte des 4 groupes typologiques créés précédemment (cf parties 2.2.1 et 3.1), qui sont importants dans la distribution des communautés phytoplanctoniques, une nouvelle exécution du modèle RF a été réalisée avec chacun des traits morpho-fonctionnels mis en évidence dans les étapes précédentes (pseudo $r^2 > 0,3$ et p-value $< 0,05$) sur chacun d'eux. Cette analyse n'a été mise en œuvre que sur les 4 premiers prédicteurs (paramètres physico-chimiques) qui sont ressortis comme ayant la plus forte influence (selon leur out-of-bag error) sur le trait (voir exemple Figure 7) pour limiter le temps de traitement et garder le plus de pertinence statistique possible.

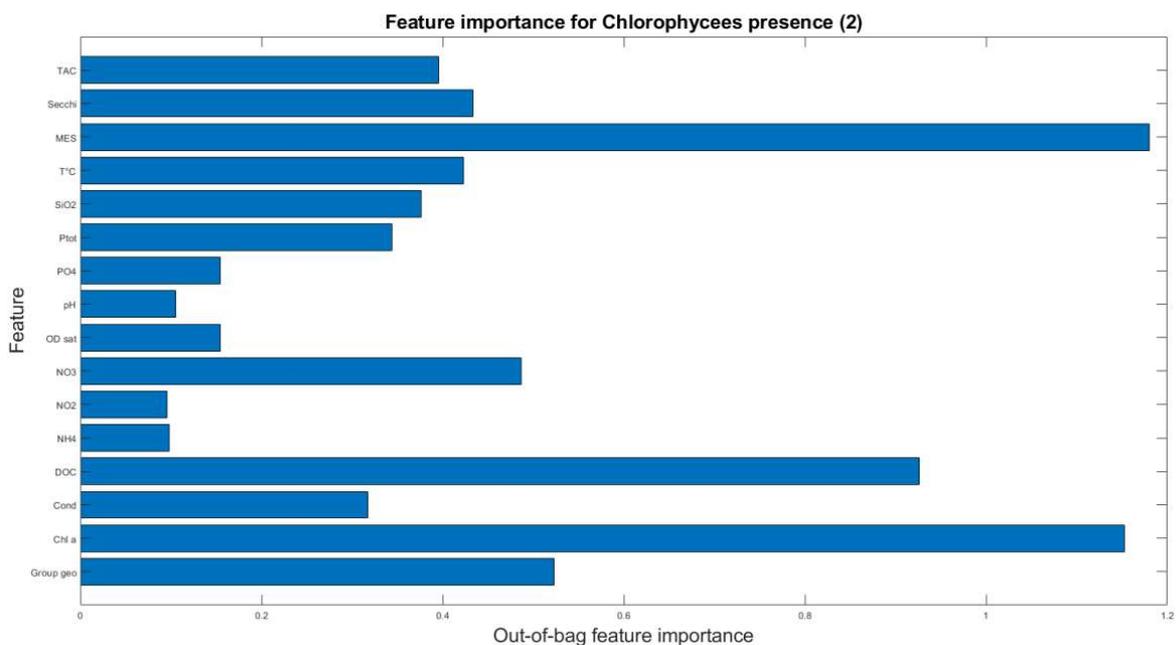


Figure 7. Exemple d'out-of-bag error pour le trait « présence de Chlorophyceae » et les 15 paramètres physico-chimiques

Finalement, l'importance des paramètres physico-chimiques, pour un trait pertinent (exprimé en biovolume), a été examinée via l'utilisation de graphiques de dépendance partielle (Partial Dependence plot – PDP) au sein de chaque groupe typologique comme utilisé par ailleurs par Friedman et al. (2001), Friedman (2001), Cutler et al., (2007) et Teicher et al. (2016). L'ensemble du cheminement peut être résumé graphiquement selon la Figure 8.

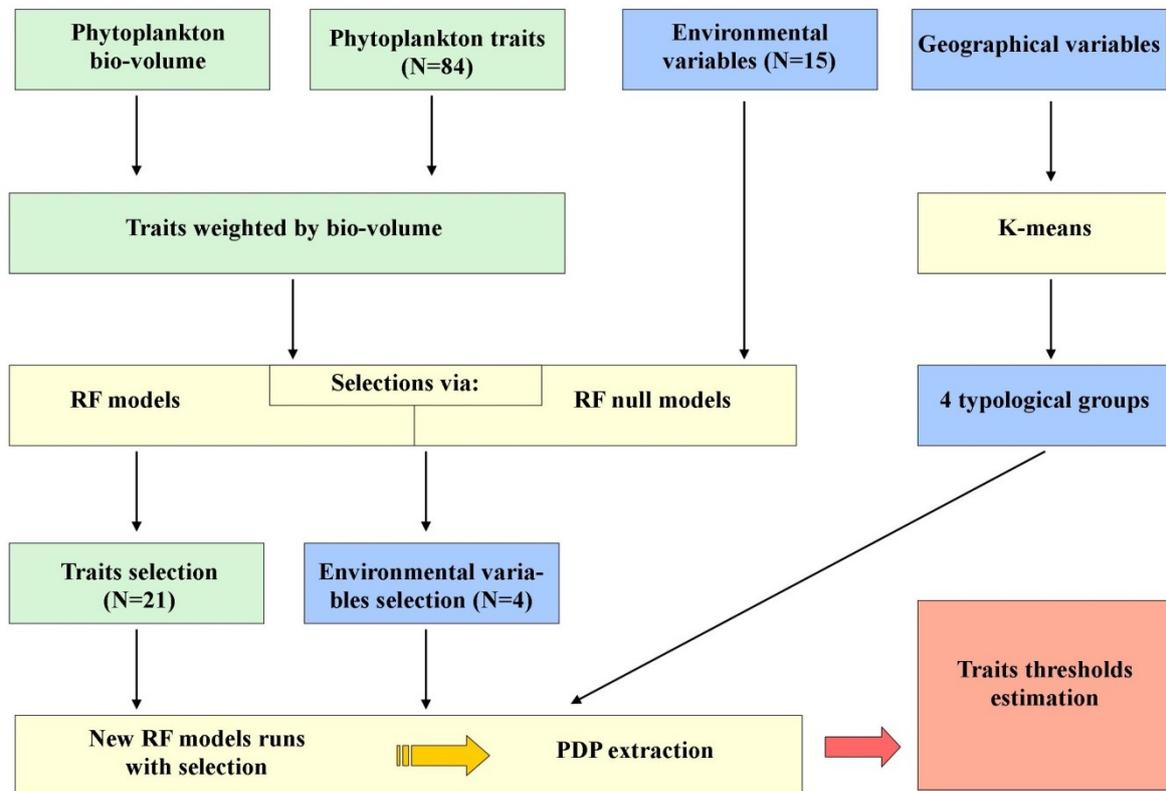


Figure 8. Diagramme conceptuel présentant la méthode de sélection des traits et des seuils physico-chimiques correspondants. Les boîtes vertes sont liées aux traits cibles, les bleues aux paramètres physico-chimiques, les jaunes aux étapes de calculs numériques et la orange aux résultats

Comme pour les modèle RF en mode régression, il n'est pas possible d'extraire l'équation qui lie le trait au paramètre physico-chimique dans un modèle basé sur une structure en arbre (Wagenhoff et al., 2017). C'est pourquoi nous avons utilisé les PDP pour déterminer visuellement les seuils écologiques (Cutler et al., 2007, Roubeix et al., 2016, Wagenhoff et al., 2017) encadrant le gradient de pression. Ces seuils se situent au niveau des deux ruptures de pente de la courbe. Sur le PDP plot, l'axe x correspond au paramètre physico-chimique et l'axe y au trait sélectionné. L'orientation du gradient de pression (augmentant ou se réduisant avec l'augmentation du paramètre physico-chimique) permet de connaître le sens de variation du trait en fonction de la pression contrairement à la méthode d'extraction des seuils de chimie employée dans le chapitre 2.3. Ce type de gradient permet de représenter les différents états de santé de l'écosystème (Holling 1973, Odum et al., 1979, Wagenhoff et al., 2012). Ainsi pour chaque trait pertinent, nous avons gardé les 4 paramètres physico-chimiques les plus influents d'après le modèle RF. Pour chaque paramètre chimique nous avons donc 2 seuils soit au total 8 seuils par trait sélectionné. Toutes les déterminations visuelles de seuils ont été basées sur des graphes PDP en échelle de puissance.

Dans le chapitre 3 ci-dessous, nous allons voir l'application et les résultats obtenus avec ces différentes méthodes, ainsi que les limites à leurs mises en œuvre.

3 Résultats, réorientations méthodologiques et discussion

Dans ce chapitre, nous allons commencer par voir que les relations traits-milieu au sein des écosystèmes lacustres sont difficilement modélisables par des méthodes numériques standards, au vu des mécanismes complexes et des interactions multiples qui régissent ces derniers. La méthode des Gradient Forest a été conçue pour un questionnement spécifique (analyse des distributions de communauté) qui ne correspond pas aux problématiques de cette étude. Mais comme nous l'avons précisé dans la partie 2.3, nous avons réutilisé une partie de ce concept, afin d'extraire des seuils physico-chimiques. Les résultats complets des analyses sont consignés dans l'annexe 1 (cf. tableaux A1-1, A1-2, A1-3 et A1-4), pour les 4 groupes typologiques décrits dans le chapitre 3.2.

Puis dans un second temps, nous allons essayer de remédier à ce problème en utilisant un algorithme d'intelligence artificielle. Il faut garder à l'esprit que pour qu'un modèle de ce type fournisse des résultats fiables, il est important de ne pas négliger certains facteurs. Tout d'abord, la « force » et la pertinence des prédicteurs sont des éléments importants. Le nombre de données disponibles est aussi un facteur qui va impacter la qualité de l'apprentissage. Dans une étude récente utilisant un modèle RF pour effectuer des prédictions de chlorophylle a dans différents types de lacs, il a été montré que lorsqu'on réduisait artificiellement le nombre de données utilisées lors de la phase d'apprentissage, on réduisait aussi la performance du modèle (Yajima et Derot, 2017). Dans ce contexte, nous avons décidé d'effectuer nos analyses via le modèle RF sur l'ensemble de la base de données sans utiliser le système de groupes précédemment présenté, et en intégrant finalement les paramètres typologiques parmi les prédicteurs. Nous avons donc intégré à cette section ces choix méthodologiques, au vu des résultats des méthodes initialement appliquées aux quatre groupes typologiques.

Enfin les résultats obtenus avec la méthode des PDP plots sont détaillés car ils s'avèrent les plus prometteurs pour le développement de métriques de diagnostics.

3.1 Groupes typologiques établis par la méthode des K-means

Pour établir la typologie des sites par la méthode des K-means et sélectionner le nombre optimal de groupes typologiques, il était important de présélectionner les paramètres géographiques les plus représentatifs de cette variabilité spatiale. Par conséquent, nous avons réalisé une ACP (Analyse en Composantes Principales) à partir des données « géographiques » des lacs suivantes : l'altitude, la latitude, la profondeur maximale, la superficie, le périmètre et la longitude. Sur la Figure 9 correspondante, le volume et la profondeur maximale sont les deux paramètres les plus explicatifs le long de l'axe 1 (43 % d'explication de la variance), et l'altitude le paramètre le plus discriminant sur l'axe 2 (29% d'explication de la variance). Dans un souci de continuité avec les travaux précédemment effectués sur l'IPLAC, nous avons décidé de sélectionner la profondeur maximale au détriment du volume comme critère de séparation associée à l'altitude.

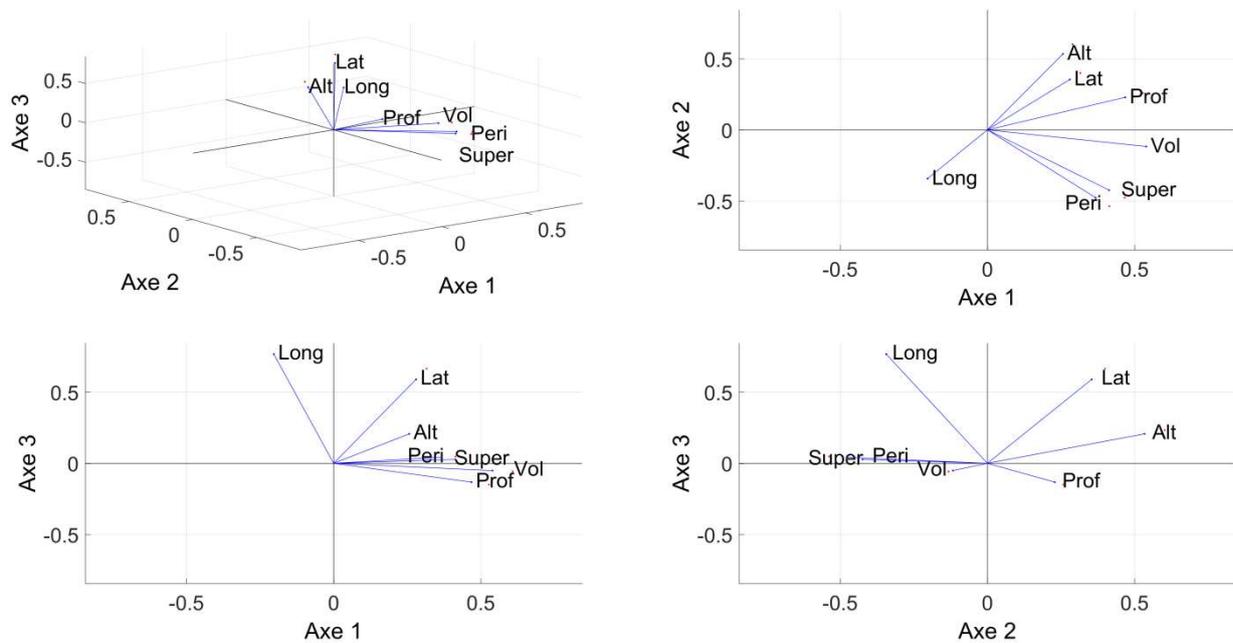


Figure 9. Analyse en composantes principales (ACP) à partir des données « géographiques » des lacs. L'altitude (Alt), la latitude (Lat), la profondeur maximale (Prof), la superficie (Super), le périmètre (Peri) et la longitude (Long)

Dans la Figure 10 est présentée la séparation des données en 4 groupes ce qui permet de conserver au moins 250 points par groupe. Dans l'annexe 2 (cf. figures A2-1, A2-2, A2-3 et A2-4), nous présentons la même analyse, mais avec un nombre de groupes différents variant entre 2 et 6. Ces autres découpages sont moins pertinents car ils ne permettent pas une répartition équitable des points ou échantillons au sein des différents groupes.

Il est important de noter que dans la Figure 10, nous avons aussi préalablement transformé les données en log. Cela nous permet de réduire l'impact de l'écart important présent au sein des données de profondeur et d'altitude des lacs. Afin de faciliter la lecture de cette figure, nous avons rajouté la correspondance des bornes des groupes en échelle normale.

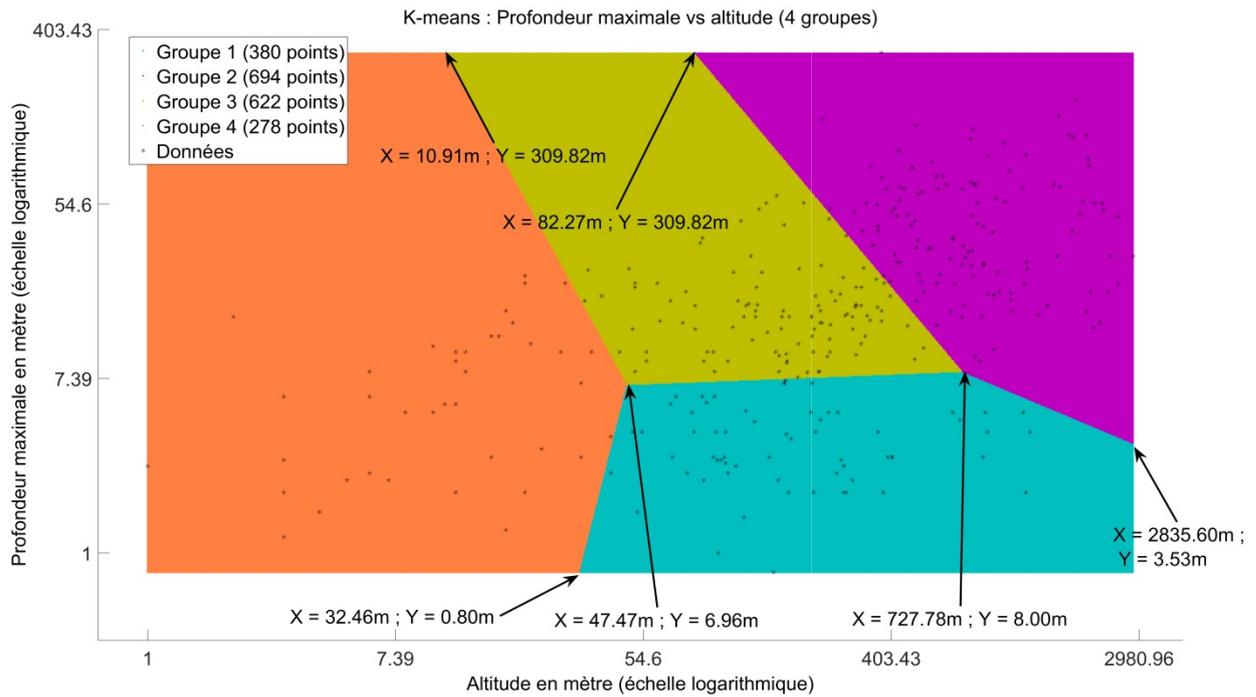


Figure 10. Séparation en 4 groupes géographiques via la méthode de K-means basée sur la profondeur maximale et l'altitude des lacs du jeu de donnée phytoplanctonique. Chacune des zones de couleurs représentent un groupe créé par l'algorithme

Finalement, cette méthode nous a permis de diviser notre base de données initiale en 4 groupes qui sont fonction de l'altitude et de la profondeur maximale des lacs, en conservant un nombre de données suffisant par groupe pour effectuer des analyses numériques robustes. Le premier groupe (dans la zone bleue en bas à droite sur la figure correspond à des lacs peu profonds situés en moyenne altitude, souvent situés près du bassin méditerranéen. Le deuxième groupe (dans la zone pourpre au milieu en haut) correspond aux lacs profonds de moyenne altitude, représentés essentiellement par les lacs de barrages artificiels. Le troisième groupe (dans la zone verte à gauche) correspond aux lacs de plaine relativement peu profonds, représentés par des lacs généralement naturels situés près de la côte. Le quatrième groupe (dans la zone orange en haut à droite) correspond aux lacs profonds de haute altitude, principalement représentés par des lacs naturels alpins/montagneux.

3.2 Corrélations directes

Disposant de 4 groupes géographiques distincts (cf. partie 3.1), les différentes corrélations directes entre traits morpho-fonctionnels et paramètres physico-chimiques ont été testées (voir un exemple en Figure 11). La distribution gaussienne des résidus a également été vérifiée systématiquement.

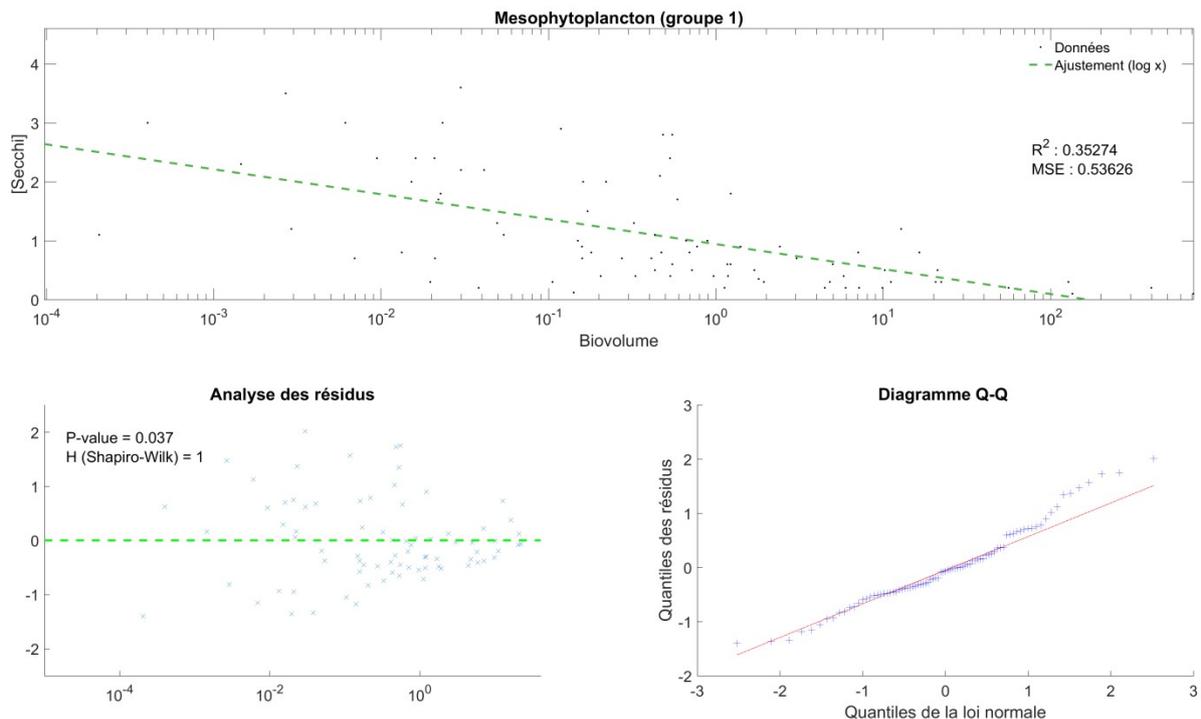


Figure 11. En haut : test de corrélation entre la pénétration de la lumière dans l'eau (Secchi) et l'occurrence binaire de la présence du trait « Mésophytoplancton » exprimé en biovolume, dans le groupe 1. En bas à gauche : représentation graphique des résidus et test de Shapiro-Wilk. En bas à droite : Diagramme Q-Q des résidus

Au final, les coefficients de corrélation les plus élevés que l'on trouve dans ces 4 groupes sont de l'ordre de 0,4 tous paramètres confondus. Il est aussi important de remarquer que si l'on applique la même procédure sur notre base de données non partitionnée (i.e. par biotype), cela n'améliore pas les coefficients de corrélation. Une des corrélations directes les plus importantes dans le groupe 1 concerne le trait « Mésophytoplancton » et la pénétration de la lumière dans l'eau exprimée par la mesure de la transparence au disque de Secchi (cf. Figure 11). Il est étonnant de ne retrouver aucune relation significative entre un trait et le Carbone Organique Dissous ou DOC qui peut, quant à lui, témoigner d'une pression anthropique. Contrairement aux mesures au disque de le Secchi nous ne retrouvons aucune autre corrélation significative dans ces 4 groupes (cf. Figure 12) quelle que soit l'échelle utilisée (normale, loi de puissance, ...). Il est fortement probable que les limites de détection de différentes mesures physico-chimiques et la fréquence d'échantillonnage faible, ainsi que les biais internes inhérents aux données des traits, affecte négativement ce type d'analyse de corrélations directes. Dans ce contexte, la robustesse d'un grand nombre de corrélations que nous avons trouvées est discutable. Il est donc impératif de trouver une méthodologie utilisant des outils numériques capables de pallier ces limites de notre jeu de données. Dans le chapitre suivant, nous allons utiliser un modèle de « machine learning » prédictif, afin de voir si l'on obtient de meilleurs résultats.

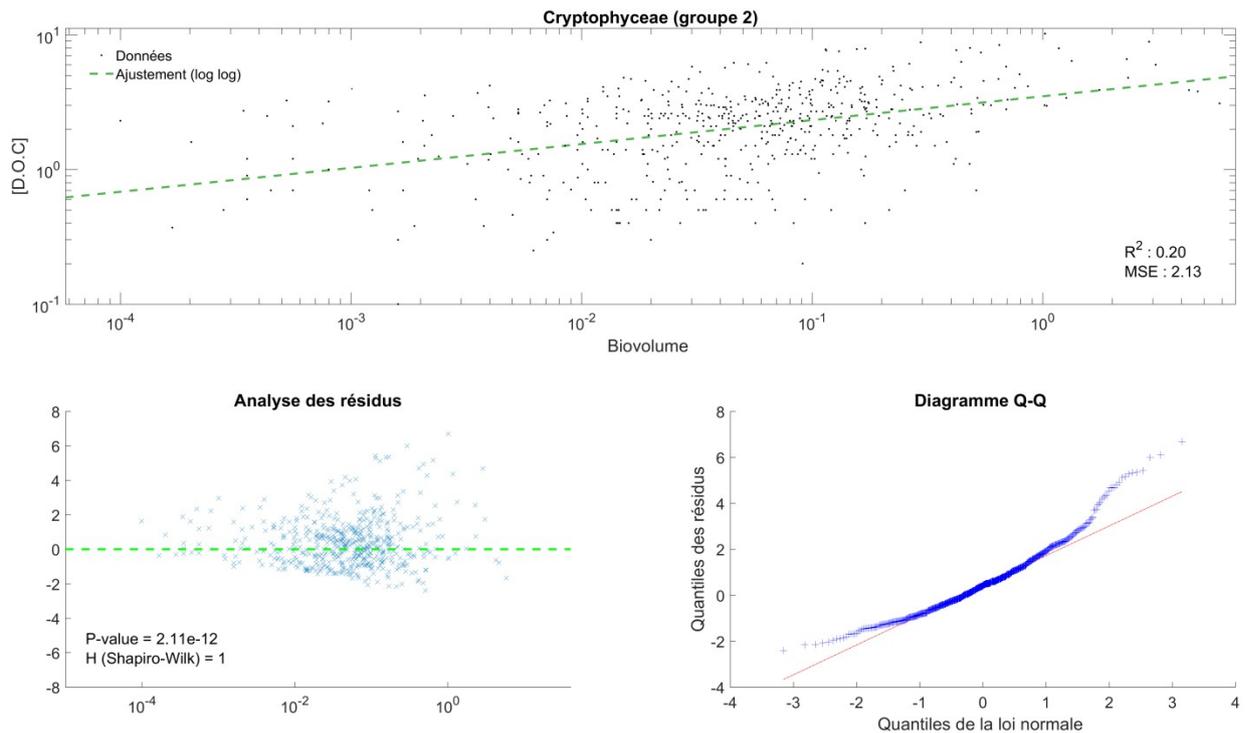


Figure 12. En haut : test de corrélation entre carbone organique dissous (DOC) et l'occurrence binaire de la présence du trait « cryptophyceae » exprimé en biovolume, dans le groupe 2. En bas à gauche : représentation graphique des résidus et test de Shapiro-Wilk. En bas à droite : Diagramme Q-Q des résidus

3.3 Limites de classes et transformation Boxcox

Des travaux antérieurs ont permis de définir des « sites de référence », afin de pouvoir caractériser des lacs considérés comme « non perturbés » ou « très faiblement perturbés » par des pressions anthropiques. Il est donc possible de partitionner notre base de données en deux groupes : un groupe contenant uniquement les lacs de référence (noté « ref » dans les Figure 13, A3-3 et A3-5 de l'Annexe 3) et un autre groupe constitué de tous les autres lacs (noté « other » dans les Figure 13, A3-3 et A3-5 de l'Annexe 3). Il est important de noter que dans ce chapitre nous avons majoritairement effectué nos analyses sur des données « log-transformées ». Une transformation de ce type a l'avantage de réduire les écarts importants qui sont dus à l'hétérogénéité de notre base de données qui s'étend sur l'ensemble du territoire français métropolitain. Par exemple, si l'on prend le cas de la variable altitude beaucoup de lacs vont être localisés en plaine : plus la valeur de cette variable augmente, plus le nombre de lacs diminue.

