

**THÈSE DE DOCTORAT**  
PRÉSENTÉE À  
**L'UNIVERSITÉ BORDEAUX I**

**ÉCOLE DOCTORALE**  
« *Sciences et environnements* »

par **Marius BOTTIN**

pour obtenir le grade de

*DOCTEUR en*  
*Écologie évolutive, fonctionnelle et des communautés*

**Structure des assemblages de diatomées benthiques en rivière :  
l'environnement explique-t-il tout ?  
Processus écologiques et développement méthodologiques.**

Sous la direction de **Michel COSTE** et **Sovan LEK**

Soutenue le **28 juin 2012**,  
devant le jury composé de

**Didier Alard**

Professeur, UMR BIOGECO, Université Bordeaux I

Président  
Rapporteur de soutenance

**Jean-Pierre Descy**

Professeur, FUNDP Namur

Rapporteur

**Isabelle Lavoie**

Chargée de recherche, INRS Centre Eau Terre et Environnement, Québec

Rapporteur

**Janne Soininen**

Professeur, Department of Geoscience and Geography, University of Helsinki

Examineur

**Juliette Tison-Rosebery**

Chargée de recherche, Irstea, Unité REBX

Examineur

**Sovan Lek**

Professeur, Laboratoire EDB, Université Toulouse Paul Sabatier

Examineur

**Michel Coste**

Directeur de recherche émérite, Irstea Unité REBX

Examineur



## RÉSUMÉ

Les diatomées sont des algues microscopiques qui sont largement utilisées pour évaluer la qualité écologique des cours d'eau. Les méthodes utilisées se basent sur des modèles simplifiés de biologie des communautés, dans lesquels seules les réponses individuelles des espèces à l'environnement sont prises en compte. Le test de l'importance de processus complémentaires a montré un impact fort des dynamiques de colonisation des espèces, mais un impact négligeable des phénomènes de compétition ou de facilitation. Ces processus impliquent une structure des assemblages bien plus complexe que celle habituellement assumée par les méthodologies de bioindication. L'adaptation et la mise en oeuvre de méthodes de réseaux de neurones et de logique floue nous ont permis de redéfinir des éco-régions françaises et de décrire des relations générales entre les traits biologiques des espèces et l'environnement, tout en prenant mieux en compte cette complexité.

## ENGLISH TITLE

Structure of benthic diatom assemblages in rivers : is environment the only explanation ?

## ABSTRACT

Diatoms are microscopic algae which are widely used to monitor the ecological quality of streams and rivers. The regular methodologies are based on simplified community models. In these models, only the individual species responses to environment are accounted for. Testing the importance of complementary processes showed a significant effect of colonization dynamics, but only a slight effect of biotic relationships. These processes led us to considerate a more complex assemblage structure than the one usually assumed by the biomonitoring methodologies. Therefore we implemented both neural networks models and fuzzy logic methodologies, in order to refine French ecoregions and to describe relationships between species traits and environment.

## MOTS-CLEFS

Diatomées ; Assemblage ; Dispersion ; Relations biotiques ; Carte auto-organisatrice de Kohonen ; Pattern trees

## KEYWORDS

Diatoms ; Assemblage ; Dispersal ; Biotic relationships ; Self organizing map ; Artificial neural network ; Pattern trees



Cette thèse a été financé par Irstea (qui s'appelait alors CEMAGREF) grâce des contrats avec l'ONEMA. Merci à ces deux organismes pour cela, ainsi qu'aux citoyens français et européens pour avoir participé à ce projet par leurs impôts!

Une thèse a cela de particulier qu'elle demande tellement d'implication personnelle que se retrouvent alors imbriqués les mondes professionnels et privés, pour le meilleur et pour le pire! Je parlerai ici surtout du meilleur...

Je tiens à remercier tous les gens qui ont participé, de près ou de loin à cette aventure, à tous ceux qui ont été là au bureau, dans les réunions professionnelles et colloques plus ou moins éloignés, dans mes successives colocations, ou même dans les bars ou à la plage!

Voici quelques mentions spéciales, rituelles et non moins sincères :

Merci aux rapporteurs de cette thèse, *Jean-Pierre Descy* et *Isabelle Lavoie*, pour avoir relu, commenté ma thèse et s'être déplacé jusqu'à Bordeaux pour ma soutenance. Vos remarques m'ont été très utile et j'essaierai de m'en souvenir pour la suite de mon travail. Merci aussi à *Janne Soininen* d'être venu à cette soutenance, ainsi qu'à *Didier Alard* de l'avoir présidé.

*Juliette* l'extraordinaire humanité, l'humilité<sup>1</sup> le dévouement aux autres, l'humour sont quelques unes de tes qualités qui m'ont facilité la vie pendant ces quatre ans. Malgré tes réticences, il faut le dire, ces qualités font de toi une chef/manager extraordinaires!!! A cela il faut rajouter une aide concrète, de tous les jours, sans laquelle je n'aurais jamais réussi à affirmer quoique ce soit... tout cela ne serait rester qu'hypothèses assorties de contre-hypothèses tout aussi probables, un fouilli informe. Bref, je ne saurai jamais te remercier assez!

*Sovan* et *Michel*, je vous suis très reconnaissant d'avoir accepter d'être mes directeurs de thèses. Votre rôle n'a pas été de m'accompagner tous les jours sur les petites questions, doutes et choix à faire c'est *Juliette* qui a assumé cela, mais les discussions avec vous m'ont toujours été utiles, notamment lors de mes comités de thèse ou juste dans le couloir pour *Michel*. Chacun à votre manière, vous êtes et resterez des exemples pour moi, *Michel* par ton humilité<sup>1</sup> et ton humour malgré une connaissance énorme et reconnue par tous, *Sovan* pour ton sérieux et ton exigence. Merci à vous deux d'avoir cru en moi et de m'avoir soutenu jusqu'au bout!

KIITOS *Janne*, Aion kirjoittaa Englanti, koska luulen, että minun huono Suomen kielen olisivat todellisuudessa vähemmän ymmärrettävää sinulle kuin minun Englanti. Thank you for coming from so far at every Ph.D. comitee meeting, for having invited me in your lab, for teaching me how to write a paper, and for helping so efficiently and quickly to write them. I won't go into details here, but I will also remember you showing me the "underground part" of Helsinki, and I'm grateful for that too.

*François*, merci de m'avoir pris en stage avant cette thèse, d'en avoir rédigé le sujet avec Juliette et de l'avoir défendu. Merci aussi pour ta vision globale des écosystèmes, et je ne peux pas faire ces remerciements sans mentionner, c'est maintenant une tradition à laquelle je ne peux déroger) tes débits et puissances

---

1. je pense que j'ai eu une chance extraordinaire de faire ma thèse avec vous pour cela en particulier, il y a, à mon avis, peu de gens qui ont pu marquer tant de fois « humilité » dans leurs remerciements de thèse, vu l'ambiance générale du milieu ...

de voix légendaires !

*Jean-Luc*, merci pour ton aide régulière, en particulier sur le sujet des SOM. Les réunions avec toi m'ont toujours fait avancé, et le code de Emuann a été une base sans laquelle la partie SOM de cette thèse n'aurait pu exister.

*Muriel*, mille fois merci pour m'avoir montré ce métier et de me l'avoir fait aimé, et de m'avoir tant appris lors de ce stage que j'ai eu la chance de faire à tes côtés. Merci aussi pour m'avoir présenté *François* quand il est venu à Toulouse, pour le stage qui a débouché sur cette thèse ! J'espère que nous aurons l'occasion de travailler ensemble, tes qualités humaine, gentillesse, humilité<sup>1</sup>, motivation sans faille rendent le travail tellement plus agréable !

*Didier*, merci pour nos discussions très intéressantes sur la biologie théorique des communautés et sur la partie « cooccurrences » de cette thèse.

*Isa*, aujourd'hui je sais à quel point il est agréable de travailler avec toi. Toi-aussi, tes attentions, ton humanité, ton humilité<sup>1</sup> et ton beurre d'arbre me rassurent grandement sur le monde de la recherche.

Merci à *Nicolas* et *Julie* pour leur travail sur les SOM et les pattern-trees, les très nombreuses lignes de codes écrites dans une fort bonne humeur malgré ma disponibilité parfois insuffisante.

Merci aux gens que j'ai eu l'occasion de rencontrer pendant les colloques et congrès, ADLAF, IDS, ECOVEG, ISEI, pour ces moments aussi stimulants qu'agréables.

Merci aussi à tous ceux qui m'ont accompagné dans les parties « moins recherche » de mon travail : *Jacques Puiroux* et l'équipe de biologie animale pour les TP de dissection, *Gilles Durieux*, ainsi que *Virgil Fievet*, *Hugues Blanchet* et l'ensemble de l'équipe pédagogique de statistiques pour ces enseignements de statistiques dont le seul intitulé fait déjà frémir les étudiants (que je remercie aussi, j'espère qu'ils ne m'en veulent pas trop de représenter ce domaine!). Je remercie aussi *Richard Michalet* et l'ensemble du conseil scientifique de l'équipe doctorale, dont mes collègues représentants des doctorants

Je suis vraiment heureux d'avoir pu faire ma thèse dans des conditions formidables, grâce à l'ambiance agréable que soutienne mes fabuleux collègues. Je remercie donc toute l'équipe CARMA, ses permanents, CDD ou stagiaires pour les bons moments passés, pour le travail ou les moments moins formels, repas ou pause-café. Un merci particulier à *Daniel Poulain* qui y est pour beaucoup dans les bonnes relations qui persistent entre les gens de l'unité REBX, grâce à une finesse toute particulière dont il fait preuve dans la gestion des personnalités (parfois fortes) présentes dans l'équipe !

Certains sont devenus de vrais amis bien au delà des bâtiments (quand je vous disais que ça mélange sec...).

*Soizic* merci pour ton engagement sans faille et sans attente de retour pour l'équipe (et pour les autres d'une façon plus générale), l'humilité<sup>1</sup>, les nombreuses corrections de fond ou de forme mais aussi les grandes discussions, l'écoute et les pauses bières de fin de journée !

*Soizic*, *Sandra*, *Nicolas P.* (ou *C.* selon ...) *Vincent* et *Nicolas A.*, vous avez été mes proches compagnons pendant ces années de thèse. Merci pour vos corrections, pour l'organisation qu'il y a eu autour de ma soutenance de thèse, mais aussi et surtout pour tout le reste ! C'était tellement bon de savoir qu'aller au travail, c'était aussi venir vous voir, et malgré tout ce temps passé ensemble,

il était encore très motivant de vous retrouver le soir, le week-end ou pendant les vacances ! Vous êtes des amis, et j'espère ne jamais vous perdre de vue (j'espère qu'on se verra à Helsinki, Bogotà, Rio, Bordeaux, Gazinet, ou n'importe quel endroit du monde). Un remerciement particulier à *Sandra* pour m'avoir aidé à affronter une épreuve énorme : les magasins de vêtements avant la thèse !

Le mélange des genres atteint ici son paroxysme, mais je ne peux écrire ces remerciements sans penser à tous les gens qui m'ont aidé d'une façon plus personnelle, à Bordeaux ou avant. . .

Tous les amis bordelais, compagnons de collocation, de musique, ou de « Sainmichelade ». Un merci particulier à ceux qui sont venu à ma soutenance, c'était vraiment une belle journée en votre compagnie, et puis comme je ne pense pas me marier de si tôt, c'était l'occaz, hein !

Mes parents : mon père, ma mère, mais aussi mes deux « autres parents » *Eric* et *Marie*.

Mes frères et soeurs (prenez une respiration. . .) *Mathilde*, *Azilis*, *Etienne*, *Hannah*, *Élie* et *Loïc*, et leurs p'tis gars *Dario*, *Oscar* et *Sidney*. Un pensée particulière à *Élie* et *Loïc*, ce qui s'est passé le jour de ma thèse m'a profondément touché !



# Table des matières

<b>Introduction</b>	<b>xi</b>
<b>1 Contexte</b>	<b>1</b>
1.1 Quelques notions sur les diatomées . . . . .	1
1.1.1 Caractéristiques et fonctionnement biologique . . . . .	1
1.1.2 Evolution, phylogénie et classification . . . . .	2
1.1.3 Écologie, habitats et importance des diatomées . . . . .	5
1.2 Les diatomées en tant qu'indicateurs biologiques . . . . .	7
1.2.1 Une forte sensibilité aux conditions du milieu . . . . .	7
1.2.2 Un bon bioindicateur ? . . . . .	7
1.2.3 Échantillonnage . . . . .	8
1.2.4 Préparation du matériel collecté et identification . . . . .	9
1.2.5 Calculs des indices . . . . .	9
1.2.6 Bioindication diatomique sur le territoire français . . . . .	12
1.2.7 Sources de variabilité . . . . .	13
1.3 Biologie des communautés . . . . .	15
1.3.1 Avant de commencer . . . . .	15
1.3.2 Super-organisme ou association d'organismes indépendants ?	15
1.3.3 Les niches écologiques . . . . .	16
1.3.4 Biogéographie et théorie neutraliste . . . . .	18
1.3.5 Les méta-communautés . . . . .	19
1.3.6 Nestedness . . . . .	20
1.3.7 Les règles d'assemblage . . . . .	20
1.3.8 Spécificités des micro-organismes . . . . .	21
1.3.9 La bioindication diatomique : un modèle simplifié . . . . .	22
1.4 Écologie numérique . . . . .	24
1.4.1 Les données biologiques . . . . .	24
1.4.2 Les mesures de la biodiversité . . . . .	24
1.4.3 Les bio-statistiques . . . . .	26
1.4.4 Les méthodes alternatives aux statistiques classiques . . . . .	29
1.5 Données . . . . .	30
1.5.1 Base de données « historique » d'Irstea . . . . .	30
1.5.2 Base de données « récente » . . . . .	31

<b>I</b>	<b>Test de l'importance des grands processus de biologie des communautés sur les assemblages de diatomées</b>	<b>41</b>
<b>2</b>	<b>Structuration spatiale</b>	<b>43</b>
2.1	Buts et contexte de l'étude . . . . .	43
2.2	Principaux résultats . . . . .	44
2.3	Spatial structure in assemblages . . . . .	45
2.3.1	Introduction . . . . .	45
2.3.2	Material and methods . . . . .	47
2.3.3	Results . . . . .	51
2.3.4	Discussion . . . . .	61
<b>3</b>	<b>Cooccurrence patterns</b>	<b>71</b>
3.1	Buts et contexte de l'étude . . . . .	71
3.2	Principaux résultats . . . . .	72
3.3	Species-by-species prediction . . . . .	73
3.3.1	Introduction . . . . .	73
3.3.2	Material and methods . . . . .	75
3.3.3	Results . . . . .	80
3.3.4	Discussion . . . . .	85
<b>II</b>	<b>Quelles méthodes pour analyser la structure établie des assemblages ?</b>	<b>97</b>
<b>4</b>	<b>Les SOM</b>	<b>99</b>
4.1	Présentation générale . . . . .	99
4.1.1	Un réseau de neurones artificiels . . . . .	99
4.1.2	Fonctionnement des SOM . . . . .	100
4.1.3	Avantages des SOM . . . . .	105
4.2	Adaptation aux diatomées . . . . .	107
4.2.1	Problèmes posés . . . . .	107
4.2.2	Solutions proposées . . . . .	108
4.3	Biotypologie . . . . .	110
4.3.1	Buts de l'étude . . . . .	110
4.3.2	Informations complémentaires sur les données et les méthodes . . . . .	110
4.3.3	Résultats . . . . .	112
4.3.4	Discussion . . . . .	114
<b>5</b>	<b>Les pattern trees</b>	<b>117</b>
5.1	Présentation . . . . .	117
5.1.1	Généralités . . . . .	117
5.1.2	Paramétrage et algorithme des pattern trees . . . . .	118
5.2	Pattern trees et guildes des diatomées . . . . .	123
5.2.1	Matériel et méthode . . . . .	123
5.2.2	Résultats et discussion . . . . .	124
5.2.3	Conclusions . . . . .	126

<b>6 Conclusions et perspectives</b>	<b>131</b>
6.1 Structure des assemblages et dispersion des espèces de diatomées	131
6.2 Nestedness . . . . .	132
6.3 Présence et abondances des espèces . . . . .	132
6.4 Le modèle de la bioindication . . . . .	133
<b>A Biotypologie</b>	<b>137</b>
<b>B Pattern trees</b>	<b>159</b>



# Introduction

Les diatomées sont des algues brunes unicellulaires, très diversifiées, et largement répandues à l'échelle du globe dans les océans et dans les eaux de surfaces continentales. Nous nous intéresserons ici aux diatomées dites "benthiques", c'est à dire se développant sur le substrat dur du fond des rivières, car celles-ci forment des assemblages d'espèces dont la distribution est fortement liée aux conditions environnementales. Nous entendons par conditions environnementales les conditions « naturelles », mais aussi des marqueurs de pollutions tels que les concentrations en nutriments. Ainsi, les diatomées benthiques sont utilisées pour décrire les niveaux de qualité écologique des cours d'eau depuis plus d'une vingtaine d'années.

La Directive Cadre européenne sur l'Eau (DCE) impose l'utilisation d'indicateurs biologiques tels que les diatomées pour évaluer la qualité écologique des eaux de surface. Sa mise en œuvre s'est accompagnée d'une dynamisation des réseaux de surveillance et d'un échantillonnage systématique à large échelle des cours d'eau. Ainsi, de nombreuses données ont été récoltées, permettant la construction de relations entre les espèces de diatomées et l'environnement, sur lesquelles se basent des indices biologiques. Ce modèle de relation entre les espèces et l'environnement repose sur une vision simplifiée des communautés biologiques dans laquelle les assemblages locaux résultent directement et uniquement des réponses indépendantes de chacune des espèces à l'environnement. Or, d'autres processus sont connus pour structurer les communautés : les phénomènes de compétition et de facilitation, ou les limitations de la dispersion des espèces, sont des phénomènes couramment évoqués concernant les végétaux terrestres.

Une des avancées récentes majeures en écologie est la théorie neutraliste. Cette théorie prend à revers les connaissances préalablement acquises sur les relations entre les espèces et leur environnement. Ainsi ses défenseurs ont montré qu'en considérant chacune des espèces comme équivalente en termes de préférences environnementales et soumise à une dispersion limitée et aléatoire, on pouvait expliquer une large part de la distribution de la biodiversité à l'échelle planétaire. Il ne s'agit pas d'affirmer que l'environnement n'a aucune importance pour la distribution des espèces, mais de proposer des processus complémentaires expliquant une part supplémentaire de la structuration des communautés. C'est dans ce cadre que s'inscrit mon travail de thèse. Il s'agit de vérifier si la prise en compte de ces facteurs supplémentaires peut aider à la compréhension de la répartition des espèces et des assemblages de diatomées, et permettre grâce à cela d'affiner les méthodes actuelles d'évaluation biologique.

Chacun des processus agissant sur les communautés se traduit par des patrons particuliers dans la structure des communautés obtenues et par des « signa-

tures » caractéristiques dans la distribution de leurs espèces. Il est donc possible, en étudiant la distribution de la diversité des assemblages de diatomées, de déduire les processus qui ont agi sur leur formation. Les micro-organismes ont des propriétés particulières, dont une formidable biodiversité, des capacités de développement rapide et de dispersion à très large échelle. Aussi, il est possible que les processus qui agissent sur ces communautés soient différents de ceux, largement étudiés, qui forment les communautés de plantes macroscopiques. Pourtant, il existe des phénomènes communs à l'ensemble du vivant, il est donc important de tester ces mêmes théories, afin de déterminer les convergences et divergences de fonctionnement. Dans ce cadre, l'étude des diatomées offre des possibilités particulièrement intéressantes, car celles-ci ont été étudiées dans un cadre relativement proche de celui des macro-organismes. En effet, la notion d'espèce de diatomées s'est basée sur des considérations morphologiques, comme pour les organismes dits « supérieurs ». Les diatomées offrent donc une possibilité inédite de comparaison entre les micro-organismes et les macro-organismes.

*Ce mémoire est construit en six chapitres :*

*Le **premier chapitre** introduit le contexte scientifique dans lequel se place ce travail de thèse, situé entre les disciplines de biologie des communautés et de biostatistiques, toutes deux appliquées dans un but d'amélioration des pratiques de bioindication par les assemblages de diatomées benthiques des cours d'eau.*

*Une **première partie**, composée des chapitres deux et trois, consiste en l'analyse de la structure et de la composition des assemblages de diatomées, dans le but de comprendre l'importance de facteurs complémentaires au seul environnement, influençant la répartition et la diversité des espèces.*

*Le **chapitre deux** est une description de la structuration spatiale des assemblages de diatomées français, mettant en lumière une possible importance de phénomènes liés à la dispersion limitée des espèces.*

*Le **chapitre trois** est une analyse des patrons de ségrégation ou d'aggrégation entre espèces, grâce à des modèles nuls permettant de discuter de l'importance des mécanismes de relations biotiques, de migration et de l'influence de la structure de la communauté sur la distribution des espèces.*

*L'étude des assemblages de diatomées dans la première partie révèle une structuration plus complexe que celle prise en compte dans les méthodologies classiques d'analyse numérique. Ainsi, le développement de méthodes biomathématiques innovantes, et leurs applications possibles pour la bioindication sont présentées dans une **seconde partie** comprenant les chapitres quatre et cinq.*

*Le **chapitre quatre** présente une adaptation des Cartes auto-organisatrices de Kohonen, et son application pour un découpage spatial de la distribution des assemblages de diatomées en France.*

*Le **chapitre cinq** présente le développement d'une technique issue des théories de la logique floue : les pattern trees. Cette méthode est utilisée pour décrire l'influence des facteurs environnementaux sur la structure tridimensionnelle des biofilms.*

*Enfin, un **sixième chapitre** dresse les conclusions générales et perspectives de ces travaux de thèse. Des pistes méthodologiques seront proposées pour de futures recherches sur l'importance relative des processus agissant sur les assemblages de diatomées et les développements possibles de méthodes de bio-indication prenant en compte leur structure complexe.*

La problématique générale de ces travaux peut-être résumée par trois questions interdépendantes :

Existe-t-il des facteurs autres que les préférences environnementales des espèces de diatomées agissant sur la structure des assemblages ?

Si ces facteurs existent, impliquent-ils une structuration particulière de la diversité diatomique à l'échelle de la France ?

Est-il possible de prendre en compte cette répartition de la diversité dans nos pratiques de bioindication ?



# Chapitre 1

## Contexte

### 1.1 Quelques notions sur les diatomées

#### 1.1.1 Caractéristiques et fonctionnement biologique

Les diatomées sont des algues microscopiques unicellulaires. Eucaryotes, elles présentent des organites différenciés et structurés au sein de leur cellule (noyau, mitochondries, chloroplastes etc.). Une caractéristique importante des diatomées est de posséder un frustule, exosquelette siliceux composé de deux valves, l'épivalve et l'hypovalve. Ce frustule présente des formes et des ornements variés, notamment des pores permettant aux cellules d'excréter des polymères. C'est à partir de l'observation de ces ornements que s'est construite la classification des diatomées et la reconnaissance des différentes espèces. La diversité de forme des frustules a été très tôt un objet d'admiration (notamment celle de Charles Darwin, dans l'Origine des Espèces 1859) sûrement autant pour des critères esthétiques que par leur illustration de la formidable diversité des micro-organismes.

Les diatomées sont en grande majorité autotrophes, bien que quelques rares espèces aient développé des stratégies leur permettant de s'affranchir de l'énergie lumineuse (Hellebust et Lewin, 1977). Leurs pigments synthétiques, chlorophylles a, c et différents caroténoïdes, confèrent à leurs chloroplastes une couleur brun-jaune et leur assure un large spectre d'absorption de la lumière.

Les cycles biologiques des diatomées sont dits « monogénétiques » car ils sont théoriquement complétés en une seule génération (en considérant la reproduction sexuée), et diplophasiques car la phase diploïde (à  $2n$  chromosomes) est plus longue que la phase haploïde (à  $1n$  chromosomes). Ce cycle, simple, est donc relativement exempt de caractères archaïques en comparaison aux autres algues, et place les diatomées parmi les micro-organismes les « plus évolués ». Les cycles de vie des diatomées sont relativement rapides (cycle de développement allant de quelques heures à quelques jours ; Baars, 1983). La reproduction par voie végétative excède largement la reproduction sexuée. Lors de la reproduction asexuée par mitose, épivalve et hypovalve se séparent, puis forment une cellule complète par génération d'une nouvelle hypovalve (l'hypovalve de la cellule-mère devient alors l'épivalve d'une des cellules filles, voir figure 1.1). Ce processus implique donc une diminution de la taille moyenne des populations lors du développement de celles-ci par reproduction végétative. Lorsque la taille

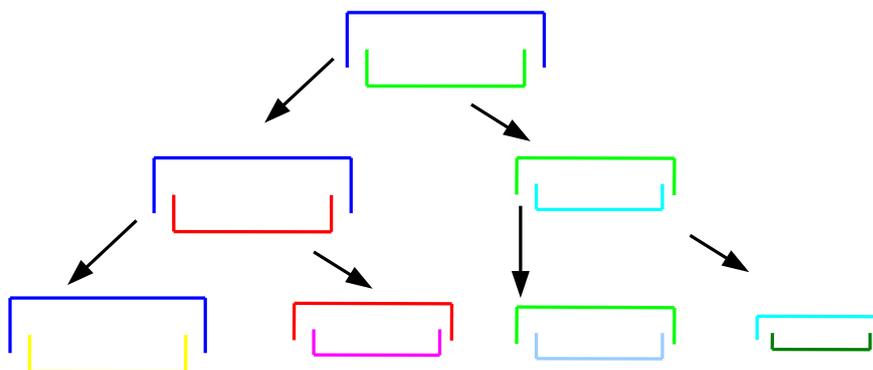


FIGURE 1.1 – Illustration de la diminution de taille des diatomées lors des reproductions végétatives successives. Chaque valve devient l'épivalve de la cellule fille.

des cellules filles devient critique, démarre une phase de reproduction sexuée qui permettra l'obtention d'un nouvel individu souche de taille « normale » (voir figure 1.2). La formation des gamètes est appelée auxosporulation, phase durant laquelle 4 noyaux haploïdes sont formés, parmi lesquels seuls un ou deux produiront finalement des gamètes. Le cycle de vie des diatomées peut comprendre des formes de résistance appelées hypnozoïdes (Edlund et Stoermer, 1997).

### 1.1.2 Evolution, phylogénie et classification

On estime que l'ancêtre commun des diatomées est apparu il y a environ 200 millions d'années (Medlin et al., 1997; Medlin, 2011). Ces estimations proviennent à la fois d'études de marqueurs moléculaires et des plus anciennes observations stratigraphiques, permettant à Julius et Theriot (2010) de considérer ces hypothèses fiables. Les diatomées semblent s'être particulièrement diversifiées à la suite de la crise du Permien-Trias grâce aux nombreuses niches laissées libres dans les océans, puis pendant le Jurassique (Kooistra et Medlin, 1996), colonisant divers milieux et substrats.

Les diatomées sont des algues brunes (embranchement des Chromophytes), caractérisées entre autre par la production de chlorophylle *c*. La place plus précise des diatomées dans la classification du vivant est assez variable selon les auteurs. Le terme « diatomées » est le plus souvent considéré comme un synonyme de la classe *Bacillariophyceae* ou de la super-classe *Bacillariophyta*. Il semble que le groupe des diatomées soit à placer quelque part entre la classe

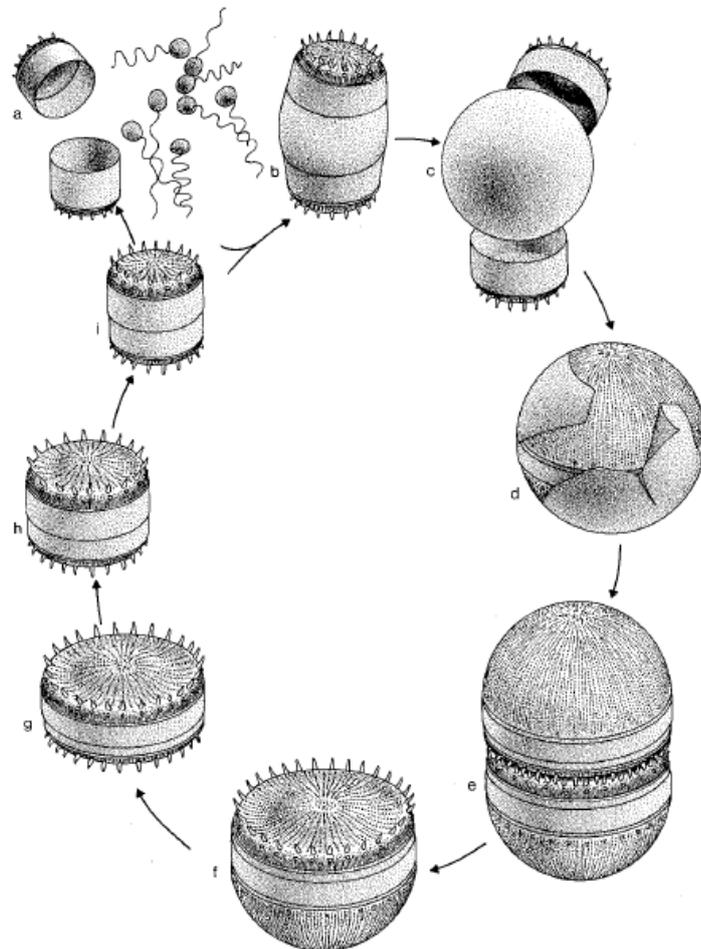


FIGURE 1.2 – Cycle biologique des diatomées, exemple d'une diatomée centrique du genre *Stephanodiscus*. a-f : phase de reproduction sexuée. g-i : phase de reproduction végétative classique. a : gamètes mobiles. b : oosphère. c : développement de l'auxospore. d : cassure de l'auxospore révélant la cellule initiale. e : la première division de la cellule initiale forme deux hypovalves normales. f : chacune des deux cellules ainsi formées présente une hypovalve et une valve dite initiale. g : la multiplication végétative permet le développement de cellules normales. h-i : la reproduction végétative entraîne une diminution de taille et aboutit à une nouvelle phase de reproduction sexuée. Tiré de Round et al. (1990).

et l'embranchement.