Dans l'IPLAC, les auteurs utilisaient déjà une transformée logarithmique classique en base 10 au niveau du phosphore total. Dans notre cas, nous avons choisi d'utiliser la transformée logarithmique en base λ (lambda) via la méthode des « Boxcox » (Box et Cox, 1964). Ce coefficient lambda est calculé automatiquement par cette méthode, afin que la distribution des données transformées soit le plus proche possible d'une loi normale. Par conséquent, ce type de transformation va nous permettre d'un côté de réduire l'écart entre les données et d'un autre côté de pouvoir utiliser efficacement les outils statistiques.

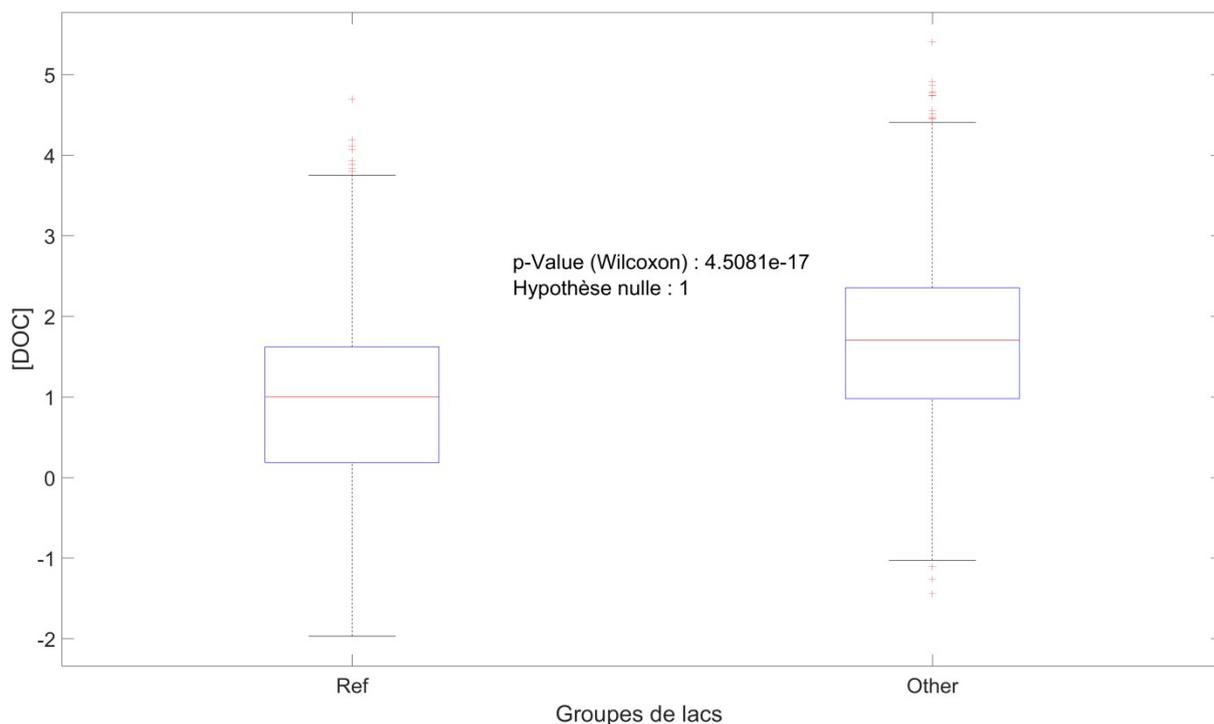


Figure 13. Visualisation des distributions entre le groupe des lacs de références et le groupe des autres lacs via des boxplots, pour le DOC (transformé avec la méthode des Boxcox). Et p-Value provenant du test de Wilcoxon réalisée à partir de ces deux distributions

Dans la Figure 13, nous avons regardé via un test statistique non paramétrique de Wilcoxon-Mann-Whitney (Wilcoxon, 1945) si la distribution du DOC entre ces deux groupes de lacs était identique ou non. Que les données soient log-transformées ou non, leurs distributions restent analogues (cf. figure A3-1 dans l'annexe 3). Les résultats montrent que la distribution des valeurs de carbone organique dissous diffère de manière statistique entre le groupe référence et le groupe des autres lacs (p -value $\ll 0,01$). On peut aussi remarquer que les données provenant des sites de référence ont tendance à contenir des valeurs de DOC plus faible que les autres lacs. De plus, dans le cadre de la DCE, le DOC fait partie des paramètres sensibles à surveiller, car des fortes concentrations de ce dernier sont bien souvent associées à des milieux anthropisés. Comme nous l'avons vu précédemment (cf. chapitre 3.2 et Figure 12), bien que faible, il existe tout de même un lien direct entre certains traits morpho-fonctionnels et le DOC. Conséquemment, dans le reste de ce chapitre nous avons essayé de voir s'il était possible d'améliorer la corrélation entre les traits morpho-fonctionnels et le DOC via l'utilisation d'un modèle de « machine learning ». Dans la continuité de nos travaux de recherche, cette méthodologie a été aussi appliquée aux autres paramètres physico-chimiques présents dans la base de données PLAN_DEAU mais aucun résultat significatif n'est ressorti.

Dans la Figure 14 (graphique du bas), on peut voir via un histogramme que la distribution des données transformées suit bien une forme de « cloche » caractéristique d'une loi normale (cf. figure A3-2 de l'annexe 3 pour visualiser l'histogramme avec les données non transformées). Quant à l'axe des ordonnées du graphique du haut de la Figure 14, il s'agit d'une échelle théorique qui va être utilisée comme signal cible lors de la phase d'apprentissage du modèle RF. Il est primordial de bien

comprendre cette notion d'« échelle théorique » : l'idée est de maximiser le nombre de données issues des lacs de référence dans l'intervalle qui est supérieur à 0,8 (seuil fixé arbitrairement pour ce test, mais qui sera calé par la suite sur la distribution des valeurs de DOC dans les lacs de référence), afin que ces lacs soient classifiés dans un état dit « très bon ».

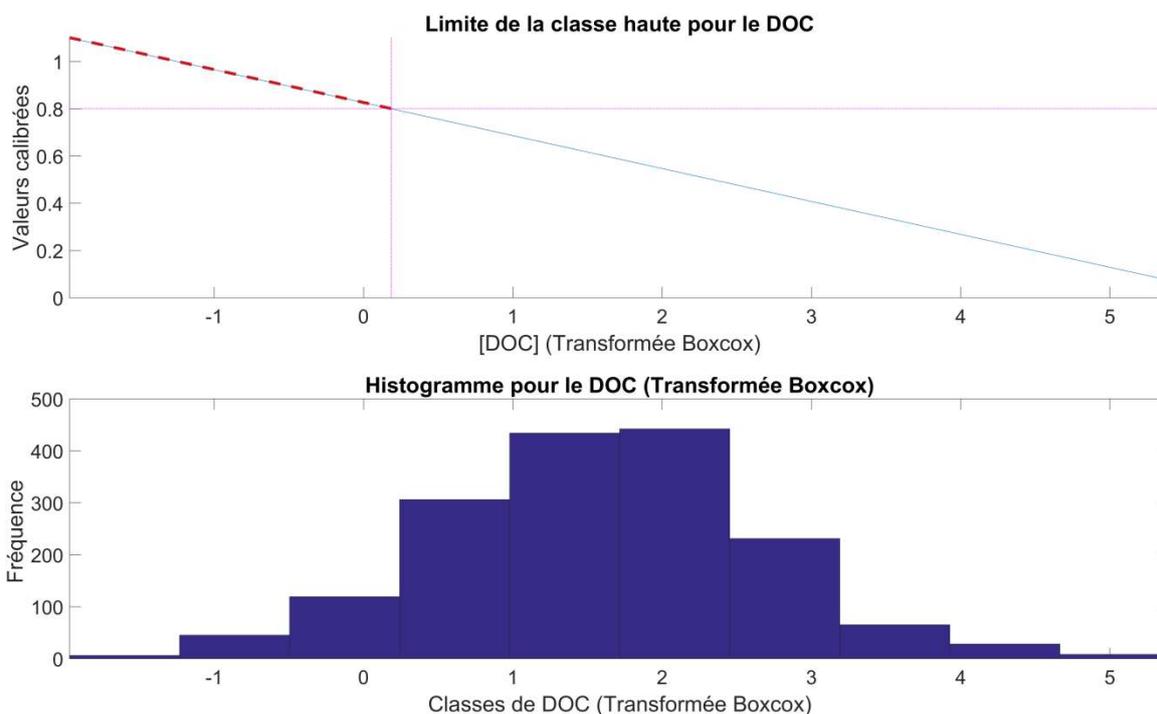


Figure 14. Sur les deux graphiques les axes des x correspondent à la concentration en DOC transformée par la méthode des Boxcox. En haut : Représentation graphique de la notion d'échelle théorique qui va être utilisée comme signal cible en entrée de notre modèle RF. La ligne en pointillé rouge correspond à la classe DCE « très bon ». En bas : histogramme représentant la distribution des données de DOC transformées

Dans la figure A3-3 en annexe 3, les croix rouges représentent les données de DOC, réparties sur leurs échelles théoriques respectives (non transformées sur le graphique du haut et transformées sur le graphique du bas). On remarque sur cette figure que la transformation via la méthode Boxcox permet bien de changer la forme de la distribution des données (logarithmiques vers gaussienne). Il faut garder à l'esprit qu'il y a un biais interne dans les données provenant de ces sites de références. C'est-à-dire que pour certains de ces sites nous disposons uniquement d'un seul d'échantillonnage, alors que pour d'autres nous disposons de plusieurs échantillonnages, mais avec une forte variabilité entre les saisons et les années de prélèvement. C'est une information importante à garder à l'esprit pour l'interprétation de ces données.

3.4 Modèle RF en mode régression

Le mode régression d'un modèle RF permet d'effectuer des prédictions de données continues. Ici, nous testons la prédiction de la métrique transformée DOC présentée ci-dessus à partir du jeu de traits morpho-fonctionnels. Dans un premier temps nous avons séparé aléatoirement notre base de données en deux jeux de données, en y incluant des lacs de référence. Le premier groupe (jeu de données d'apprentissage) est composé de 90 % des données et va être utilisé lors de la phase d'apprentissage du modèle. Le deuxième groupe (jeu de données test) va nous servir à tester les performances du modèle lors de la phase de contrôle. Dans la Figure 15, nous pouvons voir l'importance de chacun des traits morpho-fonctionnels qui ont été utilisés en entrée de notre modèle. On voit clairement qu'il y a une forte variabilité dans l'importance de ces traits pour la prédiction de notre signal cible. Il sera donc important d'effectuer une pré-sélection dans de futurs travaux, afin de diminuer le temps de calcul, d'éviter les problèmes de sur-apprentissage (« overfitting ») et de potentiellement augmenter les performances de prédictions. Dans le cadre de cette étude, nous avons limité la taille de la « forêt » à 200 arbres, car on voit clairement sur la figure A3-4 en annexe 3 que l'erreur de classification dans le cas du DOC uniquement devient assez faible et stable au-dessus du seuil de 80 arbres. Cela nous permet d'optimiser la vitesse de calcul lors de la phase d'apprentissage.

Dans la Figure 16, les données de l'axe des y (métrique test) proviennent des prédictions effectuées par le modèle avec les traits. Les ronds rouges sur cette figure représentent les données prédites par le modèle RF, provenant uniquement du jeu de données test. Il n'y a donc pas de différence de performance du modèle selon les groupes de lacs (référence et autres) dans ce test. Les croix bleues, quant à elles, représentent uniquement les lacs de référence. Vu que les données utilisées pour ce test ont été sélectionnées de manière aléatoire entre le groupe de test et d'apprentissage, il n'est donc pas anormal que dans certains cas nous trouvions une superposition de croix et de cercles. On remarque que les coefficients de corrélations (classiques et ajustés) entre nos valeurs calibrées prédites et les données de DOC log-transformées sont proches de 0,6. Et que la pente de régression est significative avec une p-value faible (très inférieure à 0,05). La gamme de variation de notre métrique test est comprise entre 0,8 et 0,3.

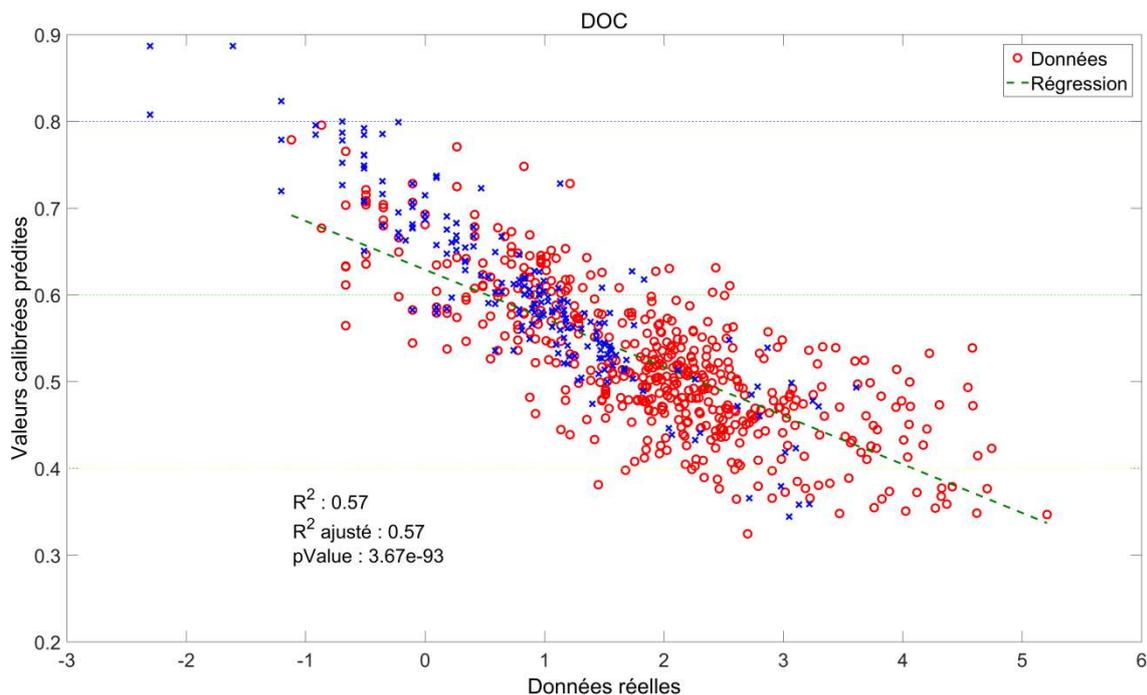


Figure 16. Test des valeurs calibrées prédites au niveau du DOC à partir d'un modèle RF ayant appris avec uniquement les traits morpho-fonctionnels. Les ronds rouges sont issus du jeu de données de test. Et les croix bleues sont calculées exclusivement à partir des lacs de références. La ligne verte en pointillé montre la droite de régression calculée à partir du jeu de données test (ronds rouges).

Nous avons aussi étudié le comportement des résidus de la régression linéaire dans la Figure 17 au moyen d'un diagramme quantile-quantile et d'un test de Shapiro-Wilk. Par l'intermédiaire de cette figure, on remarque que les résidus issus de notre métrique test sont proches d'une distribution gaussienne, traduisant la qualité de l'ajustement de cette régression linéaire. A l'exception de quelques valeurs élevées (supérieures à 2,5) dans le diagramme quantile-quantile, on voit que la majorité des points bleus jouxtent la ligne rouge qui représente une distribution normale théorique (cf. Figure 17 en bas).

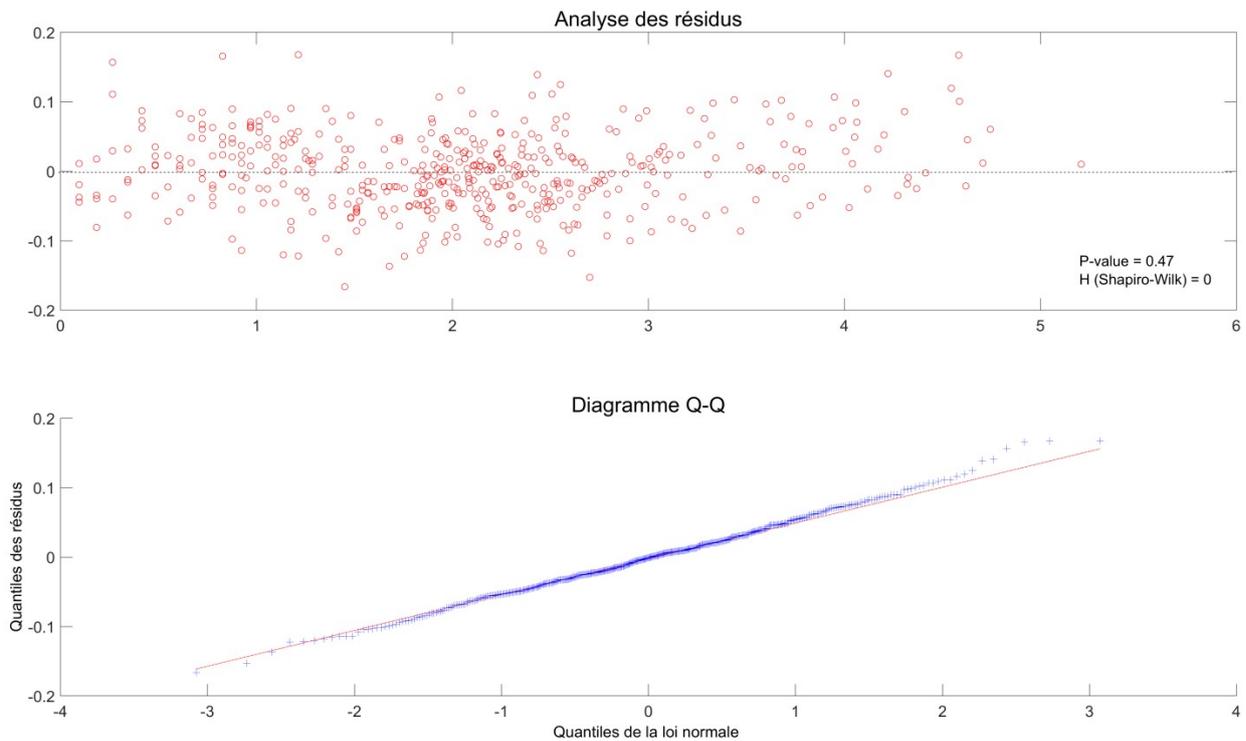


Figure 17 Test d'homoscédasticité des résidus de la régression linéaire présentée dans la Figure 16. En haut : visualisation des résidus, test de Shapiro et p-Value. En bas : représentation de ces résidus via un diagramme quantile-quantile.

3.5 Modèle RF en mode classification

Il est aussi possible d'utiliser un modèle RF en mode classification, afin d'effectuer des prédictions pour des données discrètes. La finalité d'un indice de bio-indication est de classer une masse d'eau donnée, dans notre cas un lac, dans l'une des 5 classes états définies par la DCE (High, Good, Moderate, Poor, Bad). Dans ce contexte, nous avons essayé dans ce chapitre de nous affranchir de l'étape de la régression linéaire présentée dans la partie 3.4 (cf. Figure 16). Comme le modèle RF est capable de prédire des données catégorielles, nous avons directement et simplement alloué un des 5 états à chaque échantillonnage via notre échelle théorique précédemment présentée (cf. partie 3.4). C'est-à-dire que toutes les données qui étaient supérieures à 0,8 ont été placées dans la classe High, puis les données entre 0,8 et 0,6 dans la classe Good et ainsi de suite.

Il est important de noter qu'afin de se rapprocher de la DCE compatibilité, nous avons rajouté dans les prédicteurs qui étaient jusqu'à présent uniquement constitués des traits, les données d'altitude et de profondeur des lacs. Ceci permet d'intégrer la variabilité naturelle des flores selon la typologie des lacs, en s'affranchissant d'une typologie préalablement établie qui affaiblissait la pertinence des résultats (nombre trop restreint d'échantillons par type). Il est important de garder à l'esprit que lors de sa phase d'apprentissage, le modèle RF place les données dans des groupes ou des « sacs ». C'est d'ailleurs pour cela que l'on parle de « out-of-bag error ». Par conséquent, comme nous l'avons vu dans le chapitre 2, les prédicteurs ayant un facteur d'influence important sont majoritairement utilisés pour créer le nœud racine, ou ses nœuds enfants directs.

Dans la Figure 18, on voit que les deux prédicteurs géographiques que nous avons rajoutés par rapport à la partie précédente ont une très forte influence sur la qualité de la prédiction. Ils ont une importance supérieure à 1 (cf. les deux barres bleues les plus en haut sur le graphique). A première vue, cela peut sembler anodin, cependant ce comportement est très intéressant du point de vue de notre problématique. On peut interpréter ce phénomène de la manière suivante : l'altitude et la profondeur des lacs sont sélectionnées prioritairement par le modèle RF pour créer le nœud racine ou/et ses nœuds enfants direct. Par conséquent, les données des traits morpho-fonctionnels sont préalablement séparées en interne dans des « sacs disjoints », lors de la phase d'apprentissage par ces deux paramètres géographiques. Ce qui correspond à l'une des attentes de la DCE, en ce qui concerne la séparation des données en groupes géographiques, et qui permet de s'affranchir de l'étape préliminaire des K-moyennes présentée dans le chapitre 2. Les modèles de Machine Learning fonctionnant en « boîte noire », il est difficile de savoir de manière précise comment le modèle a segmenté les groupes via la topographie des lacs lors de sa phase d'apprentissage. Mais comme nous l'avons vu dans la partie 2.2, il est possible d'extraire une valeur de séparation pour chaque prédicteur. Dans ce contexte, des tests ont été réalisés, afin de déterminer s'il était possible de récupérer cette typologie que le modèle RF crée en interne lors de sa phase d'apprentissage. Cependant cela n'a pas été concluant (résultats non montrés ici).

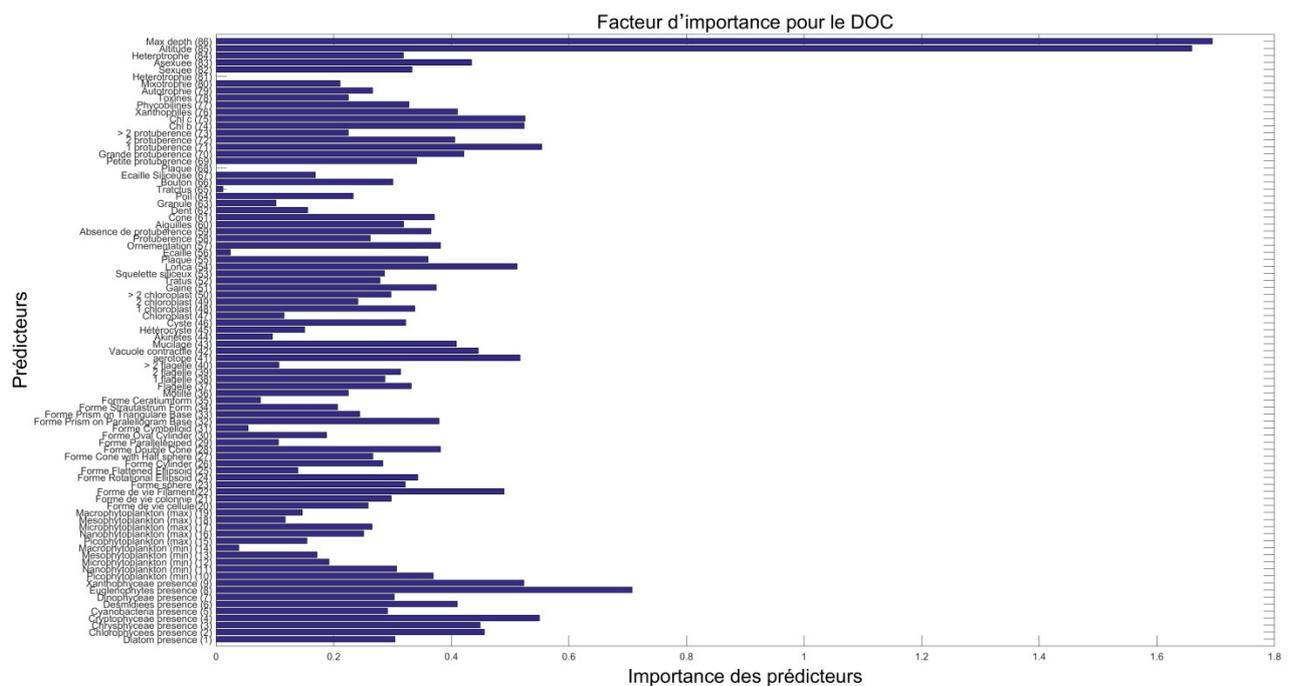


Figure 18. Importances des prédicteurs (traits morpho-fonctionnels) calculées à partir de la phase d'apprentissage du modèle RF en mode classification.

Dans le but de contrôler les performances du modèle, nous avons utilisé une matrice de confusion (cf. Figure 19). Nous avons conservé les deux groupes précédemment présentés : jeu de données d'apprentissage (90 %), jeu de données test (10 %). La matrice de confusion a été construite avec les sorties du modèle que l'on compare à notre groupe test. Cette matrice de confusion nous permet de voir comment sont réparties les erreurs de classification. Par exemple pour la classe « High » on peut voir que 73,3 % des données dans cette classe ont été correctement prédites (cf. le cadre gris en bas à gauche de la figure). Toujours pour cette même classe, 11 lacs ont été bien

classifiés et 4 ont été mal classés (cf. le cadre vert et rouge en haut à gauche de la figure). On peut aussi voir qu'aucun lac n'a été mal classifié dans les 3 autres classes (Moderate, Poor et Bad). Le pourcentage en dessous dans chaque cadre rouge et vert indique le pourcentage par rapport au total. En d'autres termes, si l'on reprend le cas du cadre vert en haut à gauche de la Figure 19, 11 lacs ont été classés dans l'état « High » ce qui représente 6,6 % des 166 lacs qui composent notre jeu de données test. Sur la globalité, on peut voir que le modèle RF arrive à classer correctement 66,9 % des données et que sur les 33,1 % de données mal classifiées, 30,7 % sont classées dans des classes directement adjacentes. Ces dernières données pourraient correspondre à des échantillons en « limite de classe ». Ainsi, 97,6 % des données sont classifiées de manière satisfaisante, confirmant la robustesse de l'approche RF proposée.

Confusion Matrix

Output Class	Class H	Class G	Class M	Class P	Class B	Mean
Class H	11 6.6%	1 0.6%	0 0.0%	0 0.0%	0 0.0%	91.7% 8.3%
Class G	4 2.4%	55 33.1%	14 8.4%	0 0.0%	0 0.0%	75.3% 24.7%
Class M	0 0.0%	13 7.8%	36 21.7%	13 7.8%	4 2.4%	54.5% 45.5%
Class P	0 0.0%	0 0.0%	2 1.2%	7 4.2%	3 1.8%	58.3% 41.7%
Class B	0 0.0%	0 0.0%	0 0.0%	1 0.6%	2 1.2%	66.7% 33.3%
Mean	73.3% 26.7%	79.7% 20.3%	69.2% 30.8%	33.3% 66.7%	22.2% 77.8%	66.9% 33.1%
	Class H	Class G	Class M	Class P	Class B	Mean

Target Class

Figure 19. Matrice de confusion entre les classes de DOC prédites par le modèle et leurs classes réelles.