Au sein des diatomées, on distingue classiquement deux grands groupes (fig. 1.3) : les *centriques* et les *pennées*, principalement différenciées en termes de symétrie (radiale chez les premières, bilatérale chez les secondes), et de formes des gamètes (petites gamètes flagellées chez les premières, larges gamètes non

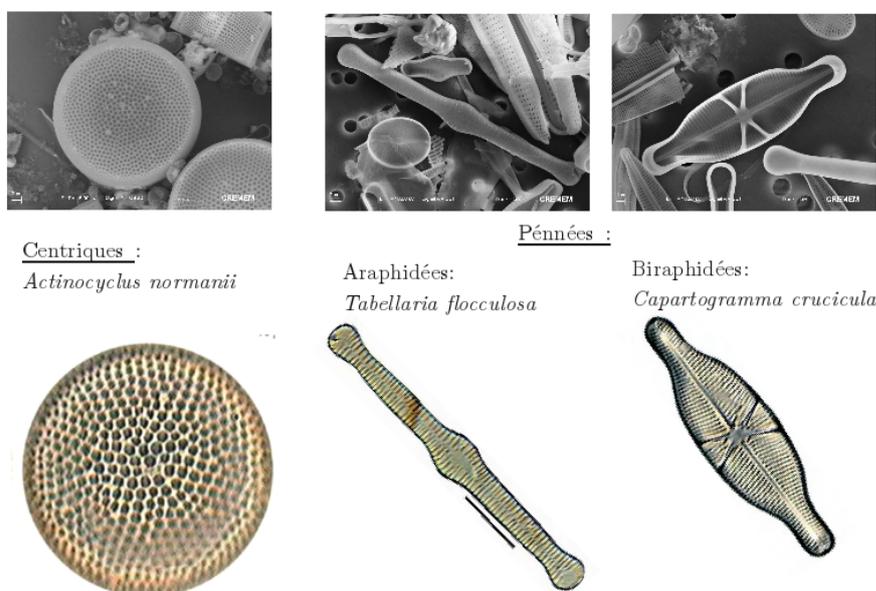


FIGURE 1.3 – Photos de diatomées en microscopie électronique à balayage (en haut) et en microscopie optique (en bas).

mobiles chez les secondes). Round et al. (1990) ont ainsi proposé une séparation des diatomées en 3 ordres : (i) les *Coscinodiscophyceae*, les centriques, (ii) les *Fragilariophyceae*, les pennées araphidées (qui ne présentent pas de raphé : fentes longitudinales sur leur squelette siliceux) et (iii) les *Bacillariophyceae*, diatomées pennées raphidées.

La classification des diatomées repose essentiellement sur des critères morphologiques, fait assez rare parmi les micro-organismes. Aujourd'hui celle-ci est largement bouleversée par l'emploi de techniques de plus en plus avancées, à la fois en microscopie et en biologie moléculaire (Mann et al., 2010). Une classification rigoureuse devrait être composée exclusivement de groupes monophylétiques, c'est à dire de groupes comprenant un ancêtre commun, tous ses descendants, et uniquement ses descendants.

Il ne semble pas que ce soit le cas pour nombre de classifications proposées pour les diatomées (Williams et Kociolek, 2007), car la recherche du sens évolutif des classifications semble être assez récente pour ces organismes (Julius et Theriot, 2010). De plus en plus d'études cherchent à conduire des analyses phylogénétiques grâce à des données de biologie moléculaire (ex : Bruder et Medlin, 2007; Medlin, 2010). Malheureusement, ces études ne montrent que rarement des résultats concordant avec la classification plus « classique » des diatomées (Cox, 2009). Il ne semble pas que les incertitudes quant à la classification puissent être résolues rapidement et par la simple application des méthodes moléculaires, car ces méthodes présentent elles aussi des difficultés et ne donnent pas de résultats incontestables (Medlin, 2010). De plus, la diversité des diatomées ne permet pas pour l'instant une analyse exhaustive des différents marqueurs moléculaires utilisables pour un nombre suffisant d'espèces.

Ainsi, si les noms d'espèces sont relativement stables, les genres, familles et ordres sont constamment remaniés et la mise à jour de base de données taxonomiques est extrêmement difficile. Pour cette raison, aucune de nos analyses ne s'est attachée à la diversité taxonomique des diatomées pour des niveaux supérieurs au genre.

### 1.1.3 Écologie, habitats et importance des diatomées

#### 1.1.3.1 Diversité et production

La diversité des diatomées est impressionnante : il n'existe pas d'estimation précise du nombre d'espèces, mais il semble que celui-ci soit de l'ordre de 200 000 espèces (Mann et Droop, 1996). Parmi celles-ci, 24 000 avaient en 2009 un nom validé Fourtanier et Kociolek (2009) et de très nombreuses espèces nouvelles sont décrites tous les ans.

Cette grande diversité s'accompagne d'une forte abondance dans la majorité des écosystèmes aquatiques, et d'une importance majeure dans les cycles bio-géo-chimiques à l'échelle planétaire. En 1995, on estimait que la part de la production primaire attribuable aux diatomées était d'environ 40% (Nelson et al., 1995). Cette estimation a pu évoluer depuis, mais il n'en reste pas moins que les diatomées participent à une part majeure de la productivité primaire mondiale, en particulier dans les zones eutrophes de l'océan, les mers froides des pôles et les zones côtières. Dans les écosystèmes aquatiques continentaux, elles sont aussi à la base des chaînes alimentaires et assurent un rôle clé pour l'écosystème.

#### 1.1.3.2 Une grande diversité de milieux colonisés

Les diatomées peuplent tous les types de milieux : océans, milieux aquatiques continentaux, air, sols. Dans les milieux aquatiques, on distingue les diatomées planctoniques (essentiellement des espèces centriques ou araphidées) qui se développent librement dans la colonne d'eau, et les diatomées périphtiques qui se développent fixées sur des substrats immergés (galets, macrophytes, etc.). Ce *périphyton* est classé en différents compartiments en fonction du substrat auquel il s'est fixé :

- *l'épipélon* et *l'endopélon* sont constitués d'espèces libres vivant dans les sédiments ou à leur surface
- *l'épipsammon* et *l'endopsammon* désigne les communautés vivant dans le sable ou à sa surface
- *l'épilithon* se forme à la surface de substrats durs inertes de types pierre-bloc-galets
- *l'épihyton* se forme à la surface des végétaux aquatiques

Dans le cadre de ce travail de thèse, nous nous intéresserons aux diatomées de l'épilithon, compartiment pris en compte en priorité dans l'évaluation de la qualité de l'eau par les indices diatomiques.

#### 1.1.3.3 Le biofilm épilithique

Les diatomées épilithiques forment un biofilm à la surface du substrat dur, rendu cohérent par sécrétion d'une matrice organique. Cette couche biologique comprend parfois d'autres organismes (algues filamenteuses, bactéries, etc.). Les

diatomées y présentent des formes de croissance et des modes de vie variés. Les cellules isolées peuvent être mobiles, attachées au bout d'une tige, ou encore collées au substrat par une substance mucilagineuse. D'autres forment des colonies en tubes, en rosettes, en étoiles, arborescentes.

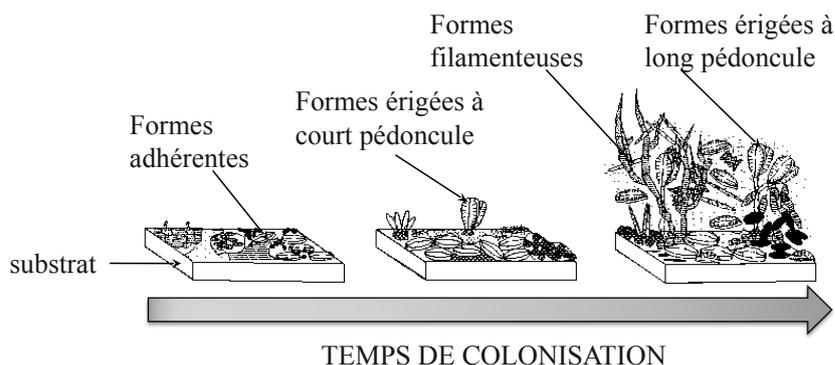


FIGURE 1.4 – Évolution de la structure tridimensionnelle des biofilms diatomiques dans le temps, à rapprocher de l'évolution temporelle des communautés de plantes supérieures (de la pelouse à la forêt). Tiré de Gold (2002), d'après un graphique original de Hoagland et al. (1982).

Le biofilm, du fait des différences de taille entre les espèces et de la forte diversité des formes de vie, a une structure irrégulière et complexe. Les biofilms peuvent donc montrer au moins autant de différences structurelles entre eux que si l'on comparait une pelouse et une forêt (fig. 1.4) L'épaisseur du biofilm et sa structure ont de fortes implications dans le développement des espèces. En effet, des études ont par exemple montré que des gradients verticaux de nutriments existaient en leur sein (Burkholder et al., 1990). De plus, la structure tridimensionnelle des biofilms évolue dans le temps, devenant de plus en plus épaisse (fig. 1.4)

Passy (2007) a défini trois « guildes » à partir des formes de vie et de croissance des diatomées :

1. *Profil haut* (High profile en anglais) : les espèces qui forment des tiges, ou des colonies leur permettant d'atteindre les couches supérieures du biofilm,
2. *Profil bas* (Low profile en anglais) : les espèces qui se maintiennent collées au substrat et vivent dans les couches inférieures du biofilm,
3. *Mobiles* (Motile en anglais) : les espèces qui peuvent se déplacer au sein du biofilm.

Cette classification simple a l'avantage de permettre une description de l'architecture tridimensionnelle du biofilm à partir de l'inventaire des espèces présentes sur un site.

## 1.2 Les diatomées en tant qu'indicateurs biologiques

### 1.2.1 Une forte sensibilité aux conditions du milieu

Les différentes espèces de diatomées présentent des préférences environnementales hétérogènes. De nombreuses études se sont attachées à caractériser les principaux facteurs environnementaux qui les distinguent. Ainsi, les diatomées se sont avérées particulièrement sensibles aux gradients liés à la balance ionique des eaux (alcalinité/acidité et conductivité; Van Dam et al., 1994; Hill et al., 2001; Potapova et Charles, 2003; Kovács et al., 2006; Tison et al., 2007), et aux concentrations de nutriments (Phosphates et nitrates; Pan et Stevenson, 1996; Winter et Duthie, 2000; Soininen et Niemela, 2002; Duong et al., 2006). Elles sont donc particulièrement sensibles aux facteurs de chimie locale des eaux. De nombreux autres facteurs, bien que moins souvent cités, semblent agir significativement sur la distribution des espèces : sans vouloir être exhaustif, on peut évoquer l'altitude (Potapova et Charles, 2002; Tison et al., 2005), la quantité d'oxygène dissous, l'intensité lumineuse (Lange et al., 2011) ou encore les substances toxiques (Morin et al., 2009).

La présence d'espèces dans un lieu donné est donc représentative des conditions environnementales du milieu, et les diatomées sont largement utilisées comme indicateurs de ces conditions. L'utilisation de caractéristiques biologiques, telles que la présence d'espèces, pour évaluer des critères environnementaux est appelée bioindication.

### 1.2.2 Un bon bioindicateur ?

De nombreux programmes de bioindication ont été mis en place depuis une trentaine d'années et les organismes utilisés sont très variés (par exemple les macroinvertébrés aquatiques pour évaluer la qualité des cours d'eau ou les mousses et lichens épiphytiques pour évaluer la pollution atmosphérique). Pour permettre une évaluation rigoureuse, les bioindicateurs doivent présenter un ensemble de caractéristiques, dont la plupart sont réunies chez les diatomées :

**Cosmopolitisme** Les bioindicateurs doivent avoir la plus large répartition géographique possible. Les diatomées sont présentes dans quasiment tous les milieux aquatiques et terrestres humides.

**Forte abondance** Cela permet un échantillonnage fiable, en nombre suffisant, d'individus de chacun des milieux que l'on souhaite évaluer. Les diatomées, en tant que micro-organismes ont la capacité de former des vastes populations sur des surfaces réduites.

**Immobilité** Les organismes sont d'autant plus indicatifs des conditions du milieu qu'ils sont fixes et doivent les subir sans avoir la possibilité d'émigrer. Il existe des diatomées mobiles, mais elles ne se déplacent qu'à une échelle extrêmement réduite. Malgré tout, les phénomènes d'immigration passive de diatomées vivantes ou mortes depuis l'amont d'une station peuvent perturber la description des conditions environnementales. Pour limiter la présence de cellules mortes dans les échantillons, ce sont les zones lotiques (à courant fort) des cours d'eau qui sont prélevées, car le dépôt de matières en suspension (y compris de diatomées mortes) y est moindre. Les

diatomées ont un cycle de vie particulièrement simple et court, qui se traduit par une réponse rapide aux variations environnementales du milieu. En revanche, elles présentent une forte variabilité saisonnière, et la plupart des programmes de bioindication ne concernent qu'une seule saison pour minimiser l'influence de cette variabilité. Notons que les organismes biologiques sans fluctuations saisonnières sont tout de même largement minoritaires.

**Échantillonnage et identification simples** Afin de permettre une mise en œuvre à large échelle ainsi qu'une utilisation « en routine », l'acquisition des données doit être relativement aisée. En ce qui concerne l'échantillonnage, les protocoles sont relativement simples pour les diatomées. De plus, la forte rémanence des frustules diatomiques n'oblige pas une identification immédiate. L'identification des espèces de diatomées nécessite un matériel conséquent, notamment des microscopes optiques performant, mais l'identification des espèces est relativement facilitée par leur grande diversité morphologique. Les erreurs induites par l'échantillonnage et l'identification des diatomées seront discutées dans le paragraphe 1.2.7, page 13.

**Capacité de survie en laboratoire** Cela permet de tester expérimentalement l'impact des pollutions. Les cultures multispécifiques de diatomées sont difficiles à mettre en place, et présentent souvent un nombre d'espèces inférieur à celui trouvé à l'état naturel. Malgré tout, de tels tests expérimentaux sont utilisés (Lange et al., 2011), notamment à Irstea pour mesurer l'effet des substances toxiques sur le fonctionnement biologique des diatomées (Morin et al., sous presse).

**Organismes clés de l'écosystème** Si l'organisme a un rôle important dans l'écosystème, son utilisation en tant que bioindicateur permettra de comprendre l'effet des éventuelles perturbations sur l'ensemble de celui-ci. Les diatomées sont des producteurs primaires, à la base de la chaîne trophique des cours d'eau, ce qui leur confère une importance toute particulière.

Les diatomées sont donc des bioindicateurs privilégiés, et sont ainsi largement utilisées dans différents programmes de surveillance, dont ceux imposés par l'actuelle Directive Cadre européenne sur l'Eau.

### 1.2.3 Échantillonnage

Les protocoles d'échantillonnage et de préparation des lames microscopiques pour les méthodes françaises de bioindication sont standardisés et définis dans la norme NFT 90-354 (AFNOR, 2007). Les échantillons sont collectés sur des galets grâce à des scalpels ou à des brosses à dents, en dehors de toute période hydrologique particulière (crue notamment). Les biofilms de plusieurs galets (au maximum 5, le tout devant représenter environ  $100\text{cm}^2$ ) provenant de zones à fort courant sont mélangés, afin d'obtenir des assemblages les plus représentatifs de l'ensemble de la station échantillonnée. Cet échantillonnage doit avoir lieu pendant la période d'étiage. Cette restriction temporelle permet d'éviter les variations naturelles saisonnières qui pourraient bruite l'évaluation, de plus la période d'étiage est une période critique pour les cours d'eau, car les éventuelles polluants y sont concentrés dans une faible quantité d'eau.



FIGURE 1.5 – Échantillonnage de biofilms diatomiques sur des galets, à l'aide d'un scalpel ou d'une brosse-à-dent.

#### 1.2.4 Préparation du matériel collecté et identification

Les échantillons subissent tout d'abord une digestion dans du peroxyde d'hydrogène porté à ébullition, qui a pour effet de détruire toute matière organique présente dans l'échantillon. Quelques gouttes d'acide chlorhydrique peuvent être ajoutées pour éliminer les résidus de carbonates de calcium. Ensuite l'échantillon est centrifugé à 4500 tr/min. Puis le culot formé est récupéré, puis dilué dans l'eau distillée, à plusieurs reprises (rinçage) ; enfin quelques gouttes de la suspension finale sont placées sur une lamelle. Lamelle et lamelle sont réunies grâce à l'application d'une goutte de résine hautement réfringente (Naphrax™,  $IR = 1,74$ ).

Les taxons sont identifiés grâce à l'observation des frustules au microscope photonique à un grossissement  $\times 1000$ . Pour aider à l'identification, Irstea publie un guide d'identification (Coste et Rosebery, 2011). 400 frustules sont comptées, ce chiffre permettant d'obtenir des richesses spécifiques et des abondances relatives stables entre plusieurs comptages sur une même lame (Prygiel et al., 2002).

#### 1.2.5 Calculs des indices

Il existe de très nombreuses méthodes pour évaluer la qualité de l'eau à partir des diatomées ou des autres organismes aquatiques dans le monde, et plus particulièrement en Europe. En effet, la Directive Cadre européenne sur l'eau (Parlement Européen, 2000), qui pose un cadre législatif dans lequel chaque état membre se doit d'évaluer la qualité de ses eaux de surface à partir d'indicateurs biologiques, a permis le recensement et le développement de méthodes anciennes et nouvelles. Ainsi, Birk et al. (2012) ont donné à leur article le titre « *Three hundred ways to assess Europe's surface waters : An almost complete overview of biological methods to implement the Water Framework Directive* ». Les 300 méthodes auxquelles les auteurs font référence ne concernent pas uniquement les diatomées, ou même les rivières, ni seulement des indices, mais cela montre la diversité des méthodologies. Parmi les indices, il existe trois grandes familles que nous allons détailler :

### 1.2.5.1 Les indices de moyennes pondérées

Ce type de méthodologie, appelée “*weighted averaging*” en anglais, est le plus fréquemment utilisé, notamment en France. Zelinka et Marvan (1961) furent les premiers à utiliser ce type de méthode, qui se base sur une formule mathématique générique :

$$ID = \frac{\sum_{j=1}^k A_j v_j i_j}{\sum_{j=1}^k A_j v_j}$$

où  $A_j$  est l’abondance du taxon sur le site à évaluer,  $v_j$  est sa valeur indicatrice,  $i_j$  est sa valeur de sensibilité, ou de valence écologique (d’autant plus forte que le taxon est présent dans des gammes faibles de variations de pollution) et  $k$  est le nombre de taxons présents sur le site.

Ainsi cette méthode implique une description préalable des préférences des espèces en termes de pollution. Ces valeurs peuvent être assignées selon des avis d’experts, ou par l’analyse numérique de bases de données.

La première mise en œuvre de ce type de méthode pour les diatomées des cours d’eau français fut l’Indice de Polluosensibilité Spécifique (IPS ; Coste dans Cemagref, 1982). Sa formule est la réplique exacte de celle de Zelinka et Marvan, dans laquelle  $v_j$  est une note allant de 1 à 5 (indice de polluosensibilité) décrivant le niveau de trophie/saprobie pour lequel l’espèce présente le plus d’affinité, et  $i_j$  est le « degré de sténoécie » de l’espèce, allant de 1 à 3. Ces valeurs sont établies à partir de l’ordination des espèces en cinq classes le long d’un gradient de pollution, et vérifiées par avis d’expert.

L’Indice Biologique Diatomées (IBD ; Lenoir et Coste, 1996 ; Prygiel et Coste, 2000 ; Coste et al., 2009) est basé sur le même principe, mais sa formule est plus complexe que la formule générique de Zelinka et Marvan :

$$F_j = \frac{\sum_{i=1}^n A_i V_i P_{ji}}{\sum_{i=1}^n A_i V_i}$$

$$IBD = \sum_{j=1}^7 F_j \times j$$

où :  $A_i$  est l’abondance du taxon  $i$ ,  $V_i$  son degré de sténoécie et  $P_{ji}$  est l’affinité de l’espèce  $i$  avec la classe  $j$ .

Cet indice requiert donc le calcul de l’affinité des taxons envers les sept classes prédéfinies de qualité de l’eau. La définition de ces probabilités de présence provient de descriptions de « profils écologiques » des taxons (Fig. 1.6), grâce au calcul des proportions de présence des taxons dans les différentes classes. Ces profils sont définis pour chacun des taxons inclus dans l’IBD (espèces, variétés, ou formes morphologiques). L’indice IBD est normalisé et défini dans la norme NFT 90-354 (AFNOR, 2007)

Les deux indices français présentés utilisent des valeurs spécifiques pour les espèces (voire les variétés de certaines espèces, ou même certaines formes morphologiques), mais il existe aussi des indices basés sur des valeurs attribuées au

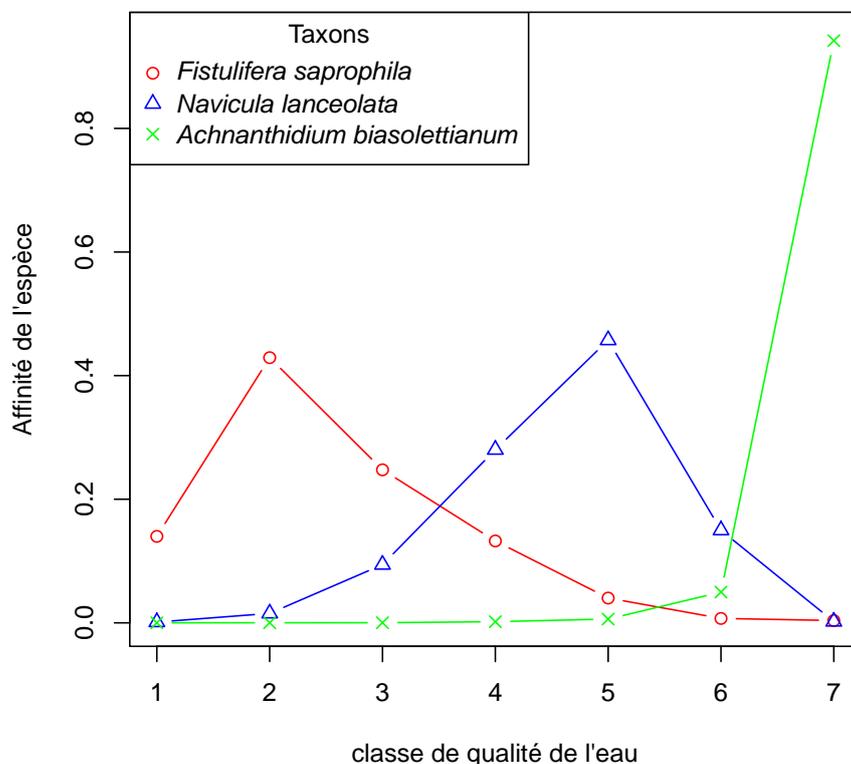


FIGURE 1.6 – Trois exemples de profils écologiques définis pour le calcul de l'IBD : *Achnantheidium biasolettianum* est une espèce sténoèce très sensible présente essentiellement dans des sites de très bonne qualité, *Navicula lanceolata* est une espèce euryèce ne présentant pas de sensibilité particulière et *Fistulifera saprophila* est une espèce résistante, principalement présente dans des sites pollués.

niveau taxonomique du genre, notamment au Royaume-Uni (Kelly et Whitton, 1995).

### 1.2.5.2 Métriques et indices multi-métriques

Ces méthodes consistent en la mesure de descripteurs généraux des assemblages de diatomées, et ne requièrent pas forcément l'identification de chacune des espèces présentes. Par exemple, la proportion d'individus morts mesurée à partir de matériel frais (Gillett et al., 2011), les indices de dominance ou de diversité spécifique sont de tels indices (Wang et al., 2005). En Amérique et en Asie, ces métriques sont très souvent utilisées et compilées dans des indices uniques par site appelés « Indices of Biotic Integrity » (IBIs; Hill et al., 2000,

2003; Zalack et al., 2010). Ces indices sont donc dit « multimétriques » car ils résultent de la compilation de plusieurs métriques calculées sur les assemblages.

### 1.2.5.3 Méthodes de classification et prédiction intégrées des assemblages

Ces méthodes ont principalement été développées pour les macroinvertébrés des cours d'eaux. Leur mise en œuvre la plus connue est le « River InVertebrates Prediction and Classification System » (RIVPACS; Moss et al., 1987), au point que les anglophones nomment « RIVPACS-like » les méthodes qui se basent sur les mêmes principes.

La méthodologie générique de ces indices se déroule en trois étapes :

1. Une classification statistique permet de définir plusieurs groupes selon la composition d'assemblages échantillonnés sur des sites de bonne qualité.
2. Des modèles, construits sur ces même sites, permettent, à partir de variables physico-chimiques mesurées, de déterminer quels groupes correspondraient aux sites que l'on veut évaluer s'ils n'avaient pas subi de perturbations anthropiques.
3. La composition de l'assemblage du site impacté est comparé avec la composition des sites du groupe de références de bonne qualité déterminé lors de l'étape précédente. Cette comparaison peut se faire de plusieurs manières, mais c'est le ratio  $O/E$  (Observed/Expected) qui est le plus souvent utilisé.  $O$  est, pour chaque taxon, une valeur de 0 ou de 1 selon que celui-ci est observé ou non dans l'échantillon à évaluer.  $E$  est, pour chaque taxon, la proportion de sites du groupe de référence de bonne qualité dans lequel le taxon a été observé. On calcule pour chaque taxon la différence entre ces deux valeurs, puis la moyenne de ces différences est utilisée pour attribuer directement un indice de qualité de l'eau variant de 0 à 1 au site impacté.

Ces méthodes ont été testées et mises en œuvre pour les diatomées (Chessman et al., 1999; Cao et al., 2007). Elles restent malgré tout largement minoritaires devant les deux autres présentées dans les paragraphes précédents. Par contre, le principe de classifier les communautés de référence (bonne qualité) et de comparer leurs indices ou leurs compositions avec ceux des assemblages à évaluer prédits comme appartenant au même groupe par un modèle est très répandu (par exemple : Tison et al., 2007; Grenier et al., 2010).

### 1.2.6 Bioindication diatomique sur le territoire français

Lors de l'application du calcul de l'IBD sur le territoire français, on constate que les notes présentent des différences selon les régions. Les indices dans les sites de montagnes sont souvent supérieurs à ceux calculés dans les plaines, ceux calculés dans les rares régions présentant des eaux acides n'ont pas la même variabilité que les autres. Ceci est lié aux fait que les altérations dues aux activités anthropiques agissent sur des paramètres qui peuvent varier naturellement. Il faut donc prendre en compte les conditions naturelles des différentes régions avant d'attribuer une note à la qualité de l'eau. Pour ce faire, des sites de bonne qualité sont échantillonnés régulièrement. Les sites sur lesquels ces échantillons sont prélevés sont appelés *sites de référence*. Les échantillons prélevés sur ces

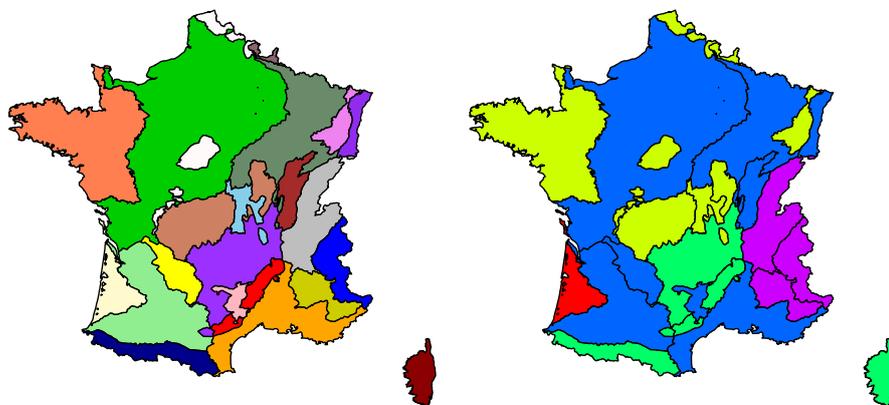


FIGURE 1.7 – Régions représentatives de la variabilité naturelle des conditions environnementales sur le territoire français. La carte de gauche représente les 22 HER telles que définies par Wasson et al. (2002), celle de droite représente les cinq régions suffisant à décrire les variations de conditions naturelles principales pour les diatomées, telles que définies par Tison et al. (2005).

sites vont permettre de décrire la variabilité naturelle des assemblages, et de comparer les indices des sites à évaluer et ceux des sites de référence.

Afin de choisir les sites de référence auxquels on compare les autres, il est important de délimiter des régions homogènes en termes de conditions naturelles. Wasson et al. (2002) a défini 22 Hydro-Éco-Régions (HER), qui décrivent les variations climatiques, géologiques et les gradients altitudinaux français (fig. 1.7). Afin de séparer uniquement les régions présentant des différences de conditions environnementales qui affectent réellement les diatomées, Tison et al. (2005) ont redéfini cinq régions pour lesquelles les assemblages naturels de référence sont homogènes (fig. 1.7).

Une fois les régions définies, la comparaison entre les assemblages diatomiques de références et les assemblages impactés se fait grâce au calcul d'*Ecological Quality Ratio*, simple ratio entre l'indice calculé et l'indice de référence de la région à laquelle appartient le site étudié.

où :

$$EQR_{IBD} = \frac{IBD - IBD_{min}}{IBD_{REF} - IBD_{min}}$$

$IBD$  est la note d'IBD calculée,  
 $IBD_{min}$  est la note minimale observée dans la région et  $IBD_{REF}$  est la médiane des notes d'IBD des sites de référence.

### 1.2.7 Sources de variabilité

Les nombreuses sources de variabilité lors de la détermination des assemblages de diatomées caractéristiques d'une station sont régulièrement étudiées (Prygiel et al., 2002; Roussel et al., 2010). Premièrement, la façon de prélever les

échantillons frais peut avoir une importance non-négligeable dans la composition finale de l'échantillon. En effet, on sait que les différentes zones d'une rivière ne présentent pas exactement les mêmes assemblages. La méthode de traitement des échantillons, le montage de la lame, les moyens d'observation et la méthode de comptage peuvent aussi induire de la variabilité. Malgré tout, ces paramètres sont relativement bien maîtrisés grâce à la normalisation des protocoles, Prygiel et al. (2002) ont estimé que cette variabilité correspondait à un point d'IBD, lorsque les protocoles étaient rigoureusement observés.

Une autre source de variabilité est la détermination des espèces. Les bases de données sont construites par de nombreuses personnes, qui peuvent parfois attribuer des noms différents au même taxon. Les noms d'espèces, bien que plus stables que les niveaux plus haut de la classification, sont en évolution constante et de nombreuses nouvelles espèces (rares le plus souvent) sont décrites chaque année. Une des difficultés majeures réside dans le fait que la résolution taxonomique demandée pour le calcul des indices peut varier selon les taxons, allant de l'espèce à la variété. La détermination des variétés demande souvent des efforts à la fois plus fins et subjectifs. Prygiel et al. (2002) ont montré qu'une part significative de la variabilité des indices pouvait être attribuée à la reconnaissance des différentes variétés de l'espèce *Achnantheidium minutissimum* (à l'époque *Achnanthes minutissima*).

Aussi, lors des traitements numériques n'impliquant pas le calcul d'indices, nous avons choisi de remplacer chaque variété et forme par leur espèce respective.

## 1.3 Biologie des communautés

### 1.3.1 Avant de commencer

L'écologie, elle-même souvent considérée comme une branche de la biologie, comprend de très nombreuses spécialités. La Biologie des communautés (ou Écologie des communautés) est une de ces disciplines. La définition classique de la communauté est : *l'ensemble des populations d'organismes vivant dans une aire ou un habitat donné* (Krebs, 1978). Parmi les différents niveaux d'organisation des structures vivantes (allant des molécules biochimiques à la biosphère), la biologie des communautés cherche à comprendre comment et pourquoi certains organismes vivent à proximité les uns des autres. Par rapport à la biologie des populations (ensemble des individus d'une même espèce vivant sur un endroit donné), qui s'intéresse au niveau d'organisation inférieur, cette science étudie les relations entrecroisées des différentes espèces avec leur environnement, mais aussi les interactions des espèces entre elles. Ainsi, il est du ressort de cette discipline de s'intéresser aux relations de prédation, aux réseaux trophiques, ou aux relations de compétition entre différentes espèces.

Cette définition très large est en fait rarement utilisée, car il est très difficile d'échantillonner l'ensemble des organismes vivants, ne serait-ce que sur une aire de 1 m<sup>2</sup>. On désigne donc classiquement par le terme communauté un assemblage d'organismes d'un groupe défini, par exemple les communautés de macroinvertébrés d'un lac, ou les communautés de micro-algues benthiques d'une rivière. Peut-on vraiment considérer que les diatomées benthiques forment des communautés ? L'expression « communauté de diatomées benthiques » est souvent utilisée, mais il est difficile de justifier la séparation des diatomées d'autres organismes vivant dans le même biofilm, telles que les algues filamenteuses. Il est certainement plus rigoureux de parler d'« assemblages de diatomées benthiques ». La biologie des communautés apporte malgré tout des éclairages applicables sur ces assemblages. En effet, si les diatomées ne constituent pas l'intégralité des communautés de micro-algues benthiques des cours d'eau, elles en sont une composante majeure. La seule étude des diatomées permet déjà d'aborder un grand nombre de processus et d'appréhender des relations complexes, entre les diatomées et avec l'environnement. De plus, Soininen et al. (2009) ont montré que les autres organismes constituant les communautés du biofilm présentaient des variations très proches de celles des communautés de diatomée. Ainsi les diatomées peuvent être considérées comme des représentants possibles de l'ensemble des communautés algales benthiques. C'est pourquoi nous utiliserons les méthodes et les théories développées sur des communautés complètes. Il faut tout de même garder à l'esprit que prédateurs (broueteurs) des diatomées et autres organismes des biofilms sont considérés au même titre que des facteurs environnementaux externes, dans ces travaux de thèse.

### 1.3.2 Super-organisme ou association d'organismes interdépendants ?

Deux visions extrêmes des communautés ont dominé la biologie des communautés au début du 20<sup>e</sup> siècle. La première considère que les communautés sont des « super-organismes » au sein desquels les espèces sont fortement dépendantes les unes des autres, du fait notamment de relations biotiques structurantes. Ce

fut la thèse défendue par Clements (1916). Dans l'entité organisée qu'est la communauté, les espèces n'ont alors plus de relations directes avec l'environnement, c'est la communauté qui porte cette relation et les espèces ne dépendent que de la communauté. La seconde est celle de Gleason (1926). Diamétralement opposée, elle considère que seules les réponses individuelles et indépendantes des espèces à l'environnement sont à l'origine d'associations désordonnées. Les communautés ne sont alors que des représentations arbitraires, dont il est difficile de trouver les limites dans l'espace. Les défenseurs de cette théorie mettent en avant les relations des traits des espèces avec l'environnement et leurs dynamiques spatio-temporelles.

À l'heure actuelle, les questions ne se posent plus tout à fait dans ces termes, et nous avons hérité d'une vision plutôt individualistique (proche de celle de Gleason) des relations des espèces à l'environnement. Ce débat n'est tout de même pas tout à fait résolu, et ces deux visions ont l'avantage de permettre le placement des associations observées sur une échelle allant de communautés fortement structurées par les relations biotiques et des facteurs internes à des réponses totalement individualisables des espèces. C'est la balance entre différents processus qui agissent sur les communautés qui va nous permettre de placer les communautés sur cette échelle. D'un côté, une forte importance des réponses environnementales directes des espèces aux paramètres environnementaux nous mènerait à considérer les communautés comme « Gleasoniennes ». De l'autre, une forte influence des paramètres internes aux communautés (compétition, facilitation, richesse, diversité) sur chacune des espèces nous pousserait à les définir comme « Clementsiennes ».

### 1.3.3 Les niches écologiques

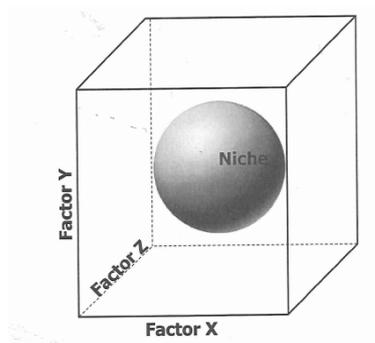


FIGURE 1.8 – Représentation de la niche écologique comme un hypervolume (Hutchinson, 1957, ici à trois dimensions;). Le cube représente la totalité des ressources ou des facteurs environnementaux disponibles. La sphère à l'intérieur de ce cube représente les besoins de l'espèce vis-à-vis de ces mêmes facteurs environnementaux. Figure tirée de Chase et Leibold (2003).

La niche écologique est un des concepts centraux de l'écologie. Le terme « niche écologique » fut défini par Grinnell en 1917 comme l'ensemble des besoins environnementaux d'une espèce (incluant l'absence des ennemis naturels de l'espèce). La redéfinition plus tardive de ce concept par Hutchinson (1957) lui

a donné son sens le plus connu. Hutchinson a défini la niche écologique comme un hypervolume contenu dans un espace à  $n$  dimensions,  $n$  étant le nombre de facteurs pouvant affecter la distribution de l'espèce (fig 1.8).

### 1.3.3.1 Filtres environnementaux et traits biologiques

Le modèle sur lequel se basent les relations entre les conditions environnementales et les espèces est un modèle de filtres successifs. Ce sont ces filtres successifs qui vont définir les limites de l'hyper-volume d'Hutchinson. Une espèce, pour pouvoir s'installer dans un habitat donné, doit avoir la capacité de survivre et de se reproduire malgré chacune des contraintes de cet habitat. Afin de s'adapter aux conditions de leur environnement, les espèces ont développé des caractéristiques biologiques particulières que l'on appelle les *traits*. Les traits biologiques importants vont concerner l'utilisation des ressources, la reproduction ou la capacité de survie au stress et aux perturbations induites par l'environnement.

On distingue souvent les traits biologiques (mode de reproduction, taille des individus etc.) des traits écologiques qui correspondent aux préférences environnementales. Les traits écologiques sont en fait liés à un ensemble de traits biologiques qui vont définir les besoins de l'espèce en termes de conditions environnementales.

### 1.3.3.2 Niches fondamentales et réalisées : l'importance des relations biotiques

Les lieux et les conditions environnementales nécessaires à l'installation possible d'une espèce ne présente pas toujours sa présence, notamment du fait de l'exclusion compétitive, ou des relations de prédation. La niche écologique comprend donc deux composantes : la *niche fondamentale* est la part de la niche qui serait occupée par l'espèce si seuls ses besoins étaient déterminants (notamment en l'absence d'autres espèces), et la *niche réalisée* est la part, souvent plus faible, de la niche que l'espèce occupe réellement. De cette distinction entre la niche fondamentale et la niche réalisée a découlé une grande activité de recherche ayant pour objet l'explication du nombre limité d'espèces vivant dans un lieu donné, sachant que de très nombreuses espèces présentent des niches fondamentales appropriées. Il semble que le phénomène qui explique cette relative pauvreté des systèmes locaux est le fait que des espèces présentant des caractéristiques trop proches ont tendance à s'exclure mutuellement par compétition (MacArthur et Levins, 1967).

La compétition est donc une composante essentielle des niches écologiques, et un facteur important dans la répartition des espèces. Il existe aussi des relations biotiques positives entre les espèces, qui peuvent avoir un effet sur les différences entre niches fondamentales et réalisées. Celles-ci ont été largement sous-estimées dans les théories écologiques (Bruno et al., 2003). Parmi les relations biotiques positives, on distingue deux phénomènes :

- La *facilitation* est une relation positive entre deux espèces, d'une façon directe ou indirecte,
- La *complémentarité de niche* correspond au fait que les communautés qui présentent des individus avec une grande diversité de traits, auront tendance à exploiter les ressources de façon plus efficace, et à augmenter leur

productivité totale. Dans de telles communautés variées, la compétition intraspécifique deviendra alors plus forte que la compétition interspécifique.

L'environnement va avoir tendance à filtrer des traits particuliers, son effet sera donc de faire converger les traits biologiques au sein d'une communauté. À l'inverse, les espèces qui ont des traits trop proches seront soumises à des relations de compétition forte, menant souvent à la disparition d'espèces (exclusion compétitive). De plus le phénomène de complémentarité de niche aura tendance à favoriser les espèces présentant des traits différents. Ainsi, l'analyse de la convergence et de la divergence des traits biologiques permet d'étudier l'importance relative des relations biotiques et de l'effet de l'environnement (Watkins et Wilson, 2003; Herzog et Kessler, 2006; Freschet et al., 2011)

### 1.3.3.3 L'effet des espèces sur l'environnement

À la suite de la publication de la *théorie neutraliste* par Hubbell (2001, ; voir paragraphe 1.3.4), qui remet en cause l'importance du concept de niche, Chase et Leibold (2003) ont publié un livre dans lequel ils redéfinissent ce concept. Pour ces auteurs, la niche écologique comprend deux phénomènes indissociables : le premier concerne l'effet de l'environnement sur la distribution des espèces, le second concerne l'effet des espèces sur l'environnement. En effet, les espèces vivantes ont la capacité de modifier l'environnement. Par exemple, pour les diatomées, les espèces de la guilde « high profile », en se développant au dessus de celles de la guilde « low profile », vont avoir pour effet de réduire l'accès à la lumière et aux nutriments de ces dernières. Selon Chase et Leibold, cette définition permet de mieux appréhender la complexité des relations des espèces entre elles et avec l'environnement. Par contre, cette définition est plus difficile à modéliser, car elle implique un système complexe où les effets de l'environnement, les rétro-effets des espèces et les effets des espèces les unes sur les autres vont induire des relations non-linéaires, difficilement prédictibles.

### 1.3.4 Biogéographie et théorie neutraliste

La biogéographie a pour but de décrire la répartition des espèces à très large échelle. Alors que la théorie des niches se focalise sur la sélection des espèces du pool régional (ensemble des espèces présentes dans une région) vers les communautés locales, la biogéographie va particulièrement s'intéresser aux déterminants du pool régional lui-même. MacArthur et Wilson (1967) ont largement contribué au développement de la biogéographie, notamment en travaillant sur la richesse spécifique des îles (fig. 1.9). Dans les îles, ce sont les probabilités d'immigration et d'extinction qui vont déterminer en premier lieu le pool d'espèces. Ces facteurs sont donc aléatoires, contrairement aux facteurs considérés dans la théorie des niches, qui impliquent une relation mécanistique entre les espèces et l'environnement.

Hubbell (2001) s'est inspiré des travaux de MacArthur et Wilson pour développer une théorie plus large : la *théorie neutraliste*. Hubbell y réfute l'importance du concept de niche dans les variations à grande échelle de la biodiversité. En effet, si l'on considère que les espèces sont toutes équivalentes en termes de traits biologiques (et donc de niches), on s'aperçoit que la seule influence des

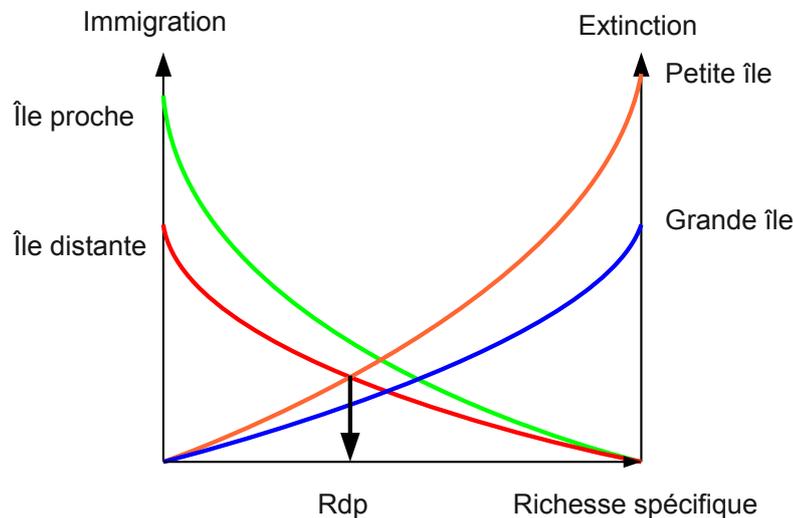


FIGURE 1.9 – Graphique théorique décrivant la richesse spécifique des îles en fonction des phénomènes d'immigration et d'extinction, eux même dépendants de la taille et de l'isolement des îles. La richesse spécifique de l'île se trouve à l'intersection des courbes : ainsi  $Rdp$  représente la richesse spécifique d'une île petite et distante. Inspiré de MacArthur et Wilson (1967).

phénomènes aléatoires de dispersion, d'extinction et de spéciation peut expliquer une grande part de la répartition des espèces.

### 1.3.5 Les méta-communautés

Le terme « méta-communauté » désigne un ensemble de communautés reliées entre elles par des phénomènes de dispersion des espèces entre les communautés (Leibold et al., 2004). La dynamique de ces méta-communautés est describable selon quatre conceptions principales qui décrivent les relations entre les communautés locales (Leibold et al., 2004; Holyoak et al., 2005) :

**Dynamique neutre** Cette conception dérive de la théorie neutraliste de Hubbell (2001). La composition des communautés locales dépendra alors des taux d'extinction, des taux de dispersion, et sur une plus large échelle de temps, des taux de spéciation des espèces.

**Dynamique de patches** Les communautés sont chacune sur des aires limitées appelées « patches », qui dans cette conception ne diffèrent que par leur positionnement géographique et leur taille. Comme dans la dynamique neutre, les phénomènes de dispersion et d'extinction vont jouer un rôle majeur, mais aussi l'exclusion compétitive entre les espèces. On considère que les traits biologiques des espèces dérivent d'un compromis entre les capacités de dispersion et la compétitivité. Ainsi, certaines espèces seront

pionnières (les premières à s'installer sur un patch vide, ou après une perturbation) car elles auront de fortes capacités de dispersion. D'autres seront compétitives, et seront capables d'exclure les espèces pionnières pour s'installer sur les patches. Les communautés montrent donc une évolution temporelle caractéristique à la suite d'une perturbation.

**Dynamique de tri des espèces (species sorting)** Dans cette conception, ce sont les niches qui seront particulièrement importantes, et l'on ne considérera pas les phénomènes aléatoires de dispersion et d'extinction (Chase et Leibold, 2003). Ce sont donc les conditions biotiques et abiotiques locales qui vont déterminer la composition en espèces des communautés, en favorisant des espèces différentes selon les patches du fait de l'hétérogénéité environnementale.

**Effet de masse** Cette conception intègre les dynamiques de patches et les dynamiques de tri des espèces. Chaque espèce se développe particulièrement dans les sites dont l'environnement est favorable et forme des populations sources qui lui permettent de disperser vers d'autres sites. Ainsi les espèces dont les préférences environnementales correspondent aux conditions les plus fréquentes auront la possibilité d'alimenter régulièrement des patches en individus même si ceux-ci ne sont pas favorables.

### 1.3.6 Nestedness

La nestedness (qui se traduirait en français par emboîtement) est un patron de distribution d'espèces, elle correspond au fait que les communautés les plus pauvres sont des sous-ensembles des communautés les plus riches (Patterson et Atmar, 1986; Brualdi et Sanderson, 1999; Ulrich et Gotelli, 2007; Ulrich et al., 2009). Ainsi, dans un tableau de données représentant la présence et l'absence des espèces dans différentes communautés « emboîtées », si l'on trie les espèces selon leur nombre d'occurrences et les communautés selon leur richesse spécifique, la grande majorité des présences se retrouve dans le coin supérieur gauche du tableau (fig. 1.10). Ce phénomène a été très largement décrit dans les communautés biologiques, notamment chez les micro-organismes aquatiques (Soininen, 2008; Schneck et al., 2011, Heino et al., sous presse).

Différentes explications ont été formulées pour ce patron de distribution des présences d'espèces. Elles impliquent principalement une forte influence des dynamiques de colonisation et d'extinction des espèces (Ulrich et al., 2009).

### 1.3.7 Les règles d'assemblage

L'introduction du livre intitulé « Ecological assembly rules : Perspectives, advances, retreats » est volontairement provocante. Keddy et Weiher (1999) y dépeignent l'écologie des communautés comme une science n'ayant que rarement des buts affirmés, et qui se perd trop souvent dans des descriptions de petits écosystèmes locaux, qui bien que passionnantes, ne permettent aucune réelle généralisation. Ils proposent donc une méthode générale et des buts bien définis pour définir des *règles d'assemblage*. Ces règles d'assemblage doivent permettre de poser des principes simples d'associations entre les espèces, ayant trait aux facteurs de niches, à la compétition ou aux facteurs de dispersion entre les espèces. Selon McGill et al. (2006), il est nécessaire d'énoncer de telles règles afin

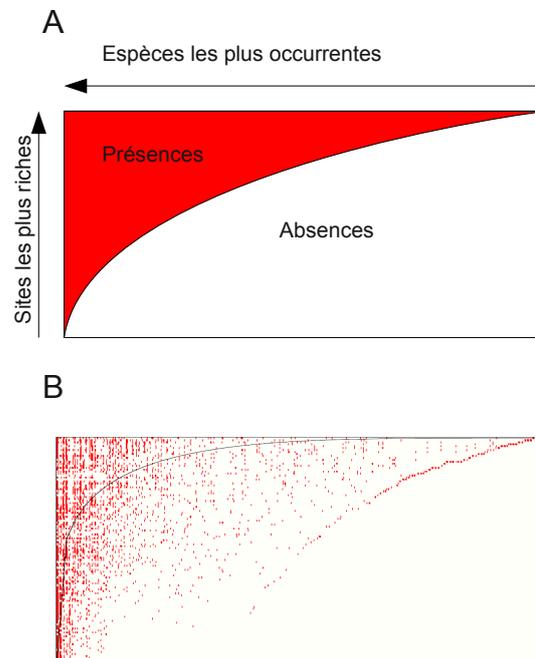


FIGURE 1.10 – Nestedness : quand les communautés les plus pauvres sont des sous-ensembles des communautés les plus riches (communautés emboîtées) et que l'on trie les tableaux de présence des espèces, toutes les présences sont placées dans le coin supérieur gauche du tableau. A : distribution théorique. B : distribution des présences de diatomées dans les sites de référence à l'échelle de la France.

d'améliorer la prédictabilité des compositions ou des propriétés fonctionnelles des communautés. En effet, la modélisation de la distribution des espèces est essentiellement réalisée grâce à la modélisation des niches des espèces, ainsi ajouter des règles simples comprenant les facteurs de relations biotiques et de dispersion pourrait s'avérer très utile, notamment pour la bioindication.

### 1.3.8 Spécificités des micro-organismes

La distribution des micro-organismes a été largement sous-étudiée par rapport à celle des macro-organismes, tels que les plantes supérieures. L'hypothèse de Baas-Becking (1934) stipule que les micro-organismes, du fait de leurs densités locales de population très élevées et de leurs très fortes capacités de dispersion passive, ne seraient soumis qu'aux pressions de l'environnement. La vision défendue est donc la suivante : « *Les micro-organismes sont partout et l'environnement sélectionne* ». Cette thèse a été défendue plus récemment par Finlay (2002) et Fenchel et Finlay (2004).

Pourtant, de nombreuses études ont montré que les micro-organismes (en

particulier les diatomées) étaient soumis à des contraintes similaires aux macro-organismes (Astorga et al., 2011), liées aux relations biotiques (McCormick, 1996), aux contraintes historiques sur le pool régional d'espèces (évolution, spéciation et dispersion inter-continentale : Vyverman et al., 2007; Vanormelingen et al., 2008), ou aux limites de dispersion des espèces à l'échelle régionale (Martiny et al., 2006; Soininen et al., 2007; Verleyen et al., 2009; Heino et Soininen, 2010). Ainsi, l'étude des principaux mécanismes agissant sur la répartition des diatomées reste importante, et peut être menée en utilisant les méthodes et les modèles développés pour les macro-organismes.

### 1.3.9 La bioindication diatomique : un modèle simplifié

La bioindication se base sur l'étude des relations entre les organismes et l'environnement, et ne considère que rarement les facteurs tels que les limites de dispersion et les relations biotiques entre les espèces (fig. 1.11). Sa pratique sous-tend donc souvent un modèle simplifié de biologie des communautés, que nous allons détailler selon les méthodes décrites dans le paragraphe 1.2.5.

**Les indices basés sur les moyennes pondérées** Ces indices, tirés d'une relation directe et indépendante de chaque espèce avec son environnement, sous-tendent un modèle de communauté purement gleasonien, dans lequel l'importance des relations biotiques est totalement négligée, ainsi que celle des limites de dispersion.

**L'utilisation de métriques** Les métriques utilisées pour la bioindication s'attachent souvent à décrire des facteurs concernant la communauté dans son ensemble, comme le pourcentage d'espèces mortes ou la richesse spécifique. La vision des communautés induite par l'utilisation de ces métriques est donc plus en adéquation avec la vision de Clements, mais l'impact des relations biotiques ou des limites de dispersion n'est pas directement prise en compte.

**La prise en compte d'une stratification régionale** La prise en compte des régions (HER ou éco-régions diatomiques) permet de se placer à une échelle plus fine, et donc de limiter l'effet des limites de dispersion en diminuant la distance entre les sites qui sont comparés. Mais les indices utilisés en France sont tout-de-même ajustés sur les données à l'échelle nationale, ainsi les valeurs spécifiques des différentes espèces ne prennent pas en compte les limites de dispersion ou les relations biotiques.

Ainsi, relations biotiques et limites de dispersion sont négligées lors des pratiques de bioindication. Si ces phénomènes s'avéraient importants pour les diatomées, ils pourraient être des facteurs de bruit dans l'estimation que l'on fait de la qualité écologique des eaux. La prise en compte de ces facteurs pourrait alors permettre d'améliorer significativement la qualité de ces estimations.

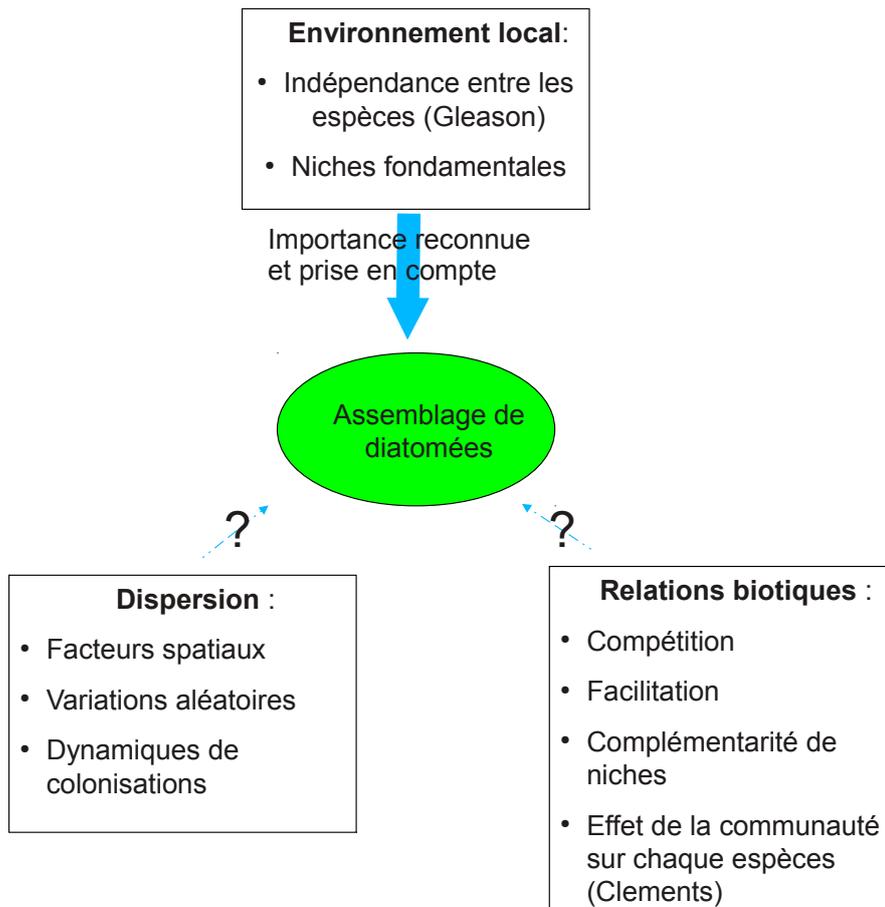


FIGURE 1.11 – Bioindication diatomique et biologie des communautés, quels sont les processus pris en compte ?

## 1.4 Écologie numérique

*Cette partie consiste en une présentation globale des méthodes d'écologie numérique. Des précisions méthodologiques seront apportées dans chacun des chapitres suivants.*

### 1.4.1 Les données biologiques

Les statistiques de base utilisées en écologie et en biologie des communautés sont les descripteurs bruts des communautés observées. Elles sont composées des observations d'individus biologiques échantillonnés. Trois types de données sont majoritairement rencontrés :

- abondances absolues : il s'agit du nombre total d'individus échantillonnés sur l'ensemble de l'espace échantillonné, répartis dans les différents taxons. Du fait du processus d'échantillonnage des diatomées, on a rarement accès à de telles données.
- abondances relatives : il s'agit de la proportion d'individus de chaque taxon que l'on a échantillonné. Pour les diatomées, ce sont souvent les données les plus complètes auxquelles on peut avoir accès.
- présence/absence : il s'agit des listes de taxons échantillonnés. Ces listes sont souvent présentées sous forme de tableaux dans lesquelles les lignes représentent les unités d'échantillonnage, les colonnes représentent les taxons, et l'intérieur du tableau contient des "0" quand l'espèce est absente et des "1" quand l'espèce est présente.

Ces données biologiques sont le plus souvent présentées sous forme de tableaux, on parle alors de « matrices biologiques » ou de « tableaux biologiques ».

### 1.4.2 Les mesures de la biodiversité

La notion de diversité a pour but de représenter la variabilité des structures biologiques. Elle tient un rôle central en biologie des communautés. En effet elle permet d'appréhender l'ensemble des organismes vivant aux échelles considérées, et permet de formuler des questions clés de cette discipline : comment se répartit la diversité biologique, et quels sont les facteurs qui expliquent sa répartition et ses variations ?