Lorsque l'on regarde la répartition des lacs en fonction des classes via leurs pourcentages dans la matrice de confusion (cf. Figure 19), on s'aperçoit que plus de la moitié des données sont distribuées entre les classes « Good » et « Moderate ». Afin de tester le comportement du modèle, nous avons modifié la façon d'affecter les classes ; afin que ces dernières comportent un nombre de données quasiment équivalentes par classe (cf. figure A3-5 en annexe 3). Par conséquent, dans ce test, chacune

des classes est composée d'environ 30 points. Comme l'on peut le constater par l'intermédiaire de la matrice de confusion (cf. figure A3-6 en annexe 3), le pourcentage de prédictions bien classées est légèrement plus faible (63,9 %) que précédemment (66,9 %), mais reste acceptable. Ce dernier test a uniquement pour vocation de mettre en avant la grande capacité adaptative du modèle RF.

3.6 Résultats de l'utilisation des PDP Plots

3.6.1 Traits associés avec un paramètre physico-chimique

Parmi les 84 traits morpho-fonctionnels du phytoplancton définis initialement, 21 traits semblent être significativement influencés par les paramètres physico-chimiques (valeur $p < 0,05$ et pseudo- $R^2 > 0,3$). Ces 21 traits (voir Tableau 2) englobent de nombreux types de traits :

- phylogénie (présence de Chlorophyceae et présence d'Euglenophyceae) ;
- taille des cellules (nano-plancton [taille minimale] et micro-plancton [taille maximale]) ;
- forme de vie (solitaire ou coloniale) ;
- forme (forme de double cône) ;
- motilité ;
- présence de vacuoles contractiles ;
- nombre de chloroplastes (absence ou présence d'un chloroplaste) ;
- ornementation (présence d'écailles, présence de protubérances, présence d'aiguille, présence d'une grande taille de protubérance) ;
- type de pigments (présence de chlorophylle b, présence de xanthophylle) ;
- source d'apport en carbone (autotrophie) ;
- et type de reproduction (asexuée ou sexuelle / asexuée).

Tableau 2. Vue synthétique des 21 traits sélectionnés par la méthode des PDP et leurs réponses aux paramètres physico-chimiques. II (Impact Initiation) correspond au premier seuil et IC (Impact Cessation) correspond au second. La flèche indique le sens de variation de la relation du trait par rapport au paramètre physico-chimique

Trait	Chlorophylle-a			Carbone Organique dissous (COD)			Nitrates (NO3)			Matières en Suspension (MES)			Transparence au disque de Secchi			Titre Alcalimétrique Complet (TAC)			Température		
	II	IC	direction	II	IC	direction	II	IC	direction	II	IC	direction	II	IC	direction	II	IC	direction	II	IC	direction
Présence de Chlorophyceae (2)	1,12	92,76	↗	0,83	4,06	↗	0,25	4,95	↘	3,49	42,52	↗									
Présence de Euglenophyceae (8)	1,70	40,45	↗	2,01	13,60	↗	0,44	27,94	↘	2,14	17,64	↗									
Classe de taille nano-plankton (minimale) (11)	1,03	43,82	↗	1,28	12,81	↗				4,66	62,18	↗	0,68	7,03	↘						
Classe de taille micro-plankton (maximale) (17)	1,57	36,60	↗	0,24	8,17	↗				1,99	62,18	↗	0,56	4,71	↘						
Forme de vie solitaire (20)	1,12	47,47	↗	0,36	9,97	↗				1,51	62,18	↗				0,90	18,73	↗			
Forme de vie filamenteuse (22)	2,18	60,95	↗	1,00	8,25	↗	0,34	5,37	↗	4,06	25,03	↗									
Forme double cône (28)	1,12	129,02	↗	0,93	8,25	↗				1,51	40,85	↗	0,59	7,61	↘						
Motilité (36)	0,95	47,47	↗	0,41	8,25	↗				3,25	57,97	↗	0,37	4,95	↘						
Présence de pseudovacuoles contractiles (42)	0,42	151,41	↗	0,88	22,87	↗				3,78	94,63	↗	0,42	12,18	↘						
absence de chloroplaste (47)	1,21	36,97	↗	0,36	9,97	↗				1,51	62,18	↗	0,96	22,20	↘						
Présence d'un chloroplaste (48)	1,21	151,41	↗	0,34	25,79	↗				1,99	40,85	↗	0,84	9,97	↘						
Présence d'écailles (57)	1,12	36,97	↗	1,57	27,66	↗				1,51	40,85	↗	1,11	16,44	↘						
Présence de protubérances (58)	1,21	47,47	↗	1,57	12,94	↗	0,29	4,95	↘	1,51	38,09	↗									
Absence de protubérance (59)	1,03	43,82	↗	0,83	24,29	↗	0,82	15,49	↗	1,99	61,56	↗									
Type de protubérance: aiguille (60)	1,68	165,67	↗	1,21	22,87	↗				1,51	50,40	↗				1,06	10,80	↘			
Taille de protubérance : grosse (70)	1,03	151,41	↗	0,32	11,36	↗	0,64	10,91	↘	1,51	40,85	↗									
Présence de chlorophylle b (74)	1,03	28,79	↗	0,83	9,39	↗	0,34	14,88	↘	1,51	40,85	↗									
Présence de xanthophylle (76)	1,03	28,79	↗	0,83	22,87	↗	0,32	17,46	↘	1,40	40,85	↗									
Autotrophie (79)	1,32	28,79	↗	0,23	9,97	↗	0,32	13,46	↘	1,51	49,40	↗									
Reproduction asexuée (83)	1,99	40,04	↗	0,41	8,76	↗				1,99	48,91	↗							7,32	17,99	↗
Reproduction sexuée/asexuée (84)	1,68	151,41	↗	0,57	25,03	↗				3,25	66,02	↗	1,11	10,70	↘						

Parmi ces 21 traits sélectionnés, la plupart sont communs à ceux utilisés dans les études antérieures visant à identifier un nombre restreint de groupes fonctionnels (Reynolds et al. 2002; Padišák et al. 2009; Kruk & Segura 2012). Par exemple, Kruk et ses collaborateurs ont proposé en 2012 une classification basée sur 7 groupes. Le second groupe de leur étude correspondait aux «Petits flagellés siliceux» (Kruk & Segura 2012) dans lesquels ils regroupaient 4 caractères également sélectionnés dans notre étude (par exemple, la présence de exosquelette siliceux, présence de flagelle, classe de taille pico-plancton (dimension maximale) et classe de taille nano-plancton (dimension maximale)).

Malgré cette apparente similitude avec les études précédentes, nous avons néanmoins identifié deux traits rarement utilisés pour la création de groupes fonctionnels : le type de reproduction (Litchman & Klausmeier 2008) et la présence de protubérances. Ces traits semblent être associés de manière significative aux variables environnementales et doivent donc révéler une pression sélective sur les espèces de phytoplancton. Dans des conditions favorables, la reproduction asexuée du phytoplancton permet une augmentation rapide de la population (Boero et al. 2008), ce qui permet à ce type d'organisme de tirer des avantages indéniables par rapport aux autres espèces de phytoplancton (Valiela 2013).

La présence de protubérances est probablement liée au développement de mécanismes de défense dans certaines conditions environnementales, qui influent sur les interactions trophiques et en particulier sur l'intensité du pâturage (Agrawal, 1998; Litchman et al., 2010). Selon nos résultats (Tableau 2), la présence de grandes protubérances et d'aiguilles apparaît comme un trait particulièrement pertinent en tant que mécanisme de défense sélective. Fait intéressant, la production de toxines, qui est généralement le seul mécanisme de défense pris en compte dans la classification classique basée sur les groupes morpho-fonctionnels, n'est pas associée de manière significative aux variables environnementales dans notre étude. Dans ce contexte, le type et la taille des protubérances devraient probablement être approfondis dans des études de classification portant sur l'utilisation de groupes fonctionnels.

3.6.2 Paramètres physico-chimiques expliquant les traits phytoplanctoniques

Pour chacun des 21 traits significativement associés à l'environnement, nous avons sélectionné les 4 prédicteurs les plus importants d'après le modèle RF (exemple pour la motilité en Figure 20 et ensemble des résultats en Annexe 4).

Globalement nous constatons que 6 paramètres physico-chimiques (sur 15) semblent avoir un impact significatif sur la sélection des traits (voir Tableau 2). De plus, trois de ces paramètres sont observés pour tous les traits pertinents : chlorophylle-a, carbone organique dissous (COD) et matière en suspension (MES). Le quatrième paramètre prédictif est soit la concentration en nitrates et/ou la transparence mesurée au disque de Secchi. Dans quelques rares cas, nous trouvons le titre d'alcalimétrique complet (TAC) et la température de l'eau. Compte tenu de la faible occurrence de ces deux derniers paramètres, il est difficile de tirer des conclusions générales. Par conséquent, notre discussion portera davantage sur les cinq autres prédicteurs les plus influents.

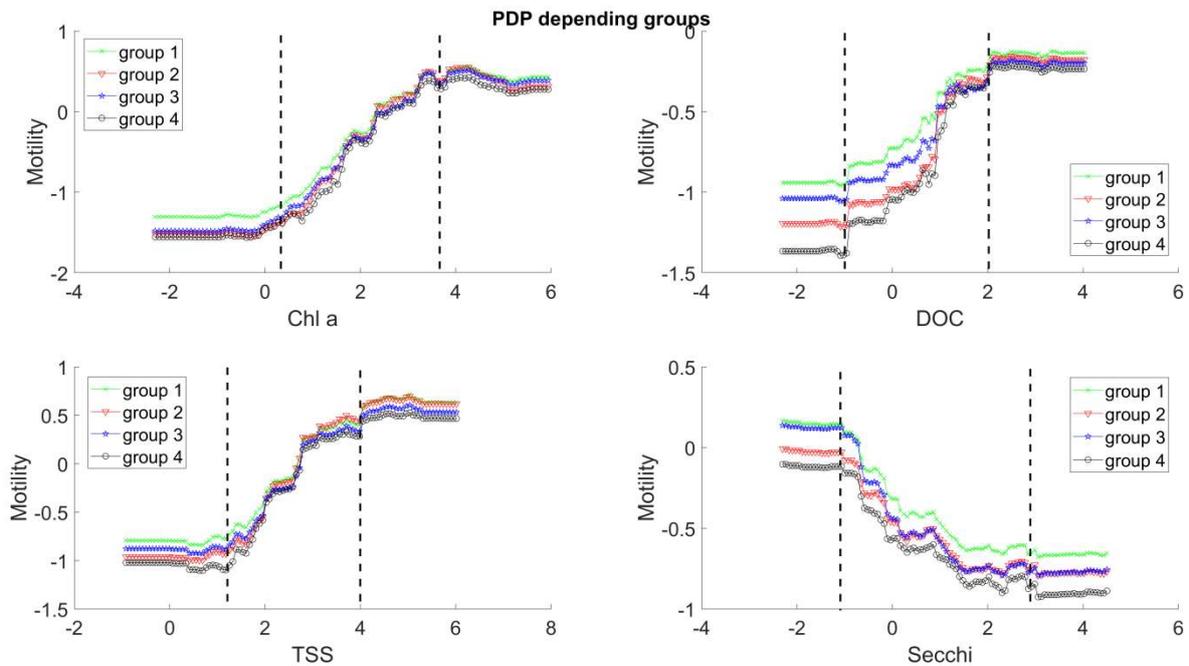


Figure 20. Partial Dependence Plot (PDP) pour les 4 premiers paramètres physico-chimiques influençant la motilité selon les 4 groupes typologiques. Dans l'ordre chlorophylle-a, Carbone Organique Dissous (DOC), matières en suspension (TSS) et transparence au disque de Secchi. Les lignes en pointillées représentent les seuils inférieur et supérieur

Chlorophylle-a, COD et les MES ont toujours une influence positive sur les concentrations de biomasse des traits, tandis que la transparence au disque de Secchi semble être négativement corrélée à la biomasse des traits (voir la direction des flèches dans le Tableau 2). Ces variations sont cohérentes dans la mesure où les eaux à forte teneur en COD, chlorophylle-a et MES sont plus troubles et ont donc une transparence moindre (voir la matrice de corrélations entre les paramètres physico-chimiques en Annexe 5).

La concentration en chlorophylle-a est toujours la variable explicative la plus importante dans tous les modèles. Comme ce paramètre est un indicateur de la biomasse phytoplanctonique, son influence dominante est tout à fait logique.

La réponse positive observée entre les traits et le COD et les MES pourrait être due à un effet indirect et liée à la forte corrélation entre toutes ces variables (COD, MES et Chlorophylle-a, voir Annexe 5). En effet, la réponse positive observée dans nos modèles pourrait probablement en être la conséquence plutôt que la cause : plus la biomasse est importante plus la quantité de MES devrait être élevée. Pour ces deux paramètres environnementaux, ainsi que pour la transparence au disque de Secchi, il semble donc difficile de distinguer l'effet direct de ces paramètres de l'effet indirect dû au compartiment biologique lui-même.

Cependant, la réponse positive des deux caractères phylogéniques, à savoir la présence de Chlorophyceae et d'Euglenophyceae, pourrait être considérée comme une réponse directe à une concentration plus élevée de COD et de MES. Les autres groupes phylogéniques n'ont pas présentés ce type de relation. La relation entre la présence des Euglenophyceae et le COD est conforme à la littérature car il est bien connu que ce groupe phylogénique est composé principalement d'espèces hétérotrophes ou phagotrophes habitant des habitats riches en matières organiques (Reynolds 2006). Comme le DCE (2000/60 / DCE) encourage l'utilisation du COD pour définir le statut sanitaire de l'eau (Koprivnjak et al. 1995; Aiken et al. 2002; Gruau et al. 2007), nos résultats suggèrent que la présence d'espèces de ce groupe phylogénique pourrait être une métrique potentielle de diagnostic pour mesurer la pression anthropique dans les lacs français.

Plus intéressant encore est probablement la réponse de certains traits aux concentrations en nitrates (Tableau 2), telle que la relation positive observée avec la présence de formes filamenteuses. Le développement d'algues filamenteuses en lien avec l'apport de nutriments a déjà été suggéré dans la littérature notamment leur capacité à utiliser des nitrates (Moss, 2005). La plupart des formes filamenteuses sont également présentes dans le groupe des Cyanobactéries, réputé pour être très compétitives en présence d'un rapport N/P élevé (McCarthy et al. 2009; Posch et al. 2012). Mais une concentration élevée en nitrates semble également être associée à des protubérances plus rares et plus petites. Cela s'explique aisément par le fait que les filaments ne présentent généralement pas de protubérance.

Étant donné que les indicateurs phytoplanctoniques actuels sont principalement basé sur une réponse au gradient de phosphore (De Hoyos et al. 2014; Phillips et al. 2014; Laplace-Treyture & Feret 2016, Wolfram et al., 2014), nos résultats suggèrent que l'utilisation de traits fonctionnels spécifiques, tels que la forme filamenteuse, pourrait apporter un éclairage spécifique pour quantifier la pression anthropique issue des nitrates.

3.6.3 Impact de la typologie sur la diversité fonctionnelle du phytoplancton

Afin de tester dans quelle mesure la typologie influence la distribution des traits fonctionnels, ou plus largement la diversité fonctionnelle du phytoplancton, les graphiques de PDP (Annexe 4) ont été construits en tenant compte des 4 groupes typologiques construits au chapitre 3.1.

Les réponses des traits (exprimés en biovolume) aux paramètres physico-chimiques sur les graphes de PDP sont très similaires quel que soit le groupe typologique considéré (Figure 20 et Annexe 4) car les ruptures des pentes (II et IC représentées en traits pointillés noirs sur les figures) sont positionnées aux mêmes valeurs seuils et cela pour l'ensemble des 21 traits sélectionnés.

Par conséquent, la typologie (avec ses 4 groupes) semble avoir un impact limité sur les traits. Ce phénomène est confirmé par les figures d'out-of-bag error dans lesquels le paramètre typologique est presque toujours le prédicteur le moins influent sur les modèles (figures non représentées dans le rapport). Cela suggère que la diversité fonctionnelle du phytoplancton est davantage influencée par les conditions physico-chimiques des eaux que par les caractéristiques typologiques des plans d'eau. Ces résultats sont cohérents avec une étude canadienne (Beisner et al. 2006), qui a montré que la typologie des lacs avait principalement un impact sur les niveaux plus élevés du réseau trophique, à savoir les communautés de zooplancton et de poissons, et que les communautés de phytoplancton et de bactéries sont faiblement structurées par des variables typologiques. Cependant, dans un contexte de bioindication, la directive-cadre sur l'eau (DCE) utilise généralement des groupes typologiques tels que l'hydro-éco-région pour évaluer l'état écologique des masses d'eau. L'indicateur phytoplancton en plan d'eau (IPLAC, Laplace-Treyture et Feret, 2016) a utilisé l'altitude et la profondeur moyenne des plans d'eau pour définir les conditions de référence et calculer les métriques de l'indice. Par rapport à cette approche taxonomique, il apparaît que l'utilisation de traits morpho-fonctionnels permet de s'affranchir de l'effet dominant de la typologie tout en garantissant la similarité des réponses des traits aux changements environnementaux. L'approche basée sur les traits semble donc tout à fait appropriée voire mieux adaptée que l'approche taxonomique pour une application à large échelle géographique (les métriques sont plus comparables à l'échelle nationale par exemple).

Conclusions et perspectives

Ce rapport s'inscrit dans la continuité du développement de l'indice IPLAC pour l'évaluation de l'état écologique des plans d'eau. Il vise à terme à compléter cette évaluation par des métriques dites de diagnostic permettant de rendre compte d'autres pressions et des éléments de biodiversité. Pour cela nous avons employé une approche morpho-fonctionnelle à large échelle (territoire national) des communautés phytoplanctoniques, basée sur la définition et l'utilisation de traits morpho-fonctionnels du phytoplancton comme descripteur des communautés.

Dans les systèmes aquatiques, les relations et interactions entre les traits phytoplanctoniques et les paramètres physico-chimiques et les variables environnementales sont souvent assez complexes. Afin de tenir compte de cette complexité, différents outils statistiques et numériques tel que le Random Forest ont été testés.

Une matrice de 84 traits morpho-fonctionnels du phytoplancton a tout d'abord été créée. Elle couvre l'ensemble des groupes taxonomiques et repose sur les taxons rencontrés dans au moins 15 relevés phytoplanctoniques. Associée aux relevés phytoplanctoniques de la base d'Aix-en-Provence, elle a permis de mettre en œuvre des méthodes de Random Forest. Ces dernières ont montré que les facteurs typologiques (profondeur du plan d'eau, altitude) étaient des facteurs importants. Cela a conduit à la création par la méthode des K-means de quatre groupes typologiques, fonction de ces critères, pour en soustraire l'effet. Cette pré-zonation géographique, pouvait être un moyen adapté pour prendre en compte la variabilité naturelle des flores, afin de se rapprocher de la DCE compatibilité. Cependant, comme nous l'avons vu dans les résultats, une diminution du nombre de données utilisées en entrée d'un modèle RF, a tendance à avoir un effet négatif sur les performances de ce dernier. Par la suite, une multitude de tests de corrélations directes, sur chacun des groupes typologiques, a été exécutée mais il en ressort des résultats plutôt décevants avec des coefficients de corrélation faibles à très faibles et peu de paramètres physico-chimiques pertinents. Dans un second temps, des modèles RF en mode régression puis classification, afin d'essayer de pallier la complexité des relations entre les taxons (traits) et l'environnement présente au sein de notre base de données ont été utilisés. Cependant leurs utilisations nous éloignent de la DCE compatibilité car elles n'ont pas encore permis de dériver des relations et des métriques de diagnostic (portant sur un trait ou combinaison de traits pertinents) exploitables.

Dans un dernier temps, une approche exploitant une structure arborescente et les Partial Dependence Plots (PDP) a été mise en œuvre. Cette approche semble bien adaptée à ce type d'études dans des écosystèmes lacustres. En effet, sur les 84 traits construits au début, 21 ressortent des analyses avec des liens directs ou indirects avec des paramètres physico-chimiques des eaux. Il s'agit notamment des traits : forme de vie (solitaire, filamenteuse), motilité, présence de protubérances, de pseudo-vacuoles ou encore de la présence de Chlorophyceae ou d'Euglenophyceae. Le carbone organique dissous, les matières en suspensions, les nitrates ou encore la transparence mesurée au disque de Secchi sont les paramètres physico-chimiques qui présentent des relations avec un ou plusieurs traits morpho-fonctionnels. L'utilisation des PDP permet de surcroit de connaître le sens de variation de la relation existante entre le trait et le paramètres physico-chimiques. En revanche, l'influence des prédicteurs typologiques (altitude et profondeur) semble pour sa part beaucoup plus limitée car les valeurs de seuils ne dépendent pas des groupes typologiques créés.

L'indice IPLAC étant principalement conçu pour répondre à la pression anthropique via les apports de phosphore, la mise en exergue, par ce travail, d'autres paramètres de qualité des eaux en lien avec des traits morpho-fonctionnels devrait permettre la définition de nouvelles métriques de diagnostic. Cela permettra d'affiner la compréhension de l'impact des pressions anthropiques sur les écosystèmes lacustres. En effet, nous avons identifié plusieurs traits, candidats potentiels, pour de futures métriques de bioindication, en particulier pour les gradients de COD et de nitrates.

Les résultats de ce travail seront valorisés par l'écriture de deux articles dans des revues scientifiques à caractère international. Un premier avec une teinte méthodologique, qui aura pour vocation d'expliquer notre démarche avec les modèles RF. Et un second article de type « data paper », afin de valoriser la création de la table des traits morpho-fonctionnels du phytoplancton.

Sigles & Abréviations

ACP : Analyse en Composantes Principales

CART : Classification And Regression Tree

DCE : Directive Cadre sur l'Eau

DOC : Dissolved Organic Carbon (Carbone organique dissous)

EQB : Eléments de Qualité Biologique

EQR : Ecological Quality Ratios (ratios de qualité écologique)

HER : HydroEcoRégions

IPLAC : Indice Phytoplancton LACustre.

MES : Matière En Suspension

MSE : Mean Squared Error

OD : Oxygène Dissous

Q-Q : Quantile- Quantile

RF : Random Forest

TAC : Titre Alcalimétrique Complet

Bibliographie

- Agrawal, A.A., 1998. Algal defense, grazers, and their interactions in aquatic trophic cascades. *Acta Oecologica*, 19, 331-337.
- Aiken, G., Kaplan, L.A. & Weishaar, J., 2002. Assessment of relative accuracy in the determination of organic matter concentrations in aquatic systems. *Journal of Environmental Monitoring*, 4, 70-74.
- Beisner, B.E., Peres-Neto, P.R., Lindström, E.S., Barnett, A. & Longhi, M.L., 2006. The role of environmental and spatial processes in structuring lake communities from bacteria to fish. *Ecology*, 87, 2985-2991.
- Boero, F., Bouillon, J., Gravili, C., Miglietta, M.P., Parsons, T. & Piraino, S., 2008. Gelatinous plankton: irregularities rule the world (sometimes). *Marine Ecology Progress Series*, 356, 299-310.
- Box, G.E. et D.R. Cox. 1964. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*: 211-252.
- Breiman, L. 2001. Random forests. *Machine learning* 45: 5-32.
- Breiman, L., J. Friedman, C.J. Stone, et R.A. Olshen. 1984. *Classification and regression trees* CRC press. 358 p.
- Coesel, P.F. et K. Meesters. 2007. *Desmids of the lowlands: Mesotaeniaceae and Desmidiaceae of the European lowlands* KNNV Publishing Zeist. pp.
- Crutzen, P.J. (2006) The 'Anthropocene'. In *Earth System Science in the Anthropocene*, eds Ehlers E and Krafft T, pp. 13-18. Springer, Berlin, Heidelberg.
- Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J. et al., 2007. Random forests for classification in ecology. *Ecology*, 88, 2783-2792.
- de Hoyos, C., Catalan, J., Dörflinger, G., Ferreira, J., Kemitzoglou, D., Laplace-Treytore, C., Pahissa López, J., Marchetto, A., Mihail, O., Morabito, G., Polykarpou, P., Romão, F., Tsiaoussi, V., 2014. *Water Framework Directive Intercalibration: Mediterranean Lake Phytoplankton ecological assessment methods (technical Report)*, Scientific and Technical Research series. Joint Research Center, Luxembourg, LUX.
- Ettl, H. 1978. *Xanthophyceae: 3/1. Teil*. Gustav Fischer Verlag: 530 p.
- European Parliament, 2000. Directive 2000/60/CE du Parlement européen et du conseil du 23 octobre 2000 établissant un cadre pour une politique communautaire dans le domaine de l'eau. *Journal Officiel, Communauté Européenne, Bruxelles*, 72 p.
- Feret, T., Laplace-Treytore, C., 2013. *IPLAC: l'indice Phytoplancton Lacustre: Méthode de développement, description et application nationale 2012*. Rapport final. Irstea.
- Friedman, J., Hastie, T. & Tibshirani, R., 2001. *The elements of statistical learning*. Springer series in statistics New York, NY, USA.
- Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.
- Gruau, G., Petitjean, P. & Panaget, T., 2007. Evaluation de la qualité matière organique des eaux destinées à la consommation humaine Eléments en faveur d'une évolution réglementaire. *Groupe d Etude dur la Pollution des Eaux par les Matières organiques (GEPMo). Note technique*.
- Hartigan, J.A. et M.A. Wong. 1979. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28: 100-108.
- Holling, C.S., 1973. Resilience and stability of ecological systems. *Annual review of ecology and systematics*, 4, 1-23.
- Huber-Pestalozzi, G., 1955. *Das Phytoplankton des Süßwassers-4. Teil Euglenophyceen*. E. Schweizerbart'sche Verlagsbuchhandlung. Stuttgart.
- Huber Pestalozzi, G. 1974. *Das phytoplankton des süsswassers systematik und biologie: 5 teil, Chlorophyceae (Grünalgen) ordnung: volvocales*. 1300.
- John, D.M., B.A. Whitton, et A.J. Brook. 2011. *The freshwater algal flora of the British Isles: an identification guide to freshwater and terrestrial algae*. Second Edition. Cambridge University Press. 878 p.
- Komarek, J. et K. Anagnostidis. 2005. *Cyanoprokaryota 19/2. teil: Oscillatoriales*. Elsevier GmbH 759 p.