Habituellement, la diversité biologique est séparée en trois éléments, les diversités  $\alpha$ ,  $\beta$  et  $\gamma$  dont chacune se traduit par des mesures numériques différentes (Whittaker, 1972).

#### 1.4.2.1 Diversité $\alpha$

Elle représente la diversité biologique en un lieu donné, à l'échelle de la communauté. Sa mesure la plus simple est la richesse spécifique (le nombre d'espèces identifiées en un lieu donné), mais celle-ci est souvent considérée incomplète car elle ne prend pas en compte la dominance entre les espèces. En effet, une même

mesure de richesse spécifique peut cacher des situations très différentes selon que la majorité des individus appartiennent à la même espèce ou qu'ils sont répartis équitablement entre les différentes espèces. Cette notion de diversité  $\alpha$  est très proche de celle de l'entropie physique : elle représente le désordre (le contraire de la pureté). D'ailleurs elle est très souvent mesurée grâce à une métrique issue de la théorie de l'information appelée « entropie de Shannon » (Shannon, 1948). La mesure appliquée à l'écologie qui en découle est appelé indice de Shannon-Wiener ( $H$ ) :

$$H = - \sum_{i=1}^R p_i \log p_i$$

où :  
 $p_i$  est l'abondance relative (proportion) de l'espèce  $i$   
 $R$  est la richesse spécifique.

L'indice de Shannon-Wiener est donc dépendant de la richesse spécifique et de la dominance des espèces (plus il y aura de dominance de la part d'un petit nombre d'espèces, plus l'indice sera faible). L'indice  $J$  de Pielou fournit une mesure plus exclusive de cette notion de dominance (en anglais on parle d'evenness : régularité, équitabilité). Il correspond au ratio de l'indice de Shannon-Wiener  $H$  sur la valeur maximale  $H_{max}$  que celui-ci aurait pu prendre pour une même richesse spécifique :

$$H_{max} = - \sum_{i=1}^R \frac{1}{R} \log \frac{1}{R} = \log R \quad J = \frac{H}{\log R}$$

#### 1.4.2.2 Diversité $\beta$

La diversité  $\beta$  est la mesure de la différence entre deux communautés. C'est une composante importante de la diversité biologique, car elle permet de savoir si une communauté locale est particulière, ou homogène avec les communautés environnantes. Il existe des communautés très peu riches (faible diversité  $\alpha$ ), mais qui peuvent par exemple justifier des programmes de conservation de la biodiversité parce qu'elles présentent de fortes différences avec les autres (forte diversité  $\beta$ ). Numériquement, cette notion de dissimilarité est transposable à celle de distance. Pour les données d'abondance absolue, la distance euclidienne représente la distance minimale entre deux points, dans un espace où chaque espèce représente une dimension.

$$D_{eucl}(A, B) = \sqrt{\sum_{i=1}^R (A_i - B_i)^2}$$

où  $A_i$  et  $B_i$  sont respectivement les abondances de l'espèce  $i$  dans les communautés  $A$  et  $B$

La distance euclidienne n'est pas particulièrement adaptée aux mesures de dissimilarités entre des communautés ; en écologie des communautés, on l'utilise surtout pour mesurer les différences entre les conditions environnementales (paramètres physico-chimiques). La mesure de distance de Manhattan semble plus adaptée pour les dissimilarités entre communautés car elle accorde le même

poids à chaque espèce. Il s'agit de la somme des différences d'abondances de chaque espèce entre les deux communautés.

$$D_{manh}(A, B) = \sum_{i=1}^R |A_i - B_i|$$

Enfin, la distance de Bray-Curtis (Bray et Curtis, 1957) est souvent considérée comme la mieux adaptée aux mesures de dissimilarité entre communautés biologiques. Elle correspond à la distance de Manhattan, mais prend en compte l'abondance totale des deux communautés.

$$D_{Bray}(A, B) = \frac{\sum_{i=1}^R |A_i - B_i|}{\sum_{i=1}^R A_i + B_i}$$

Ainsi, les distances de Bray-Curtis et de Manhattan sont équivalentes lorsqu'appliquées à des données exprimées en abondance relative.

#### 1.4.2.3 Diversité $\gamma$

La diversité  $\gamma$  correspond à la diversité sur un ensemble d'habitats, ou à la diversité régionale. Elle se mesure généralement de la même manière que la diversité  $\alpha$ . Elle est moins utilisée que les diversités  $\alpha$  et  $\beta$ , auxquelles elle est souvent comparée pour étudier l'importance relative des facteurs de petite échelle sur la diversité  $\alpha$  locale.

### 1.4.3 Les bio-statistiques

#### 1.4.3.1 Les tests statistiques

Les tests statistiques permettent d'analyser la variabilité des données lorsqu'il est nécessaire de statuer sur d'éventuelles différences entre des groupes d'échantillons. Ces tests se basent sur quelques « lois de distribution statistique », qui ont pour but de décrire la variabilité générale des jeux de données. Ces lois de distribution décrivent la probabilité d'obtenir des valeurs particulières, selon des formules mathématiques.

Les lois statistiques reposent sur l'observation empirique de la variabilité naturelle. La loi la plus fréquemment utilisée et observée dans la nature est la loi normale (fig. 1.12). Cette loi est adaptée à beaucoup de phénomènes mesurables dont les valeurs se répartissent autour d'une valeur moyenne (par exemple, les répartitions de taille des poissons ou les erreurs de mesure d'un appareil).

Les tests statistiques se déroulent généralement en quatre étapes :

1. La définition d'une hypothèse nulle et d'une hypothèse alternative. Par exemple, si l'on veut tester si les valeurs obtenues dans deux échantillons sont différentes, l'hypothèse nulle est l'égalité moyenne (ou égalité des moyennes) des deux échantillons, l'hypothèse alternative est l'inégalité des deux moyennes.
2. Le calcul d'une statistique sur l'échantillon que l'on veut tester. Les tests statistiques portent sur des métriques particulières, dont on connaît la distribution théorique sous l'hypothèse nulle.

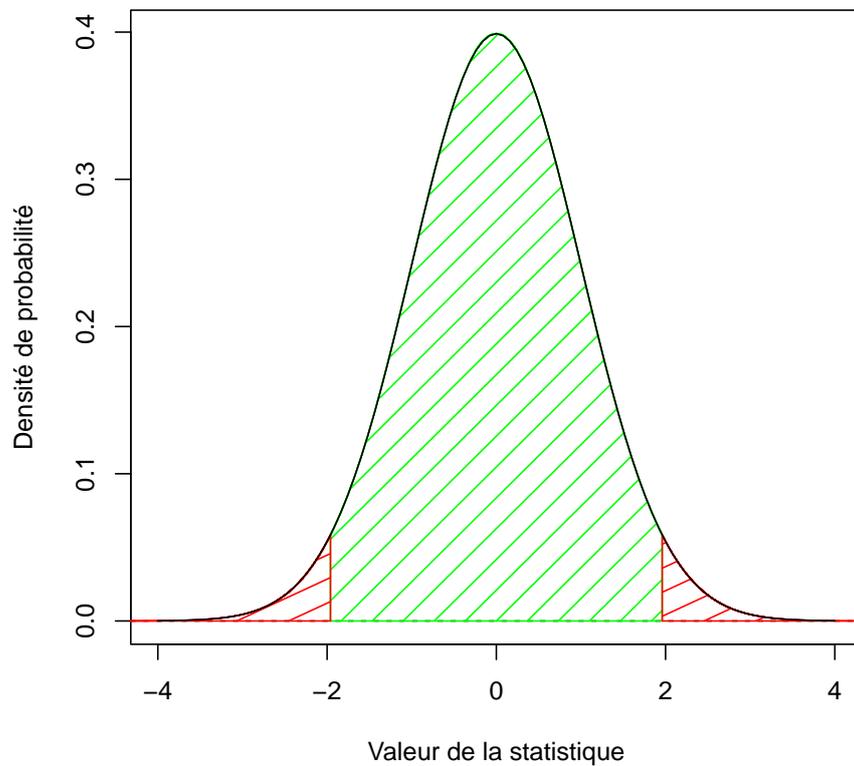


FIGURE 1.12 – Les lois de distribution statistique : exemple de la loi normale centrée (moyenne de 0) réduite (variance de 1). L'aire sous la courbe représente la probabilité d'obtenir une valeur dans un intervalle. Ainsi, l'aire hachurée en vert représente une probabilité de 0,95, et les deux aires hachurées en rouge représentent une probabilité de 0,05.

3. La comparaison entre la statistique calculée et la distribution théorique de la statistique sous l'hypothèse nulle. Sous l'hypothèse nulle, la statistique de test calculée dans l'étape précédente suit une loi statistique connue, ainsi, on peut mesurer la probabilité théorique d'obtenir la valeur calculée si l'hypothèse nulle est vraie. Cette probabilité est appelée valeur  $P$  ( $P$ -value en anglais)
4. Le choix entre l'hypothèse nulle et l'hypothèse alternative. Ce choix se fait grâce à l'observation de la valeur  $P$  : si celle-ci est très faible (généralement on utilise le seuil  $< 5\%$  : fig 1.12), on considère que c'est la probabilité que l'hypothèse nulle soit vraie qui est très faible, on choisit donc l'hypothèse alternative. En revanche, si cette valeur est plus forte, cela veut dire que la probabilité d'obtenir une valeur de la statistique de test comparable à celle calculée est forte, même si l'hypothèse nulle est vraie, il n'y a donc pas de raison objective de rejeter l'hypothèse nulle.

**Modèles nuls** Les variables mesurées en biologie des communautés sont rarement distribuées selon les lois de distribution statistique classiques. Il est parfois difficile d'appliquer ces tests sur les données biologiques, car la distribution théorique des variables mesurées (étape 3) est inconnue. Ainsi, il est souvent avantageux de modéliser la réponse des variables sous l'hypothèse nulle. On parle alors de « modèles nuls ». Par exemple, le mélange des valeurs d'une matrice de présence/absence d'espèces va permettre de modéliser des « pseudo-communautés » dans lesquelles ni les relations biotiques, ni les facteurs environnementaux n'ont contraint la présence des espèces (Gotelli, 2000). L'application répétée de ce mélange fournira donc des pseudo-communautés dont la comparaison avec les communautés réelles pourra permettre de déterminer si ces dernières ont subi les contraintes de l'environnement ou celles des relations biotiques. La principale difficulté liée à l'application de ces modèles nuls est de s'assurer que les différences entre les pseudo-communautés et les communautés réelles sont bien liées au phénomène que l'on veut tester (Gotelli et Ulrich, 2012).

#### 1.4.3.2 Les méthodes multivariées de représentation des données

Ces méthodes sont très largement utilisées pour l'analyse des données de biologie des communautés. Elles sont souvent appelées méthodes d'ordination. Parmi les plus utilisées, on compte l'Analyse en Composantes Principales (ACP), l'analyse des correspondances (CA), la « Detrended Correspondance Analysis » (DCA) ou le « Non-metric Multi-dimensional scaling » (NMDS). Sans rentrer dans le détail de chacune de ces méthodes, on peut expliquer leur fonctionnement général.

Deux étapes sont importantes et récurrentes dans ces différentes techniques :

1. *Calcul des relations* entre les lignes (échantillons) et/ou les colonnes (espèces) du tableau de données biologiques. Pour l'ACP, la mesure de ces relations se fera en calculant les corrélations entre toutes les lignes et les colonnes du tableaux. Il en résultera deux matrices carrées, contenant autant de lignes et de colonnes que d'éléments comparés, représentant la variabilité et les covariations contenues dans le tableau. Pour les méthodes CA et DCA, c'est le  $\chi^2$  qui est utilisé, une mesure statistique de la dépendance entre des variables. Enfin, la mesure utilisée dans la NMDS pourra

être une des mesures de dissimilarité présentées page 25.

2. *Représentation linéaire* des relations calculées dans l'étape précédente. Les matrices carrées calculées représentent les covariations dans un espace comprenant autant de dimensions qu'elles ont de lignes (ou de colonnes). Grâce à des calculs complexes issus de l'algèbre linéaire, il est possible de redéfinir cet espace de façon à ce qu'une partie majeure de l'information soit représentée sur un nombre réduit de dimensions. D'une façon imagée, cela revient à tourner autour d'un objet de façon à trouver l'angle de vue qui permet d'en distinguer le maximum de détails.

Ces techniques d'ordination des données ont donc un désavantage majeur : en utilisant des techniques d'algèbre linéaire, elle sont aptes à représenter les relations existantes uniquement de façon linéaire. Or, en écologie, les relations complexes des espèces entre elles et avec l'environnement résultent bien souvent de variations non-linéaires. De plus, en dehors des mesures de diversité  $\beta$  utilisées par la NMDS, le calcul des relations, dans la première étape, utilise des mesures souvent inadaptées aux communautés biologiques.

#### 1.4.4 Les méthodes alternatives aux statistiques classiques

Les statistiques classiques présentent de nombreuses techniques permettant le test d'hypothèses, la représentation de données et l'analyse de relations entre les éléments d'une base de données. Malheureusement, ces méthodes présentent de nombreux défauts : la prépondérance des notions de variance (plutôt que de diversité) et des relations linéaires rend souvent leur utilisation inappropriée pour l'analyse des communautés biologiques.

De nombreuses techniques, issues de la recherche en intelligence artificielle, présentent des alternatives intéressantes aux statistiques classiques. En effet, ces techniques peuvent permettre la représentation de relations non-linéaires ou l'utilisation de notions plus proches de la biologie des communautés. Plusieurs de ces techniques ont été développées et utilisées dans ce travail de thèse. Dans la première partie, l'utilisation de modèles nuls complexes nous a permis de tester des phénomènes chiffrés dont nous ne connaissons pas la distribution, dans la partie deux le développement et l'utilisation de méthodes issues de l'intelligence artificielle nous a permis de décrire les assemblages de diatomées français, dans le but d'améliorer les pratiques de bioindication.

## 1.5 Données

### 1.5.1 Base de données « historique » d'Irstea

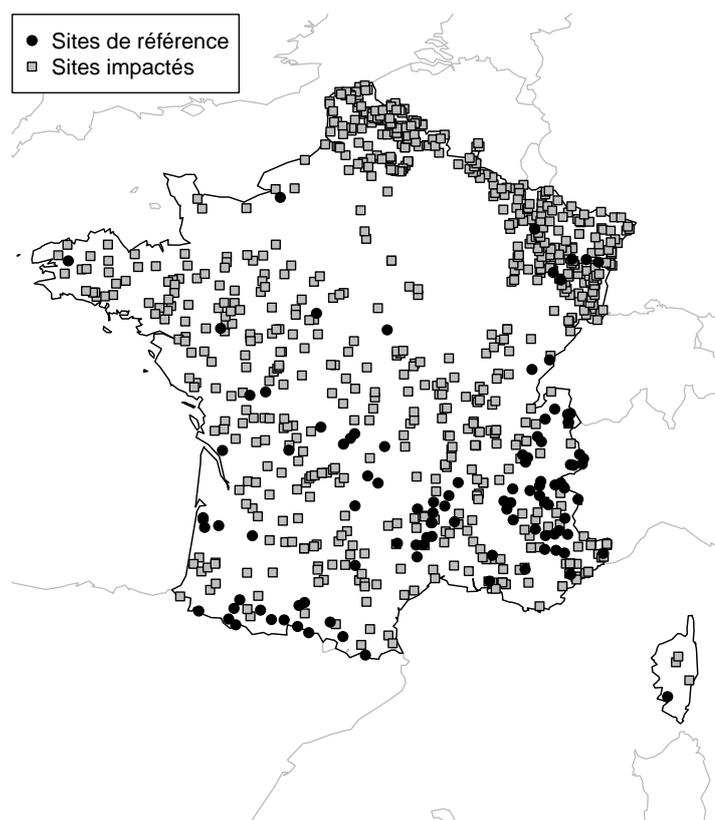


FIGURE 1.13 – Localisation des sites de la base « historique »

Cette base de données a déjà été utilisée dans le cadre de thèses et d'études ultérieures, et est décrite plus précisément dans les publications s'y rapportant (Tison et al., 2005, 2007).

Les données biologiques sont sous forme d'abondances relatives. Elles comprennent 926 taxons (variétés, formes ou espèces qui correspondent à 847 espèces selon les nouvelles dénominations) provenant de 836 échantillons dont 115 ont été récoltés sur des sites considérés comme des sites de référence pour la écologique qualité de l'eau).

Ces données sont géo-référencées (fig. 1.13). Des mesures physico-chimiques ont été réalisées sur ces mêmes sites et comprennent : l'altitude (m), la distance à la source (m), la pente (%), le pH, la conductivité à 20°C ( $\mu\text{S}/\text{cm}$ ), dissolved oxygen (mg/l), l'azote total (mg/l), les phosphates (mg/l), le calcium (mg/l), les carbonates (mg/l), la température ( $^{\circ}\text{C}$ ), les matières en suspension

(mg/l), l'ammonium(mg/l), les nitrates (mg/l), les nitrites (mg/l), et l'alcalinité (meq/l). Une moyenne des paramètres qui varient dans le temps a été calculée sur une période de deux mois (le mois du prélèvement diatomique et le mois précédent).

Les déterminations floristiques de cette base ont été réalisés par deux personnes (Luc Ector, Centre de Recherche Gabriel Lippman, et Michel Coste, Irstea). La variabilité des données attribuable aux différences entre les opérateurs est donc réduite. Aussi, cette base de données ayant servi à de nombreuses études ultérieures, notamment les travaux de Michel Coste et Juliette Tison-Rosebery, les gradients et particularités contenus dans les données sont donc connus et cette base est tout-à-fait adaptée à des développement méthodologiques. En revanche, elle présente un déséquilibre fort d'échantillonnage selon les régions, en particulier pour les sites de référence, dont une grande partie se trouve dans les montagnes des Alpes ou des Pyrénées.

### 1.5.2 Base de données « récente »

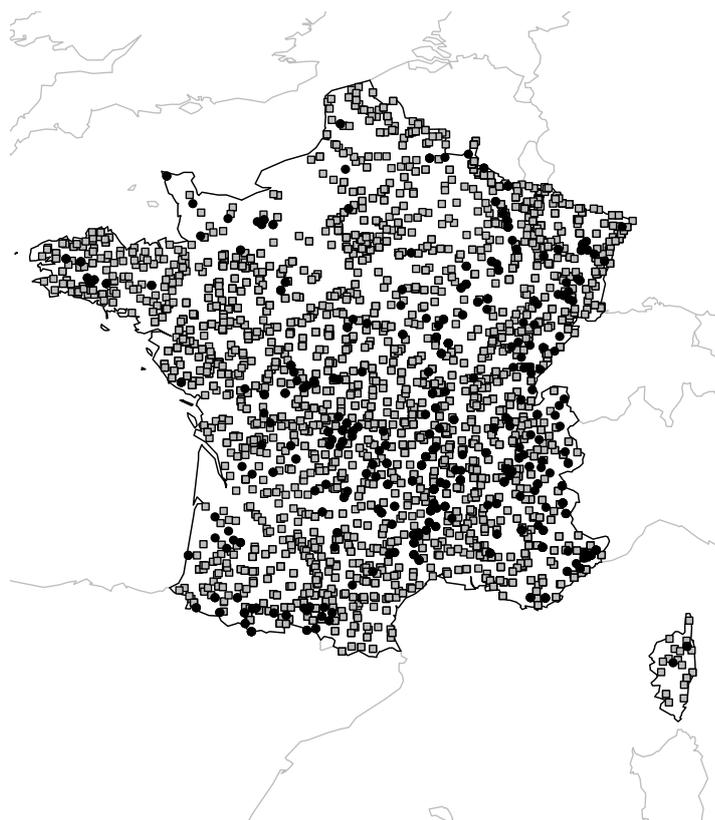


FIGURE 1.14 – Localisation des sites de la base de données récente

Cette base de données est disponible à Irstea depuis 2011. Elle provient des réseaux de surveillance de la qualité des eaux, qui ont été fortement dynamisés par l'application de la Directive Cadre sur l'Eau. Elle résulte de la compilation de données brutes provenant de 21 organismes différents (agences régionales ou départementales de l'environnement), ce qui implique une forte hétérogénéité. Un premier travail de tri et de mise en forme a résulté en la suppression de près de 40% de ces données brutes, afin d'obtenir un jeu plus homogène.

Les données biologiques comprennent l'abondance relative de 1272 espèces. 5027 échantillons ont été récoltés sur 2782 sites entre les années 1977 et 2007, dont 600 sur des sites de référence (fig 1.14).

Les données physico-chimiques récoltées ont été moyennées sur une période de 60 jours (45 jours avant l'échantillon de diatomées à 15 jours après cet échantillonnage). Elles comprennent la température de l'eau, le pH, l'altitude, l'oxygène dissous, les matières en suspension, la demande biologique en oxygène sur cinq jours (DBO5), l'alcalinité, la conductivité, les concentrations en nitrate, nitrite, ammonium, azote total, orthophosphate, phosphore total et carbone organique.

Le principal défaut de cette base de données est le nombre d'opérateurs ayant participé à récolter les informations. En revanche, les très nombreuses données que cette base contient sont bien réparties sur le territoire français, et permettent une description détaillée et équilibrée de la répartition de la biodiversité diatomique à l'échelle nationale.

## Bibliographie

- AFNOR (2007). Qualité de l'eau. détermination de l'indice biologique diatomées (ibd) - norme nf t90-354 -.
- Astorga, A., J. Oksanen, M. Luoto, J. Soininen, R. Virtanen, et T. Muotka (2011). Distance decay of similarity in freshwater communities : do macro- and microorganisms follow the same rules? *Global Ecology and Biogeography*.
- Baars, J. (1983). Autoecological investigations on freshwater diatoms. 1. generation times of some species. *Archiv für Hydrobiologie - Supplement* 67, 11–18.
- Baas-Becking, L. (1934). *Geobiologie of inleiding tot de milieukunde*, Volume 18. WP Van Stockum & Zoon.
- Birk, S., W. Bonne, A. Borja, S. Brucet, A. Courrat, S. Poikane, A. Solimini, W. van de Bund, N. Zampoukas, et D. Hering (2012). Three hundred ways to assess europe's surface waters : An almost complete overview of biological methods to implement the water framework directive. *Ecological Indicators* 18(0), 31 – 41.
- Bray, J. R. et J. T. Curtis (1957). An ordination of the upland forest communities of southern wisconsin. *Ecological Monographs* 34, 77–87.
- Brualdi, R. A. et J. G. Sanderson (1999). Nested species subsets, gaps, and discrepancy. *Oecologia* 119(2), 256–264.
- Bruder, K. et L. Medlin (2007). Molecular assessment of phylogenetic relationships in selected species/genera in the naviculoid diatoms (bacillariophyta). i. the genus placoneis. *Nova Hedwigia* 85(3-4), 331–352.
- Bruno, J., J. Stachowicz, et M. Bertness (2003). Inclusion of facilitation into ecological theory. *Trends in Ecology and Evolution* 18(3), 119–125.
- Burkholder, J. M., R. G. Wetzel, et K. L. Klomparens (1990). Direct comparison of phosphate uptake by adnate and loosely attached microalgae within an intact biofilm matrix. *Applied and Environmental Microbiology* 56(9), 2882–2890.
- Cao, Y., C. P. Hawkins, J. Olson, et M. A. Kosterman (2007). Modeling natural environmental gradients improves the accuracy and precision of diatom-based indicators. *Journal of the North American Benthological Society* 26(3), 566–585.
- Cemagref (1982). Etude des méthodes biologiques d'appréciation quantitative de la qualité des eaux. Technical report, Rapport Q.E. Lyon A.F. - Bassin Rhône-Méditerranée-Corse.
- Chase, J. M. et M. A. Leibold (2003). *Ecological niches : linking classical and contemporary approaches*. Chicago, USA : University of Chicago press.
- Chessman, B., I. Gowns, J. Currey, et N. Plunkett-Cole (1999). Predicting diatom communities at the genus level for the rapid biological assessment of rivers. *Freshwater Biology* 41(2), 317–331.

- Clements, F. E. (1916). *Plant Succession : An Analysis of the Development of Vegetation*. Number 242. Washington, DC : Carnegie institution of Washington.
- Coste, M., S. Boutry, J. Tison-Rosebery, et F. Delmas (2009). Improvements of the biological diatom index (BDI) : Description and efficiency of the new version (BDI-2006)mo. *Ecological Indicators* 9(4), 621–650.
- Coste, M. et J. Rosebery (2011). Guide iconographique pour la mise en oeuvre de l'indice biologique diatomée 2007.
- Cox, E. (2009). What's in a name? - diatom classification should reflect systematic relationships. *Acta Botanica Croatica* 68(2), 443–454.
- Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. John Murray.
- Duong, T., M. Coste, A. Feurtet-Mazel, D. Dang, C. Gold, Y. Park, et A. Boudou (2006). Impact of urban pollution from the hanoi area on benthic diatom communities collected from the red, nhue and tolích rivers (vietnam). *Hydrobiologia* 563, 201–216.
- Edlund, M. B. et E. F. Stoermer (1997). Ecological, evolutionary, and systematic significance of diatom lifehistories. *Journal of Phycology* 33(6), 897–918.
- Fenchel, T. et B. Finlay (2004). The ubiquity of small species : Patterns of local and global diversity. *BioScience* 54(8), 777–784.
- Finlay, B. (2002). Global dispersal of free-living microbial eukaryote species. *Science* 296(5570), 1061–1063.
- Fourtanier, E. et J. Kociolek (2009). Catalogue of diatom names. Occasional papers of the California Academy of Sciences.
- Freschet, G. T., A. T. Dias, D. D. Ackerly, R. Aerts, P. M. van Bodegom, W. K. Cornwell, M. Dong, H. Kurokawa, G. Liu, V. G. Onipchenko, J. C. Ordoñez, D. A. Peltzer, S. J. Richardson, I. I. Shidakov, N. A. Soudzilovskaia, J. Tao, et J. H. C. Cornelissen (2011). Global to community scale differences in the prevalence of convergent over divergent leaf trait distributions in plant assemblages. *Global Ecology and Biogeography* 20(5), 755–765.
- Gillett, N., Y. Pan, K. Manoylov, et R. Stevenson (2011). The role of live diatoms in bioassessment : A large-scale study of western us streams. *Hydrobiologia* 665(1), 79–92.
- Gleason, H. A. (1926). The individualistic concept of the plant association. *Bulletin of the Torrey Botanical Club* 53(1), pp. 7–26.
- Gold, C. (2002). *Etude des effets de la pollution métallique (Cd/Zn) sur la structure des communautés de diatomées périphytiques des cours d'eau. Approches expérimentales in situ et en laboratoire*. Doctorat en écotoxicologie n°2487.
- Gotelli, N. J. (2000). Null model analysis of species co-occurrence patterns. *Ecology* 81(9), 2606–2621.

- Gotelli, N. J. et W. Ulrich (2012). Statistical challenges in null model analysis. *Oikos* 121(2), 171–180.
- Grenier, M., S. Lek, M. Rodríguez, A. Rousseau, et S. Campeau (2010). Algae-based biomonitoring : Predicting diatom reference communities in unpolluted streams using classification trees, random forests, and artificial neural networks. *Water Quality Research Journal of Canada* 45(4), 413–425.
- Grinnell, J. (1917). The niche-relationships of the california thrasher. *The Auk* 34(4), 427–433.
- Heino, J., M. Grönroos, J. Soininen, R. Virtanen, et T. Muotka. Context dependency and metacommunity structuring in boreal headwater streams. Sous presse.
- Heino, J. et J. Soininen (2010). Are common species sufficient in describing turnover in aquatic metacommunities along environmental and spatial gradients? *Limnology and Oceanography* 55(6), 2397–2402.
- Hellebust, J. et J. Lewin (1977). Heterotrophic nutrition. In D. Werner (Ed.), *The Biology of Diatoms. Bot. Monogr.* 13., pp. 169–197. Oxford. : Blackwell Sci. Publ.
- Herzog, S. K. et M. Kessler (2006). Local vs. regional control on species richness : a new approach to test for competitive exclusion at the community level. *Global Ecology and Biogeography* 15(2), 163–172.
- Hill, B., A. Herlihy, P. Kaufmann, S. DeCelles, et M. Vander Borgh (2003). Assessment of streams of the eastern united states using a periphyton index of biotic integrity. *Ecological Indicators* 2(4), 325–338.
- Hill, B., A. Herlihy, P. Kaufmann, R. Stevenson, F. McCormick, et C. Burch Johnson (2000). Use of periphyton assemblage data as an index of biotic integrity. *Journal of the North American Benthological Society* 19(1), 50–67.
- Hill, B., R. Stevenson, Y. Pan, A. Herlihy, P. Kaufmann, et C. Johnson (2001). Comparison of correlations between environmental characteristics and stream diatom assemblages characterized at genus and species levels. *Journal of the North American Benthological Society* 20(2), 299–310.
- Hoagland, K., S. Roemer, et J. Rosowski (1982). Colonization and community structure of 2 periphyton assemblages, with emphasis on the diatoms (bacillariophyceae). *American Journal of Botany* 69(2), 188–213.
- Holyoak, M., M. Leibold, et R. Holt (2005). *Metacommunities : Spatial Dynamics And Ecological Communities*. University of Chicago Press.
- Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton, USA : Princeton University Press.
- Hutchinson, G. E. (1957). Concluding remarks. In *Cold Spring Harbor Symposia on Quantitative Biology*, Volume 22, pp. 415–427.

- Julius, M. L. et E. C. Theriot (2010). The diatoms : a primer. In J. P. Smol et E. F. Stoermer (Eds.), *The diatoms : applications for the environmental and earth science*, Chapter 2, pp. 8–22. Cambridge University Press.
- Keddy, P. et E. Weiher (1999). Introduction : The scope and goals of research on assembly rules. In E. Weiher et P. Keddy (Eds.), *Ecological assembly rules : perspectives, advances, retreats*, Chapter Introduction, pp. 1–20. Cambridge, GBR : Cambridge university press.
- Kelly, M. et B. Whitton (1995). The trophic diatom index : a new index for monitoring eutrophication in rivers. *Journal of Applied Phycology* 7, 433–444.
- Kooistra, W. et L. K. Medlin (1996). Evolution of the diatoms (bacillariophyta). reconstruction of their age from small subunit rna coding regions and the fossil record. *Molecular Phylogenetics and Evolution* 6(3), 391–407.
- Kovács, C., M. Kahlert, et J. Padisák (2006). Benthic diatom communities along ph and tp gradients in hungarian and swedish streams. *Journal of Applied Phycology* 18, 105–117.
- Krebs, C. (1978). *Ecology : the experimental analysis of distribution and abundance*. Harper and Row.
- Lange, K., A. Liess, J. Piggott, C. Townsend, et C. Matthaei (2011). Light, nutrients and grazing interact to determine stream diatom community composition and functional group structure. *Freshwater Biology* 56(2), 264–278.
- Leibold, M. A., M. Holyoak, N. Mouquet, P. Amarasekare, J. M. Chase, M. F. Hoopes, R. D. Holt, J. B. Shurin, R. Law, D. Tilman, M. Loreau, et A. Gonzalez (2004). The metacommunity concept : A framework for multi-scale community ecology. *Ecology Letters* 7(7), 601–613.
- Lenoir, A. et M. Coste (1996). Development of a practical diatom index of overall water quality applicable to the french national water board network. In B. A. Whitton et E. Rott (Eds.), *Use of Algae for monitoring rivers II*, pp. 29–43. Innsbruck Austria 17-19 Sept. 95 : Studia Student. G.m.b.H.
- MacArthur, R. et R. Levins (1967). The limiting similarity, convergence, and divergence of coexisting species. *The American Naturalist* 101(921), pp. 377–385.
- MacArthur, R. et E. Wilson (1967). *The theory of Island biogeography*. New Jersey, Etats-Unis : Princeton University Press.
- Mann, D. G. et S. J. M. Droop (1996). Biodiversity, biogeography and conservation of diatoms. *Hydrobiologia* 336(1-3), 19–32.
- Mann, D. G., D. W. Williams, J. P. Kociolek, et E. J. Cox (2010). Discovering diatom species : is a long history of disagreements about species-level taxonomy now at an end ? towards a comprehensive diatom classification and phylogeny (bacillariophyta) morphogenetic information and the selection of taxonomic characters for raphid diatom systematics. *Plant Ecology and Evolution* 143, 251–264.

- Martiny, J., J. C. R. Bohannan, B.J.M. and Brown, J. Fuhrman, J. Green, M. Horner-Devine, M. Kane, J. Krumins, C. Kuske, P. Morin, S. Naeem, L. Ø vreås, A.-L. Reysenbach, V. Smith, et J. Staley (2006). Microbial biogeography : Putting microorganisms on the map. *Nature Reviews Microbiology* 4(2), 102–112.
- McCormick, P. V. (1996). Resource competition and species coexistence in freshwater benthic algal assemblages. In *Algal Ecology : Freshwater Benthic Ecosystems*. Academic press.
- McGill, B., B. Enquist, E. Weiher, et M. Westoby (2006). Rebuilding community ecology from functional traits. *Trends in Ecology and Evolution* 21(4), 178–185.
- Medlin, K., W. Kooistra, R. Gersonde, P. Sims, et U. Wellbrock (1997). Is the origin of the diatoms related to the end-permian mass extinction? *Nova Hedwigia* 65(1-4), 1–11.
- Medlin, L. (2011). The permian-triassic mass extinction forces the radiation of the modern marine phytoplankton. *Phycologia* 50(6), 684–693.
- Medlin, L. K. (2010). Pursuit of a natural classification of diatoms : An incorrect comparison of published data. *European Journal of Phycology* 45(2), 155 – 166.
- Morin, S., M. Bottin, N. Mazzella, F. Macary, F. Delmas, P. Winterton, et M. Coste (2009). Linking diatom community structure to pesticide input as evaluated through a spatial contamination potential (phytopixal) : A case study in the neste river system (south-west france). *Aquatic Toxicology* 94(1), 28–39.
- Morin, S., S. Pesce, S. Kim-Tiam, X. Libert, M. Coquery, et N. Mazzella. Use of polar organic chemical integrative samplers to assess the effects of chronic pesticide exposure on biofilms. *Ecotoxicology*, 1–11.
- Moss, D., M. T. Furse, J. F. Wright, et P. D. Armitage (1987). The prediction of the macro-invertebrate fauna of unpolluted running-water sites in great britain using environmental data. *Freshwater Biology* 17(1), 41–52.
- Nelson, D., P. Treguer, M. Brzezinski, A. Leynaert, et B. Queguiner (1995). Production and dissolution of biogenic silica in the ocean : revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Global Biogeochemical Cycle* 9(3), 359–372.
- Pan, Y. et R. Stevenson (1996). Gradient analysis of diatom assemblages in western kentucky wetlands. *Journal of Phycology* 32(2), 222–232.
- Parlement Européen (2000). Directive 2000/60/CE du parlement européen et du conseil du 23 octobre 2000 établissant un cadre pour la politique communautaire dans le domaine de l'eau, J.O.C.E. du 22.12.2000, L327/1-L327/72. Technical report.
- Passy, S. I. (2007). Diatom ecological guilds display distinct and predictable behavior along nutrient and disturbance gradients in running waters. *Aquatic Botany* 86(2), 171–178.

- Patterson, B. D. et W. Atmar (1986). Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biological Journal of the Linnean Society* 28(1-2), 65–82.
- Potapova, M. et D. F. Charles (2003). Distribution of benthic diatoms in u.s. rivers in relation to conductivity and ionic composition. *Freshwater Biology* 48(8), 1311–1328.
- Potapova, M. G. et D. F. Charles (2002). Benthic diatoms in usa rivers : distributions along spatial and environmental gradients. *Journal of Biogeography* 29(2), 167–187.
- Prygiel, J., P. Carpentier, S. Almeida, M. Coste, J. C. Druart, L. Ector, D. Guillard, M. A. Honoré, R. Iserentant, P. Ledeganck, C. Lalanne-Cassou, C. Lesniak, I. Mercier, P. Moncaut, M. Nazart, N. Nouchet, F. Peres, V. Peeters, F. Rimet, A. Rumeau, S. Sabater, F. Straub, M. Torrisi, L. Tudesque, B. Van de Vijver, H. Vidal, J. Vizinnet, et N. Zydek (2002). Determination of the biological diatom index *ibdnft90* – 354 Results of an intercomparison exercise. *Journal of Applied Phycology* 14(1), 27–39.
- Prygiel, J. et M. Coste (2000). *Guide méthodologique pour la mise en œuvre de l'indice biologique diatomées NF T90-354. Agences de l'Eau, MATE, Cemagref Bordeaux. 134 p. + 1 cédérom.*
- Round, F., R. Crawford, et D. Mann (1990). *The Diatoms. Biology & morphology of the genera.* Cambridge Univ.Press Ed.
- Roussel, F., J. Tison-Rosebery, et M. Coste (2010). Etude de la variabilité inter et intra opérateur de la note d'indice biologique diatomée 2007 : Résultats des exercices d'intercomparaison. Technical report, Convention ONEMA-Cemagref / Action 14.
- Schneck, F., A. Schwarzbald, et A. Melo (2011). Substrate roughness affects stream benthic algal diversity, assemblage composition, and nestedness. *Journal of the North American Benthological Society* 30(4), 1049–1056.
- Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal* 27(3), 379–423.
- Soininen, J. (2008). The ecological characteristics of idiosyncratic and nested diatoms. *Protist* 159(1), 65–72.
- Soininen, J., R. McDonald, et H. Hillebrand (2007). The distance decay of similarity in ecological communities. *Ecography* 30(1), 3–12.
- Soininen, J. et P. Niemela (2002). Inferring the phosphorus levels of rivers from benthic diatoms using weighted averaging. *Archiv Fur Hydrobiologie* 154(1), 1–18.
- Soininen, J., R. Paavola, J. Kwandrans, et T. Muotka (2009). Diatoms : unicellular surrogates for macroalgal community structure in streams? *Biodiversity and Conservation* 18(1), 79–89.

- Tison, J., Y. S. Park, M. Coste, J. G. Wasson, L. Ector, F. Rimet, et F. Delmas (2005). Typology of diatom communities and the influence of hydroecoregions : A study on the french hydrosystem scale. *Water Research* 39(14), 3177–3188.
- Tison, J., Y. S. Park, M. Coste, J. G. Wasson, F. Rimet, L. Ector, et F. Delmas (2007). Predicting diatom reference communities at the french hydrosystem scale : A first step towards the definition of the good ecological status. *Ecological Modelling* 203(1-2), 99–108.
- Ulrich, W., M. Almeida-Neto, et N. J. Gotelli (2009). A consumer's guide to nestedness analysis. *Oikos* 118(1), 3–17.
- Ulrich, W. et N. J. Gotelli (2007). Null model analysis of species nestedness patterns. *Ecology* 88(7), 1824–1831.
- van Dam, H., A. Mertens, et J. Sinkeldam (1994). A coded checklist and ecological indicator values of freshwater diatoms from the netherlands. *Netherlands Journal of Aquatic Ecology* 28(1), 117–133.
- Vanormelingen, P., E. Verleyen, et W. Vyverman (2008). The diversity and distribution of diatoms : From cosmopolitanism to narrow endemism. *Biodiversity and Conservation* 17(2), 393–405.
- Verleyen, E., W. Vyverman, M. Sterken, D. A. Hodgson, A. De Wever, S. Juggins, B. Van de Vijver, V. J. Jones, P. Vanormelingen, D. Roberts, R. Flower, C. Kilroy, C. Souffreau, et K. Sabbe (2009). The importance of dispersal related and local factors in shaping the taxonomic structure of diatom meta-communities. *Oikos* 118(8), 1239–1249.
- Vyverman, W., E. Verleyen, K. Sabbe, K. Vanhoutte, M. Sterken, D. A. Hodgson, D. G. Mann, S. Juggins, B. V. De Vijver, V. Jones, R. Flower, D. Roberts, V. A. Chepurinov, C. Kilroy, P. Vanormelingen, et A. De Wever (2007). Historical processes constrain patterns in global diatom diversity. *Ecology* 88(8), 1924–1931.
- Wang, Y.-K., R. J. Stevenson, et L. Metzmeier (2005). Development and evaluation of a diatom-based index of biotic integrity for the interior plateau ecoregion, usa. *Journal of the North American Benthological Society* 24(4), 990–1008.
- Wasson, J. G., A. Chandresris, H. Pella, et L. Blanc (2002). Definition des hydroécorégions françaises métropolitaine. approche régionale de la typologie des eaux courantes et éléments pour la définition des peuplements de référence d'invertébrés - rapport final. Technical report, Ministère de l'Ecologie et du Développement Durable - Cemagref GMA. Biol. des Syst. Aquat. Lab. Hydroécol. Quantitative Lyon.
- Watkins, A. J. et J. B. Wilson (2003). Local texture convergence : a new approach to seeking assembly rules. *Oikos* 102(3), 525–532.
- Whittaker, R. (1972). Evolution and measurement of species diversity. *Taxon*, 213–251.

- Williams, D. et J. Kociolek (2007). Pursuit of a natural classification of diatoms : History, monophyly and the rejection of paraphyletic taxa. *European Journal of Phycology* 42(3), 313–319.
- Winter, J. G. et H. C. Duthie (2000). Epilithic diatoms as indicators of stream total n and total p concentration. *Journal of the North American Benthological Society* 19(1), 32–49.
- Zalack, J., N. Smucker, et M. Vis (2010). Development of a diatom index of biotic integrity for acid mine drainage impacted streams. *Ecological Indicators* 10(2), 287–295.
- Zelinka, M. et P. Marvan (1961). Zur präzisierung der biologischen klassifikation der reinheit fließender gewässer. *Arch. Hydrobiol* 57(3), 389–407.

## Première partie

# Test de l'importance des grands processus de biologie des communautés sur les assemblages de diatomées



## Chapitre 2

# Structures spatiales dans les assemblages de diatomées benthiques

*Ce chapitre présente un article soumis Freshwater Science.*

### 2.1 Buts et contexte de l'étude

Afin de déterminer quels sont les processus qui ont une influence significative sur les assemblages de diatomées, il est important de décrire la structure de ces assemblages. La structuration spatiale est une composante importante des assemblages, elle résulte principalement de deux phénomènes : (i) la structuration spatiale de l'environnement, et (ii) les patrons bio-géographiques des espèces, liés entre autres aux limites de dispersion de ces dernières.

Les limites de dispersion des micro-organismes font depuis quelques années l'objet d'un débat. En effet, du fait de leurs fortes densités de population et de capacités de dispersion passive supposées fortes des espèces, il a été proposé que, pour les micro-organismes, contrairement aux macro-organismes, « *everything is everywhere, but the environment selects* » (« tous sont partout, et l'environnement sélectionne » ; Finlay, 2002; Fenchel et Finlay, 2004). Selon cette théorie, les micro-organismes ne devraient pas montrer de patrons bio-géographiques semblables à ceux, largement décrits, des plantes supérieures ou des animaux macroscopiques. Pourtant de nombreuses études se sont attachées à décrire de tels patrons chez les diatomées, montrant de fortes structures spatiales (Martiny et al., 2006; Soininen, 2007; Astorga et al., 2012), ou même de l'endémisme (Vyverman et al., 2007; Vanormelingen et al., 2008).

De nombreuses techniques ont été utilisées pour décrire les structures « purement spatiales » des assemblages, exemptes de l'effet spatialisé de l'environne-

ment. Parmi celles-ci, les tests de Mantel (méthode basée sur l'étude de corrélations partielles entre des matrices de distances) ont rarement été utilisés pour analyser les assemblages diatomiques des rivières, et encore plus rarement en utilisant des bases de données aussi complètes que la base de données disponible à Irstea (voir page 31).

Cette étude concerne donc l'application des tests de Mantel pour décrire la composante spatiale des assemblages de diatomées françaises, en relation avec les grandes régions diatomiques connues, la variabilité temporelle des assemblages et la qualité de l'eau. De plus, nous avons cherché à déterminer l'implication de nos résultats pour l'utilisation des diatomées en bioindication, car les structures spatiales des communautés sont insuffisamment prises en compte dans les programmes d'évaluation biologique.

## 2.2 Principaux résultats

Cette étude a montré que la principale force structurant les assemblages était l'environnement. Malgré cela, nous avons pu montrer une forte structuration purement spatiale des assemblages de diatomées, à l'échelle régionale et nationale :

- sans influence de la qualité de l'eau
- liée à la topographie des différentes régions
- liée à leur variabilité inter-annuelle

Ces résultats sont compatibles avec l'effet de limites de dispersion des espèces, et doivent attirer notre attention sur la faible prise en compte de tels phénomènes dans les pratiques de bioindication.

Aussi nous proposons de ré-évaluer la stratification régionale, en particulier d'affiner celle-ci dans les régions de montagne, de décrire la variabilité naturelle des assemblages grâce à des données biologiques pluri-annuelles et d'étudier la possibilité d'utiliser des données régionales pour la détermination des valeurs spécifiques utilisées dans le calcul des indices diatomiques.

## 2.3 Spatial structure in benthic diatom assemblages - patterns across regions, years and for different water quality levels

Marius BOTTIN<sup>1\*</sup> Janne SOININEN<sup>2</sup>, Martial FERREOL<sup>3</sup> and Juliette TISON-ROSEBERY<sup>1</sup>

<sup>1</sup> Irstea, UR REBX, 50 avenue de Verdun, 33612 Cestas cedex, France

<sup>2</sup> Department of Geosciences and Geography, PO Box 64, University of Helsinki, Finland

<sup>3</sup> Irstea, UR MALY, 3 bis quai Chauveau - CP 220, 69336 Lyon, France

### Abstract

1. Numerous analyses of benthic diatoms have shown clear spatial patterns in assemblage structure. Besides examining the underlying processes, it is important to describe the factors which affect those patterns, in order to define good practices for community studies and applications.
2. We analysed correlations between assemblage dissimilarity and geographical and environmental distances with Mantel and partial Mantel tests using a large French database, subdivided into regions, years and water quality levels.
3. Although environmental control of assemblages was stronger than spatial structure in explaining assemblage gradients, pure spatial patterns were also significant, both at the national scale and within some of the regions. When we averaged the data over three years, assemblage-environment relationships were stronger than within a single year data.
4. Spatial processes seem to act strongly on assemblages, even at smaller spatial scales than expected (i.e. , within regions). Study scale, year or water quality level did not notably affect the spatial patterns that emerged. Assemblages in mountain regions showed particularly strong spatial patterns, perhaps due to dispersal barriers.
5. Due to the strong spatial structure in diatom data, we encourage researchers to divide ecoregions into even smaller areas especially in mountainous landscapes when studying environment-assemblage relationships. We also recommend the use of averaged biological and environmental data (if available) for biotypologies using diatom assemblages.

### Keywords

Diatom; River; Spatial; Barrier; Dispersal

#### 2.3.1 Introduction

Biotic communities are assembled by the interplay of local (i.e. biotic interactions and environmental filtering) and regional (i.e. history, climate and dispersal) forces. However, macro- and microorganisms may show different responses

to such factors. The Baas-Becking hypothesis postulates that dispersal abilities and local abundances of microorganisms are so high that one can overlook the effects of dispersal limitation and historical factors (Finlay, 2002; Fenchel et al., 2004). Hence, for example, river diatom assemblages had mostly been studied through the prism of: “everything is everywhere, but, the environment selects”. From this point of view, one can consider that diatoms are purely driven by environmental factors and that observed noise in diatom-environment relations is mostly due to inaccuracies in environmental measurements and biotic samplings.

Diatoms have been widely used as water quality indicators of continental lakes and rivers (Stoermer et al., 1999; Hill et al., 2000). Ionic balance and alkalinity are typically the main environmental drivers of diatom species distributions (Potapova et al., 2003) and the trophic status of the water has also been shown to be important (Lange-Bertalot, 1979; van Dam et al., 1994; Pan et al., 1996; Winter et al., 2000; Soininen et al., 2002). Diatoms are used in numerous water quality monitoring programs worldwide, and diatom-based indices have been developed in many countries mostly to be used for river monitoring purposes (e.g. Kelly et al., 1995; dell Uomo, 2004; Lavoie et al., 2006; Hürlimann et al., 2007; Coste et al., 2009). Large diatom datasets have been used to construct relationships between species and environment, and this wide use of diatom assemblages in bioassessment programs might have strengthened the idea of a direct relationship between assemblages and local environment. However, evidence for the involvement of processes other than environmental filtering acting on diatom distributions has emerged. For example, biotic relationships (Burkholder et al., 1990; McCormick, 1996), historical events such as long-term diversification (evolution), dispersion between continents (Vyverman et al., 2007; Vanormelingen et al., 2008) and dispersal limitation at a finer scale (Martiny et al., 2006; Soininen, 2007; Verleyen et al., 2009; Heino et al., 2010; Astorga et al., 2012) may account for some of the noise in environment-species relationships.

Several studies about the relationships between diatom assemblages and environment have shown strong geographical patterns (Potapova et al., 2002), and also stressed the importance of considering the spatial structure of the assemblages in bioassessment programs (Pan et al., 2000; Potapova et al., 2002; Tison et al., 2005; Potapova et al., 2007). Smucker et al. (2011) showed that recognizing the potential influence of spatial factors on diatoms can help to improve bioassessment methodologies. Debate remains especially concerning the processes underlying such patterns - it is difficult to separate the unique role of dispersal limitation (i.e., neutrality: random variation of composition of assemblages due to random dispersion of species) from spatially structured variation in environmental variables.

Here, our aim is (i) to assess spatial patterns in diatom assemblages, (ii) to examine the effect of spatial scale on such patterns, (iii) to examine whether reference sites and impacted sites are spatially structured to the same degree, and (iv) to examine the inter-annual variations of the spatial structure in the assemblages. Specifically, we test three main hypotheses:

1. Assemblages are more homogeneous across sites in nutrient rich ecosystems because of higher productivity compared with sites with lower levels of nutrients (Soininen, 2004; Donohue et al., 2009). However, some earlier

studies have shown the opposite patterns for diatoms (Pan et al., 2000; Chase et Leibold, 2002; Soininen et al., 2004).

2. The influence of spatial factors on assemblages decreases with decreasing spatial scale (Soininen, 2004; Verleyen et al., 2009), and assemblages are only weakly spatially structured within a region, i.e., at the metacommunity scale (Donohue et al., 2009; Soininen et al., 2011). This is because assemblages at larger scales are subject to stronger dispersal limitation and longer environmental gradients (Soininen et al., 2011).
3. Assemblages data averaged over a longer period of time may show less spatial structure than “assemblage snapshots” that reflect more stochastic dispersal processes. However, averaged assemblages show stronger associations with environmental conditions. We emphasize though that the effect of dispersal limitations during the process when communities shift from one successional state to another might also enhance spatial patterns in a complex way (Hoagland et al., 1982; Peterson et Stevenson, 1990; Lavoie et al., 2003). Indeed, each step toward a more mature biofilm should imply the ability of new species to become established in the system.

In order to test these hypotheses, we used a very large French diatom dataset. Such a large amount of data allowed us to conjointly tackle the spatial patterns in the assemblages and potential underlying factors at multiple scales (i.e., across the whole study area as well as within separate subsets of the data). The spatial patterns were also examined separately at “reference” (near-pristine) sites and at “impacted” sites. Lastly, we tested for annual variability of the spatial structure of the assemblages by comparing spatial patterns for the three consecutive years at the same sampling sites as well as examining spatial structure in assemblages averaged over three years. During these analyses, we also kept in mind practical issues related to bioassessment, i.e., we suggest methodological avenues to decrease the noise in spatio-temporal diatom data and to separate appropriately regions in which assemblages are compared for biomonitoring purposes.

## 2.3.2 Material and methods

### 2.3.2.1 Data

We used a large diatom dataset from France, collected by the Water Agencies between 2007 and 2009. The databases include local physico-chemical data from the sampling sites and diatom counts for numerous geo-referenced sites. The samples were collected from stones with a toothbrush and scalpel during the low flow period, in order to avoid seasonal variability in biotic assemblages as well as local physico-chemical variables. All diatom samples were collected according to a standardized method (NFT 90-354) (AFNOR, 2007).

Before diatom composition analysis, samples were digested in boiling  $H_2O_2$  (30%) and HCl (35%) and mounted in a high-refractive-index medium (Naphrax, Northern Biological Supplies Ltd, UK; RI = 1.74). Diatom species were identified at 1000 x magnification (Leitz DMRD light microscope) (400 valves per slide) according mainly to Krammer et Lange-Bertalot (1991) by examining permanent slides of cleaned diatom frustules. Due to such a large dataset, diatom species were identified by different analysts. To mitigate taxonomic problems, we lumped together several under-varieties into their respective higher taxonomic units (i.e., species or varieties). The resulting biological data matrix,

expressed in relative abundance of species, comprised 1091 species in 2920 biological samples collected from 1493 different sites. Diatom relative abundance data were arcsine-transformed ( $Y = \arcsine(\sqrt{Y})$ ) in order to reduce variations in abundance and to enhance the information brought by rare species.

Physico-chemical data consisted of average values of each variable for the period between 45 days before diatom sampling and 15 days after sampling. We included 15 variables: water temperature, pH, altitude, dissolved oxygen, suspended matter, biological oxygen demand averaged over 5 days ( $BOD_5$ ), alkalinity, conductivity, nitrate, nitrite, ammonium, Kjeldahl nitrogen, orthophosphate, total phosphorus and organic carbon. Physico-chemical data were log-transformed ( $Y = \log(Y+1)$ ) and scaled to obtain a more normal distribution of variables with a similar weight for each variable.

The year 2009 included the largest dataset of all the years. A first large dataset comprising all data from 2009 was used as a “total dataset”. Our data analyses were separated into three categorical variables: region, water quality classes and years.

**Variation among regions** Our data included information on different Hydro-Eco-Regions (HER), based on natural characteristics in geology, relief and climate (Wasson et al., 2002). HERs have been described as a general geographical framework for all aquatic organisms, but these are not the most appropriate when considering diatoms. Based on the relationships between assemblage composition and the driving environmental gradients (mainly natural alkalinity and altitude), Tison et al. (2005) defined five main diatom biotypes across all the HERs. Based on these biotypes, we simplified the HER classification, resulting in a final set of 12 regions. Therefore, data from year 2009 were divided into 12 regions for studying patterns within regions (Fig 2.1 and Table 2.1). In addition to geology, relief and climate, we also characterized the topographic variation of the regions by calculating their interquartile ranges of altitude.

**Reference versus impacted sites** The influence of water quality on the spatial structure of the assemblages was assessed by separating reference sites (near-pristine sites used as references in European bioassessment programs) and impacted sites in the total dataset (year 2009). To facilitate comparison, the same number of reference and impacted sites were randomly selected for each region (Fig 2.2)

**Temporal variation** A fraction of the sites was sampled repeatedly in three consecutive years (2007, 2008 and 2009) (Fig 2.3). The corresponding data allowed us to separate the influence of temporal variability on diatom assemblage structure. To avoid a greater influence of the much more numerous low-water quality sites, the same number of reference and impacted sites was included for each region in these analyses. Data from year 2009 were then compared to those of 2007 and 2008 by creating three sets corresponding to the three consecutive years and including data from the same sites. We also used averaged values of diatom and physico-chemical data calculated over these three sampling years in order to test for the possible stronger relationship between diatoms and environment when taking a longer period of variability into account.

Table 2.1: Environmental characteristics of the different regions. Values of environmental variables are in the form: mean  $\pm$  standard deviation of the mean.

	Number of sites		Altitude Interquartile range	pH	Mean and Standard error of the mean of main variables				
	Reference	impacted			Alkalinity (mg.L-1)	Conductivity ( $\mu$ S.cm-1)	Total phosphorus (mg.L-1)	Nitrates (mg.L-1)	
<b>Highly acidic plains</b>									
Landes	0	11	17	7.15 $\pm$ 0.16	7.03 $\pm$ 1.55	204.47 $\pm$ 29.09	0.06 $\pm$ 0.01	9.88 $\pm$ 2.46	
<b>Circum-neutralid-altitude context</b>									
Brittany	6	140	54	7.52 $\pm$ 0.03	13.84 $\pm$ 0.36	331.11 $\pm$ 11.65	0.15 $\pm$ 0.01	20.63 $\pm$ 1.17	
Vosges	3	11	180.5	7.38 $\pm$ 0.07	3.34 $\pm$ 0.44	117.46 $\pm$ 8.74	0.1 $\pm$ 0.02	4.06 $\pm$ 0.42	
Limousin and Auvergne	10	107	134	7.56 $\pm$ 0.04	7.92 $\pm$ 0.45	202.75 $\pm$ 16.56	0.1 $\pm$ 0.01	5.89 $\pm$ 0.46	
Centre	1	18	20	7.89 $\pm$ 0.15	11.57 $\pm$ 1.06	247.21 $\pm$ 21.37	0.15 $\pm$ 0.03	8.1 $\pm$ 0.94	
Ardennes and Flanders	0	15	109	8.11 $\pm$ 0.05	25.62 $\pm$ 1.52	763.56 $\pm$ 61.68	0.35 $\pm$ 0.09	17.46 $\pm$ 2.07	
<b>Crystalline Highland</b>									
Pyrenees	2	28	352.75	7.87 $\pm$ 0.05	8.63 $\pm$ 0.52	170.45 $\pm$ 17.46	0.04 $\pm$ 0	1.89 $\pm$ 0.32	
Massif Central	10	101	479	7.68 $\pm$ 0.05	7.76 $\pm$ 0.4	137.14 $\pm$ 10.24	0.05 $\pm$ 0.01	4.06 $\pm$ 0.31	
Corsica	2	16	257	7.83 $\pm$ 0.07	8.34 $\pm$ 0.88	194.89 $\pm$ 33.42	0.1 $\pm$ 0	2.06 $\pm$ 0.06	
<b>Calcareous Alcaline plains</b>									
South	9	209	122.25	7.99 $\pm$ 0.02	15.86 $\pm$ 0.39	469.62 $\pm$ 19.41	0.08 $\pm$ 0.01	10.16 $\pm$ 0.76	
North	19	328	128	7.96 $\pm$ 0.01	18.98 $\pm$ 0.32	575.07 $\pm$ 15.49	0.12 $\pm$ 0.01	17.8 $\pm$ 0.64	
<b>Calcareous Highland</b>									
Alps	12	71	479	8.1 $\pm$ 0.02	14.12 $\pm$ 0.41	428.08 $\pm$ 15.56	0.05 $\pm$ 0.01	4.29 $\pm$ 0.65	

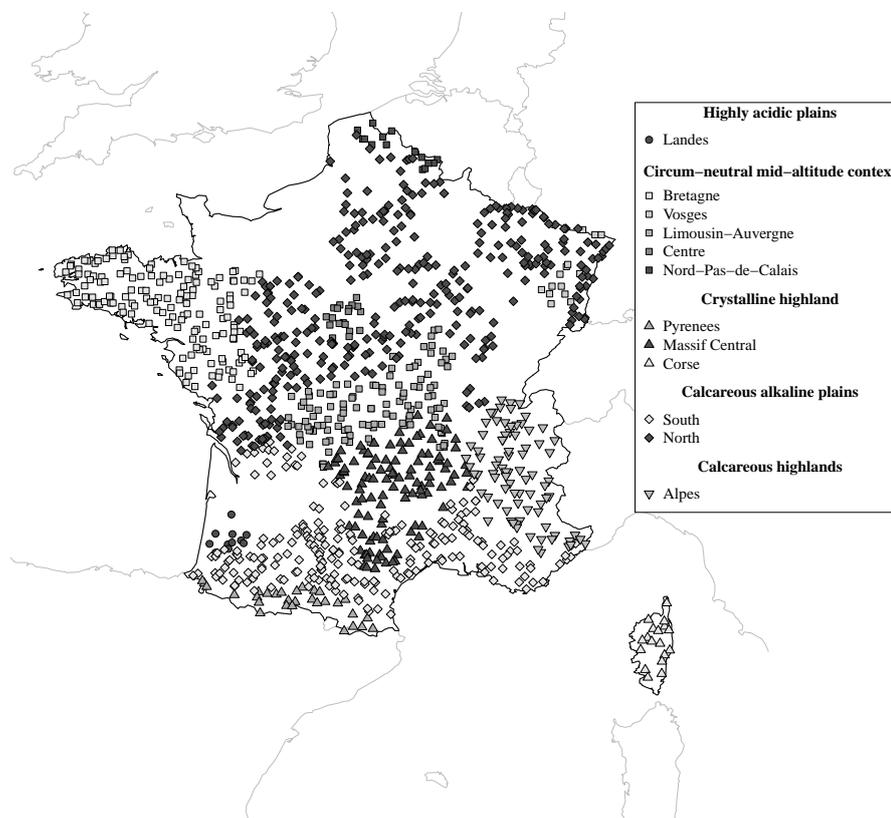


Figure 2.1: Map of all sites sampled in year 2009. Symbols correspond to the 12 regions identified as different subsets.