- Komárek, J. 2013. Cyanoprokaryota II. Teil/Part 3: Heterocytous Genera. Süßwasserflora von Mitteleuropa 19: 31.
- Komárek, J. et K. Anagnostidis. 2000. Cyanoprokaryota-1. Teil: Chroococcales Berlin: Spektrum Akademischer Verlag Heidelberg; ISBN 3-8274-0890-3. pp.
- Komárek, J., B. Fott, et G. Huber-Pestalozzi. 1983. Das Phytoplankton des Süßwassers-7. Teil, 1. Hälfte. Die Binnengewässer 16.
- Koprivnjak, J., Blanchette, J., Bourbonniere, R., Clair, T., Heyes, A., Lum, K. *et al.*, 1995. The underestimation of concentrations of dissolved organic carbon in freshwaters. *Water research*, 29, 91-94.
- Krammer, K. 1988. Bacillariophyceae Teil 2, Bacillariaceae, Epithemiaceae, Surirellaceae. Süßwasserflora von Mitteleuropa 2: 595.
- Krammer, K. et H. Lange-Bertalot. 1986. Bacillariophyceae. 1. Teil: Naviculaceae.
- Krammer, K. et H. Lange-Bertalot. 1991a. Bacillariophyceae, Teil 3. Centrales, Fragilariaceae, Eunotiaceae, Achnanthaceae. Süßwasserflora von Mitteleuropa. VEB Gustav Fisher Verlag, Stuttgart, Germany.
- Krammer, K. et H. Lange-Bertalot. 1991b. Süßwasserflora Von Mitteleuropa: Bacillariophyceae, Achnanthaceae and Index to Parts 1-4 Volume 2 Spektrum Akademischer Verlag. pp.
- Krammer, K. et H. Lange-Bertalot. 2000. Bacillariophyceae: English and French translation of the keys Spektrum Akademischer Verlag GmbH. pp.
- Kruk, C. & Segura, A.M., 2012. The habitat template of phytoplankton morphology-based functional groups. In: *Phytoplankton responses to human impacts at different scales*. Springer, pp. 191-202.
- Laplace-Treyture, C. et T. Feret. 2016. Performance of the Phytoplankton Index for Lakes (IPLAC): A multimetric phytoplankton index to assess the ecological status of water bodies in France. *Ecological Indicators* 69: 686-698.
- Laplace-Treyture C., Barbe J., Dutartre A., Druart J.C., Rimet F., Anneville O., 2009. Protocole standardisé d'échantillonnage, de conservation, d'observation et de dénombrement du phytoplancton en plan d'eau pour la mise en œuvre de la DCE : version 3.3.1. Cemagref, 44 p.
- Large, S.I., Fay, G., Friedland, K.D. & Link, J.S., 2015. Quantifying patterns of change in marine ecosystem response to multiple pressures. *PLoS One*, 10 (3) : e0119922.
- Larras, F., R. Coulaud, E. Gautreau, E. Billoir, J. Rosebery, et P. Usseglio-Polatera. 2017. Assessing anthropogenic pressures on streams: A random forest approach based on benthic diatom communities. *Science of The Total Environment* 586: 1101-1112.
- Litchman, E., de Tezanos Pinto, P., Klausmeier, C.A., Thomas, M.K. & Yoshiyama, K., 2010. Linking traits to species diversity and community structure in phytoplankton. In: *Fifty years after the "Homage to Santa Rosalia": Old and new paradigms on biodiversity in aquatic ecosystems*. Springer, pp. 15-28.
- Litchman, E. & Klausmeier, C.A.; 2008. Trait-based community ecology of phytoplankton. *Annual review of ecology, evolution, and systematics*, 39, 615-639.
- McCarthy, M.J., James, R.T., Chen, Y., East, T.L. & Gardner, W.S., 2009. Nutrient ratios and phytoplankton community structure in the large, shallow, eutrophic, subtropical Lakes Okeechobee (Florida, USA) and Taihu (China). *Limnology*, 10, 215-227.
- Møller, A. & Jennions, M.D., 2002. How much variance can be explained by ecologists and evolutionary biologists? *Oecologia*, 132, 492-500.
- Moss, B., 2005. A filamentous green algae-dominated temperate shallow lake: Variations on the theme of clear-water stable states? *Archiv für Hydrobiologie*, 163, 25-47.
- North, B.V., Curtis, D. & Sham, P.C., 2002. A note on the calculation of empirical P values from Monte Carlo procedures. *American journal of human genetics*, 71, 439.
- Odum, E.P., Finn, J.T. & Franz, E.H., 1979. Perturbation theory and the subsidy-stress gradient. *Bioscience*, 29, 349-352.
- Padisák, J., Crossetti, L.O. & Naselli-Flores, L., 2009. Use and misuse in the application of the phytoplankton functional classification: a critical review with updates. *Hydrobiologia*, 621, 1-19.
- Phillips, G., Free, G., Karotki, Y., Laplace-Treyture, C., Mischke, U., Ott, I., Pasztaleniec, A., Portielje, R., Søndergaard, M., Trodd, W., Van Wichelen, J., 2014. Water Framework Directive

- Intercalibration: Central Baltic Lake Phytoplankton ecological assessment methods (technical Report), Scientific and Technical Research series. Join Research Center, Luxembourg, LUX.
- Popovsky, J. et L.A. Pfiester. 1990. Dinophyceae (Dinoflagellida) G. Fischer. 272 p.
- Posch, T., Köster, O., Salcher, M.M. & Pernthaler, J., 2012. Harmful filamentous cyanobacteria favoured by reduced water turnover with lake warming. *Nature Climate Change*, 2, 809.
- Prévost, E., 2014. Caractérisation des traits morpho-fonctionnels des communautés phytoplanctoniques des grands lacs du littoral aquitain. Rapport de stage. Ecole Polytechnique de l'université François Rabelais de Tours. Spécialité MSc IMACOF.
- Reynolds, C.S., Huszar, V., Kruk, C., Naselli-Flores, L. & Melo, S., 2002. Towards a functional classification of the freshwater phytoplankton. *Journal of plankton research*, 24, 417-428.
- Reynolds, C.S. 2006. The ecology of phytoplankton Cambridge University Press. 535 p.
- Rieth, A. 1980. Xanthophyceae: 4/2. Teil. Gustav Fischer Verlag: 147 p.
- Ricklefs, R.E., 1987. Community diversity: relative roles of local and regional processes. *Science*, 235, 167-171.
- Roubeix, V., P.-A. Danis, T. Feret, et J.-M. Baudoin. 2016. Identification of ecological thresholds from variations in phytoplankton communities among lakes: contribution to the definition of environmental standards. *Environmental monitoring and assessment* 188: 246.
- Shapiro, S.S. et M.B. Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52: 591-611.
- Starmach, K. 1985. Chrysophyceae und haptophyceae Fischer. 515 p.
- Steffen, W., P.J. Crutzen, et J.R. McNeill. 2007. The Anthropocene: are humans now overwhelming the great forces of nature. *AMBIO: A Journal of the Human Environment* 36: 614-621.
- Teichert, N., Borja, A., Chust, G., Uriarte, A. & Lepage, M., 2016. Restoring fish ecological quality in estuaries: implication of interactive and cumulative effects among anthropogenic stressors. *Science of the Total Environment*, 542, 383-393.
- Valiela, I., 2013. *Marine ecological processes*. Springer Science & Business Media.
- Wagenhoff, A., Liess, A., Pastor, A., Clapcott, J.E., Goodwin, E.O. & Young, R.G., 2017. Thresholds in ecosystem structural and functional responses to agricultural stressors can inform limit setting in streams. *Freshwater Science*, 36, 178-194.
- Wagenhoff, A., Townsend, C.R. & Matthaei, C.D., 2012. Macroinvertebrate responses along broad stressor gradients of deposited fine sediment and dissolved nutrients: a stream mesocosm experiment. *Journal of Applied Ecology*, 49, 892-902.
- Weiherr, E., Freund, D., Bunton, T., Stefanski, A., Lee, T. & Bentivenga, S., 2011. Advances, challenges and a developing synthesis of ecological community assembly theory. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 366, 2403-2413.
- Wehr, J.D., R.G. Sheath, et J.P. Kociolek. 2015. *Freshwater algae of North America: ecology and classification* Elsevier. pp.
- Wilcoxon, F. 1945. Individual comparisons by ranking methods. *Biometrics bulletin* 1: 80-83.
- Wilk, M.B. et R. Gnanadesikan. 1968. Probability plotting methods for the analysis for the analysis of data. *Biometrika* 55: 1-17.
- Wolfram, G., Buzzi, F., Dokulil, M., Friedl, M., Hoehn, E., Laplace-Treytore, C., Menay, M., Marchetto, A., Morabito, G., Reichmann, M., Remec-Rekar, Š., Riedmüller, U., Urbanič, G., 2014. Water Framework Directive Intercalibration: Alpine Lake Phytoplankton ecological assessment methods (technical Report), Scientific and Technical Research series. Join Research Center, Luxembourg, LUX.
- Yajima, H. et J. Derot. 2018. Application of the Random Forest model for chlorophyll-a forecasts in fresh and brackish water bodies in Japan, using multivariate long-term databases. *Journal of Hydroinformatics* 20 : 206-220.

Annexes

Annexe 1 : seuils physico-chimiques en fonction des groupes établis par la méthode des K-means

Tableau A1-1. Seuil pour groupe 1 (altitudes intermédiaires et faibles profondeurs, cf. 3.2). Légende : Conductivité ; Carbone organique dissous (DOC) ; Nitrates (NO₃), Nitrite (NO₂) ; Ammonium (NH₄) ; Oxygène dissout à saturation (ODsat) ; Potentiel hydrogène (pH) ; Phosphate (PO₄), Phosphate total (Ptot) ; Silice (SiO₂) ; Température (T°C) ; Matière en suspension (MES) ; Secchi et Titre alcalimétrique complet (TAC).

Signal cible	Coefficient de corrélation pour la classification binaire	1^{ère} influence, seuil et tendance	2^{ème} influence, seuil et tendance	3^{ème} influence, seuil et tendance
Présence de diatomées	0,83	NO ₃ (10,58/26,46)(-)	T°C (11,81)(+)	MES (51,38)(-)
Présence de chlorophycées	0,85	Ptot (0,38/0,62)(-)	MES (71,53/107,8)(-)	DOC (PS)(-)
Présence de chrysophycées	0,97	PO ₄ (0,08)(-)	MES (55,41)(-)	Secchi (1,48/2,36)(-)
Présence de cryptophycées	0,91	MES (49,36)(-)	NO ₃ (7,28)(-)	Secchi (2,36/3,15)(-)
Présence de cyanobactéries	0,92	DOC (14,12/35,06)(-)	Ptot (0,07/0,39)(-)	Secchi (0,57/1,66)(-)
Présence de desmidiées	0,92	pH (7,06/8,32)(?)	Secchi (0,65/1,79)(-)	Ptot (0,17)(-)
Présence de dinophycées	0,97	Conductivité (203,2/339,5)(-)	ODsat (95,11)(+)	NO ₂ (0,04)(-)
Présence d'euglenophycées	0,98	Ptot (0,22/0,41)	PO ₄ (0,20)(-)	TAC (5,72)(-)
Taille minimale : Pico-phytoplancton	0,98	DOC (19,63)(+)	Ptot (0,12)(-)	PO ₄ (1,28)(-)
Taille minimale : Nano-phytoplancton	0,66	ODsat (92,42)(+)	DOC (14,8)(-)	NO ₃ (2,19/8,68)(-)
Taille minimale : Microphytoplancton	0,76	pH (6,44)(-)	NO ₃ (5,08)(-)	ODsat (97,43)(+)
Taille minimale : Méso-phytoplancton	0,97	DOC (11,37/29,28)(-)	MES (23,17/121,9)(-)	pH (8,27)(-)
Taille minimale : Macro-phytoplancton	0,99	Ptot (0,14)(-)	pH (9,17)(-)	NO ₃ (1,90)(-)
Taille maximale : Pico-phytoplancton	0,99	NO ₃ (2,79)(-)	PO ₄ (0,08)(-)	Ptot (0,21)(-)
Taille maximale : Nano-phytoplancton	0,92	Ptot (0,47/0,63)(-)	Secchi (0,75/2,71)(-)	NO ₃ (19,53)(-)
Taille maximale : Micro-phytoplancton	0,70	pH (7,06/8,99)(-)	Secchi (1,35/)(-)	Ptot (0,26)(-)
Taille maximale : Méso-phytoplancton	0,86	DOC (10,76)(-)	Secchi (P,S,)(-)	TAC (4,23/10,21)(-)
Forme de vie cellule	0,74	MES (45,33/127,9)(-)	pH (8,56/10,13)(-)	DOC (5,01/34,19)(-)
Forme de vie colonie	0,85	Conductivité (249,3/447,2)(-)	SiO ₂ (9,24)(-)	Ptot (0,19/0,62)(-)
Forme de vie filament	0,89	Ptot (0,18/0,39)(-)	DOC (14,93)(-)	MES (121,9)(-)
Forme sphère	0,96	NO ₃ (5,80)(-)	DOC (18,53)(-)	Ptot (0,23/0,50)(-)
Forme ellipsoïde rotationnelle	0,93	Conductivité (122,2/460,6)(-)	pH (6,88/8,38)(-)	DOC (14,78)(-)
Forme ellipsoïde aplatie	0,86	MES (11,08)(-)	Secchi (1,17/2,59)(-)	Ptot (0,04)(-)

Forme cylindrique	0,85	Ptot (0,27)(-)	DOC (15,32)(-)	NH4 (0,64/1,87)(-)
Forme double cône	0,92	pH (6,52/8,49)(-)	Ptot (0,15)(-)	NO3 (1,24/7,54)(-)
Forme de cylindre ovale	0,99	Ptot (0,11)(-)	MES (16,52)(-)	NO2 (0,06/0,24)(-)
Forme prisme sur une base de parallélogramme	0,99	PO4 (0,04)(-)	NO3 (0,34)(-)	Ptot (0,02)(-)
Forme <i>Ceratium</i>	0,98	pH (7,95)(-)	T°C (24,14)(+)	NH4 (0,09)(-)
Motilité	0,56	Secchi (1,10/2,10)(-)	SiO2 (9,13)(-)	ODsat (50,53/109,6)(+)
Flagelle	0,67	Secchi (0,67/2,04)(-)	MES (29,21/79,59)(-)	Ptot (0,14/0,43)(-)
2 Flagelles	0,71	MES (117,9)(-)	Secchi (0,89/1,92)(-)	Ptot (0,88)(-)
Aérotape	0,95	Ptot (0,21)(-)	NH4 (0,17/0,67)(-)	TAC (5,46)(?)
Vacuole contractile	0,84	Conductivité (210,4/457,1)(-)	NO3 (5,44)(-)	SiO2 (6,03/12,1)(-)
Mucilage	0,88	DOC (10,54/32,31)(-)	Ptot (0,10/0,49)(-)	NO3 (3,10)(-)
Akinètes	0,97	Ptot (0,13)(-)	DOC (7,24/22,11)(-)	Conductivité (182,7/415,4)(-)
Hétérocyste	0,98	Ptot (0,13)(-)	PO4 (0,08)(-)	DOC (9,44/22,11)(-)
Cyste	0,95	pH (6,52/7,67)(-)	PO4 (0,08)(-)	Ptot (0,12/0,42)(-)
Chloroplaste	0,93	PO4 (0,27)(-)	Ptot (0,24)(-)	Secchi (1,41)(-)
1 Chloroplaste	0,62	Conductivité (190,80/330,2)(-)	DOC (14,05/23,09)(-)	PO4 (0,27/1,29)(-)
2 chloroplastes	0,83	MES (57,42/144,1)(-)	Secchi (0,82/2,97)(-)	DOC (12,75)(-)
> 2 chloroplastes	0,87	TAC (10,96)(-)	Secchi (0,75/1,76)(-)	DOC (14,95/38,37)(-)
Tractus	0,98	MES (9,06)(-)	Conductivité (261,2)(-)	NH4 (0,29)(-)
Squelette siliceux	0,84	MES (53,39)(-)	NO3 (10,23/27,62)(-)	T°C (10,15)(+)
Lorica	0,97	SiO2 (6,51)(-)	Ptot (0,16/0,41)(-)	NH4 (0,23/1,09)(-)
Plaque	0,97	PO4 (0,06)(-)	Secchi (0,50/1,32)(-)	MES (67,5)(-)
Ecailles	0,99	DOC (20,19)(-)	Ptot (0,09)(-)	MES (15,11)(-)
Ornementation	0,64	T°C (19,32)(+)	NO3 (8,43/21,33)(-)	MES (25,63/59,12)(-)
Protubérance	0,64	Conductivité (94,21/383,8)(-)	DOC (12,61)(-)	T°C (8,20/22,34)(+)
Absence de protubérance	0,67	Conductivité (99,3/637,7)(-)	T°C (11,26/18,18)(+)	DOC (12,51/25,09)(-)
Aiguilles	0,87	T°C (23,70)(+)	DOC (15,80)(-)	MES (51,38)(-)
Cône	0,96	T°C (23,03)(+)	PO4 (0,05)(-)	MES (11,08/57,42)(-)
Dent	0,98	Conductivité	NO3	NH4

		(99,3)(-)	(0,39)(-)	(0,16)(-)
Granule	0,98	MES (147,6)(-)	Ptot (0,21)(-)	NH4 (0,55)(-)
Poile	0,99	DOC (20,42)(-)	NO3 (2,19)(-)	Ptot (0,09)(-)
Petite protubérance	0,92	Conductivité (231,6/378,6)(-)	NO3 (3,99/31,34)(-)	MES (23,17/218,6)(-)
Grande protubérance	0,87	NO3 (4,77)(-)	pH (6,57/9,71)(?)	Secchi (1,39)(-)
1 protubérance	0,94	MES (35,5/68,28)(-)	TAC (10,6)(-)	Ptot (0,21)(-)
2 protubérances	0,99	NO3 (0,34)(-)	pH (7,10)(-)	NO2 (0,03)(-)
> 2 protubérances	0,87	Ptot (0,20)(-)	TAC (1,73/19,85)(-)	NO3 (1,54/23,42)(-)
Chlorophylle b	0,61	Conductivité (109,7/255,6/505,7)(-)	MES (62,2)(-)	NO3 (5,74)(-)
Chlorophylle c	0,65	MES (120)(-)	NO3 (5,74/9,03)(-)	Secchi (2,18)(-)
Xanthophylles	0,81	MES (106,2/202,8)(-)	Secchi (2,10/3,25)(-)	Ptot (0,15)(-)
Phycobilines	0,82	Secchi (0,92/2,68)(-)	PO4 (1,935)(-)	MES (55,41/99,73)(-)
Toxines	0,96	DOC (7,59/22,36)(-)	Ptot (0,48)(-)	Conductivité (170,4/383)(-)
Autotrophie	0,68	MES (41,3)(-)	Secchi (1,17/2,18)(-)	SiO2 (9,13)(-)
Mixotrophie	0,69	Secchi (0,85)(-)	MES (37,27/127,9)(-)	SiO2 (9,13)(-)
Reproduction : sexuée	0,88	pH (PS)(-)	NO3 (8,49)(-)	Secchi (2,64)(-)
Reproduction : asexuée	0,60	NO3 (8,31)(-)	Secchi (1,57)(-)	T°C (13,61/23,03)(+)

Tableau A1-2. Seuil pour groupe 2 (altitude et profondeur élevées, cf. 3.2). Légende identique au tableau A1-1.

Signal cible (Target)	Coefficient de corrélation pour la classification binaire	1 ^{ère} influence, seuil et tendance	2 ^{ème} influence, seuil et tendance	3 ^{ème} influence, seuil et tendance
Présence de diatomées	0,83	NO3 (10,58/26,46)(-)	T°C (11,81)(+)	MES (51,38)(-)
Présence de chlorophycées	0,85	Ptot (0,38/0,62)(-)	MES (71,53/107,8)(-)	DOC (PS)(-)
Présence de chrysophycées	0,97	PO4 (0,08)(-)	MES (55,41)(-)	Secchi (1,48/2,36)(-)
Présence de cryptophycées	0,91	MES (49,36)(-)	NO3 (7,28)(-)	Secchi (2,36/3,15)(-)
Présence de cyanobactéries	0,92	DOC (14,12/35,06)(-)	Ptot (0,07/0,39)(-)	Secchi (0,57/1,66)(-)
Présence de desmidiées	0,92	pH (7,06/8,32)(?)	Secchi (0,65/1,79)(-)	Ptot (0,17)(-)
Présence de dinophycées	0,97	Conductivité (203,2/339,5)(-)	ODsat (95,11)(+)	NO2 (0,04)(-)
Présence Euglenophycées	0,98	Ptot (0,22/0,41)	PO4 (0,20)(-)	TAC (5,72)(-)
Taille minimale : Pico-phytoplancton	0,98	DOC (19,63)(+)	Ptot (0,12)(-)	PO4 (1,28)(-)
Taille minimale : Nano-phytoplancton	0,66	ODsat (92,42)(+)	DOC (14,8)(-)	NO3 (2,19/8,68)(-)
Taille minimale : Microphytoplancton	0,76	pH (6,44)(-)	NO3 (5,08)(-)	ODsat (97,43)(+)
Taille minimale : Méso-phytoplancton	0,97	DOC (11,37/29,28)(-)	MES (23,17/121,9)(-)	pH (8,27)(-)
Taille minimale : Macro-phytoplancton	0,99	Ptot (0,14)(-)	pH (9,17)(-)	NO3 (1,90)(-)
Taille maximale : Picophytoplancton (14)	0,99	NO3 (2,79)(-)	PO4 (0,08)(-)	Ptot (0,21)(-)
Taille maximale : Nano-phytoplancton	0,92	Ptot (0,47/0,63)(-)	Secchi (0,75/2,71)(-)	NO3 (19,53)(-)
Taille maximale : Micro-phytoplancton	0,70	pH (7,06/8,99)(-)	Secchi (1,35/)(-)	Ptot (0,26)(-)
Taille maximale : Méso-phytoplancton	0,86	DOC (10,76)(-)	Secchi (P.S.)(-)	TAC (4,23/10,21)(-)
Forme de vie cellule	0,74	MES (45,33/127,9)(-)	pH (8,56/10,13)(-)	DOC (5,01/34,19)(-)
Forme de vie colonie	0,85	Conductivité (249,3/447,2)(-)	SiO2 (9,24)(-)	Ptot (0,19/0,62)(-)
Forme de vie filament	0,89	Ptot (0,18/0,39)(-)	DOC (14,93)(-)	MES (121,9)(-)
Forme sphère	0,96	NO3 (5,80)(-)	DOC (18,53)(-)	Ptot (0,23/0,50)(-)
Forme ellipsoïde rotationnelle	0,93	Conductivité (122,2/460,6)(-)	pH (6,88/8,38)(-)	DOC (14,78)(-)
Forme ellipsoïde aplatie	0,86	MES (11,08)(-)	Secchi (1,17/2,59)(-)	Ptot (0,04)(-)
Forme cylindrique	0,85	Ptot (0,27)(-)	DOC (15,32)(-)	NH4 (0,64/1,87)(-)
Forme double cône	0,92	pH (6,52/8,49)(-)	Ptot (0,15)(-)	NO3 (1,24/7,54)(-)
Forme de cylindre ovale	0,99	Ptot	MES	NO2

		(0,11)(-)	(16,52)(-)	(0,06/0,24)(-)
Forme prisme sur une base de parallélogramme	0,99	PO4 (0,04)(-)	NO3 (0,34)(-)	Ptot (0,02)(-)
Forme <i>Ceratium</i>	0,98	pH (7,95)(-)	T°C (24,14)(+)	NH4 (0,09)(-)
Motilité	0,56	Secchi (1,10/2,10)(-)	SiO2 (9,13)(-)	ODsat (50,53/109,6)(+)
Flagelle	0,67	Secchi (0,67/2,04)(-)	MES (29,21/79,59)(-)	Ptot (0,14/0,43)(-)
2 Flagelles	0,71	MES (117,9)(-)	Secchi (0,89/1,92)(-)	Ptot (0,88)(-)
Aérotape	0,95	Ptot (0,21)(-)	NH4 (0,17/0,67)(-)	TAC (5,46)(?)
Vacuole contractile	0,84	Conductivité (210,4/457,1)(-)	NO3 (5,44)(-)	SiO2 (6,03/12,1)(-)
Mucilage	0,88	DOC (10,54/32,31)(-)	Ptot (0,10/0,49)(-)	NO3 (3,10)(-)
Akinètes	0,97	Ptot (0,13)(-)	DOC (7,24/22,11)(-)	Conductivité (182,7/415,4)(-)
Hétérocyste	0,98	Ptot (0,13)(-)	PO4 (0,08)(-)	DOC (9,44/22,11)(-)
Cyste	0,95	pH (6,52/7,67)(-)	PO4 (0,08)(-)	Ptot (0,12/0,42)(-)
Chloroplaste	0,93	PO4 (0,27)(-)	Ptot (0,24)(-)	Secchi (1,41)(-)
1 Chloroplaste	0,62	Conductivité (190,80/330,2)(-)	DOC (14,05/23,09)(-)	PO4 (0,27/1,29)(-)
2 chloroplastes	0,83	MES (57,42/144,1)(-)	Secchi (0,82/2,97)(-)	DOC (12,75)(-)
> 2 chloroplastes	0,87	TAC (10,96)(-)	Secchi (0,75/1,76)(l)	DOC (14,95/38,37)(-)
Tractus	0,98	MES (9,06)(-)	Conductivité (261,2)(-)	NH4 (0,29)(-)
Squelette siliceux	0,84	MES (53,39)(-)	NO3 (10,23/27,62)(-)	T°C (10,15)(+)
Lorica	0,97	SiO2 (6,51)(-)	Ptot (0,16/0,41)(-)	NH4 (0,23/1,09)(-)
Plaque	0,97	PO4 (0,06)(-)	Secchi (0,50/1,32)(-)	MES (67,5)(-)
Ecailles	0,99	DOC (20,19)(-)	Ptot (0,09)(-)	MES (15,11)(-)
Ornementation	0,64	T°C (19,32)(+)	NO3 (8,43/21,33)(l)	MES (25,63/59,12)(-)
Protubérance	0,64	Conductivité (94,21/383,8)(-)	DOC (12,61)(-)	T°C (8,20/22,34)(+)
Absence de protubérance	0,67	Conductivité (99,3/637,7)(-)	T°C (11,26/18,18)(+)	DOC (12,51/25,09)(-)
Aiguilles	0,87	T°C (23,70)(+)	DOC (15,80)(-)	MES (51,38)(-)
Cône	0,96	T°C (23,03)(+)	PO4 (0,05)(-)	MES (11,08/57,42)(-)
Dent	0,98	Conductivité (99,3)(-)	NO3 (0,39)(-)	NH4 (0,16)(-)
Granule	0,98	MES (147,6)(-)	Ptot (0,21)(-)	NH4 (0,55)(-)
Poil	0,99	DOC (20,42)(-)	NO3 (2,19)(-)	Ptot (0,09)(-)

Petite protubérance	0,92	Conductivité (231,6/378,6)(-)	NO3 (3,99/31,34)(-)	MES (23,17/218,6)(-)
Grande protubérance	0,87	NO3 (4,77)(-)	pH (6,57/9,71)(?)	Secchi (1,39)(-)
1 protubérance	0,94	MES (35,5/68,28)(-)	TAC (10,6)(-)	Ptot (0,21)(-)
2 protubérances	0,99	NO3 (0,34)(-)	pH (7,10)(-)	NO2 (0,03)(-)
> 2 protubérances	0,87	Ptot (0,20)(-)	TAC (1,73/19,85)(-)	NO3 (1,54/23,42)(-)
Chlorophylle b	0,61	Conductivité (109,7/255,6/505,7)(-)	MES (62,2)(-)	NO3 (5,74)(-)
Chlorophylle c	0,65	MES (120)(-)	NO3 (5,74/9,03)(-)	Secchi (2,18)(-)
Xanthophylles	0,81	MES (106,2/202,8)(-)	Secchi (2,10/3,25)(-)	Ptot (0,15)(-)
Phycobilines	0,82	Secchi (0,92/2,68)(-)	PO4 (1,935)(-)	MES (55,41/99,73)(-)
Toxines	0,96	DOC (7,59/22,36)(-)	Ptot (0,48)(-)	Conductivité (170,4/383)(-)
Autotrophie	0,68	MES (41,3)(-)	Secchi (1,17/2,18)(-)	SiO2 (9,13)(-)
Mixotrophie	0,69	Secchi (0,85)(-)	MES (37,27/127,9)(-)	SiO2 (9,13)(-)
Reproduction : sexuée	0,88	pH (PS)(-)	NO3 (8,49)(-)	Secchi (2,64)(-)
Reproduction : asexuée	0,60	NO3 (8,31)(-)	Secchi (1,57)(-)	T°C (13,61/23,03)(+)

Tableau A1-3. Seuil pour groupe 3 (altitudes intermédiaires et profondeurs élevées, cf. 3.2). Légende identique au tableau A1-1.

Signal cible	Coefficient de corrélation pour la classification binaire	1 ^{ère} influence, seuil et tendance	2 ^{ème} influence, seuil et tendance	3 ^{ème} influence, seuil et tendance
Présence de diatomées	0,82	T°C (8,66/16,21)(+)	MES (21,96)(-)	NO3 (23,39)(-)
Présence de chlorophycées	0,88	Secchi (1,89/4,18)(-)	DOC (6,66)(-)	MES (9,97)(-)
Présence de chrysophycées	0,97	NO2 (0,12)(-)	DOC (3,66)(-)	NO3 (28,34)(-)
Présence de cryptophycées	0,83	NO3 (12,95/39,88)(-)	Secchi (3,38)(-)	ODsat (91,04)(-)
Présence de cyanobactéries	0,93	DOC (14,21)(-)	PO4 (0,07)(-)	SiO2 (5,01/12,52)(-)
Présence de desmidiées	0,93	NO3 (6,24/43,21)(-)	TAC (15,98)(-)	MES (14,8/34,2)(-)
Présence de dinophycées	0,96	NO2 (0,10)(-)	MES (9,47)(-)	NO3 (13,5)(-)
Taille minimale : Pico-phytoplancton	0,97	SiO2 (2,04/9,65)(-)	PO4 (0,03)(-)	NO3 (11,89/30,71)(-)
Taille minimale : Nano-phytoplancton	0,64	NO3 (22,09)(-)	Ptot (0,14)(-)	DOC (5,6/10,5)(-)
Taille minimale : Microphytoplancton	0,74	Conductivité (280,7/461,6)(-)	PO4 (0,13)(-)	DOC (12,49)(-)
Taille minimale : Méso-phytoplancton	0,98	Ptot (0,07)(-)	MES (6,89/27,66)(-)	PO4 (0,05)(-)
Taille minimale : Macro-phytoplancton	0,99	SiO2 (7,11)(-)	NO2 (0,03)(-)	Conductivité (296,9)(-)
Taille maximale : Picophytoplancton	0,98	Secchi (2,38)(-)	DOC (7,21)(-)	PO4 (0,06)(-)
Taille maximale : Nano-phytoplancton	0,94	MES (10,73)(-)	NO3 (15,41/39,16)(-)	Ptot (0,08)(-)
Taille maximale : Micro-phytoplancton	0,74	NO3 (13,52)(-)	Secchi (4,45/5,81)(-)	pH (8,06/9,54)(+)
Taille maximale : Méso-phytoplancton	0,85	Secchi (5,58/13,68)(-)	MES (19,49)(-)	Ptot (0,13/0,81)(-)
Forme de vie cellule	0,75	NO2 (0,25)(-)	Conductivité (296,9/393,8)(-)	DOC (9,18)(-)
Forme de vie colonie	0,82	DOC (11,03)(-)	NH4 (0,39)(-)	TAC (13,82)(-)
Forme de vie filament	0,91	DOC (11,02)(-)	MES (18,9)(-)	NO2 (0,20/0,42)(-)
Forme sphère	0,96	SiO2 (1,62/11)(-)	NH4 (0,10)(-)	PO4 (0,14)(-)
Forme ellipsoïde rotationnelle	0,95	DOC (8,13)(-)	NO3 (28,9)(-)	ODsat (101,9)(?)
Forme ellipsoïde aplatie	0,76	Secchi (4,18)(-)	NO3 (8,55/51,98)(-)	DOC (7,08)(-)
Forme cylindrique	0,88	MES (13,16)(-)	DOC (15,14)(-)	TAC (6,67/17,44)(-)
Forme double cône	0,94	NO3 (47,58)(-)	Secchi (1,69/2,62)(-)	MES (18,24/27,94)(-)
Forme parallélépipédique	0,98	NO3 (15,42/29,16)(-)	SiO2 (3,99/10,07)(-)	PO4 (0,05)(-)
Forme de cylindre ovale	0,96	NO3 (13,5/29,44)(-)	DOC (3,13)(-)	Secchi (2,28)(-)

Forme <i>Strautastrum</i>	0,98	MES (11,98)(-)	Ptot (0,05)(-)	Secchi (3,07)(-)
Forme <i>Ceratium</i>	0,98	Ptot (0,04)(-)	PO4 (0,09)(-)	NO2 (0,06)(-)
Motilité	0,60	Secchi (3,28)(-)	MES (27,2)(-)	DOC (6,38)(-)
Flagelle	0,63	Secchi (3,09)(-)	MES (14,44)(-)	ODsat (84,51)(+)
1 Flagelle	0,99	DOC (1,95)(-)	SiO2 (1,37)(-)	NO3 (4,16)(-)
2 Flagelles	0,66	Secchi (3,62)(-)	ODsat (99,13)(+)	MES (56,88)(-)
Aérotape	0,96	PO4 (0,10)(-)	Ptot (0,13/0,77)(-)	MES (17,51)(-)
Vacuole contractile	0,80	TAC (8,33/19,68)(-)	Secchi (2,79/8,56)(-)	Ptot (0,10)(-)
Mucilage	0,89	SiO2 (4,49/12,94)(-)	NO2 (0,41)(-)	DOC (6,66/13,42)(-)
Akinètes	0,98	NO2 (0,09/0,25)(-)	Ptot (0,82)(-)	NO3 (11,3)(-)
Hétérocyste	0,98	NO2 (0,07/0,25)(-)	Conductivité (167,7/318,7)(-)	Ptot (0,83)(-)
Cyste	0,93	DOC (4,55)(-)	NO3 (17,35)(-)	Secchi (3,31)(-)
Chloroplaste	0,93	SiO2 (4,24)(-)	pH (8,32/9,62)(+)	NH4 (0,21)(-)
1 Chloroplaste	0,68	TAC (16,81)(-)	NO3 (25,4)(-)	NO2 (0,05/0,48)(-)
2 chloroplastes	0,74	Secchi (3,38/5,43)(-)	DOC (1,49)(-)	TAC (8,76/14,38)(-)
> 2 chloroplastes	0,87	DOC (1,92/10,85)(-)	Secchi (2,49/7,33)(-)	MES (17,69)(-)
Gaine	0,99	PO4 (0,08)(-)	NH4 (0,08)(-)	NO2 (0,09)(-)
Tractus	0,98	TAC (10,33/15,14)(-)	NO2 (0,12)(-)	MES (2,47/10,56)(-)
Squelette siliceux	0,83	T°C (8,76/16,79)(+)	NO3 (28,89)(-)	MES (9,89/21,32)(-)
Lorica	0,98	TAC (3,33/15,35)(-)	NO2 (0,11)(-)	SiO2 (2,8/11,93)(-)
Plaque	0,96	NO2 (0,10)(+)	MES (6,66)(-)	NO3 (13,77)(-)
Ecailles	0,99	TAC (2,24/4,96)(-)	NO2 (0,07)(-)	DOC (2,99)(-)
Ornementation	0,57	ODsat (84,51/154,2)(-)	Ptot (0,02/0,7)(-)	MES (5,34/32,62)(-)
Protubérance	0,75	TAC (7,03/14,07)(-)	PO4 (0,09)(-)	Secchi (5,23)(-)
Absence de protubérance	0,75	PO4 (0,13)(-)	Ptot (0,11)(-)	TAC (14,27)(-)
Aiguilles	0,92	TAC (4,99/10,81)(-)	Conductivité (170,9/283,9)(-)	NO2 (0,14)(-)
Cône	0,98	Ptot (0,11)(-)	SiO2 (5,68/9,65)(-)	PO4 (0,09)(-)
Dent	0,97	MES (8,85/16,99)(-)	pH (6,37/9,98)(+)	Secchi (2,38)(-)
Granule	0,99	NO3 (6,63)(-)	NO2 (0,08)(-)	NH4 (0,03)(-)

Poil	0,99	TAC (1,83)(-)	NO2 (0,05)(-)	PO4 (0,03)(-)
Bouton	0,99	Ptot (0,04)(-)	T°C (12,96/20,04)(-)	NO2 (0,13)(-)
Petite protubérance	0,91	MES (4,23/25,28)(-)	PO4 (0,11)(-)	TAC (6,42/10,25)(-)
Grande protubérance	0,91	Conductivité (113,9/288)(-)	NO3 (8,825/35,49)(-)	pH (6,77/8,01)(?)
1 protubérance	0,95	DOC (4,18/12,39)(-)	NO2 (0,05/0,12)(-)	SiO2 (5,52)(-)
2 protubérances	0,99	DOC (2,99)(-)	TAC (4,13)(-)	Secchi (5,09)(-)
> 2 protubérances	0,87	DOC (0,92/9,06)(-)	Ptot (0,07)(-)	PO4 (0,12)(-)
Chlorophylle b	0,73	NO3 (3,64/23,49)(-)	T°C (15,05)(+)	DOC (3,35/8,38)(-)
Chlorophylle c	0,69	DOC (4,41)(-)	NO3 (8/35,49)(-)	Secchi (2,98/14,83)(-)
Xanthophylles	0,74	SiO2 (1,62/10,17)(-)	ODsat (92,36)(-)	TAC (9,39/16,92)(-)
Phycobilines	0,74	NO2 (0,08/0,35)(-)	ODsat (96,12/127,1)(-)	Ptot (0,08/0,22)(-)
Toxines	0,97	MES (15,57/26,11)(-)	NO2 (0,27)(-)	Secchi (1,78)(-)
Autotrophie	0,63	Secchi (3,17/13,57)(-)	MES (14,49)(-)	DOC (5,6)(-)
Mixotrophie	0,65	Secchi (2,99)(-)	MES (21,96/57,52)(-)	TAC (15,73)(-)
Reproduction : sexuée	0,85	TAC (11,27)(-)	Secchi (1,56/7,24)(-)	NO3 (22,84)(-)
Reproduction : asexuée	0,68	ODsat (89,8)(-)	Conductivité (171,7/433,3)(-)	NO2 (0,10/0,22)(-)

Tableau A1-4. Seuil pour groupe 4 (altitudes faibles, cf. 3.2). Légende identique au tableau A1-1.

Signal cible	Coefficient de corrélation pour la classification binaire	1 ^{ère} influence, seuil et tendance	2 ^{ème} influence, seuil et tendance	3 ^{ème} influence, seuil et tendance
Présence de diatomées	0,80	NO3 (12,9/32,2)(-)	T°C (9,58/21,87)(+)	Conductivité (369,3/658,3)(-)
Présence de chlorophycées	0,94	Secchi (5,74/13,37)(-)	pH (8,26/9,64)(+)	Ptot (0,06/0,28)(-)
Présence de chrysophycées	0,98	NO3 (9,29/30,31)(-)	NO2 (0,06)(-)	NH4 (0,13)(-)
Présence de cryptophycées	0,96	TAC (4,77/8,87)(-)	MES (16,95)(-)	SiO2 (4,75/7,71)(-)
Présence de cyanobactéries	0,88	NO3 (9,13)(-)	Secchi (0,90/3,98)(-)	NH4 (0,38)(-)
Présence de desmidiées	0,86	DOC (9,99/32,07)(-)	Secchi (0,72/3,10)(-)	NO3 (6,99)(-)
Taille minimale : Pico-phytoplancton	0,99	PO4 (0,02)(-)	pH (8,41)(?)	DOC (4,69)(-)
Taille minimale : Nano-phytoplancton	0,72	Conductivité (296,6)(-)	pH (7,71/9,73)(4)	ODsat (83,15/129,9)(?)
Taille minimale : Microphytoplancton	0,71	Conductivité (429/606,6)(-)	Ptot (0,47/073)(-)	NO2 (0,12)(-)
Taille minimale : Méso-phytoplancton	0,97	NO2 (0,15)(-)	NH4 (0,11)(-)	PO4 (0,03)(-)
Taille minimale : Macro-phytoplancton	0,98	NO2 (0,03)(-)	DOC (5,43)(-)	SiO2 (7,39)(-)
Taille maximale : Picophytoplancton	0,99	PO4 (0,03)(-)	pH (8,14)(?)	DOC (6,71)(-)
Taille maximale : Nano-phytoplancton	0,96	PO4 (0,16)(-)	Ptot (0,12)(-)	MES (25,54)(-)
Taille maximale : Micro-phytoplancton	0,81	pH (8,49/9,32)(+)	NO2 (0,22)(-)	PO4 (0,05/0,41)(-)
Taille maximale : Méso-phytoplancton	0,85	pH (6,76/8,77)(+)	SiO2 (4,01/15,11)(-)	Conductivité (339,7/617,5)(-)
Forme de vie cellule	0,79	Ptot (0,44)(-)	NH4 (0,19)(-)	NO2 (0,13/0,19)(-)
Forme de vie colonie	0,87	DOC (9,99)(-)	SiO2 (9,51)(-)	NO2 (0,21)(-)
Forme de vie filament	0,83	NO2 (0,07)(-)	Secchi (0,7/2,39)(-)	SiO2 (5,49)(-)
Forme sphère	0,98	pH (8,13)(+)	Ptot (0,05/0,12)(-)	DOC (4,68/9,88)(-)
Forme ellipsoïde rotationnelle	0,94	NO2 (0,14)(-)	DOC (6,38/11,37)(-)	NO3 (8,24/18)(-)
Forme ellipsoïde aplatie	0,92	SiO2 (4,75)(-)	MES (28,11)(-)	Ptot (0,11)(-)
Forme cylindrique	0,82	NO2 (0,07)(-)	Ptot (0,14)(-)	NH4 (0,13)(-)
Forme double cône	0,92	Secchi (0,47/2,21)(-)	NO2 (0,04/0,14)(-)	PO4 (0,09/052)(-)
Forme parallélépipédique	0,99	NO3 (0,92)(-)	NH4 (0,03)(-)	MES (5,79)(-)
Forme de cylindre ovale	0,95	Ptot (0,24/0,72)(-)	SiO2 (5,94)(-)	DOC (5,19)(-)
Forme <i>Strautastrum</i>	0,94	DOC (9,81/23,31)(-)	Secchi (2,48)(-)	PO4 (0,08)(-)
Motilité	0,65	pH	TAC	NO3

		(7,46/9,78)(+)	(4,48/14,67)(-)	(16,89)(-)
Flagelle	0,81	MES (13,27)(-)	NO3 (4,47)(-)	NO2 (0,12)(-)
2 Flagelles	0,81	MES (10,16/216,2)(-)	Secchi (2,48/6,77)(-)	DOC (14,77/30,58)(-)
Aérotape	0,89	NO3 (6,57)(-)	Ptot (0,08)(-)	NH4 (0,16)(-)
Vacuole contractile	0,92	MES (19,41)(-)	NH4 (0,38)(-)	Ptot (0,17)(-)
Mucilage	0,81	DOC (13,3)(-)	Ptot (0,78)(-)	NO3 (1,81/10,9)(-)
Akinètes	0,97	NH4 (0,13)(-)	NO3 (1,81)(-)	NO2 (0,02)(-)
Hétérocyste	0,97	NH4 (0,13)(-)	NO3 (1,81)(-)	NO2 (0,06)(-)
Cyste	0,95	NO3 (3,14/12,9)(-)	SiO2 (1,27/8,75)(-)	NO2 (0,03)(-)
Chloroplaste	0,90	Secchi (0,79/3,38)(-)	NO3 (10,9)(-)	NO2 (0,13)(-)
1 Chloroplaste	0,73	Secchi (2,98)(-)	T°C (15,16/20,65)(-)	NO3 (24,97)(-)
2 chloroplastes	0,80	Secchi (5,57)(-)	TAC (3,28/10,23)(-)	MES (36,09)(-)
> 2 chloroplastes	0,88	NO3 (1,73/10,22)(-)	NO2 (0,07)(-)	Conductivité (277,1/620,9)(-)
Tractus	0,99	Ptot (0,31)(-)	SiO2 (3,13)(-)	pH (8,15)(?)
Squelette siliceux	0,79	T°C (14,48/19,47)(+)	NO3 (18,54)(-)	Ptot (0,55)(-)
Ecailles	0,99	NO3 (1,36)(-)	NH4 (0,19)(-)	SiO2 (4,01)(-)
Ornementation	0,97	NO3 (3,80/24,88)(-)	Conductivité (302,6/373)(-)	MES (18,32/175,4)(-)
Protubérance	0,63	Secchi (2,79)(-)	TAC (9,55)(-)	NO2 (0,06/0,44)(-)
Absence de protubérance	0,63	MES (23,2/48,18)(-)	NO3 (15,78)(-)	TAC (5,13)(-)
Aiguilles	0,87	NO2 (0,12)(-)	TAC (9,52)(-)	Ptot (0,37/0,63)(-)
Cône	0,98	DOC (6,16)(-)	MES (9,18)(-)	PO4 (0,07)(-)
Dent	0,93	Secchi (2,19)(-)	NO3 (5,13)(-)	DOC (13,06/23,62)(-)
Poil	0,99	NO3 (1,58)(-)	NH4 (0,21)(-)	SiO2 (3,72)(-)
Bouton	0,97	PO4 (0,04)(-)	TAC (2,74)(-)	DOC (6,53/20,6)(-)
Petite protubérance	0,78	NO3 (29,54)(-)	DOC (13,27)(-)	Ptot (0,09/0,74)(-)
Grande protubérance	0,95	NO3 (9,79)(-)	NO2 (0,13)(-)	NH4 (0,26)(-)
1 protubérance	0,97	DOC (13,06)(-)	TAC (3,53)(-)	Ptot (0,10)(-)
> 2 protubérances	0,68	TAC (4,03/8,99)(-)	NO3 (6,91)(-)	NH4 (0,17)(-)
Chlorophylle b	0,69	Secchi	NO3	DOC

		(5,97)(-)	(11,5)(-)	(5,49/20,98)(-)
Chlorophylle c	0,65	NO3 (2,69/17,11)(-)	T°C (9,89/22,39)(+)	Secchi (2,55)(-)
Xanthophylles	0,82	NO3 (11,22)(-)	PO4 (0,06)(-)	Secchi (0,99/2,59)(-)
Phycobilines	0,83	NO3 (14,67)(-)	Secchi (0,90/3,68)(-)	TAC (5,00/14,24)(-)
Toxines	0,90	NO3 (7,57/17,78)(-)	Ptot (0,09)(-)	NH4 (0,34)(-)
Autotrophie	0,83	MES (15,36)(-)	NO3 (10,44)(-)	NO2 (0,14)(-)
Mixotrophie	0,82	NO2 (0,05)(-)	MES (14,24)(-)	SiO2 (7,56)(-)
Reproduction : sexuée	0,79	DOC (13,09)(-)	NO3 (3,03/8,34)(-)	Secchi (3,6)(-)
Reproduction : asexuée	0,77	Conductivité (439,7)(-)	DOC (11,63)(-)	PO4 (0,09)(-)

Annexe 2 : partitionnement via K-moyennes

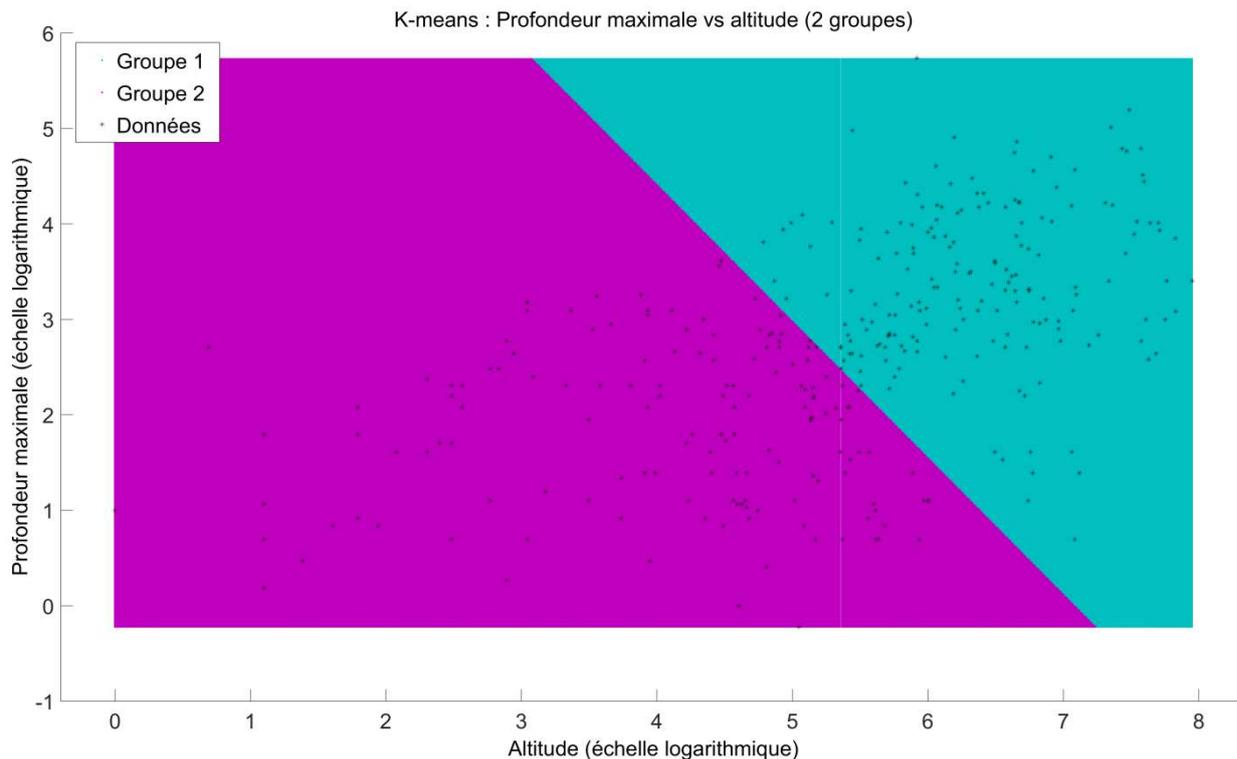


Figure A2-1. Séparation en 2 groupes géographiques via la méthode de K-means basée sur la profondeur maximale et l'altitude des lacs. Chacune des zones de couleurs représente un groupe créé par l'algorithme.

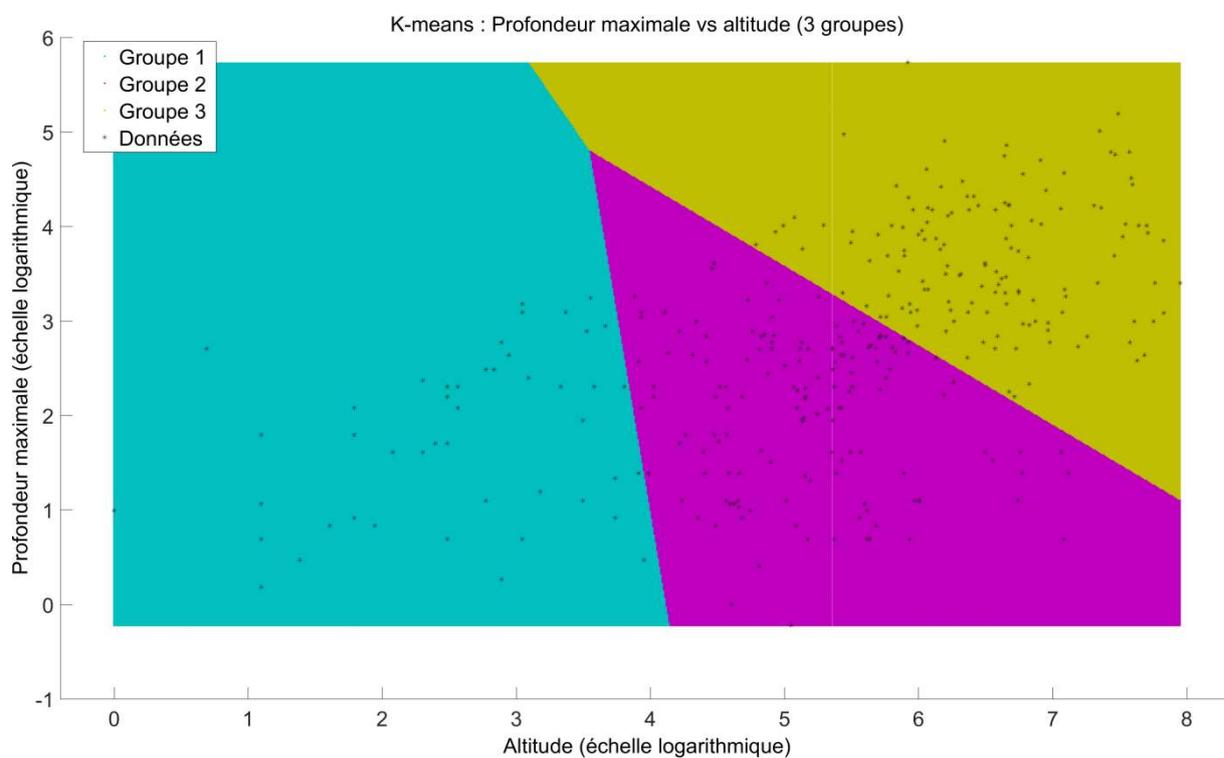


Figure A2-2. Séparation en 3 groupes géographiques via la méthode de K-means basée sur la profondeur maximale et l'altitude des lacs. Chacune des zones de couleurs représente un groupe créé par l'algorithme.