### 2.3.2.2 Data analysis

**BIO-ENV procedure** The BIO-ENV procedure (Clarke et Ainsworth, 1993) was first used to determine which subset of the environmental variables produced the highest non-parametric correlation with the assemblage dissimilarities. This method consists of calculating Spearman correlation indices between assemblage dissimilarities (Bray-Curtis) and environmental distances (Euclidean) using all possible different subsets of the variables. Indeed, including poor predictors into analyses increases noise in the relationships between environment and assemblages, leading to poorer correlations than those obtained with only the important predictors. This procedure indicates the strongest relationship possible between the assemblages and environment, and thus allows the spatial control on the assemblages to be tested more effectively.

**Dissimilarity matrices** To describe biological, environmental, spatial gradients and their correlations, three dissimilarity matrices were constructed for each set of data. We used the Bray-Curtis index on transformed diatom data between all site pairs to describe beta-diversity gradients. Environmental gradients were computed by calculating Euclidean distances on scaled, transformed and BIO-ENV-selected environmental variables. Geographical distances were calculated using Euclidean distance. One could argue that overland distances are not optimal to assess movements of biota in river-systems (Landeiro et al., 2011) because river flow moves more biological material than wind (or animals ; see Kristiansen (1996) for a more exhaustive list of dispersal processes), but Finn et al. (2006) showed that straight distance was an effective surrogate for more complex connectivity measures. Several previous studies have thus used Euclidean distances for similar purposes (e.g. Heino et Soininen, 2010; Astorga et al., 2012). Euclidean distances represent realistic connectivities especially for small unicellular organisms that disperse across sites easily via the air. Moreover, stream network distances are not suitable for data in which sites span multiple drainage basins.

**Mantel tests** We tested for the correlations between environmental, biotic and geographical distance matrices using Mantel tests. Mantel tests comprise calculations of Pearson correlations between distance (or dissimilarity) matrices, using permutation tests for an estimation of significance. Because environmental variables are typically spatially structured, environmental and geographical distances are strongly correlated. We therefore also used partial Mantel test, where either the environmental or the geographical distance matrix was the explanatory matrix, and the other one was the partial matrix. These partial Mantel tests aim to examine the influence of “pure spatial” or “pure environmental” effects on assemblages.

It has been argued that, due to inflated type I errors in Mantel tests under spatial autocorrelation, Multiple Regressions on distance Matrices (MRM) should be used instead (Lichstein, 2007; Krasnov et al., 2010). Using MRM on the total dataset, the results were very similar to those obtained with Mantel tests (coefficients for environmental and spatial distances were respectively  $4.71 * 10^{-2}$  and  $7.35 * 10^{-5}$ , both highly significant:  $P < 0.001$ , and an estimated value of R squared of 0.27). In order to facilitate the comparison with other studies, we used Mantel tests here. Up to now, the limitations of MRM are poorly known due to the fact that the method was only developed recently (Lichstein, 2007).

We ran all data analyses under R 2.12.1, with the “vegan” package (Oksanen et al., 2011).

### 2.3.3 Results

#### 2.3.3.1 Best environmental predictors (BIO-ENV procedure)

BIO-ENV analyses identified different sets of environmental variables depending on the regions and the sampling year, resulting in one to eight variables to be selected (Table 2.2).

In the total dataset, the best set of predictors included water pH, conductivity, nitrates, organic carbon, water temperature and altitude. These variables

were also the six most frequent predictors over the different data sets.

The set of best predictors varied strongly between regions. Altitude was surprisingly a poor predictor for the two regions with the widest altitudinal range: Alps and Massif-central. Indeed, only different fractions of nitrogen and phosphorus were included in the best model in the Alps. Alkalinity was identified as a sole variable in the Massif-central subset.

The best predictors did not vary notably between reference and impacted sites, and were very similar to those of the total dataset. The small differences included the fact that alkalinity was selected for both reference and impacted sites, amount of phosphates was selected for the reference sites (instead of the amount of nitrate in the total dataset) and suspended matter was identified as one of the best predictors for impacted sites.

The best predictors differed clearly between the years, although sampled sites were the same for all three consecutive years.

### 2.3.3.2 Mantel tests

**Total dataset** In 2009 for the whole of France, Mantel tests and partial Mantel tests showed a highly significant ( $P < 0.001$ ) correlation between assemblage dissimilarity and environmental or geographical distances (Table 2.3). Partial Mantel tests showed a much stronger “pure environmental” control of the assemblage (when testing the correlation between assemblage dissimilarity and environmental distance while controlling for geographical distance,  $r=0.48$ ) than “pure spatial” control (when testing the correlation between assemblage dissimilarity and geographical distance while controlling for environmental distance,  $r=0.13$ ) of assemblage.

**Among-region variation** For all 12 regions studied, assemblage dissimilarity showed significant correlation with their environment, while 8 regions correlated significantly with geographical distances (Table 2.3). Running the analyses for environmental distance using partial tests, all of the regions studied showed significant “pure” Mantel correlations, and only 5 regions for geographical distances. The only region showing a higher correlation between assemblage dissimilarity and geographical distance than between assemblage dissimilarity and environmental distance was the Alps.

Smaller regions (i.e. regions in which average pairwise distances between sites was lower than 100 km) did not show a significant effect of pure spatial factors in partial Mantel tests (i.e. when testing the effect of geographical distance while controlling for environmental distance). For larger regions, the effects of pure spatial factors were significant (except in the Pyrenees and the Northern plains).

We further tested the correlation between partial Mantel test  $r$  statistics ( $r$  for the pure spatial effect) and two regional variables: average distances within each region and altitude interquartile range within a region. Results did not show any significant correlation between average distance between sites within a region and partial Mantel test  $r$  statistics (unilateral Kendall rank correlation test:  $\tau=32$   $P=0.580$ ; see Fig 2.4a). However, we found a significant positive correlation between altitude interquartile ranges of regions and partial Mantel  $r$  statistics (unilateral Kendall rank correlation test:  $\tau=47$   $P=0.031$ ; fig. 2.4).

**Reference versus impacted sites** Results of Mantel tests and partial Mantel tests did not reveal any considerable differences between reference and impacted sites (Table 2.4) as Mantel and partial Mantel tests were significant and  $r$  statistics were comparable for both groups of sites. Results were overall similar to those obtained for the whole of France (Table 2.3).

**Temporal variation** The correlation between assemblage dissimilarity and environmental distance or between assemblage dissimilarity and geographical distance varied depending on the sampling year, particularly in the partial tests (Table 2.5). However, all correlations were significant ( $P < 0.01$ ). When biological and environmental data were averaged over the three consecutive years, the correlation between assemblage dissimilarity and environmental distance was higher than in any of the consecutive years, even if geographical distances were controlled for. By contrast, correlation between assemblage dissimilarity and geographical distance was comparable for the data averaged over the three years and for each single year analysed separately.



Figure 2.2: Map of the reference and impacted sites sampled in year 2009. For facilitating the comparison, the same number of reference and impacted sites were sampled for each region.

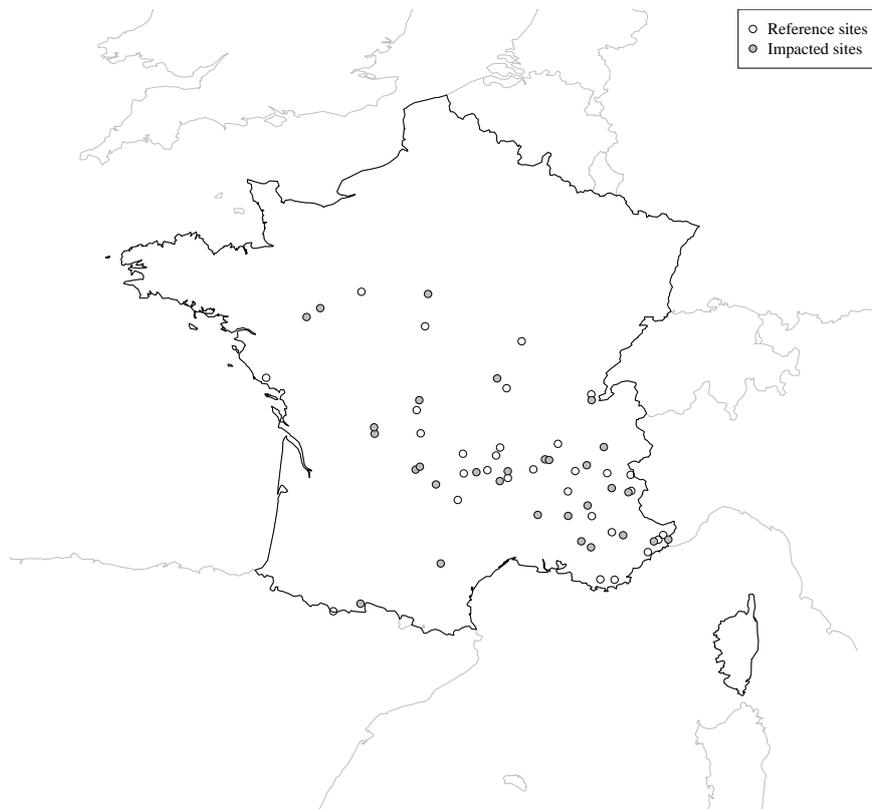


Figure 2.3: Map of the sites consecutively sampled in years 2007, 2008 and 2009.

Table 2.2: Results of Mantel tests and partial Mantel tests for community dissimilarities against geographical and environmental distances in the total dataset and in the different regions for year 2009. “r” is the Pearson statistic of Mantel tests and P is the significance estimation resulting of the permutation test.

Total dataset	Number of Sites	Number of Variables	pH	Alkalinity	Conductivity	Total N	NO <sub>2</sub>	NO <sub>3</sub>	NH <sub>4</sub>	PO <sub>4</sub>	Total P	BDO <sub>5</sub>	Organic C	Dissolved O <sub>2</sub>	Susp. Mat.	Temp.	Altitude	
<b>Regions</b>	1129	6	X		X		X						X				X	X
<b>Highly acidic plains</b>											X							
Landes	11	2	X															
<b>Circum-neutral mid-altitude context</b>																		
Brittany	146	3			X													X
Vosges	14	7	X			X	X						X					X
Limousin and Auvergne	117	3		X														X
Centre	19	4	X		X													X
Plantes and Ardennes	15	2						X				X						
<b>Crystalline Highland</b>																		
Pyrenees	30	6		X			X					X						X
Massif Central	111	1			X													
Corsica	18	6		X					X				X					X
<b>Calcareous Alkaline plains</b>																		
South	218	5			X		X	X					X					X
North	347	8	X		X		X						X					X
<b>Calcareous Highland</b>																		
Aps	83	3				X				X								
<b>Reference versus impacted sites</b>																		
Reference sites	74	7	X	X	X								X					X
Impacted sites	74	7	X		X					X								X
<b>Temporal variation</b>																		
Averaged data	60	6	X		X		X				X	X	X					
2007	60	5			X		X				X	X	X					
2008	60	7	X		X		X				X	X	X					
2009	60	3			X			X				X	X					X

Table 2.3: Results of Mantel tests and partial Mantel tests for community dissimilarities against geographical and environmental distances in the total dataset and in the different regions for year 2009. “r” is the Pearson statistic of Mantel tests and P is the significance estimation resulting of the permutation test.

	Number of sites	Maximum geographical distance (km)	Average geographical distance (km)	Environment		Geography		Environment (geography controlled for)		Geography (environment controlled for)	
				r	P	r	P	r	P	r	P
<b>Total</b>	1129	1316.74	399.63	0.51	<0.001	0.23	<0.001	0.48	<0.001	0.13	<0.001
<b>Regions</b>											
<b>Highly acidic plains</b>											
Landes	11	76.71	37.48	0.64	<0.001	-0.01	0.5	0.65	<0.001	0.17	0.149
<b>Circum-neutral mid-altitude context</b>											
Brittany	146	374.80	141.61	0.57	<0.001	0.26	<0.001	0.53	<0.001	0.09	0.004
Vosges	14	154.28	66.35	0.50	<0.001	0.29	0.016	0.46	<0.001	0.18	0.097
Limousin and Auvergne	117	327.86	130.08	0.69	<0.001	0.20	<0.001	0.68	<0.001	0.18	<0.001
Centre	19	85.94	36.73	0.61	<0.001	0.04	0.313	0.61	<0.001	-0.04	0.668
Ardennes and Flanders	15	142.49	66.57	0.59	<0.001	0.02	0.368	0.59	<0.001	0.00	0.434
<b>Crystalline Highland</b>											
Pyrenees	30	335.63	111.59	0.51	<0.001	0.18	0.021	0.49	<0.001	0.09	0.13
Massif Central	111	314.22	123.04	0.48	<0.001	0.19	<0.001	0.47	<0.001	0.15	<0.001
Corsica	18	142.94	52.23	0.31	<0.001	0.02	0.386	0.32	0.002	0.07	0.202
<b>Calcareous Alkaline plains</b>											
South	218	718.95	226.01	0.40	<0.001	0.14	<0.001	0.38	<0.001	0.05	0.036
North	347	767.66	285.04	0.44	<0.001	0.14	<0.001	0.43	<0.001	0.03	0.076
<b>Calcareous Highland</b>											
Alps	83	307.45	116.65	0.22	0.004	0.33	<0.001	0.21	0.009	0.33	<0.001

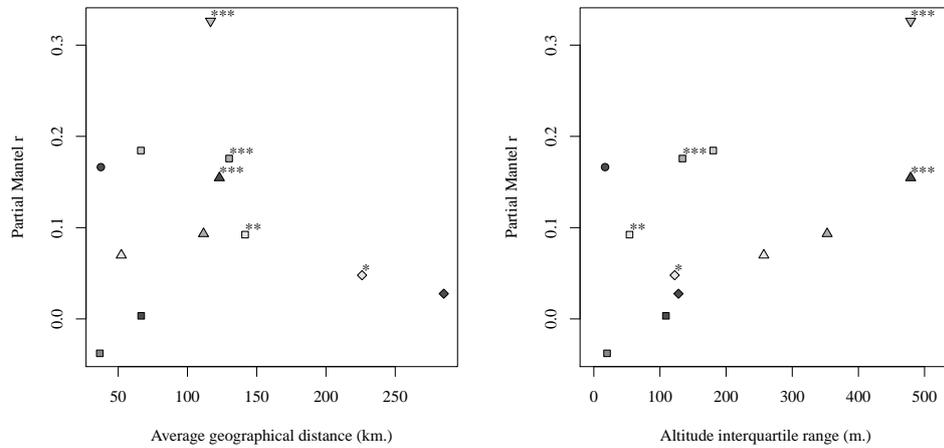


Figure 2.4: Relationship between  $r$  statistic of partial Mantel tests (Pearson correlation between community dissimilarities and geographical distances when environmental distances are controlled for) and average geographical distances of the different regions (a) or altitude interquartile ranges of the different regions (b). Symbols here are the same as those of figure 2. Stars over the symbols of regions describe the level of significance of partial Mantel tests (\*  $P < 0.05$ ; \*\*  $P < 0.01$ ; \*\*\*  $P < 0.001$ )

Table 2.4: Results of Mantel tests and partial Mantel tests for community dissimilarities against geographical and environmental distances in impacted sites and reference sites of year 2009. “r” is the Pearson statistic of Mantel tests and P is the significance estimation resulting from the associated permutation test.

	Number of sites	Maximum geographical distance (km)	Average geographical distance (km)	Environment		Geography		Environment (geography controlled for)		Geography (environment controlled for)	
				r	P	r	P	r	P	r	P
Reference sites	74	1231.46	390.68	0.50	<0.001	0.21	<0.001	0.48	<0.001	0.15	0.005
Impacted sites	74	1255.79	408.95	0.49	<0.001	0.23	<0.001	0.46	<0.001	0.16	0.003

Table 2.5: Results of Mantel tests and partial Mantel tests for community dissimilarities against geographical and environmental distances for years 2007 to 2009. “r” is the Pearson statistic of Mantel tests and P is the significance estimation resulting from the associated permutation test.

	Number of sites	Maximum geographical distance (km)	Average geographical distance (km)	Environment		Geography		Environment (geography controlled for)		Geography (environment controlled for)	
				r	P	r	P	r	P	r	P
Averaged data	60	768.09	272.87	0.66	<0.001	0.50	<0.001	0.52	<0.001	0.17	<0.001
2007	60	768.09	272.87	0.54	<0.001	0.39	<0.001	0.41	<0.001	0.11	0.01
2008	60	768.09	272.87	0.57	<0.001	0.41	<0.001	0.46	<0.001	0.18	<0.001
2009	60	768.09	272.87	0.47	<0.001	0.41	<0.001	0.33	<0.001	0.22	<0.001

## 2.3.4 Discussion

### 2.3.4.1 Spatial and environmental control of assemblages

Our data showed notable environmental control of benthic diatom assemblages, both at the scale of the whole of France and at the regional scale. The pattern was consistent in all regions, for varying water qualities and through the study years. This control by the environment is congruent with that reported in previous studies (e.g. Passy, 2007; Soininen, 2007) and is implicitly regarded as the scientific basis for the common use of diatoms in bioassessment programs (e.g. Kelly et Whitton, 1995; dell Uomo, 2004; Lavoie et al., 2006; Hürlimann et Niederhauser, 2007; Coste et al., 2009). However, our data also support the idea that diatom assemblages exhibit a rather strong spatial component at multiple spatial scales.

Such strong spatial patterns (even when controlling for environmental variables) in lotic diatom assemblages have also been found earlier by using a large variety of methods such as Redundancy Analysis (Passy, 2007), Canonical Correspondence Analysis (Soininen, 2004; Smucker et Vis, 2011), direct ordination with variance partitioning (Potapova et Charles, 2002; Soininen et al., 2004), analysis of spatial autocorrelation (Heino et al., 2010), and Mantel tests (Astorga et al., 2012). Such a great number of studies using different methodologies all suggest that lotic diatom assemblages are spatially structured over a wide range of scales. Two alternative hypotheses may explain these spatial patterns - diatoms are not freely dispersed across sites and/or assemblages are structured by some unmeasured environmental variables or biotic processes (e.g. competition from other algae, grazing or light conditions). However, some authors have argued that due to huge population densities and great abilities to disperse, microorganisms should not exhibit biogeographical patterns similar to those of vascular plants or animals (Finlay, 2002; Fenchel et Finlay, 2004).

### 2.3.4.2 Scale dependence

The control of microbial communities by spatial factors is known to be scale-dependent – the spatial structure should be stronger at large study scales than at small scales due to stronger dispersal limitation and longer environmental gradients (Green et al., 2004; Soininen, 2004; Soininen et al., 2007; Bennett et al., 2010; Soininen et al., 2011). Our results did not however show a stronger effect of spatial factors at the national scale compared with the smaller, regional scale: although the pure spatial effects on assemblages were significant at the national scale, some regions (Alps, Massif Central and Limousin-Auvergne) showed even a stronger spatial component of assemblages, with higher Mantel  $r$  (see discussion on elevation). However, we did not find significant pure spatial patterns in the regions where the average distance between sites were smaller than 100 km. This may be partly because the smallest regions in our data set often included less than 30 sites. Due to this bias, conclusions about the weaker spatial patterns at smaller scales remain speculative at present. This is especially true as an increase in the size of the regions was not paralleled by an increase in the strength (Mantel  $r$ ) of the spatial patterns. This might be due to the distance measure used, which could be appropriate to describe spatial patterns in an environmentally homogeneous region, but fails to take into account the spatial patterns at the national scale due to the mosaic of different ecoregions and the

increase in the number of basins considered (For a more complete discussion about distance measure in lotic systems see Finn et al., 2006; Landeiro et al., 2011).

#### 2.3.4.3 Effects of water-quality

In addition to studying scale, we examined how water quality and the trophic status of the water may influence the strength of the spatial patterns in diatom assemblages. Homogenization of assemblages due to human impact such as eutrophication is a well documented phenomenon in bioassessment (Rahel, 2002; McKinney, 2006; Donohue et al., 2009). Increase of primary production due to high human input of nutrients is now considered as the most important polluting effect on river biotas and ecosystems (e.g. Mainstone et Parr, 2002). In our data, spatial patterns and environmental control of assemblages were very similar in reference and impacted sites, suggesting that human impact did not affect assemblage patterns in our study. The impacted sites studied covered a wide range of environmental conditions, and a wide range in severity of impact might have had opposing effects on assemblages, resulting in unclear patterns in our data. A better understanding of the effects of pollution would require a more precise separation of the different water-quality levels, and a coupled analysis of the effects of nutrient enrichment and other types of pollution.

#### 2.3.4.4 Temporal variation of spatial attributes

One of the most novel aims of this paper was to examine the temporal variation in diatom beta-diversity. In our data, the relative importance of spatial patterns and environmental gradients on diatom assemblage were comparable during the three consecutive years. However, averaging environmental and assemblage data from these three years yielded an even better correlation between environment and assemblages, without affecting the correlation between spatial factors and assemblages. These results suggest that spatial patterns act as constant constraints on assemblages. The contribution to noise in the environment-assemblage relationships related to this spatial constraint can be reduced by considering a longer period of time. These considerations would be consistent with the hypothesis of dispersal-related factors: presences of some species might be partly random and taking into account a longer period of time might increase the probability of finding species potentially present in a given environment. Future studies should examine spatio-temporal patterns of assemblages, as suggested by Soininen (2010) and Passy (2007), in order to increase the understanding of biogeographical structure of diatom diversity.

#### 2.3.4.5 Altitude and topography

Finally, we were interested in how elevation gradients may shape the beta-diversity patterns. Since altitude reflects a range of factors such as hydrology, light conditions, human population density and productivity, it is difficult to disentangle its “direct” effect on biological communities. However, altitude is considered a major factor to define diatom-related hydro-ecoregions (Potapova et Charles, 2002; Tison et al., 2005, i.e. regional separation allowing the description of the main natural gradients for assemblages;). Altitude *per se* has

been demonstrated to be a poor predictor of assemblage variations in mountainous regions (Rimet et al., 2007; Jüttner et al., 2010). In our data, altitude was not considered as one of the best environmental predictors in the BIO-ENV procedure, even for the two regions with highest altitudinal ranges. This might be due to the fact that altitudinal gradients are related to geographical barriers and species' dispersal limitation thus causes "unpredictable" assemblages (Hubbell, 2001). In order to analyse the effect of elevation gradient in our data, we focused on three regions (Alps, Massif Central and Pyrenees), which can be considered as particularly mountainous, with an average altitude above 500 m and an altitude interquartile range higher than 350 m. Within these regions, only the Pyrenees failed to show significant spatial patterns in diatom assemblages. Note that in our dataset, all samples from the Pyrenees region were taken on the northern slopes of the mountain, the southern slopes being Spanish. Pyrenees are East-West oriented so sites from the same side are separated by weaker geographical barriers than sites from mountainous massifs such as the Alps or Massif Central. Moreover, altitude interquartile range, as a surrogate for topographic irregularities, showed a significant positive relationship with the degree of "spatial-control" on assemblages. These results tend to show that the spatial patterns described are related to geographical barriers, i.e. dispersal-related factors.

#### 2.3.4.6 Implications for bioassessment

From a more practical perspective, the spatial patterns described in this study represent noise in the relationships between diatom assemblages and environmental variables. Fortunately, additional analyses indicated that diatom indices (e.g. the French Pollution Sensitivity Index: IPS) did not show any pure spatial structure in the study area. Indeed, the variation was mostly explained by environmental variation (35%) and the combined effect of space and environment (9 %). Moreover, the regions showing the strongest spatial structures of assemblage compositions did not show more inter-annual variation in their indices, i.e. no relationship was found between inter-annual variance of indices and altitude. Smucker et Vis (2011) showed similar results, where spatial factors explained only a small part of index variations.

European bioassessment programs have to use a methodology in which the first steps consist of describing natural variability of assemblages in order to define proper references to be compared with impacted assemblages (Parlement Européen, 2000). Such methodologies have also been used in other continents (e.g. Pan et al., 2000; Lavoie et al., 2006; Philibert et al., 2006). The spatial structure of the assemblages might still be problematic during the initial implementation of these methodologies. Several authors have advised the separation of diatom data into regions in order to mitigate the spatial patterns in the data (e.g. Potapova et Charles, 2002; Soininen et al., 2004; Soininen, 2007). Our results tend to show that depending on the characteristics of the regions (relief, topography and perhaps region size), regional stratification might be insufficient and diatom assemblages within regions may show substantial spatial patterns. This is especially true in countries which present a large range of ecoregional variations, such as in France.

Currently, specific values for the calculation of indices are often fitted using the national description of the species environmental preferences. Spatial pat-

terns described in this study raise the question of whether there are regional differences in environmental preferences. Until recently, including specific values fitted regionally in the calculation of indices was unrealistic because available data were not sufficient at regional scales, but large databases have been recently constructed. Future development of indices should assess whether such regional differentiations result in more precise quality evaluations.

Future studies and bioassessment programs should assess which size of regions would be optimal, with regard to spatial patterns. The relationship that we found between topographic irregularities and the spatial control of assemblages is especially problematic as mountainous regions may show stronger spatial structure in diatom assemblages. Indeed, in France and many other countries, a large proportion of reference sites are located in mountainous regions, due to high population densities in lower elevations. The mountainous reference assemblages should be carefully compared with other assemblages. Refining ecoregional zonation in mountains might be useful to tackle this task. Moreover, averaging the diatom data over three years resulted in stronger relationships between environmental variables and assemblages, without adding space-related noise. Such pre-treatment of data could be useful in studies which aim to describe general relationships between environment and assemblages or particular species.

#### 2.3.4.7 Perspectives

Future studies should attempt to assess which species are the most dispersal limited, as initiated by Heino et Soinen (2010). If the spatial patterns detected are related to dispersal processes, rare species should be particularly spatially structured, due to the lower local densities and lower probabilities that new sites will be occupied. Moreover, examination of which biological and ecological traits influence a species dependence on spatial effects might reveal whether species are especially prone to dispersal limitation, or if spatial patterns could be explained by unmeasured environmental factors. Such studies might also allow us to define whether a particular assemblage is structured by spatial factors, in order to reduce spatial noise in a set of data. We emphasize that species traits are overall vital for community ecology given that environmental factors typically filter species traits, not necessarily species identity. Rebuilding community ecology from traits, not species identity, may shed more light on important processes shaping local communities.

#### Acknowledgement

The authors would like to thank François Delmas and Michel Coste for their kind help concerning general discussion and taxonomic issues.

We would like to acknowledge all the regional agencies for environment, DRIEE and DREAL, and water agencies for sending the data.

We are grateful to two anonymous reviewers for their suggestions which substantially improved the manuscript.