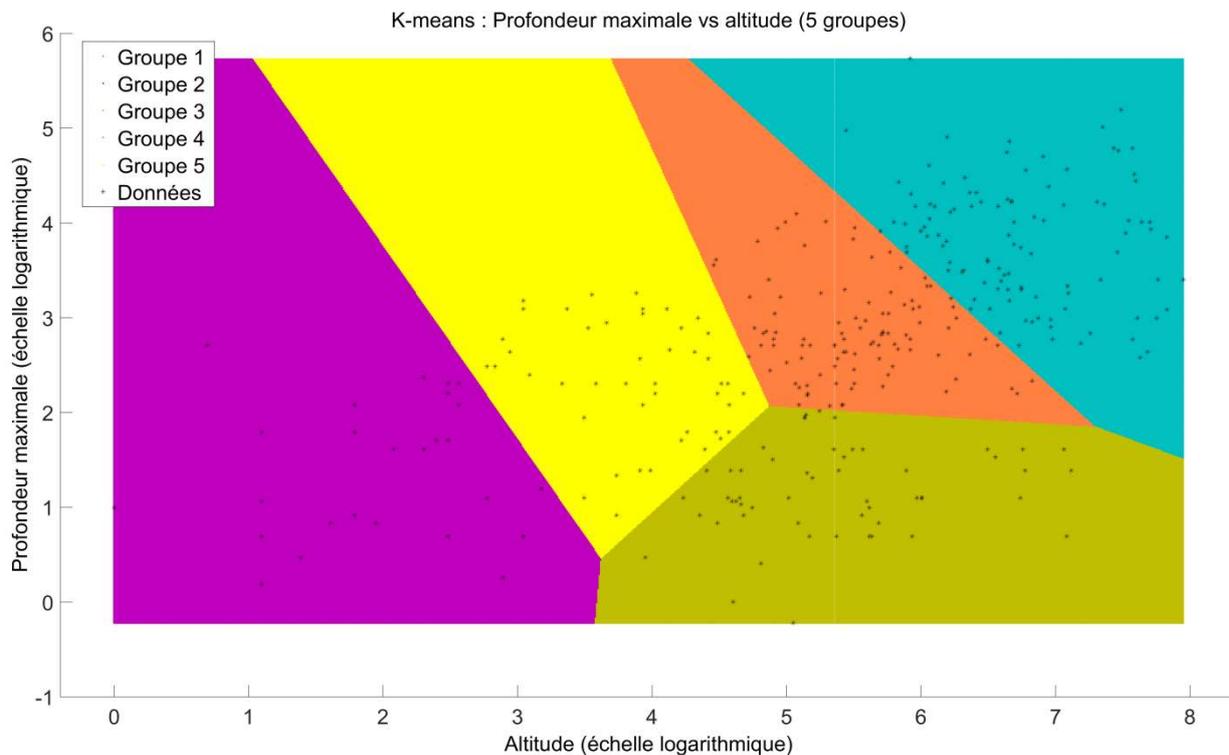


Figure A2-3. Séparation en 5 groupes géographiques via la méthode de K-means basée sur la profondeur maximale et l'altitude des lacs. Chacune des zones de couleurs représente un groupe créé par l'algorithme.

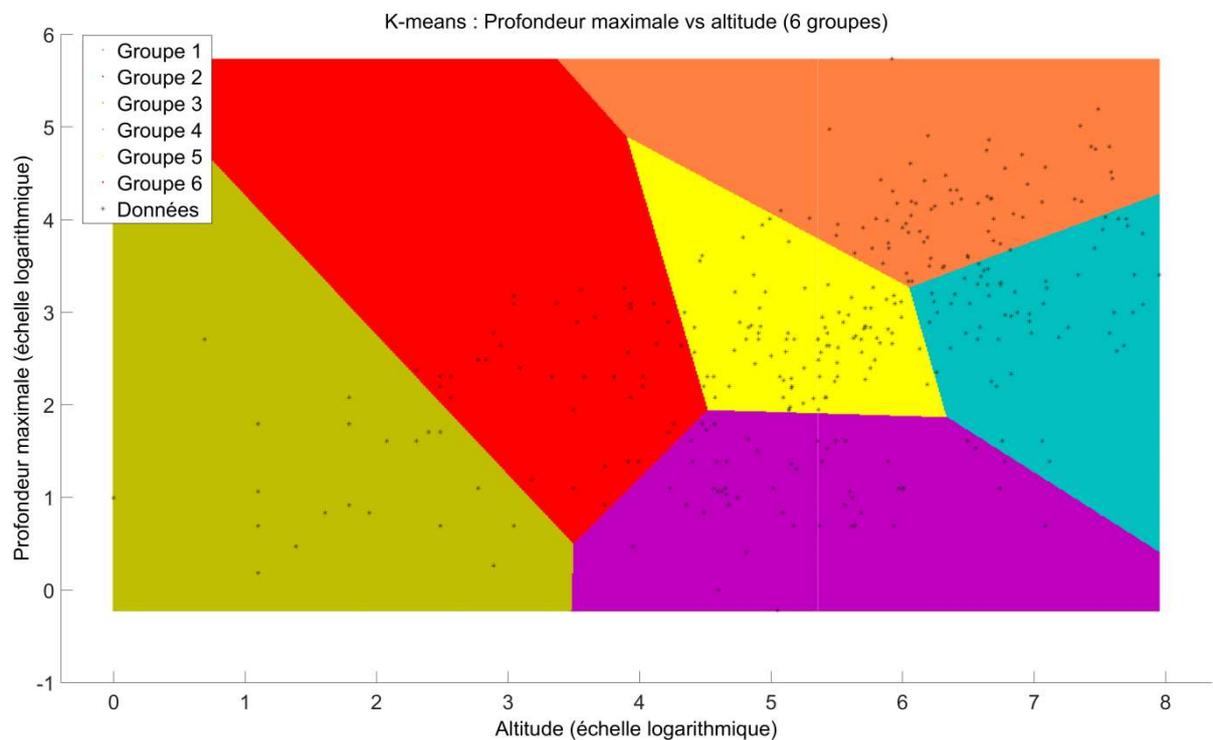


Figure A2-4. Séparation en 6 groupes géographiques via la méthode de K-means basée sur la profondeur maximale et l'altitude des lacs, Chacune des zones de couleurs représente un groupe créé par l'algorithme.

Annexe 3 : analyses complémentaires pour le paramètre Carbone Organique Dissous - DOC

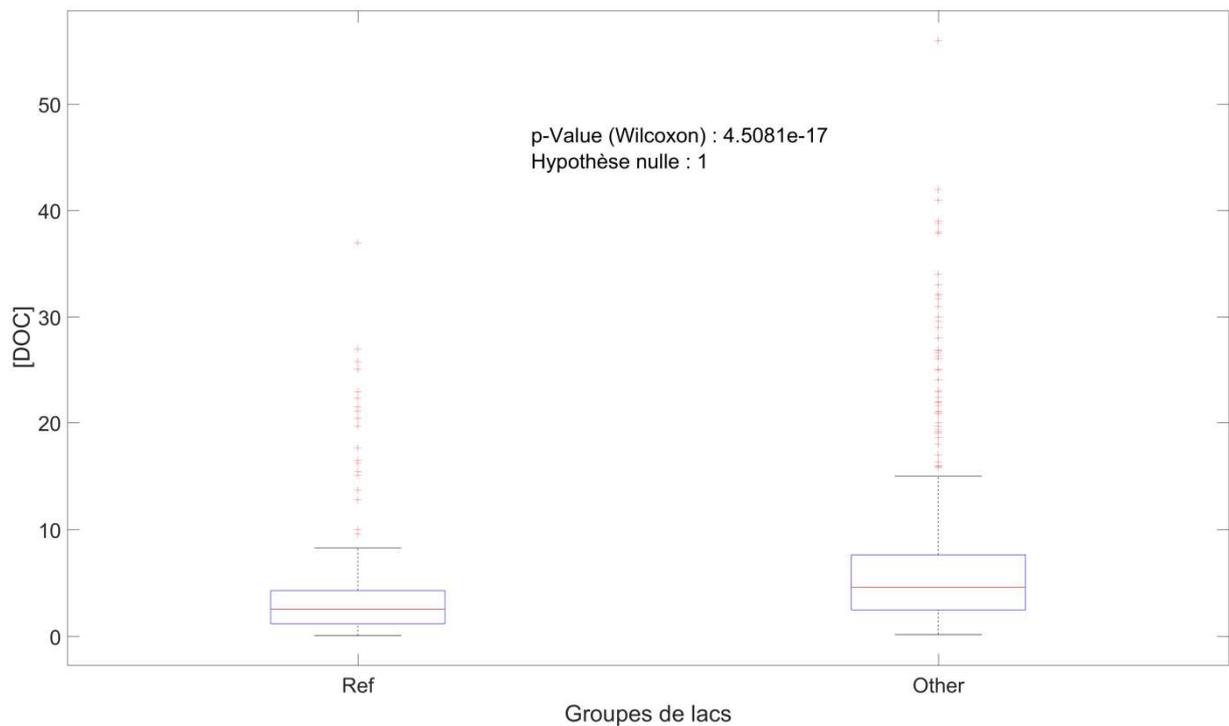


Figure A3-1. Visualisations des distributions entre le groupe des lacs références et le groupe des autres lacs via des boxplots, pour le DOC (données non-transformées). Et p-Value provenant du test de Wilcoxon réalisée à partir de ces deux distributions.

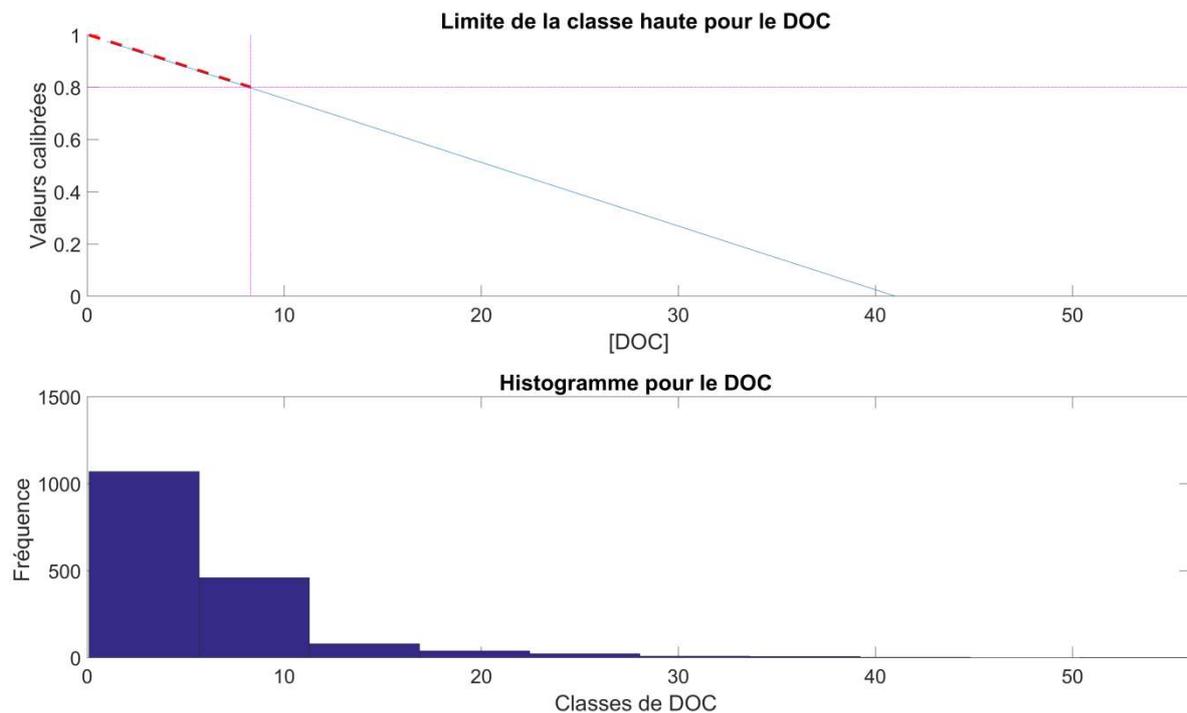


Figure A3-2. Sur les deux graphiques les axes des X correspondent à la concentration en DOC non-transformée. En haut : Présentation graphique de la notion d'échelle théorique qui va être utilisée comme signal cible en entrée de notre modèle RF. La ligne en pointillé rouge correspond à la classe DCE « très bon ». En bas : histogramme représentant la distribution des données de DOC non-transformée.

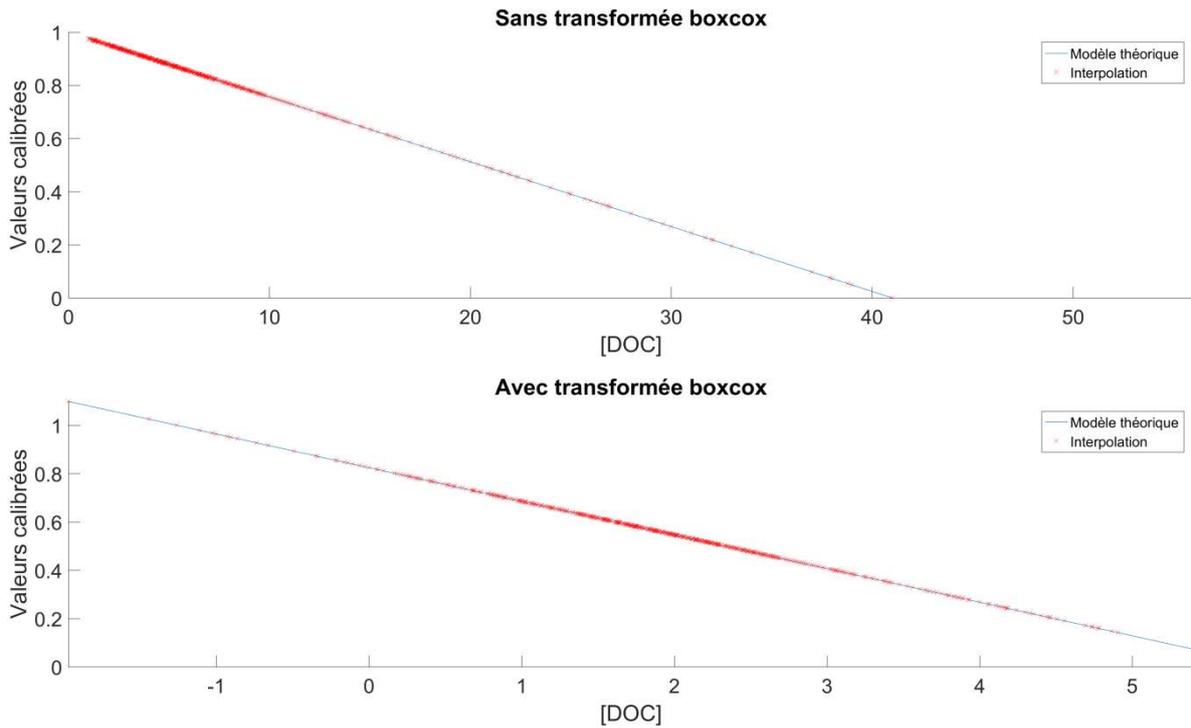


Figure A3-3. Répartition de la distribution des données de DOC en fonction de leurs échelles théoriques ; sans transformation en haut et avec une transformation Boxcox en bas. Dans les deux cas les croix rouges représentent des données de DOC et les lignes en pointillés bleus leurs échelles théoriques respectives.

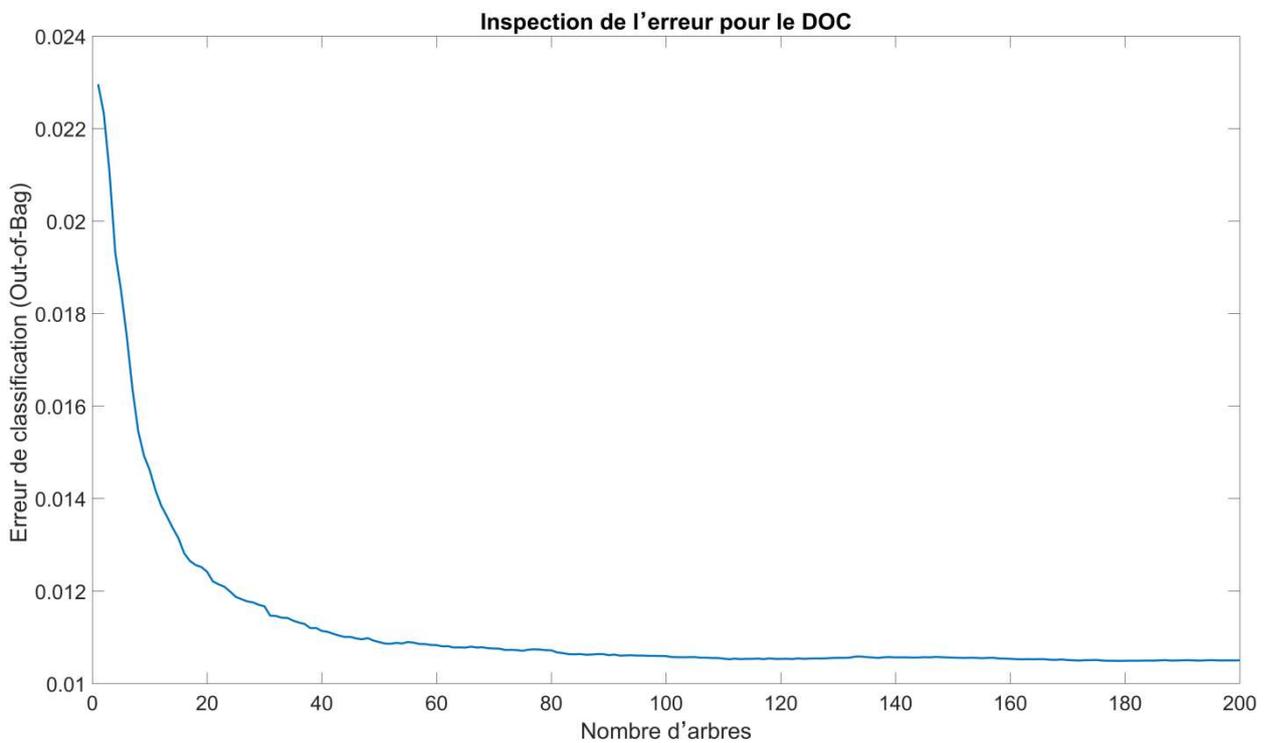


Figure A3-4. Test de sensibilité du modèle RF issu de la phase d'apprentissage du modèle en mode régression pour la prédiction du DOC.

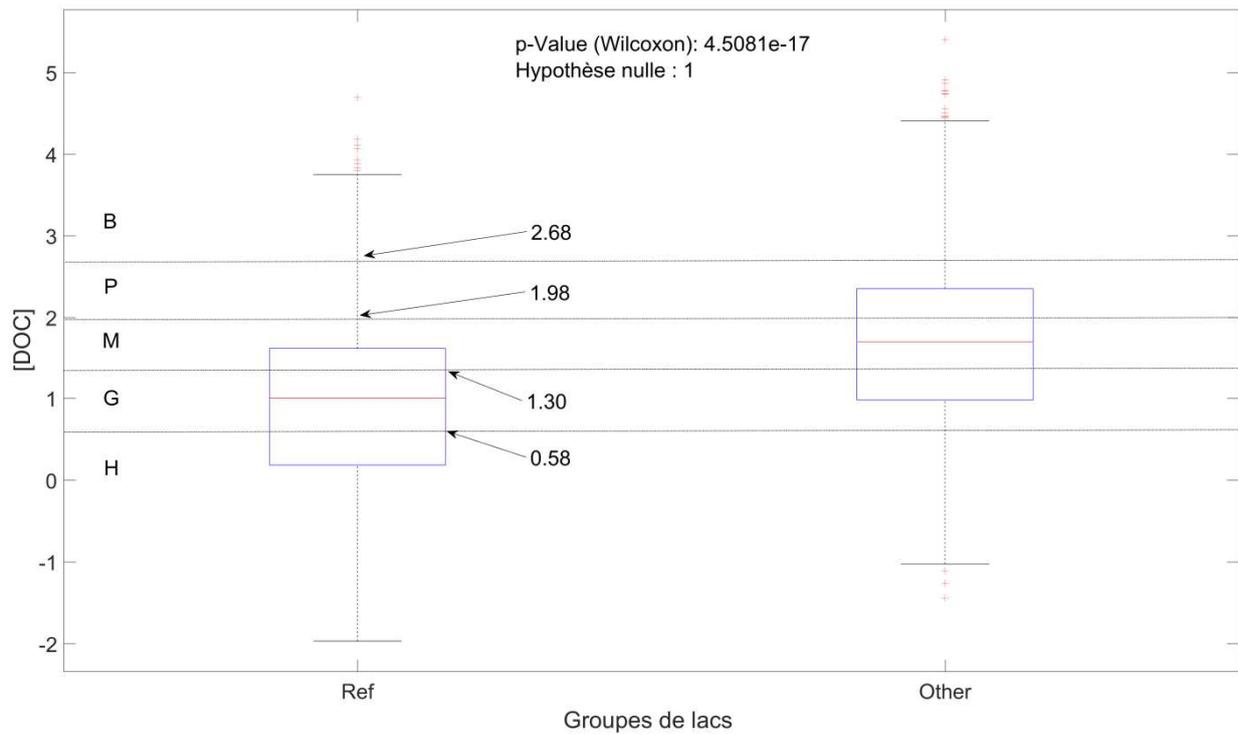


Figure A3-5. Visualisation du découpage de la distribution du DOC, pour aboutir à des classes comportant un nombre de points homogène dans le pool test. Toutes les données sont classées de la manière suivante : High (H) inférieures à 0,58 ; Good (G) comprises entre 0,58 et 1,3 ; Moderate (M) comprises entre 1,3 et 1,98 ; Poor (P) comprises entre 1,98 et 2,68 ; Bad (B) supérieures à 2,68.

Confusion Matrix

Output Class	Target Class					Mean
	Class H	Class G	Class M	Class P	Class B	
Class H	21 12.7%	2 1.2%	1 0.6%	1 0.6%	0 0.0%	84.0% 16.0%
Class G	4 2.4%	29 17.5%	8 4.8%	4 2.4%	0 0.0%	64.4% 35.6%
Class M	1 0.6%	7 4.2%	20 12.0%	6 3.6%	1 0.6%	57.1% 42.9%
Class P	0 0.0%	2 1.2%	11 6.6%	25 15.1%	9 5.4%	53.2% 46.8%
Class B	0 0.0%	0 0.0%	0 0.0%	3 1.8%	11 6.6%	78.6% 21.4%
Mean	80.8% 19.2%	72.5% 27.5%	50.0% 50.0%	64.1% 35.9%	52.4% 47.6%	63.9% 36.1%

Figure A3-6. Matrice de confusion entre les classes de DOC prédites par le modèle et leurs classes réelles, avec un découpage homogène des classes (Cf. figure A3-5).

Annexe 4 : graphiques des partial Dependant Plots (PDP) pour les 21 traits significatifs en fonction des 4 groupes typologiques

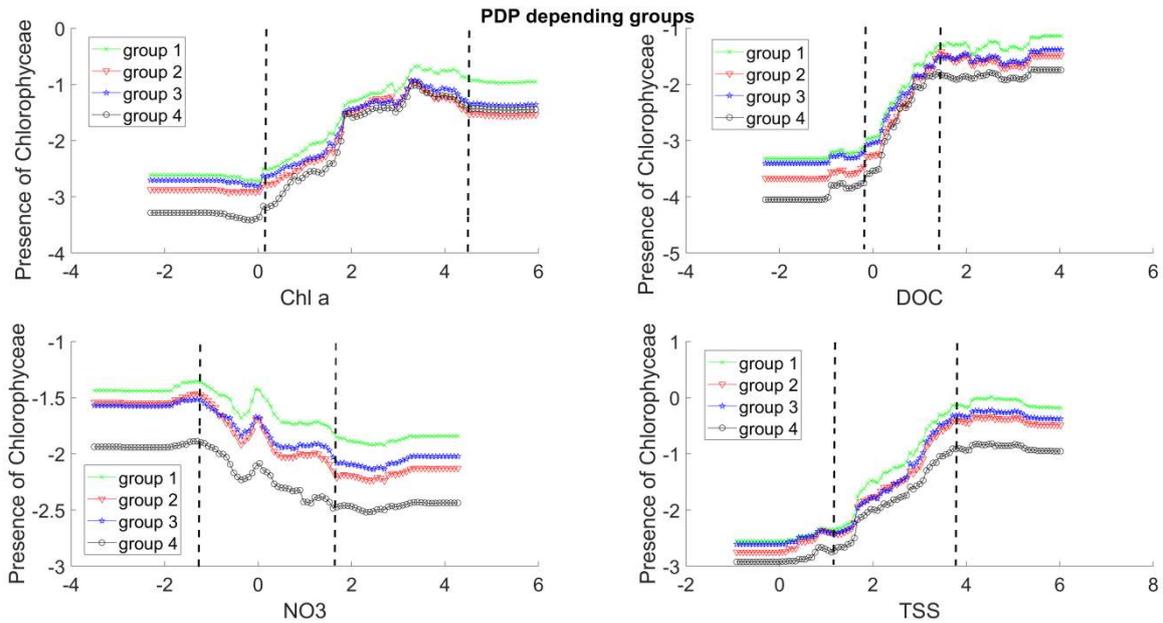


Figure A3.1. Partial Dependence Plot (PDP) pour les 4 premiers paramètres physico-chimiques influençant la présence des Chlorophyceae selon les 4 groupes typologiques. Dans l'ordre chlorophylle-a, Carbone Organique Dissous (DOC), nitrates (NO3) et matières en suspension (TSS). Les lignes en pointillées représentent les seuils inférieur et supérieur

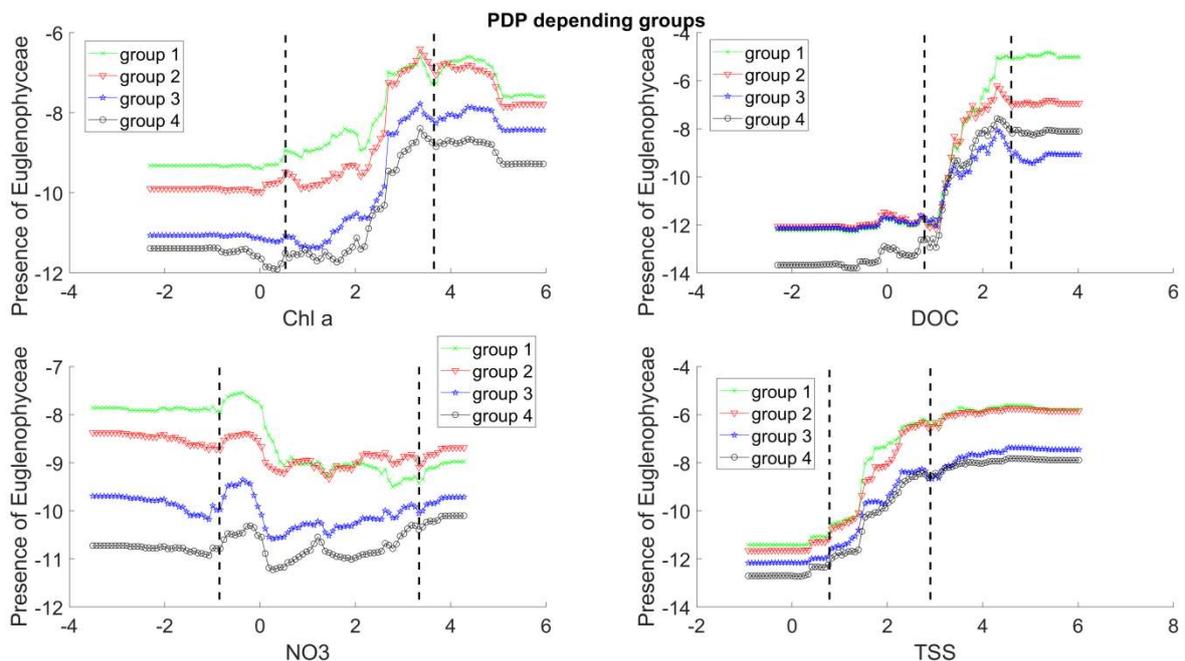


Figure A3.2. Partial Dependence Plot (PDP) pour les 4 premiers paramètres physico-chimiques influençant la présence des Euglenophyceae selon les 4 groupes typologiques. Dans l'ordre chlorophylle-a, Carbone Organique Dissous (DOC), nitrates (NO3) et matières en suspension (TSS). Les lignes en pointillées représentent les seuils inférieur et supérieur PDP for the presence of

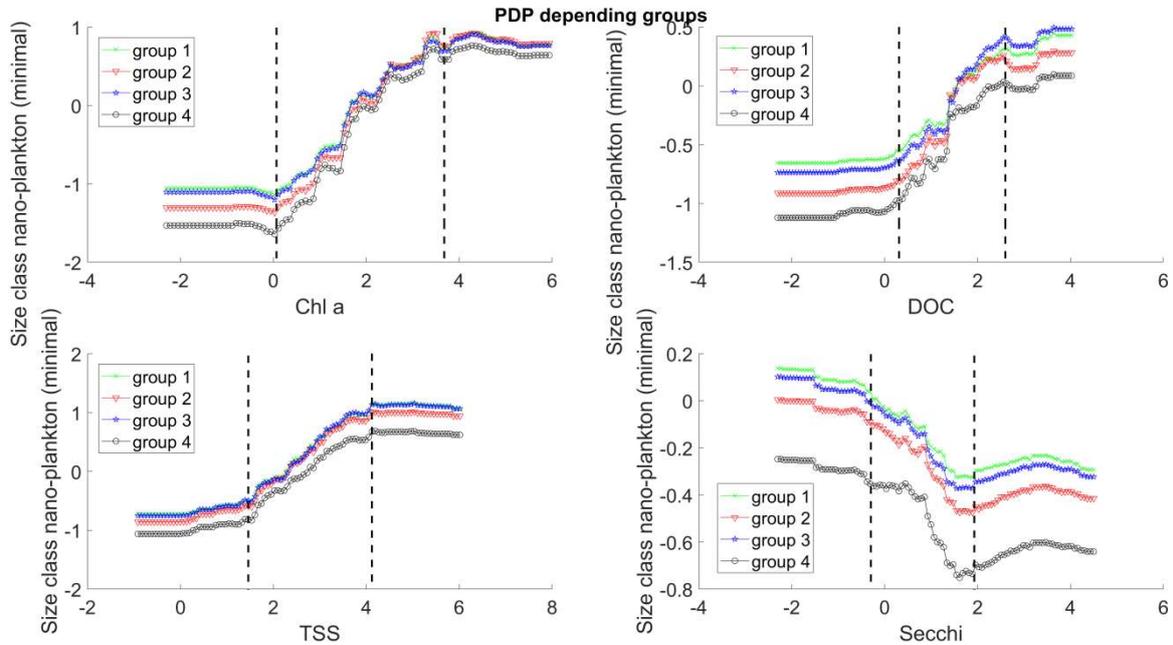


Figure A3.3. Partial Dependence Plot (PDP) pour les 4 premiers paramètres physico-chimiques influençant la classe de taille nano-plancton selon les 4 groupes typologiques. Dans l'ordre chlorophylle-a, Carbone Organique Dissous (DOC), matières en suspension (TSS) et transparence au disque de Secchi. Les lignes en pointillées représentent les seuils inférieur et supérieur

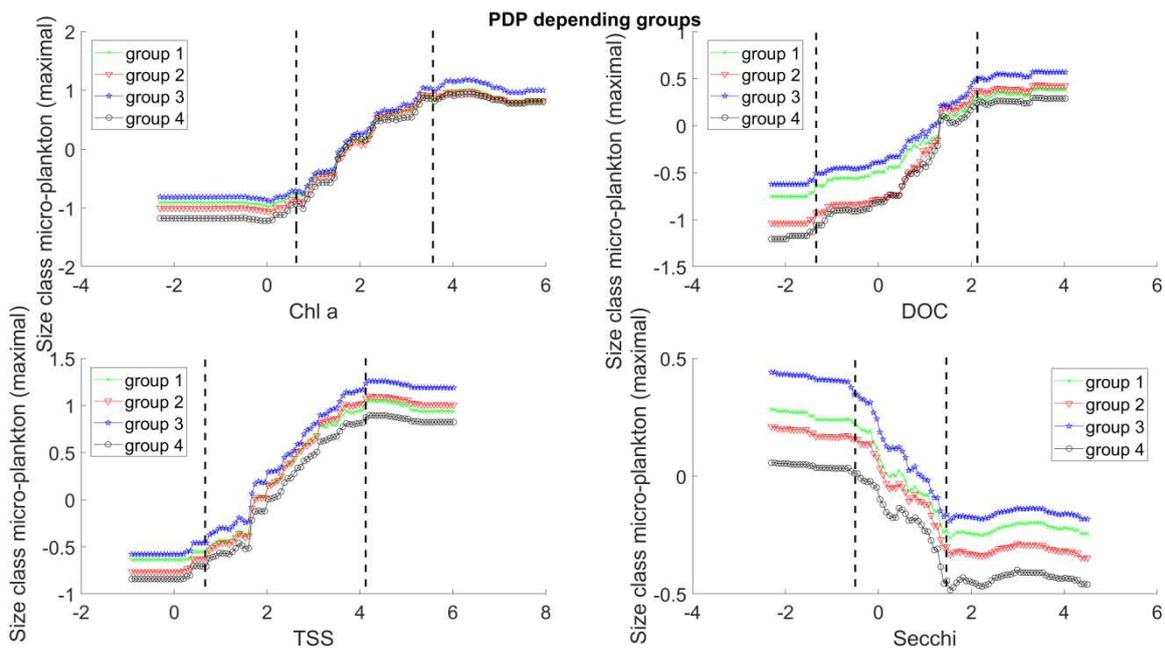


Figure A3.4. Partial Dependence Plot (PDP) pour les 4 premiers paramètres physico-chimiques influençant la classe de taille du micro-plancton selon les 4 groupes typologiques. Dans l'ordre chlorophylle-a, Carbone Organique Dissous (DOC), matières en suspension (TSS) et transparence au disque de Secchi. Les lignes en pointillées représentent les seuils inférieur et supérieur

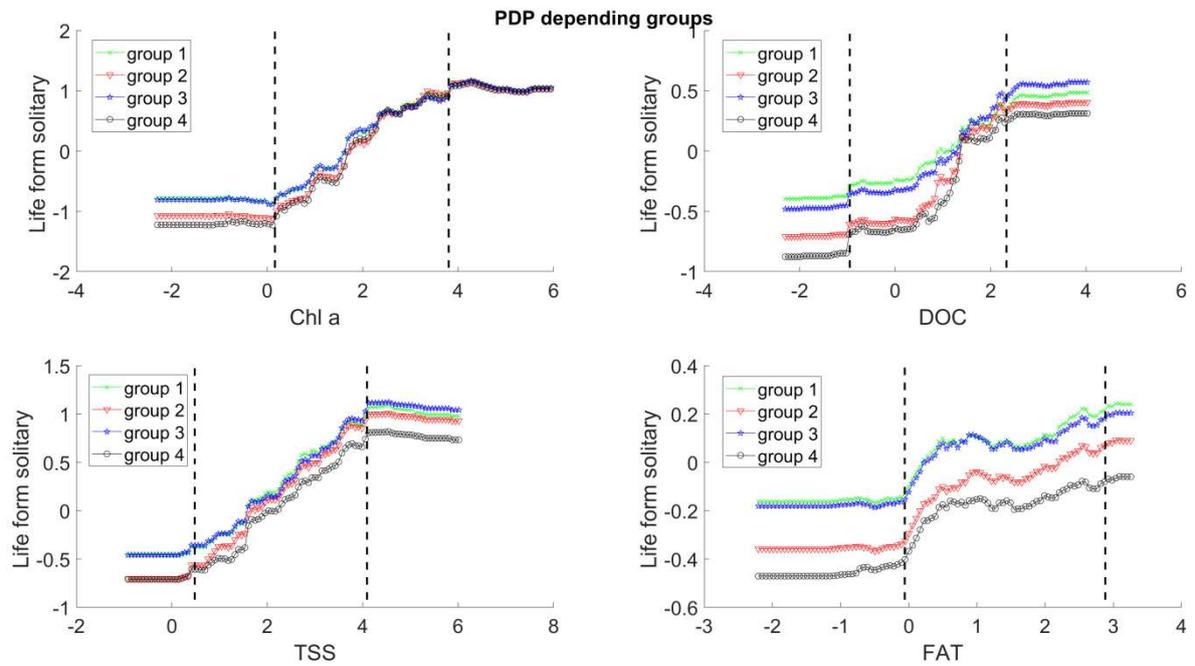


Figure A3.5. Partial Dependance Plot (PDP) pour les 4 premiers paramètres physico-chimiques influençant la forme de vie solitaire selon les 4 groupes typologiques. Dans l'ordre chlorophylle-a, Carbone Organique Dissous (DOC), matières en suspension (TSS) et le titer alcalimétrique complet (FAT). Les lignes en pointillées représentent les seuils inférieur et supérieur

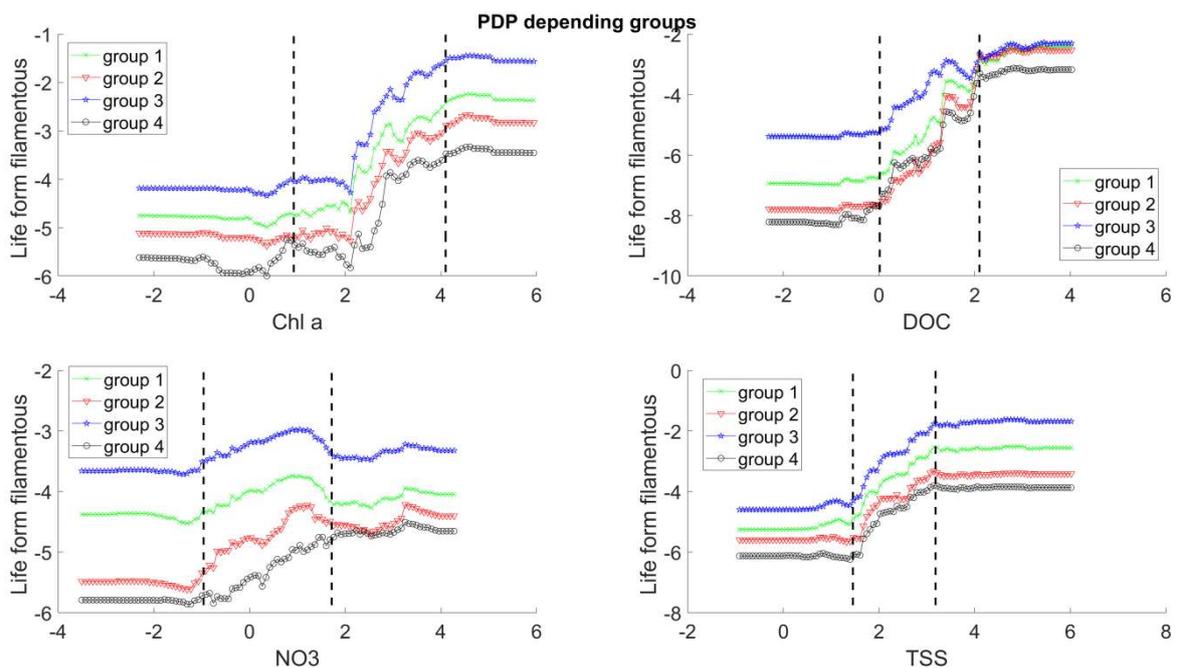


Figure A3.6. Partial Dependance Plot (PDP) pour les 4 premiers paramètres physico-chimiques influençant la forme de vie filamenteuse selon les 4 groupes typologiques. Dans l'ordre chlorophylle-a, Carbone Organique Dissous (DOC), nitrates (NO3) et matières en suspension (TSS). Les lignes en pointillées représentent les seuils inférieur et supérieur

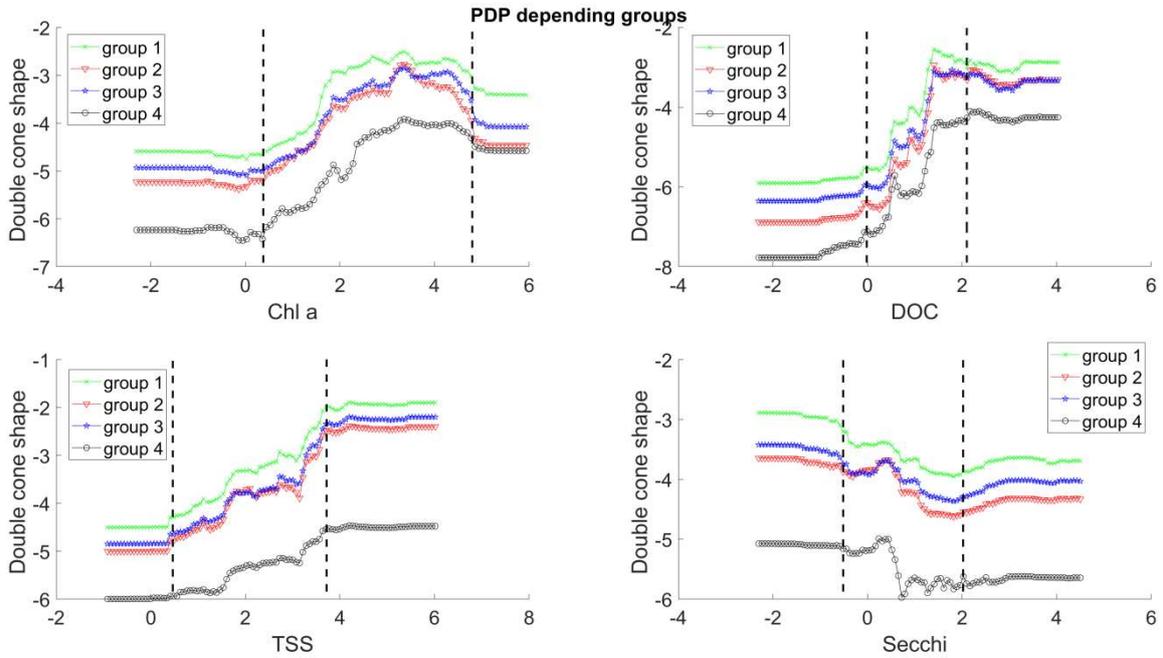


Figure A3.7. Partial Dependence Plot (PDP) pour les 4 premiers paramètres physico-chimiques influençant la forme double cône selon les 4 groupes typologiques. Dans l'ordre chlorophylle-a, Carbone Organique Dissous (DOC), matières en suspension (TSS) et transparence au disque de Secchi. Les lignes en pointillées représentent les seuils inférieur et supérieur

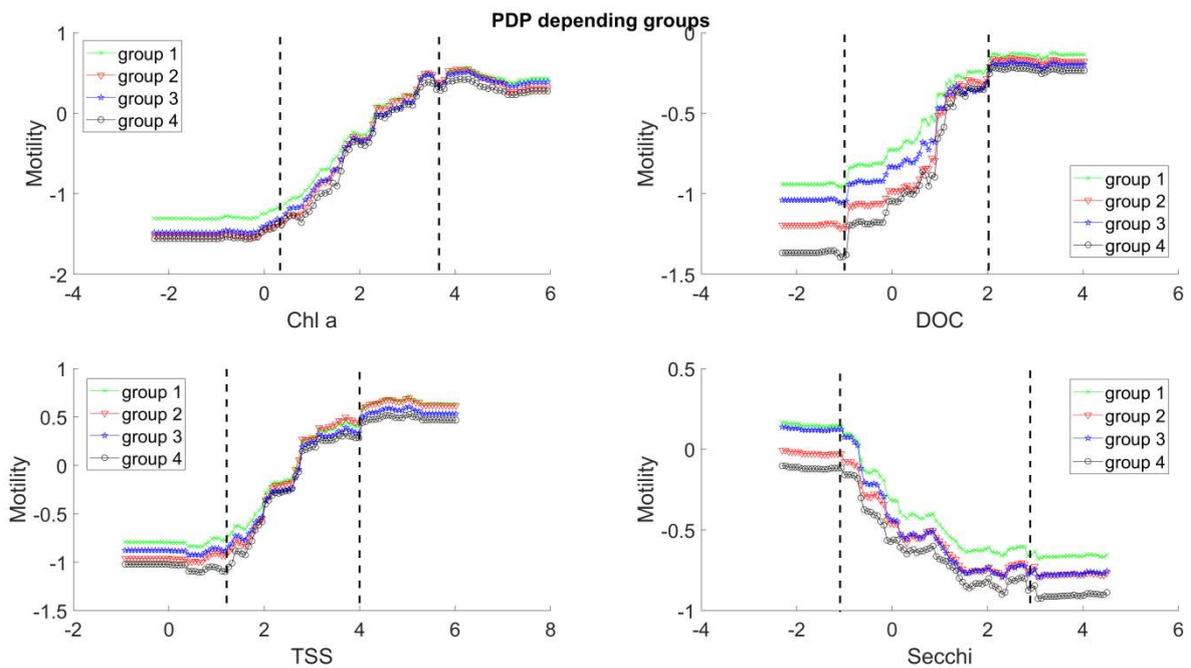


Figure A3.8. Partial Dependence Plot (PDP) pour les 4 premiers paramètres physico-chimiques influençant la motilité selon les 4 groupes typologiques. Dans l'ordre chlorophylle-a, Carbone Organique Dissous (DOC), matières en suspension (TSS) et transparence au disque de Secchi. Les lignes en pointillées représentent les seuils inférieur et supérieur

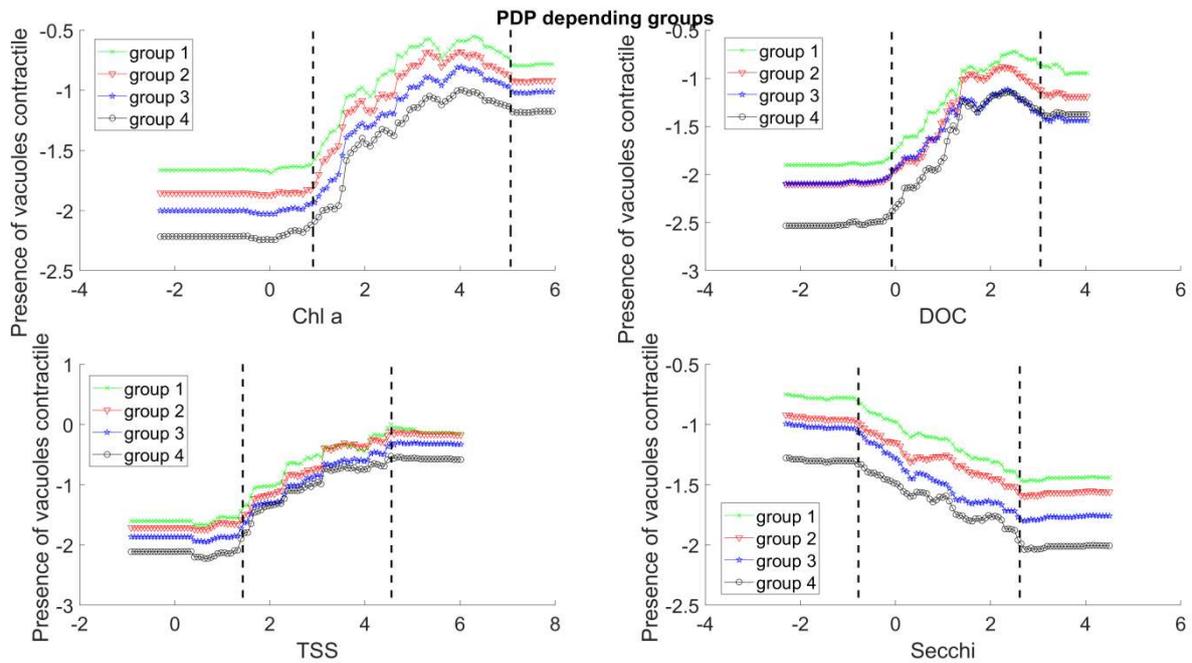


Figure A3.9. Partial Dependence Plot (PDP) pour les 4 premiers paramètres physico-chimiques influençant la présence de pseudo-vacuoles contractiles selon les 4 groupes typologiques. Dans l'ordre chlorophylle-a, Carbone Organique Dissous (DOC), matières en suspension (TSS) et transparence au disque de Secchi. Les lignes en pointillées représentent les seuils inférieur et supérieur

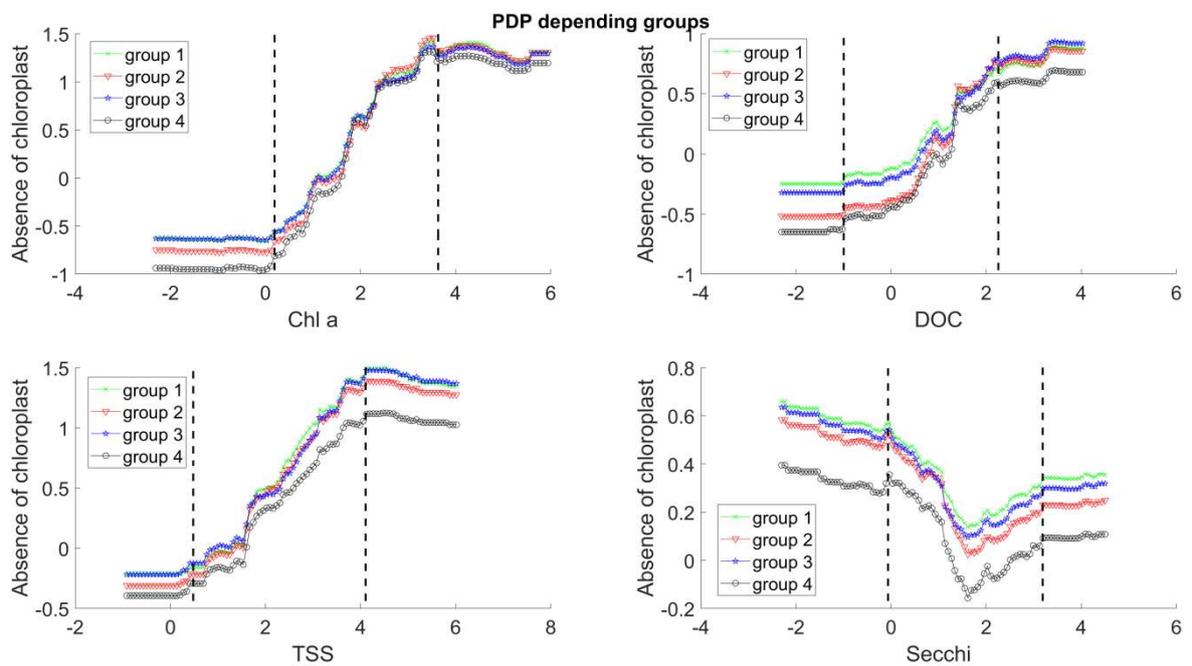


Figure A3.10. Partial Dependence Plot (PDP) pour les 4 premiers paramètres physico-chimiques influençant la présence de chloroplastes selon les 4 groupes typologiques. Dans l'ordre chlorophylle-a, Carbone Organique Dissous (DOC), matières en suspension (TSS) et transparence au disque de Secchi. Les lignes en pointillées représentent les seuils inférieur et supérieur

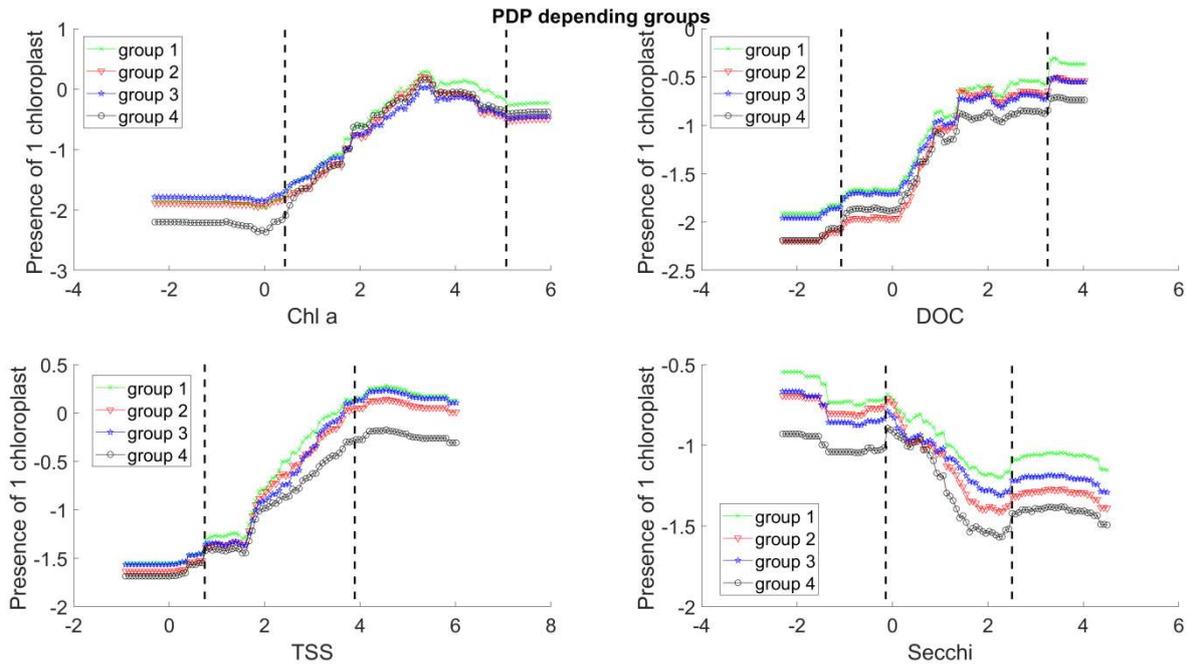


Figure A3.11. Partial Dependence Plot (PDP) pour les 4 premiers paramètres physico-chimiques influençant la présence d'un chloroplaste selon les 4 groupes typologiques. Dans l'ordre chlorophylle-a, Carbone Organique Dissous (DOC), matières en suspension (TSS) et transparence au disque de Secchi. Les lignes en pointillées représentent les seuils inférieur et supérieur

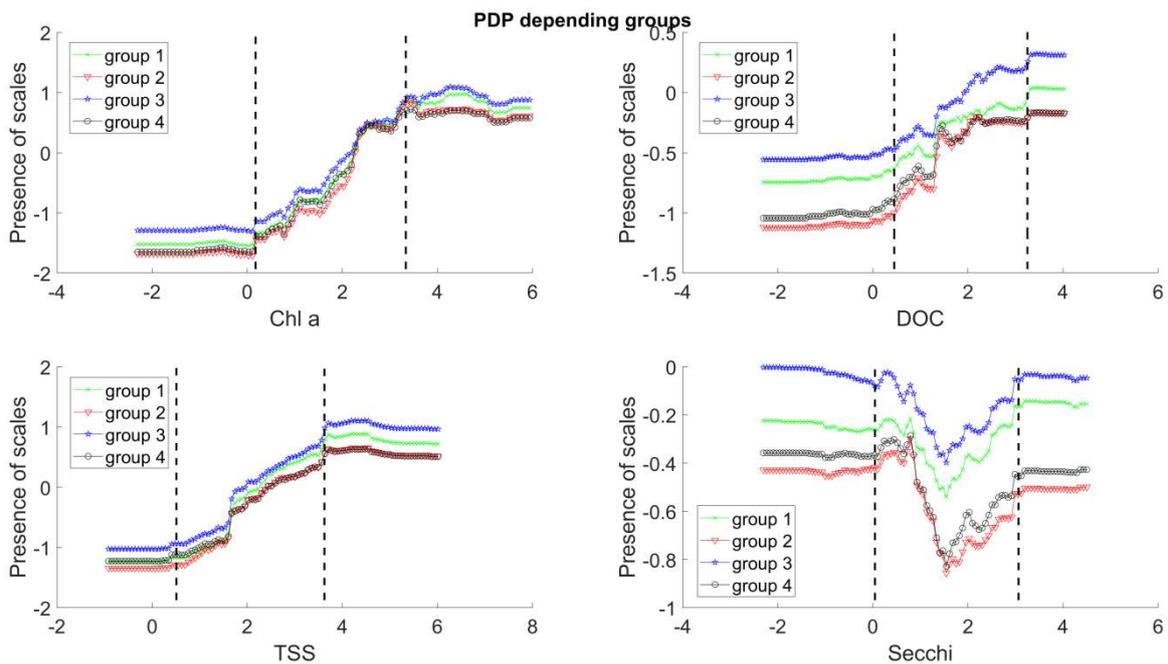


Figure A3.12. Partial Dependence Plot (PDP) pour les 4 premiers paramètres physico-chimiques influençant la présence d'écaille selon les 4 groupes typologiques. Dans l'ordre chlorophylle-a, Carbone Organique Dissous (DOC), matières en suspension (TSS) et transparence au disque de Secchi. Les lignes en pointillées représentent les seuils inférieur et supérieur