## Bibliography

- AFNOR (2007). Qualité de l'eau. détermination de l'indice biologique diatomées (ibd) - norme nf t90-354 -.
- Astorga, A., J. Oksanen, M. Luoto, J. Soininen, R. Virtanen, et T. Muotka (2012). Distance decay of similarity in freshwater communities: do macro- and microorganisms follow the same rules? *Global Ecology and Biogeography* 21, 365–375.
- Bennett, J. R., B. F. Cumming, B. K. Ginn, et J. P. Smol (2010). Broad-scale environmental response and niche conservatism in lacustrine diatom communities. *Global Ecology and Biogeography* 19(5), 724–732.
- Burkholder, J. M., R. G. Wetzel, et K. L. Klomparens (1990). Direct comparison of phosphate uptake by adnate and loosely attached microalgae within an intact biofilm matrix. *Applied and Environmental Microbiology* 56(9), 2882–2890.
- Chase, J. et M. Leibold (2002). Spatial scale dictates the productivity-biodiversity relationship. *Nature* 416(6879), 427–430.
- Clarke, K. et M. Ainsworth (1993). A method of linking multivariate community structure to environmental variables. *Marine Ecology Progress Series* 92(3), 205–219.
- Coste, M., S. Boutry, J. Tison-Rosebery, et F. Delmas (2009). Improvements of the biological diatom index (BDI): Description and efficiency of the new version (BDI-2006)mo. *Ecological Indicators* 9(4), 621–650.
- dell Uomo, A. (2004). L'indice diatomico di eutrofizzazione/polluzione (epi-d) nel monitoraggio delle acque correnti. Technical report, Linee guida, APAT, ARPAT, CTN-AIM, Roma, Firenze.
- Donohue, I., A. Jackson, M. Pusch, et K. Irvine (2009). Nutrient enrichment homogenizes lake benthic assemblages at local and regional scales. *Ecology* 90(12), 3470–3477.
- Fenchel, T. et B. Finlay (2004). The ubiquity of small species: Patterns of local and global diversity. *BioScience* 54(8), 777–784.
- Finlay, B. (2002). Global dispersal of free-living microbial eukaryote species. *Science* 296(5570), 1061–1063.
- Finn, D., D. Theobald, W. Black IV, et N. Poff (2006). Spatial population genetic structure and limited dispersal in a rocky mountain alpine stream insect. *Molecular Ecology* 15(12), 3553–3566.
- Green, J., A. Holmes, M. Westoby, I. Oliver, D. Briscoe, M. Dangerfield, M. Gillings, et A. Beattie (2004). Spatial scaling of microbial eukaryote diversity. *Nature* 432(7018), 747–750.
- Heino, J., L. M. Bini, S. M. Karjalainen, H. Mykra, J. Soininen, L. C. G. Vieira, et J. A. F. Diniz (2010). Geographical patterns of micro-organismal community structure: are diatoms ubiquitously distributed across boreal streams? *Oikos* 119(1), 129–137.

- Heino, J. et J. Soininen (2010). Are common species sufficient in describing turnover in aquatic metacommunities along environmental and spatial gradients? *Limnology and Oceanography* 55(6), 2397–2402.
- Hill, B., A. Herlihy, P. Kaufmann, R. Stevenson, F. McCormick, et C. Burch Johnson (2000). Use of periphyton assemblage data as an index of biotic integrity. *Journal of the North American Benthological Society* 19(1), 50–67.
- Hoagland, K., S. Roemer, et J. Rosowski (1982). Colonization and community structure of 2 periphyton assemblages, with emphasis on the diatoms (bacillariophyceae). *American Journal of Botany* 69(2), 188–213.
- Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton, USA: Princeton University Press.
- Hürlimann, J. et P. Niederhauser (2007). Méthodes d’analyse et d’appréciation des cours d’eau. diatomées niveau r (région). État de l’environnement n° 0740. Technical report, Office fédéral de l’environnement, Berne.
- Jüttner, I., P. Chimonides, S. Ormerod, et C. J. (2010). Ecology and biogeography of himalayan diatoms: distribution along gradients of altitude, stream habitat and water chemistry. *Fundamental and Applied Limnology* 177(4), 293–311.
- Kelly, M. et B. Whitton (1995). The trophic diatom index: a new index for monitoring eutrophication in rivers. *Journal of Applied Phycology* 7, 433–444.
- Krammer, K. et H. Lange-Bertalot (1986 - 1991). *Bacillariophyceae 1. Teil: Naviculaceae. 876 p.; 2. Teil: Bacillariaceae, Epithemiaceae, Surirellaceae, 596 p.; 3. Teil: Centrales, Fragilariaceae, Eunotiaceae, 576 p.; 4. Teil: Achnantheaceae. Kritische Ergänzungen zu Navicula (Lineolatae) und Gomphonema. 437 p.*, Volume Band 2/1-4 of *Süßwasserflora von Mitteleuropa*. Stuttgart: G. Fischer Verlag.
- Krasnov, B., D. Mouillot, G. Shenbrot, I. Khokhlova, et R. Poulin (2010). Deconstructing spatial patterns in species composition of ectoparasite communities: The relative contribution of host composition, environmental variables and geography. *Global Ecology and Biogeography* 19(4), 515–526.
- Kristiansen, J. (1996). Dispersal of freshwater algae - a review. *Hydrobiologia* 336(1-3), 151–157.
- Landeiro, V., W. Magnusson, A. Melo, H. Espírito-Santo, et L. Bini (2011). Spatial eigenfunction analyses in stream networks: Do watercourse and overland distances produce different results? *Freshwater Biology* 56(6), 1184–1192.
- Lange-Bertalot, H. (1979). Pollution tolerance of diatoms as a criterion for water quality estimation. *Nova Hedwigia* 64, 285–304.
- Lavoie, I., S. Campeau, M. A. Fallu, et P. J. Dillon (2006). Diatoms and biomonitoring: should cell size be accounted for? *Hydrobiologia* 573, 1–16.

- Lavoie, I., S. Campeau, M. Grenier, et P. J. Dillon (2006). A diatom-based index for the biological assessment of eastern canadian rivers: An application of correspondence analysis (ca). *Canadian Journal of Fisheries and Aquatic Sciences* 63(8), 1793–1811.
- Lavoie, I., W. Vincent, R. Pienitz, et J. Painchaud (2003). Effect of discharge on the temporal dynamics of periphyton in an agriculturally influenced river [effet du débit sur la dynamique temporelle des algues périphytiques dans une rivière influencée par les activités agricoles]. *Revue des Sciences de l'Eau* 16(1), 55–77.
- Lichstein, J. (2007). Multiple regression on distance matrices: A multivariate spatial analysis tool. *Plant Ecology* 188(2), 117–131.
- Mainstone, C. P. et W. Parr (2002). Phosphorus in rivers — ecology and management. *Science of The Total Environment* 282-283(0), 25 – 47.
- Martiny, J., J. C. R. Bohannan, B.J.M. and Brown, J. Fuhrman, J. Green, M. Horner-Devine, M. Kane, J. Krumins, C. Kuske, P. Morin, S. Naeem, L. Ø vreås, A.-L. Reysenbach, V. Smith, et J. Staley (2006). Microbial biogeography: Putting microorganisms on the map. *Nature Reviews Microbiology* 4(2), 102–112.
- McCormick, P. V. (1996). Resource competition and species coexistence in freshwater benthic algal assemblages. In *Algal Ecology: Freshwater Benthic Ecosystems*. Academic press.
- McKinney, M. L. (2006). Urbanization as a major cause of biotic homogenization. *Biological Conservation* 127(3), 247 – 260.
- Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, et H. Wagner (2011). *vegan: Community Ecology Package*.
- Pan, Y. et R. Stevenson (1996). Gradient analysis of diatom assemblages in western kentucky wetlands. *Journal of Phycology* 32(2), 222–232.
- Pan, Y., R. Stevenson, B. Hill, et A. Herlihy (2000). Ecoregions and benthic diatom assemblages in mid-atlantic highlands streams, usa. *Journal of the North American Benthological Society* 19(3), 518–540.
- Parlement Européen (2000). Directive 2000/60/CE du parlement européen et du conseil du 23 octobre 2000 établissant un cadre pour la politique communautaire dans le domaine de l'eau, J.O.C.E. du 22.12.2000, L327/1-L327/72. Technical report.
- Passy, S. (2007). Community analysis in stream biomonitoring: What we measure and what we don't. *Environmental Monitoring and Assessment* 127(1-3), 409–417.
- Peterson, C. et R. Stevenson (1990). Post-spate development of epilithic algal communities in different current environments. *Canadian Journal of Botany* 68(10), 2092–2102.

- Philibert, A., P. Gell, P. Newall, B. Chessman, et N. Bate (2006). Development of diatom-based tools for assessing stream water quality in south-eastern australia: assessment of environmental transfer functions. *Hydrobiologia* 572, 103–114.
- Potapova, M. et D. Charles (2007). Diatom metrics for monitoring eutrophication in rivers of the united states. *Ecological Indicators* 7(1), 48–70.
- Potapova, M. et D. F. Charles (2003). Distribution of benthic diatoms in u.s. rivers in relation to conductivity and ionic composition. *Freshwater Biology* 48(8), 1311–1328.
- Potapova, M. G. et D. F. Charles (2002). Benthic diatoms in usa rivers: distributions along spatial and environmental gradients. *Journal of Biogeography* 29(2), 167–187.
- Rahel, F. J. (2002). Homogenization of freshwater faunas. *Annual Review of Ecology and Systematics* 33, pp. 291–315.
- Rimet, F., J. Gomà, J. Cambra, E. Bertuzzi, M. Cantonati, C. Cappelletti, F. Ciutti, A. Cordonier, M. Coste, F. Delmas, J. Tison, L. Tudesque, H. Vidal, et L. Ector (2007). Benthic diatoms in western european streams with altitudes above 800 m: Characterisation of the main assemblages and correspondence with ecoregions. *Diatom Research* 22(1), 147–188.
- Smucker, N. J. et M. L. Vis (2011). Spatial factors contribute to benthic diatom structure in streams across spatial scales: Considerations for biomonitoring. *Ecological Indicators* 11(5), 1191 – 1203.
- Soininen, J. (2004). Determinants of benthic diatom community structure in boreal streams: The role of environmental and spatial factors at different scales. *International Review of Hydrobiology* 89(2), 139–150.
- Soininen, J. (2007). Environmental and spatial control of freshwater diatoms - a review. *Diatom Research* 22(2), 473–490.
- Soininen, J. (2010). Species turnover along abiotic and biotic gradients: Patterns in space equal patterns in time? *BioScience* 60(6), 433–439.
- Soininen, J., J. Heino, J. Lappalainen, et R. Virtanen (2011). Expanding the ecological niche approach: Relationships between variability in niche position and species richness. *Ecological Complexity* 8(1), 130–137.
- Soininen, J., R. McDonald, et H. Hillebrand (2007). The distance decay of similarity in ecological communities. *Ecography* 30(1), 3–12.
- Soininen, J. et P. Niemela (2002). Inferring the phosphorus levels of rivers from benthic diatoms using weighted averaging. *Archiv Fur Hydrobiologie* 154(1), 1–18.
- Soininen, J., R. Paavola, et T. Muotka (2004). Benthic diatom communities in boreal streams: community structure in relation to environmental and spatial gradients. *Ecography* 27(3), 330–342.

- Stoermer, E. F. et J. P. Smol (1999). *The diatoms: applications for the environmental and earth sciences*. Cambridge: Cambridge university press.
- Tison, J., Y. S. Park, M. Coste, J. G. Wasson, L. Ector, F. Rimet, et F. Delmas (2005). Typology of diatom communities and the influence of hydroecoregions: A study on the french hydrosystem scale. *Water Research* 39(14), 3177–3188.
- van Dam, H., A. Mertens, et J. Sinkeldam (1994). A coded checklist and ecological indicator values of freshwater diatoms from the netherlands. *Netherlands Journal of Aquatic Ecology* 28(1), 117–133.
- Vanormelingen, P., E. Verleyen, et W. Vyverman (2008). The diversity and distribution of diatoms: From cosmopolitanism to narrow endemism. *Biodiversity and Conservation* 17(2), 393–405.
- Verleyen, E., W. Vyverman, M. Sterken, D. A. Hodgson, A. De Wever, S. Juggins, B. Van de Vijver, V. J. Jones, P. Vanormelingen, D. Roberts, R. Flower, C. Kilroy, C. Souffreau, et K. Sabbe (2009). The importance of dispersal related and local factors in shaping the taxonomic structure of diatom meta-communities. *Oikos* 118(8), 1239–1249.
- Vyverman, W., E. Verleyen, K. Sabbe, K. Vanhoutte, M. Sterken, D. A. Hodgson, D. G. Mann, S. Juggins, B. V. De Vijver, V. Jones, R. Flower, D. Roberts, V. A. Chepurinov, C. Kilroy, P. Vanormelingen, et A. De Wever (2007). Historical processes constrain patterns in global diatom diversity. *Ecology* 88(8), 1924–1931.
- Wasson, J. G., A. Chandesris, H. Pella, et L. Blanc (2002). Definition des hydroécorégions françaises métropolitaine. approche régionale de la typologie des eaux courantes et éléments pour la définition des peuplements de référence d'invertébrés - rapport final. Technical report, Ministère de l'Ecologie et du Développement Durable - Cemagref GMA. Biol. des Syst. Aquat. Lab. Hydroécol. Quantitative Lyon.
- Winter, J. et H. Duthie (2000). Stream epilithic, epipellic and epiphytic diatoms: Habitat fidelity and use in biomonitoring. *Aquatic Ecology* 34(4), 345–353.



## Chapitre 3

# Les patrons de cooccurrences des diatomées françaises

*Ce chapitre présente un article en cours de rédaction, à soumettre après correction de l'anglais et concertation avec les co-auteurs.*

### 3.1 Buts et contexte de l'étude

À l'instar de l'étude du chapitre précédent, il s'agit ici d'étudier la structure des assemblages afin de déterminer quels sont les processus qui ont agi sur les diatomées. Nous nous intéresserons plus précisément aux structures de cooccurrence entre les différentes espèces. La coexistence des espèces en un lieu donné dépend principalement de la compatibilité de leurs niches écologiques fondamentales avec l'environnement, des relations biotiques et de leur capacité à s'installer dans le système (dispersion/migration et développement local).

L'étude des structures de cooccurrence entre les espèces a fait l'objet de développements de modèles nuls particuliers, qui consistent à mélanger les tableaux biologiques de présence/absence (Gotelli, 2000). Ce mélange est destiné à supprimer virtuellement l'effet des relations biotiques, mais élimine également l'effet de l'environnement sur les assemblages. C'est une des raisons pour lesquelles Gotelli et Ulrich (2012) et Götzenberger et al. (2012) ont récemment précisé les améliorations à apporter dans le domaine de ces modèles nuls.

Nous proposons dans cette étude une méthodologie originale utilisant la prédiction de chacune des espèces, indépendamment, en fonction de variables environnementales, et permettant de modéliser un fonctionnement gleason-ien des assemblages diatomiques (Gleason, 1926, ; voir page 16). Sa mise en œuvre, en association avec l'utilisation de modèles nuls classiques, nous permet de tester

si les associations entre les espèces sont explicables par les processus de niches des espèces et de déterminer de façon plus précise quels sont les processus qui ont agi sur la structure de cooccurrence des matrices biologiques.

Nous avons testé plusieurs composantes de la structure de cooccurrences entre les espèces :

1. la ségrégation ou l'agrégation globale des présences d'espèces,
2. la ségrégation ou l'agrégation des espèces selon leur similarité de traits biologiques
3. la nestedness globale (voir page 20).

## 3.2 Principaux résultats

Les assemblages de diatomées ont montré une structure globale d'agrégation entre les espèces, en particulier entre les espèces similaires en termes de traits biologiques. Cette agrégation a montré de forts liens avec la nestedness des matrices biologiques de présence/absence. Par ailleurs, les mécanismes agissant sur la coexistence des espèces semblent être différents entre les situations de bonne qualité de l'eau et celles des cours d'eau dégradés.

Il semble donc que l'occurrence des espèces de diatomées soit liée à des phénomènes impliquant la structure globale des assemblages et les dynamiques de colonisation à l'échelle nationale. La productivité locale des assemblages semble jouer un rôle important dans les dynamiques de migration passive des espèces.

### 3.3 Species-by-species prediction of the French diatom assemblages reveals the importance of non-environmental processes

Marius BOTTIN<sup>1\*</sup>, Didier ALARD<sup>2</sup>, Sovan LEK<sup>3</sup>, François DELMAS<sup>1</sup>, Michel COSTE<sup>1</sup> and Juliette TISON-ROSEBERY<sup>1</sup>

1 : Irstea, UR REBX, 50 avenue de Verdun, 33612 Cestas cedex, France

2 : UMR 1202 BIOGECO, Avenue des Facultés, 33405 Talence, France

3 : Laboratoire EDB, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse cedex 4, France.

#### 3.3.1 Introduction

Diatoms assemblages are known to react quickly and predictably to their environment (e.g. Smol et Stoermer, 2010). They have therefore been extensively used for biomonitoring. Large databases have been constructed, comprising biological assemblages and their main environmental drivers, in order to describe affinities between species and environmental conditions and to develop biomonitoring methodologies. An implicit model is underlying these methodologies, considering that environment is the only driver of species, and that response of each species is independent to the others.

However, other processes are known to shape community structure (e.g. dispersal, biotic relationships). Recent studies have revealed that diatom assemblages might show structures related to these processes (e.g. Martiny et al., 2006; Astorga et al., 2011, Bottin et al., submitted), although previously thought relevant only for macroorganisms (e.g. Fenchel et Finlay, 2004). Informations, datasets and models about the relations between species and environment should allow us to develop comprehensive models which consider not only environmental control of assemblages but also relationships between species, and the global structure of assemblages.

Assembly rules are simple comprehensive rules describing important biotic and abiotic factors leading to particular species/traits assemblages in communities (Weiher et Keddy, 1999). The main aim of assembly rules is to enunciate general principles for communities, in spite of the complex interactions acting on biological communities. Numerous studies about assembly rules have focused on the understanding of biotic relationships between species. For this aim, null models describing virtual communities without biotic interactions have been developed (Gotelli, 2000; Miklós et Podani, 2004; Ulrich et Gotelli, 2007) and widely used during the last decades.

Recently, the limits of such approaches have been discussed and directions for future research in this field have been given (Gotelli et Ulrich, 2012; Götzenberger et al., 2012). One of the major criticism about these null models is that they do not take environmental niches into account (Götzenberger et al., 2012; Gotelli et Ulrich, 2012; McGill et al., 2006). All the same, some studies have managed to assess cooccurrences between species while controlling for environmental preferences of species, by accounting for habitats (Azeria et al., 2012) or by constructing predictive models of occurrences from environmental variables (Ovaskainen et al., 2010).

Local coexistence of diatom species (and other micro-organisms) have been

largely understudied, maybe because of difficulties due to high diversity. To our knowledge, only Heino et Soininen (2005) have directly assessed competition patterns with these methods in freshwater diatom assemblages. They showed a segregative pattern in boreal headwater stream diatoms. Soininen et Heino (2007) showed that the most diverse diatom assemblages are characterised by more specialised species. This pattern could be explained by competition, but the authors argued that response of species to environmental gradients might also produce such patterns without considering biotic interactions.

Studies about species interactions have mainly focused on competitive relationships, however positive biotic interaction might be very important (Bruno et al., 2003). Positive biotic relationships are separated in two main phenomena : (i) *facilitation* is a direct or indirect influence of one species to another one and (ii) *niche complementarity* arises when the diversity of traits in a community increases the global functioning and causes the negative relationships (such as competition) to be higher between individuals of the same species than between individuals showing different functioning. Vanelslander et al. (2009) showed experimentally that facilitation and niche complementarity increased significantly the productivity of benthic marine diatom. Similar processes might also affect freshwater diatoms and result in global aggregative patterns in species occurrences.

Diatom assemblages are part of complex three-dimensional biofilms, although this property is often under-considered. This could play an important role by shaping biotic relationships in diatom assemblages. This structure is characterised by internal gradients of resources (Burkholder et al., 1990; Stevenson et Glover, 1993) due to overgrowing species which consume nutrients. Strong structure implications have already been demonstrated for successions of assemblages (Hoagland et al., 1982; Passy et Larson, 2011). In order to apprehend such processes, classification of species have been proposed depending on growth forms and location in the biofilm (Passy, 2007). These groups of species have been shown reliable to elaborate comprehensive functional relationships with environment (Passy, 2007) or to study successions (Passy et Larson, 2011).

Nestedness is an other pattern related to the cooccurrences of species. It corresponds to the degree of which poorer assemblages are proper subsets of the richest ones, and implies global mechanisms, such as ordered sequences of colonization and extinction of species on the different sites. This pattern have been shown to be strong in diatom assemblages despite a conjointly high number of species being independent from it (Heino et Soininen, 2005; Soininen, 2008). While the cooccurrences of species depend on environmental preferences of species, capabilities of species to colonize localities and biotic relationships, nestedness of diatoms seems to depend mainly on the colonization dynamic (Soininen, 2008). Analysing both cooccurrence and nestedness, while using environmental models should allow us to examine more precisely which processes drive the local coexistence of species.

Our main aim here is to unravel species coexistence patterns. Our method consists in checking whether models predicting independently species presences from environment lead to assemblage structures comparable to the real ones. If not, can differences in predicted structure be attributed to comprehensive processes like competition or complementarity? Random null models are also used to check simultaneously whether real assemblages show more segregation or aggregation between species than expected by chance. Specifically, we test

different hypotheses:

1. Concerning the global patterns of cooccurrence, we expect the real pattern to show more segregation than expected by chance, due to different environmental affinities between species. Depending on the importance of positive and negative species' interactions, real pattern would show less segregations than predicted from environmental variables, or more. If all species interactions take place at a finer scale than the grain of the study, we expect real patterns of segregation to be comparable to the predicted ones.
2. We expect the competition to be higher between species showing comparable biological traits, and facilitation and niche complementarity to promote assemblages with species presenting different biological traits. Then, if both positive interactions and negative interactions structure diatom assemblages, models should under-estimate segregations between species from the same genus or guild, and over-estimate segregations between species' pairs from different genera or guilds.
3. We expect species interactions to have a higher importance on common and abundant species, and then the patterns to be stronger on these species.
4. Since impacted sites have often a higher level of nutrients promoting high productivity, patterns of segregation or aggregation should be clearer on impacted sites. On the other hand, reference sites might show levels of nutrients more stable in time, which could allow species interactions to eventually structure assemblages, whereas assemblages from impacted sites are constantly reorganized due to changing productivity, avoiding species interaction to take place.
5. Relationships between nestedness and segregations between species have been specifically studied by Ulrich et Gotelli (2007), but remained unclear. They found both positive and negative relationships depending on the way virtual matrices were created. In the speciose diatom assemblages, we expect the eventual relationship to be negative, because nestedness results in aggregations of species in the richest sites. Comparison of nestedness and cooccurrence should allow us to discuss about the importance of the different processes driving the local coexistence of species. If colonization dynamics explain a large part of cooccurrence patterns, we expect the two patterns to be highly dependant. On the other hand, if biotic relationships rule a larger part of species cooccurrence patterns, we expect the relationship to be weaker, because biotic relationships are more likely to affect locally and independantly the richness of sites.

### 3.3.2 Material and methods

#### 3.3.2.1 Data

**Biological data** Our database consists of 836 samples collected in French rivers and streams (Fig 3.1) from 1979 to 2001 by Irstea (France) and the Centre de Recherche Gabriel Lippman (Luxembourg). All samples were collected and analysed by only two experts according to a standardized method (NFT 90-354) (AFNOR, 2000) from stones and during the low flow period, in order

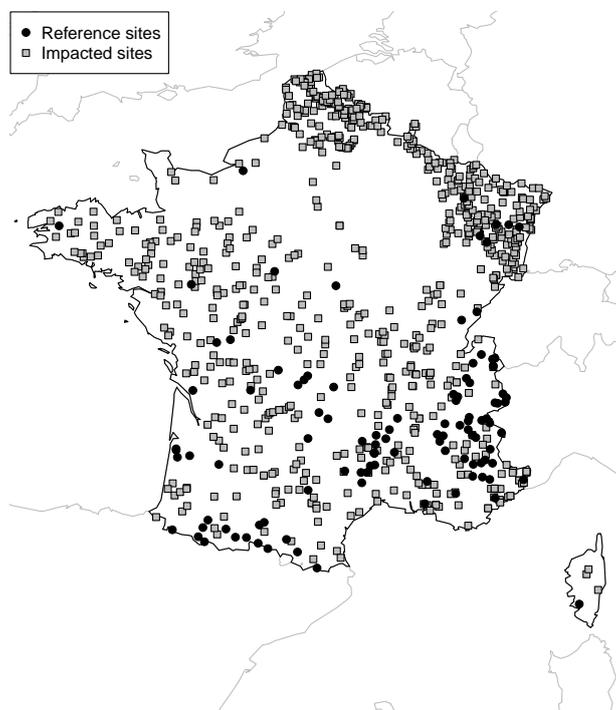


Figure 3.1: **Localization of sampling sites.**

to limit variability due to the season and to local factors like sampling support. Diatom taxa were identified at  $1000\times$  magnification (Leitz DMRD light microscope) (400 valves per slide) according to Krammer et Lange-Bertalot (1991) by examining permanent slides of cleaned diatom frustules, digested in boiling  $\text{H}_2\text{O}_2$  (30%) and  $\text{HCl}$  (35%) and mounted in a high refractive index medium (Naphrax, Northern Biological Supplies Ltd, UK;  $\text{RI} = 1.74$ ). Under-species forms (form, varieties, subspecies) were lumped together in order to obtain a biologically coherent dataset and to limit differences between analysts. Species present in less than five sites were excluded from data because of difficulties to handle them by statistical models. We also tested to exclude species presences from sampling-units when relative abundance was lower than 5 %, but since the resulting segregation patterns were similar, results were not included. The resulting biological data matrix, expressed in presence/absence (0/1) of species, was composed of [836 sites  $\times$  426 species]. The influence of water quality on cooccurrence patterns was assessed by separating reference sites (near-pristine sites used as references in European bioassessment programs) and impacted sites. After deleting species which were present in less than 5 sites the reference biological matrix was composed of [115 sites  $\times$  137 species] and the impacted biological matrix of [721 sites  $\times$  392 species].

**Physical-chemical data** Physico-chemical data for samples collected at the study sites were provided by the Water Agencies. Average values of each variable were calculated from 2 month's data: that of the diatom sampling month and of the previous one. For the sites sampled in 2001 out of the institutional network, the values of the physico-chemical parameters of water quality were determined at the sampling date with two litres of water collected in a free flowing area near the middle of the stream, preserved at 4°C then analysed within 24h, according to standardised AFNOR protocols (by the Cemagref laboratory and the Laboratoire Santé-Environnement in Lyon).

The physico-chemical data is composed of [836 sampling units  $\times$  16 variables], namely altitude (m), distance from the source (m), slope (%), pH, conductivity at 20 °C ( $\mu$ S/cm), dissolved oxygen (mg/l), total inorganic nitrogen (mg/l), phosphate (mg/l), calcium (mg/l), hydrogen carbonate (mg/l), temperature (°C), suspended solids (mg/l), ammonium (mg/l), nitrate (mg/l), nitrite (mg/l), alkalinity (meq/l).

**Groups of species** In order to disentangle the global patterns of species cooccurrences, we systematically analysed the different observed patterns for different groups of species.

**Guilds** Diatom guilds were introduced by Passy (2007), it consists of three classes of species, depending on the growth forms and of positions of species in the biofilm. The three classes are “high profile” for species which are present in overstory of the biofilm, “Low profile” for species which are limited to the understory of the biofilm, and “Motile” for species with the ability to move in the biofilm. The guilds were mainly attributed at the genus resolution, except for species known to have a particular growth form in their genera, following Passy (2007). Additional informations were included from various sources (personal communications from Coste, Morin and Passy Hoagland et al., 1982; Katoh, 1992; Eulin, 1997; Tuji, 2000; Kelly et al., 2005). 85 species were considered as “low profile”, 105 as “high profile” and 160 as “motile”. We didn't have sufficient information for attributing a guild for the 76 other species, which were grouped as “Unknown guild”.

**Scarcity-commonness types of species** Depending on occurrences and abundances of species, one can separate species in different groups. Gillett et al. (2011) adapted definitions of “core” species (highly abundant and very common) and “satellites” species (lowly abundant and rare) from Hanski (1982), and those of “rural” species (lowly abundant and common) and “urban” species (highly abundant and rare) from Söderström (1989) for diatoms. Only a very small number of both “rural” and “urban” species (about 15) were found in our dataset, which could lead to a biased analysis of the pattern of these species. Therefore we redefined only core and satellite species on our total dataset as the following:

- Core species have a maximum relative abundance greater than 10% and are present in more than 200 sites on 836 (48 species).
- Satellite species have a maximum relative abundance lower than 2.5 % and are present in less than 25 sites on 836 (131 species).
- “Others” species are neither core nor satellite (247 species).

### 3.3.2.2 Random null models

In order to verify the non-randomness of co-occurrence pattern, we used null model to create sets of random matrices. Several null models exist depending on constraints in terms of fixed sums of rows/columns (Gotelli, 2000). Null-Model with fixed row and column sums are more difficult to compute, but showed interesting statistical properties results (Connor et Simberloff, 1979; Gotelli, 2000; Gotelli et Entsminger, 2001), because it preserves differences in species richness among sites and differences in occurrence frequencies among species. Among these algorithms, the “trial swap” algorithm (Miklós et Podani, 2004) have been considered as the most suitable and the best designed because authors produced evidences that resulting matrices were equiprobable samples of all possible matrices. This type of null model is called "NullModFF" in this study.

Miklós et Podani (2004) suggested that the number of trials in each step should be set such that the expected number of swaps equals twice the number of presences in the matrix, but we found inconvenient results this way (some matrices equivalent to the preceding ones) so we used ten fold the number of presences in the matrix. Before computing, matrices were initialised with twice the number of swaps during actual computing. We used this algorithm to sequentially obtain 1000 random matrices. This algorithm was computed with an adapted version of the C-program provided in the appendixes of Miklós et Podani (2004). That code was used in R.