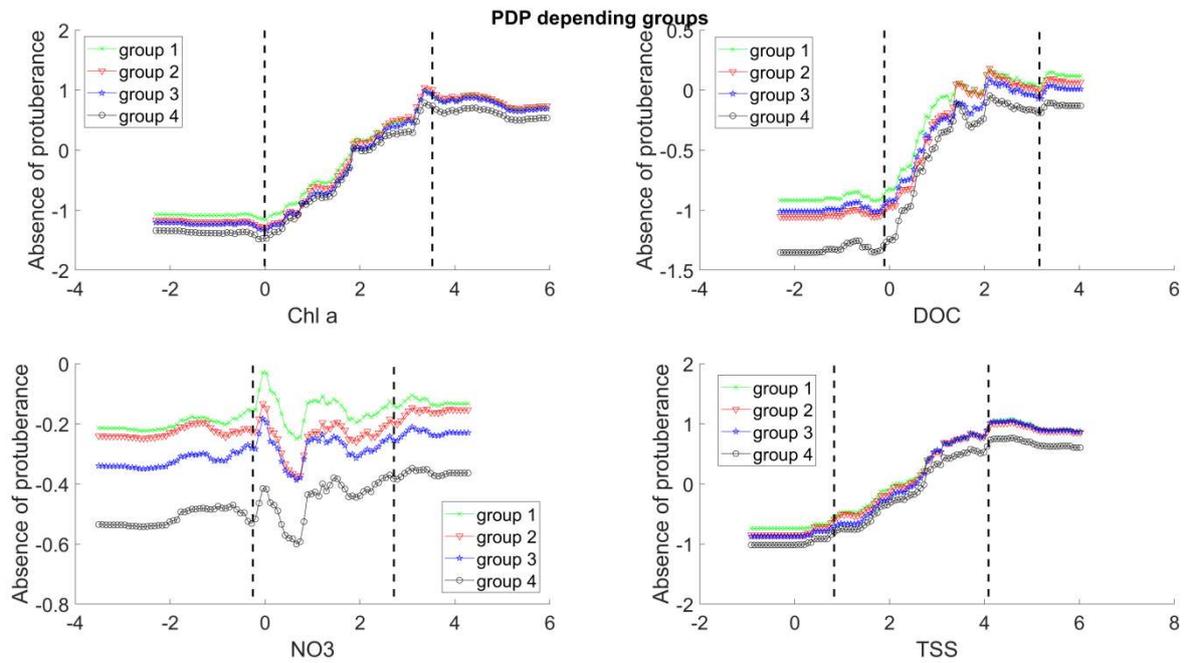


Figure A3.13. Partial Dependence Plot (PDP) pour les 4 premiers paramètres physico-chimiques influençant l'absence de protubérance selon les 4 groupes typologiques. Dans l'ordre chlorophylle-a, Carbone Organique Dissous (DOC), nitrate (NO3) et matières en suspension (TSS). Les lignes en pointillées représentent les seuils inférieur et supérieur

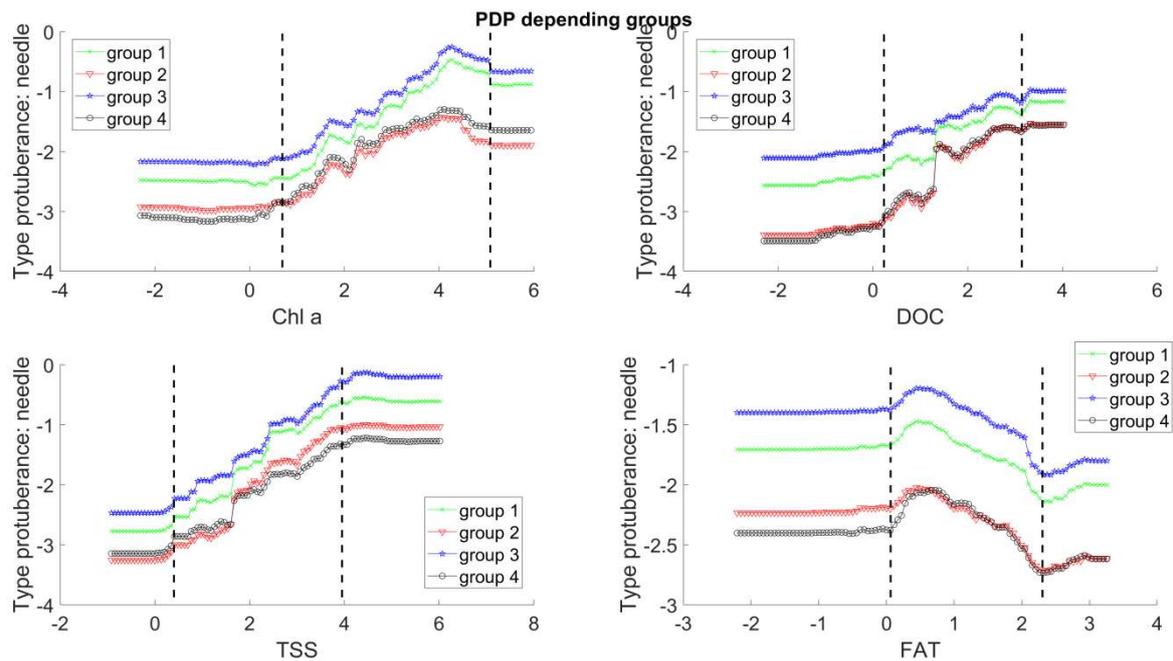


Figure A3.14. Partial Dependence Plot (PDP) pour les 4 premiers paramètres physico-chimiques influençant le type de protubérance aiguille selon les 4 groupes typologiques. Dans l'ordre chlorophylle-a, Carbone Organique Dissous (DOC), matières en suspension (TSS) et titre alcalimétrique complet (FAT). Les lignes en pointillées représentent les seuils inférieur et supérieur

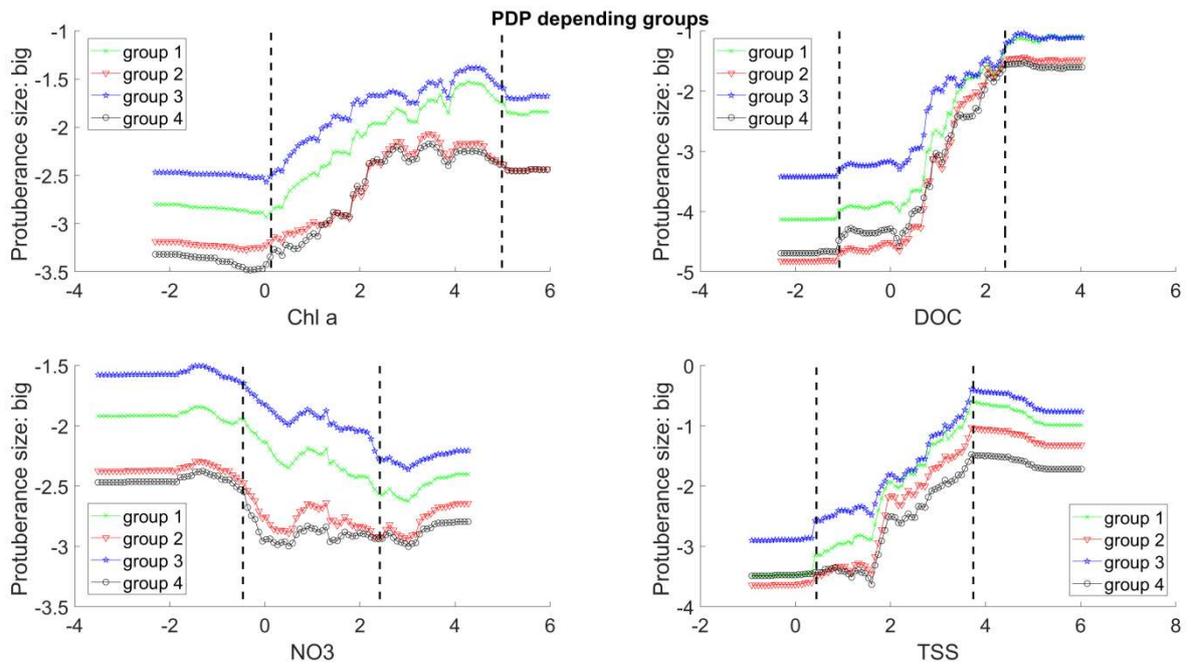


Figure A3.15. Partial Dependance Plot (PDP) pour les 4 premiers paramètres physico-chimiques influençant la taille de la protuberance : grosse selon les 4 groupes typologiques. Dans l'ordre chlorophylle-a, Carbone Organique Dissous (DOC), nitrate (NO3) et matières en suspension (TSS). Les lignes en pointillées représentent les seuils inférieur et supérieur

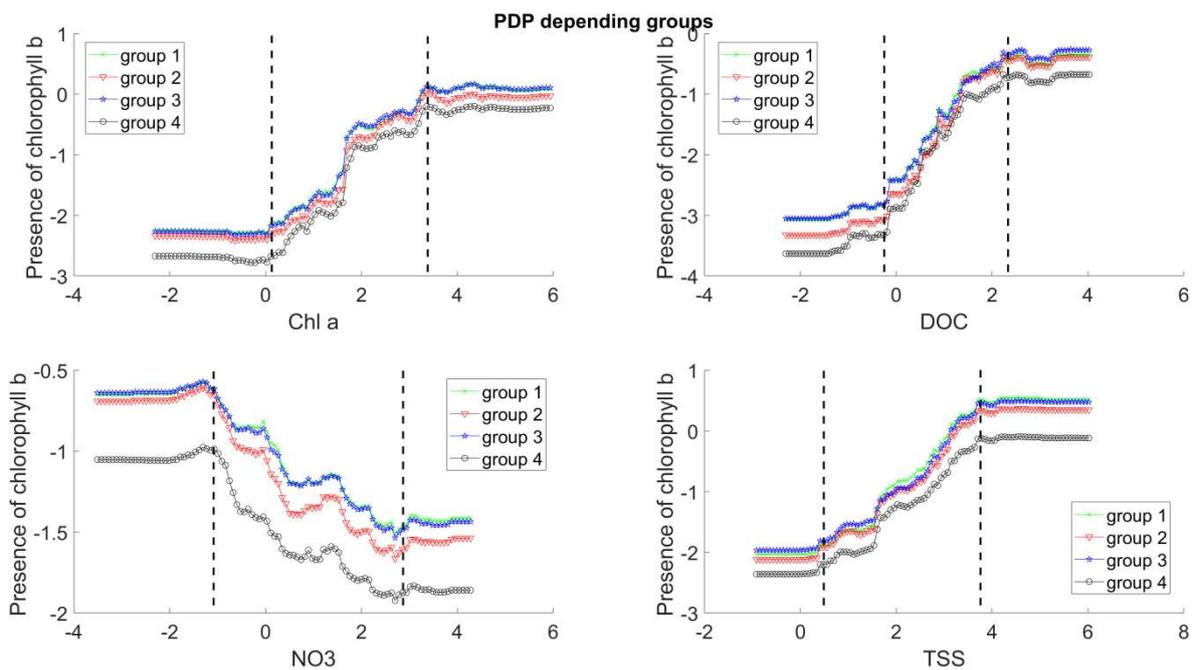


Figure A3.16. Partial Dependance Plot (PDP) pour les 4 premiers paramètres physico-chimiques influençant la présence de chlorophylle-b selon les 4 groupes typologiques. Dans l'ordre chlorophylle-a, Carbone Organique Dissous (DOC), nitrate (NO3) et matières en suspension (TSS). Les lignes en pointillées représentent les seuils inférieur et supérieur

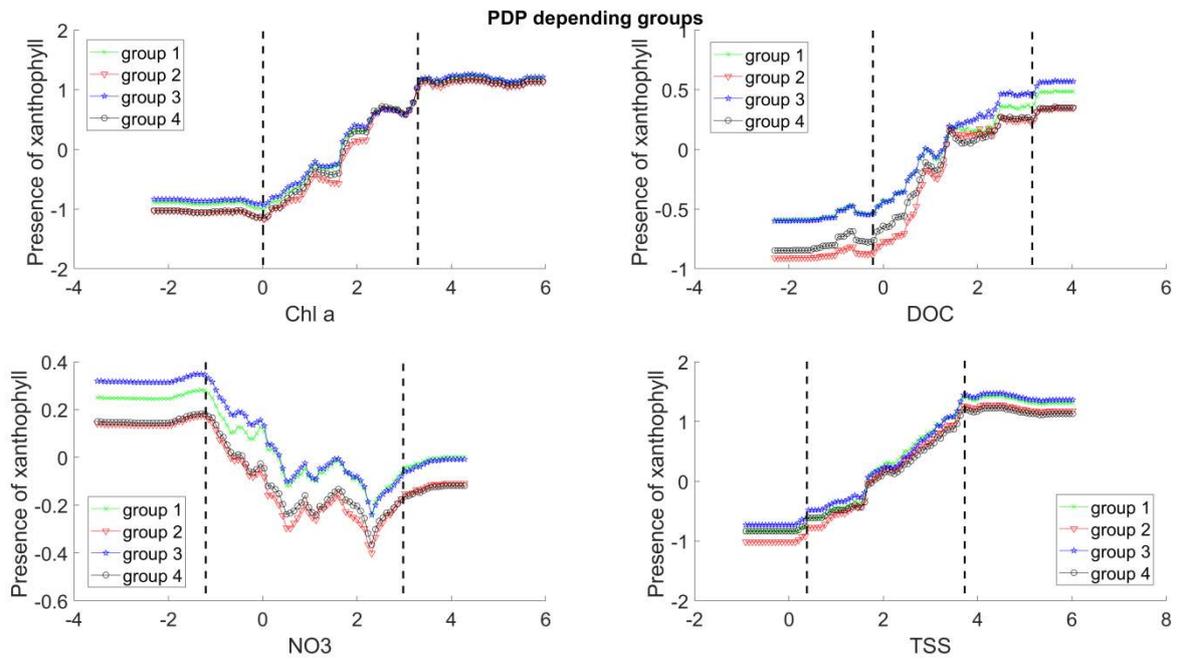


Figure A3.17. Partial Dependance Plot (PDP) pour les 4 premiers paramètres physico-chimiques influençant la présence de Xanthophylle selon les 4 groupes typologiques. Dans l'ordre chlorophylle-a, Carbone Organique Dissous (DOC), nitrate (NO3) et matières en suspension (TSS). Les lignes en pointillées représentent les seuils inférieur et supérieur

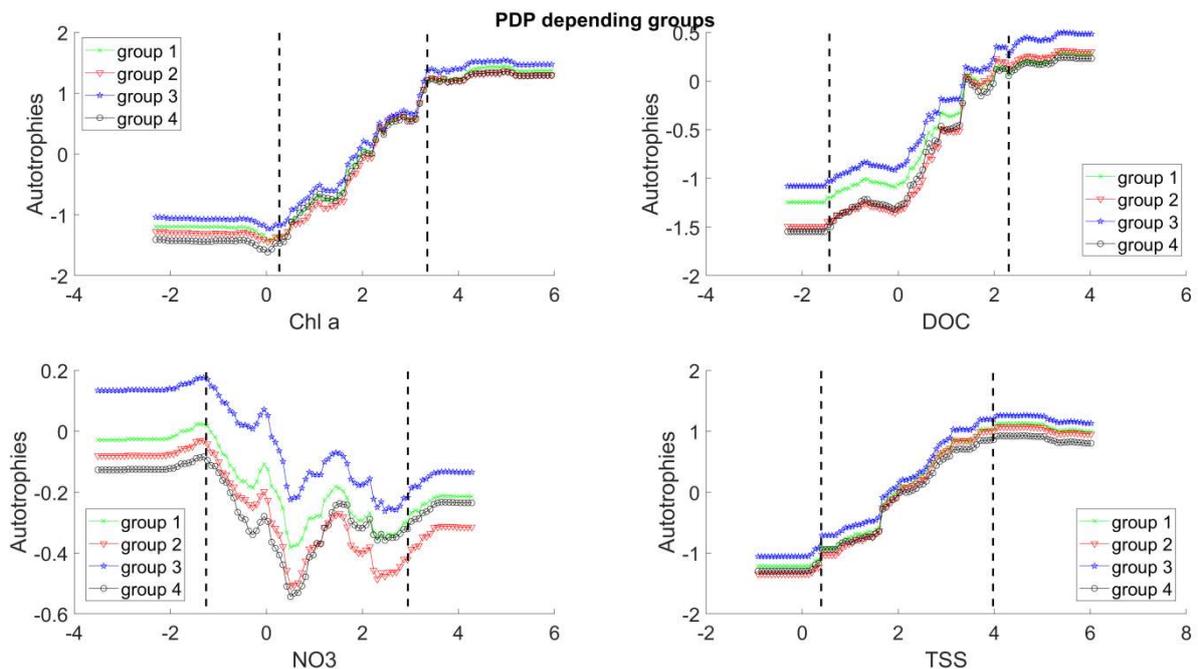


Figure A3.18. Partial Dependance Plot (PDP) pour les 4 premiers paramètres physico-chimiques influençant l'autotrophie selon les 4 groupes typologiques. Dans l'ordre chlorophylle-a, Carbone Organique Dissous (DOC), nitrate (NO3) et matières en suspension (TSS). Les lignes en pointillées représentent les seuils inférieur et supérieur

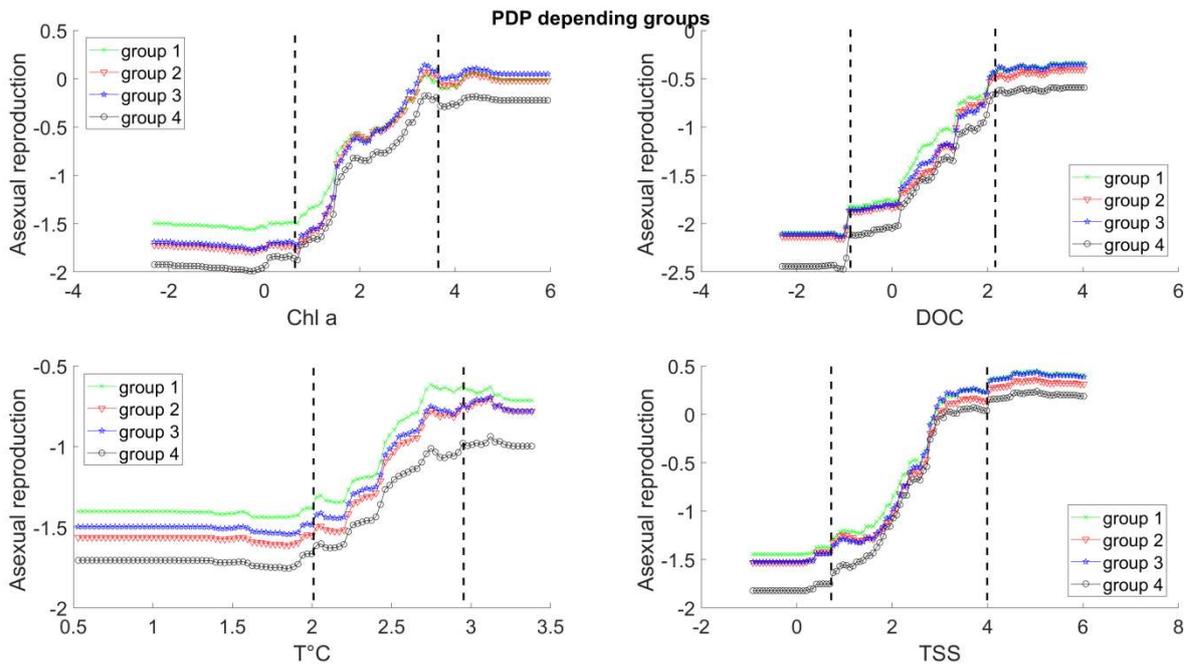


Figure A3.19. Partial Dependance Plot (PDP) pour les 4 premiers paramètres physico-chimiques influençant la reproduction asexuée selon les 4 groupes typologiques. Dans l'ordre chlorophylle-a, Carbone Organique Dissous (DOC), température (T°C) et matières en suspension (TSS). Les lignes en pointillées représentent les seuils inférieur et supérieur

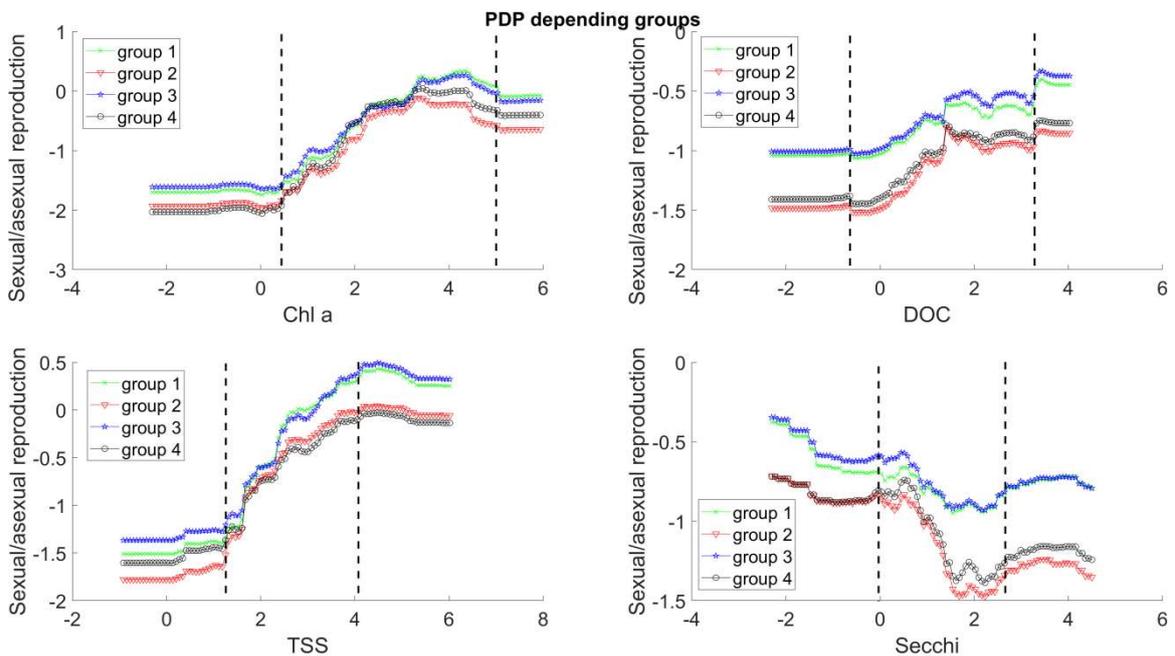
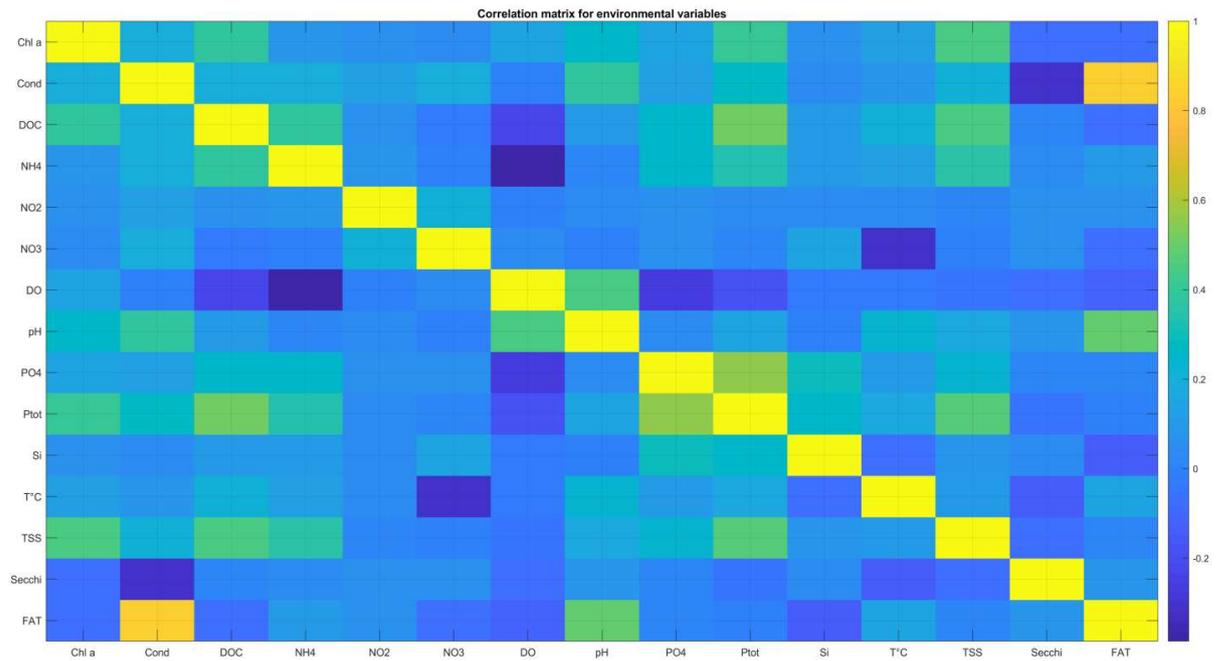


Figure A3.20. Partial Dependance Plot (PDP) pour les 4 premiers paramètres physico-chimiques influençant la reproduction sexuée/asexuée selon les 4 groupes typologiques. Dans l'ordre chlorophylle-a, Carbone Organique Dissous (DOC), matières en suspension (TSS) et transparence au disque de Secchi. Les lignes en pointillées représentent les seuils inférieur et supérieur

Annexe 5 : matrice de corrélations entre les paramètres environnementaux. Plus la couleur de la case tend vers le jaune (coefficient de corrélation de 1) et plus la corrélation est forte



Irstea

1, rue Pierre-Gilles de Gennes

CS 10030

92761 Antony Cedex

01 40 96 61 21

www.irstea.fr

Agence Française pour la Biodiversité

Hall C – Le Nadar

5, square Félix Nadar

94300 Vincennes

01 45 14 36 00

www.afbiodiversite.fr