As noted by Azeria et al. (2012) null models which fix both sums of rows and columns are considered biased towards revealing segregated cooccurrences, but conservative in detecting aggregated patterns of co-occurrences, while the ones which fix only sum of species occurrences have shown converse properties. We therefore applied both these types of null models in order to test whether real communities cooccurrences pattern differ from randomness either in an aggregative or a segregative way. The null model in which sites are equiprobable for presences of species with fixed total occurrences is called NullModFE.

### 3.3.2.3 Predictive models

**Presence predictions** In order to describe variability of indices explainable by environmental differences between sampling units, we predicted species presences based on physical-chemical data. Predictions were applied in a leave-one-out procedure, using logistic regressions and Random Forest models (Breiman, 2001). We obtained presence scores for each species and each sampling unit.

**Multiple logistic regressions** Multiple logistic regressions describes the relationship between a combination of environmental predictive variables and a binary response variable (Hosmer et Lemeshow, 2000), here : presence or absence of species. An automatic stepwise procedure based on Akaike Criterion was used to select relevant variables for each species. This procedure was applied by using *stepAIC* and *glm* functions of the *MASS* package in R (Venables et Ripley, 2002). Then predictive models using these variables were applied in a leave-one-out procedure, i.e presence score for each sampling unit is predicted using all presence data but itself. Responses from models fitted on all site but the one being predicted were arranged in a “pseudo-probability” matrix.

**Random Forest** Random Forests belong to the decision trees techniques family. Decision trees allows a data partition in order to assign objects (here : vectors of environmental variables) into classes (here: presence, or absence of a species). Random Forest consists in constructing a lot of these decision trees, each with a randomly selected amount of data. Output results of these random forests are votes : every decision tree predict a class for the input vector. The prediction of a class by one is considered as a "vote" for this class. Random Forest models were computed with the *randomForest* package in R (Liaw et Wiener, 2002). We used 500 trees in each forest. Due to robustness of random forest, we did not use any early selection of variables for these models, but directly the leave-one-out procedure. We used Random Forest as supervised classification to predict score (pseudo-probabilities) of presence and absence. We kept proportions of predicted-presence votes as output of models. These proportions were further considered as "pseudo-probabilities".

**Predicted pseudo-community matrices** Predictions leded to matrices, similar to the real presence/absence matrices, but with values comprised between 0 and 1 (instead of only 0/1 values). With these scores (pseudo-probabilities), we applied samplings taking these pseudo-probabilities into account.

In a first procedure, we drawn as much presences by sites (richness fixed) depending on species' pseudo-probabilities without replacement. This was repeated 1000 times for each site in order to produce 1000 pseudo-community presence matrices. For random forest and logistic regressions these procedures were respectively called "RandForPF" and "LogitPF".

A second procedure consisted of drawing equally distributed between 0 and 1 scores to create matrices of the same dimensions than the pseudo-probability matrices. Pseudo-probabilities higher than the score were considered as presences in a pseudo-community matrix. This was repeated 1000 times to obtain 1000 pseudo-community matrices. For random forest and logistic regressions these procedures were respectively called "RandForPP" and "LogitPP".

Table 3.1: Summary of models. Each model was used to obtain 1000 pseudo-community matrices for each dataset

	random	Logistic regression	Random Forest	richness fixed	occurrences fixed
NullModFF	×			×	×
NullModFE	×				×
LogitPF		×		×	
LogitPP		×			
RandForPF			×	×	
RandForPP			×		

### 3.3.2.4 Indices

**Checkerboard Units index and C-score** Checkerboard Units index consists in numbering species pairs forming perfect checkerboards ( $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$  or  $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ) in the overall matrix (considering all possible couples of species and all couples

of sampling units). It actually measures the importance of segregative patterns between species. Considering 2 species, searching for perfect checkerboards patterns is equivalent to calculate:

$$CU_{ij} = \sum (S_i - Q)(S_j - Q)$$

where  $S_i$  and  $S_j$  are the numbers of sites where the species  $i$  and  $j$  are respectively present, and  $Q$  is the number of sites where both species are present. The sum of  $CU_{ij}$  considering all species pairs gives the global index for a biological matrix, but it is also possible to separate the calculation between different groups of species. In the different communities and pseudo-communities, we calculated Checkerboard units between species from the same group (e.g. Motile-Motile) and between species from different groups (e.g. Motile-High profile). In order to disentangle differences of global patterns between real communities and pseudo-communities from the different models, an index of over-estimation (or under-estimation) of Checkerboard units in pseudo-communities was calculated for each pair of species' groups as follows:

$$f_{MOD} = \frac{CU_{MOD}}{CU_{REAL}}$$

where  $CU_{REAL}$  and  $CU_{MOD}$  are respectively the number of checkerboard units in the communities and pseudo-communities.

For more readability global C-scores of the matrices were presented. C-scores consists of averaged number of Checkerboard units by species and are calculated as follows:

$$C = \frac{\sum CU_{ij}}{N(N-1)}$$

where  $N$  is the number of species of the matrix.

**NODF index** The NODF index (Nestedness metric based on Overlap and Decreasing Fill; Almeida-Neto et al., 2008) has been proved to be consistent with the original concept of nestedness. It was calculated using the “nestednodf” function of the package “vegan” (Oksanen et al., 2011) in R, and the values of NODF were compared with the global C-scores of the different community and pseudo-community matrices.

### 3.3.3 Results

#### 3.3.3.1 Richness among guilds and among species' scarcity types

Richness were distributed differently among the different groups of species depending on the water-quality (figure 3.2). Reference sites showed globally lower richness than the impacted ones.

Concerning the guilds, the distribution and values of High profile species richness were similar in both impacted and reference conditions. Numbers of Low profile and motile species were lower in reference conditions. Numbers of motile species were higher than richness of High profile species in impacted sites, lower in reference sites. Richness of motile species showed a bimodal distribution in reference conditions, with a local maximum around 5 species and a local maximum around 15 species.

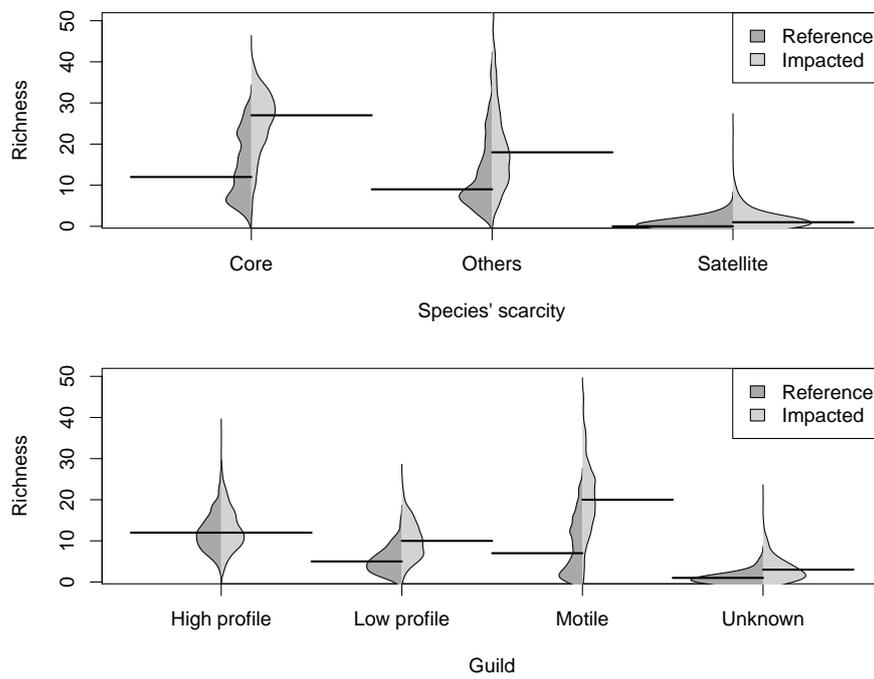


Figure 3.2: Comparison of richness distributions of the different species' groups (guilds and scarcity-related groups of species) in impacted and reference real communities. Lines crossing the distributions show the median of distributions.

Concerning the species' scarcity types, the most different density distributions of richness between reference and impacted sites were found for the core species. Core species were the richest group in both references and impacted conditions, with higher relative richness in the impacted sites. Distribution of core species richness was bimodal in the reference dataset, with a mode around 8 species and an other one around 23 species.

### 3.3.3.2 Global patterns of cooccurrences

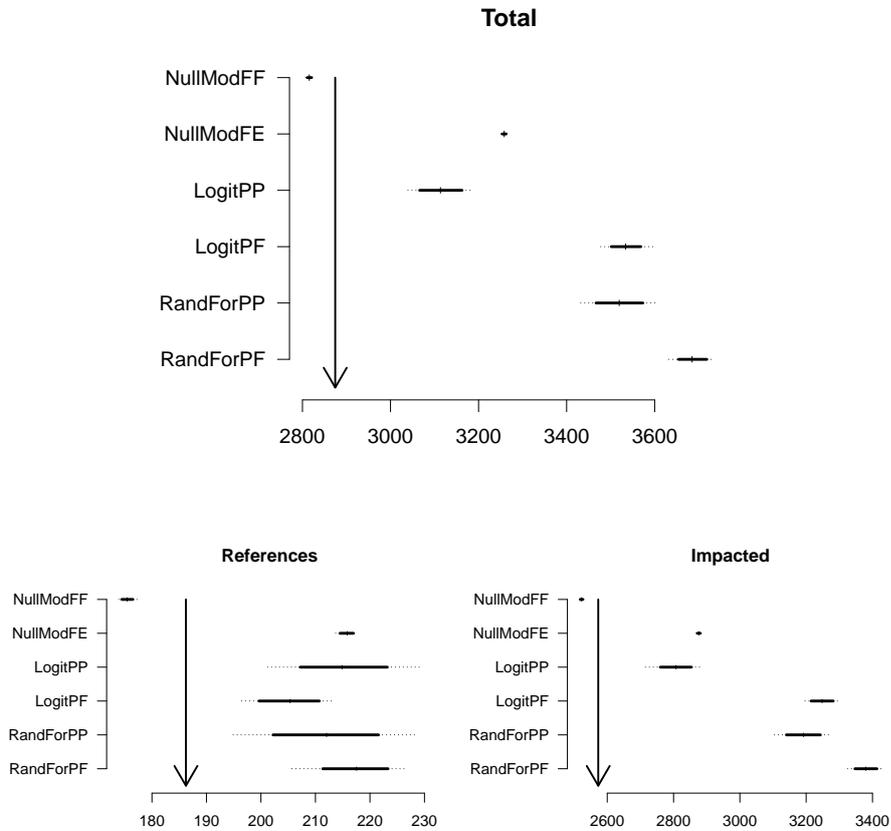


Figure 3.3: Distribution of the C-scores of pseudo-communities matrices, in comparison to the C-score of the real matrix, in the different datasets. Bold segments represent 95% of the distributions. Perpendicular marks represent the median of distributions. Arrows indicate the values of real-community matrices.

Models in which richness of sites and/or total occurrences of species departed from reality (all but NullModFF) produced patterns of cooccurrence more segregative than reality in the three datasets (figure 3.3). By contrary, NullModFF procedure led to a more aggregative pattern.

Sorting pseudo-communities models by values of resulting C-score led to the same sequence in the total dataset and the impacted dataset. In both of them, constraining richness of sites in the environmental models led to more segregative patterns (LogitFP and RandForFP) and models using logistic regressions resulted globally in more aggregative patterns than the ones using Random Forest models. In these datasets, only LogitFP pseudo-communities showed lower C-scores than NullModFE among the environment-accounting models.

Comparison of the different pseudo-communities resulted in a different pattern for the reference dataset. Pseudo-communities showed comparable C-scores, with overlapping distributions, in all models but NullModFF. No clear differences between environmental models constraining richness or not were found.

### 3.3.3.3 Deconstructing cooccurrence patterns

Averaged  $f$ -values presented in tables 3.2, 3.3 and 3.4 were characterised by very low standard errors of the means (always lower than  $1 \times 10^{-4}$ ). The ones on which we focused on this part were always different to the global  $f$ -values presented in the tables.

Table 3.2: Over-estimation (or under-estimation) of checkerboard units in the different models depending on species' guilds. Values shown are the averaged  $f$  values for the 1000 pseudo-communities.

	NullModFE	NullModFF	LogitPP	LogitPF	RandForPP	RandForPF
<b>Reference dataset</b>						
Global	<b>1.16</b>	<b>0.94</b>	<b>1.15</b>	<b>1.1</b>	<b>1.14</b>	<b>1.17</b>
Low-High	1.09	0.87	1.12	1.03	1.11	1.1
Low-Motile	1.28	1.04	1.2	1.13	1.17	1.18
High-Motile	1.13	0.92	1.13	1.1	1.13	1.15
Low-Low	1.17	0.94	1.2	1.06	1.15	1.12
High-High	1.05	0.84	1.09	1.05	1.1	1.13
Motile-Motile	1.48	1.22	1.29	1.29	1.25	1.32
<b>Impacted dataset</b>						
Global	<b>1.12</b>	<b>0.98</b>	<b>1.09</b>	<b>1.26</b>	<b>1.24</b>	<b>1.31</b>
Low-High	1.09	0.95	1.07	1.23	1.21	1.27
Low-Motile	1.11	0.97	1.07	1.21	1.2	1.25
High-Motile	1.09	0.95	1.07	1.24	1.21	1.28
Low-Low	1.14	1	1.1	1.23	1.23	1.27
High-High	1.14	0.99	1.11	1.32	1.28	1.38
Motile-Motile	1.17	1.02	1.13	1.28	1.26	1.32

**Guilds** In impacted conditions,  $f$ -values of all models showed a remarkably higher over-estimation of segregations between the species from the motile guild (Table 3.2). NullModFF under-estimated the general pattern of segregations, but still over-estimated segregations between the species of the motile guild. Reference conditions showed globally comparable  $f$ -values for the different guilds among models, with a slightly higher over-estimation of segregation between species from the high-profile guild in environment accounting models and a slightly higher over-estimation of species from the motile guilds in random models.

Globally segregations between species from the same guild were more over-estimated than segregations between species from different guilds, although this pattern was less often found for the low-profile guild.

Table 3.3: Over-estimation (or under-estimation) of checkerboard units in the different models depending on species' genera. "Among" correspond to the sum of checkerboard units of species from the same genus, "between" concerns species pairs from different genera. Values shown are the averaged  $f$  values for the 1000 pseudo-communities.

	NullModFE	NullModFF	LogitPP	LogitPF	RandForPP	RandForPF
<b>Reference dataset</b>						
Global	<b>1.16</b>	<b>0.94</b>	<b>1.15</b>	<b>1.1</b>	<b>1.14</b>	<b>1.17</b>
between	1.15	0.93	1.15	1.1	1.13	1.16
among	1.35	1.09	1.23	1.23	1.23	1.27
<b>Impacted dataset</b>						
Global	<b>1.12</b>	<b>0.98</b>	<b>1.09</b>	<b>1.26</b>	<b>1.24</b>	<b>1.31</b>
between	1.11	0.98	1.09	1.26	1.24	1.31
among	1.23	1.07	1.18	1.35	1.31	1.39

**Genera** In all models and datasets, pseudo-communities were characterised by an higher over-estimation of segregations between species from the same genus (see table 3.3).

Table 3.4: Over-estimation (or under-estimation) of checkerboard units in the different models depending on species scarcity types. Values shown are the averaged  $f$  values for the 1000 pseudo-communities.

	NullModFE	NullModFF	LogitPP	LogitPF	RandForPP	RandForPF
<b>Reference dataset</b>						
Global	<b>1.16</b>	<b>0.94</b>	<b>1.15</b>	<b>1.1</b>	<b>1.14</b>	<b>1.17</b>
Satellite-Core	1.01	0.81	1.09	0.98	1.13	1.17
Satellite-Others	1.19	1.04	1.13	1.02	1.19	1.3
Core-Others	1.1	0.88	1.14	1.08	1.11	1.12
Satellite-Satellite	1.27	1.15	1.12	0.98	1.21	1.39
Core-Core	1.52	1.19	1.27	1.34	1.28	1.34
Others-Others	1.11	0.94	1.14	1.06	1.12	1.16
<b>Impacted dataset</b>						
Global	<b>1.12</b>	<b>0.98</b>	<b>1.09</b>	<b>1.26</b>	<b>1.24</b>	<b>1.31</b>
Satellite-Core	1.15	0.96	1.15	1.41	1.48	1.61
Satellite-Others	1.09	1.03	1.09	1.44	1.38	1.61
Core-Others	1.13	0.95	1.1	1.22	1.23	1.26
Satellite-Satellite	1.04	1.03	1.08	1.54	1.49	1.83
Core-Core	1.21	1.01	1.14	1.19	1.14	1.14
Others-Others	1.08	0.99	1.06	1.28	1.22	1.33

**Species types** All the models, when applied on the reference dataset, over-estimated particularly the segregations between core-species (see table 3.4). Random forest models (RandForPF and RandForPP) and random null models (NullModFF and NullModFE) showed also a higher over-estimation of the segregations concerning the satellite species' pairs.

Concerning impacted conditions, pseudo-communities from environmental predictions showed a higher over-estimation of segregations implying satellite species (between satellite species and between satellite species and the others), except with LogitPP. Null models led to a higher segregative pattern between the core species.

### 3.3.3.4 Relationship between cooccurrence patterns and nestedness

NODF of the real matrices were higher than those calculated on the pseudo-communities except for NullModFF in impacted sites and in the total dataset. Then the nestedness can be considered higher than expected by environmental preferences of species, but lower than expected by chance when constrained richness and sum of occurrences are applied. In reference condition, real NODF was the highest.

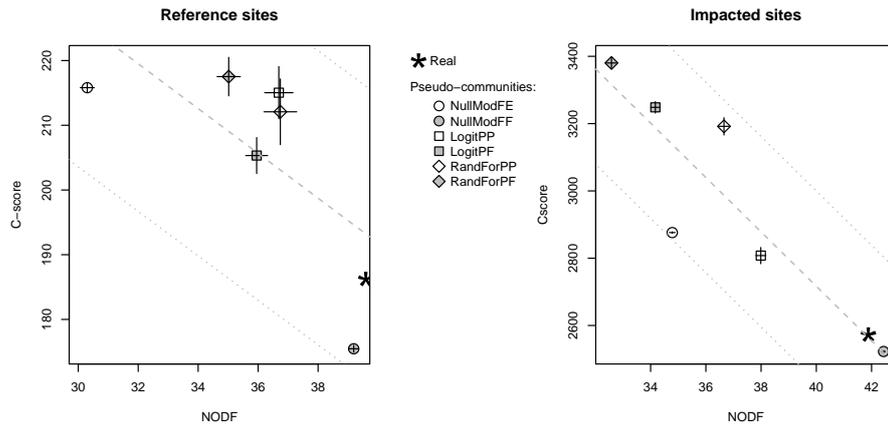


Figure 3.4: Relations between C-scores and NODF of real and pseudo-communities depending on the water-quality. Lines crossing the symbols represent standard deviations of the pseudo-community indices. Dashed grey lines represent linear models fitted on the 6000 pseudo-communities. Dotted grey lines represent predictive confidence interval of these models.

Relations between NODF and C-score indices are shown in figure 3.4. In both reference sites and impacted sites, relations between these indices were negative in pseudo-communities. However the relation appears to be stronger when pseudo-communities are constructed from impacted sites dataset ( $R^2 = 0.76$   $P < 0.001$ ) than on reference sites dataset ( $R^2 = 0.39$   $P < 0.001$ ). When linear models are fitted on pseudo-communities between NODF and C-score, real communities are both included in the predictive confidence interval. That is particularly true for the impacted condition, in which the predicted C-score (2564.38) is very close to the real one (2572.84).

## 3.3.4 Discussion

### 3.3.4.1 Competition

Our results did not confirm any of the expected results in case of structuring competition: (i) the real cooccurrence showed more aggregative patterns than the ones predicted by the environmental models, (ii) the different models over-estimated more the segregations between similar species (from the same genus or guild) than between more different species and (iii) the relationship between segregations and nestedness was strong, showing that other processes than only

competition can explain differences in cooccurrence patterns. One could argue that our models were fitted on realized niches of species which could already account for effect of competition, but we assume that if competitive exclusions had played an important role, errors of the models would have cause converse patterns. These results are congruent with the current view of microbial communities, in which interspecific exclusions have been presumed low because of different properties:

1. Due to high dispersal ability (e.g. Finlay, 2002; Fenchel et al., 2004) frequent recruitment avoid complete exclusions.
2. Unicellular microorganisms perceive their environment with a very fine resolution (Azovsky, 2002). Since most of lotic diatom studies use sampling protocols in which sampling from different stones are lumped together, species interaction might be hidden at the considered grain.
3. Disturbances are more frequent and more effective at this scale, interrupting competition before local extinctions (Reynolds, 2006)
4. More specifically for diatoms, tolerance between species has been shown during the succession process (McCormick et al., 1991; Passy et al., 2011).

These considerations are linked to the scale of studies, now yet our sampling design was constrained by standardized sampling protocols in which biological material from several stones are lumped together. Our results show that, at the usual sampling grain, competitive exclusions effect can be overlooked, or at least considered lower than converse processes. However conclusions about the absence of competitive exclusions at a finer scale would be speculative at present, and would require a particular sampling design.

#### 3.3.4.2 Positive biotic relationships and niche complementarity

The global aggregation patterns (compared to what environment predicts), associated with our knowledge of the diversity of diatom niches and the complex tridimensional structure of biofilms might look in the first place like an evidence of niche complementarity. Niche complementarity would promote locally rich and diversified assemblages, and explain the aggregation patterns. However, environmental models over-estimated more the segregations between species from the same genus, and often between species from the same guilds. These results highlight an aggregation of species from the same genus or the same guild in particular sites, more than what is predictable by the described environment. Such pattern suggests that some properties of sites or communities filters particular traits of the concerned guilds or genera. Among the properties of sites, one explanation can be that our environmental description is not complete. Indeed, some environmental descriptors are missing which have been described as important, e.g. light intensity (Lange et al., 2011), toxic contamination (Morin et al., 2009), grazers (Pan et al., 1994). However we assume that adding new predictor would have resulted in even more segregative patterns due to more precise description of the environmental preferences of species. Thus, a more probable explanation is that properties of the assemblages or species themselves cause an aggregation of similar species. Different processes could explain such patterns:

1. Leflaive et Ten-Hage (2009) reviewed diatom allelopathic mechanisms of interactions. This phenomenon is not well studied in freshwater biofilms, but several studies have shown diatom abilities to produce biologically active compounds. Such compounds could affect positively close species, and explain their aggregations, but to our knowledge, no such mechanism has been described yet in freshwater biofilms.
2. The dispersal abilities of species due to properties of genera (e.g. size) or guilds (e.g. size, abilities to loose from substrat) might explain both the global aggregation patterns and the particular aggregation between similar species. Reynolds (2006) showed that microorganisms communities were prone to frequent perturbations, which would reinforce the importance of dispersal abilities, due to constant possibilities of species to install in empty localities or reduced assemblages. This explanation is coherent with the strong relationship between cooccurrences patterns and nestedness (see discussion about nestedness).
3. In higher plant communities, successional dynamics lead to segregative patterns, due to clear replacement of species depending on the successional state. However, studies on diatoms have shown that successional dynamics were not characterized by the exclusion of earlier species, because of tolerance mechanisms (McCormick et Stevenson, 1991; Passy et Larson, 2011). Thus, the successive advantages of species depending on species traits and successional stages might play a role in the particular aggregation of similar species, without affecting the global aggregative patterns.

#### 3.3.4.3 Nestedness

Nestedness has often been described as a result of metacommunity processes (Ulrich et al., 2009). Thus, this pattern should be tested on close enough sites to permit a significant migratory flow of species between them. Limitations of diatom dispersal is still a controversial issue (Finlay, 2002; Fenchel et Finlay, 2004; Martiny et al., 2006), but a previous study (Bottin et al., submitted) showed that even when considering 12 smaller regions in France, diatom assemblages show strong spatial structures unrelated to the spatial structure of environment. The national extent of this study might then be inappropriate for assessing specifically nestedness and its causes. Here we used nestedness measurement mainly for studying the link between segregations patterns and a more global pattern of species' presence distribution. A specific analysis of nestedness pattern would require a multi-scale approach, in order to determine the appropriate metacommunity extent. In order to avoid wrong conclusions about nestedness specifically, we will only discuss the insights nestedness can provide in comparison to cooccurrence patterns.

The significant negative relationships between cooccurrences indexes and nestedness indexes in the models shows that the community structure variations implied by the models affect both patterns at the same time. Due to large population sizes and high among-sites dispersion, diatom assemblages are more colonization-dominated than extinction-dominated and Soinenen (2008) argued that diatom nestedness was stongly related with colonization dynamics. The relationship between nestedness and the aggregation patterns of assemblages

suggests that cooccurrences of species sites is at least partly related from colonization dynamics.

Nestedness is highly dependant on the marginal sums of matrices (richness of sites and sum of species' occurrences). In the total dataset (differences between reference and impacted sites will be discussed below), the different constraints applied to the pseudo-communities in terms of marginal sums resulted in different cooccurrence patterns. This result, associated with the relationship between nestedness and cooccurrence patterns emphasizes that distributions of richness between sites could explain by itself an important part of the cooccurrence patterns. Fixing both richness and sums of species' occurrences resulted in more aggregations of species than fixing only the sums of occurrences, because species are evenly distributed in the second case, resulting in lower overlaps in species distributions. On the other hand, fixing the richness on the environmental accounting pseudo-communities resulted in even more segregative patterns. Fixing richness in these models might have resulted in a selection of environmentally concordant species in pseudo-communities, whereas richness in real communities might be hold more importantly by widespread generalist species. These considerations are concordant with the hypothesis of factors independent from environment (e.g. colonization) shaping the structure of assemblages.

#### 3.3.4.4 Differences between impacted and reference conditions

C-score of the different pseudo-communities in impacted dataset showed the same pattern as in the total dataset, due to the high proportion of impacted sites in the total dataset. In reference condition however, all models but ModNulFF showed similar C-scores, even when richness of sites were fixed in environment accounting pseudo-communities. These results emphasize different processes acting on cooccurrence pattern depending on water quality. In reference conditions, richness might be hold by species more dependant on the environment, whereas impacted site richness might be due to species which are selected by other factors than environment and causing models to fail predicting them (see discussion above).

**Nestedness** In the reference dataset, the linear model between C-score and NODF lead to a quite different predicted values from the real C-score, so we can presume that the real segregation structure of communities is only slightly constrained by the nestedness pattern. Conversely, in the impacted dataset, the relationships is stronger, and the predicted C-score for real communities is very close to the real value. In impacted condition, it seems that nestedness patterns and segregation patterns are then strongly related, not only in the models, but also in the real communities.

Impacted sites are generally characterised by a high nutrient level, which might allow tolerant species to develop efficiently huge populations. Such large population sizes in the impacted sites would strengthen the passive dispersion process of diatoms and explain the differences between reference and impacted sites in terms of colonization dynamics.

On the other hand, reference condition assemblages might be characterized by more stable and lower productivity levels, allowing specialized species to install. The reference systems would probably show lower population sizes, with a more stable composition due to a higher nutrient stress. Then, the

colonization dynamics in these sites might be weaker and allow other processes to structure partly the assemblages, explaining the weaker relationships between nestedness and aggregation patterns in the reference conditions.

**Guilds and species' scarcity groups** Reference sites showed a weaker global richness, and the distribution of richness among the different groups of species was also different depending on water quality. Richness of high profile guild species did not seem to be dependant of the water-quality, while motile species and low-profile species (to a lesser extent) showed a higher richness in impacted sites. Passy (2007) and Lange et al. (2011) studied the effects of nutrient enrichment on the relative abundance of the different guilds. They found a higher development of high profile species under nutrient-rich conditions. The overgrowing of these species led to a concomitant limitation of low profile species due to lower nutrient availability and an increase of motile species because of their ability to adapt their positions depending on internal nutrients gradient in thick biofilms which forms in such conditions. Thus our results about motile species are concordant with theirs, but high and low profile have shown converse patterns with our data. Such differences might be due to presence data versus relative abundances. Our apparent equal high profile guild richness under both reference and impacted might hide the expected higher relative abundances of high profile species in impacted sites. The overgrowth of species in nutrient rich biofilms might lead to more complex internal gradients and locally more diverse nutrient availabilities, explaining the higher richness of low profile species, while their local relative abundance might decrease nevertheless.

Motile species were more aggregated between them in real communities than what was expected by environmental preferences of species in models. Passy (2007) and Lange et al. (2011) have shown that the advantage of being motile for a species is related to the three-dimensional structure of biofilm depending on nutrient level. Such indirect effect induces difficulties of prediction for species-by-species models which might have resulted in a wider distribution between these globally nutrient-poor sites.

Concerning scarcity groups, the richness of impacted sites were relatively more hold by core species, pointing out a more important relationship between regional distribution and local richness of species. This is concordant with a strengthened migration process, due to higher nutrient levels and larger population sizes. Reference sites composition of assemblages, by contrary might reflect more complex gradients.

The environmental models showed a very high overestimation of segregations between the satellite species in the impacted dataset, which emphasize a strong aggregation in real data, not explainable by environmental preferences of species. Interestingly, that strong overestimation applied also between satellite species and both the others scarcity types. That pattern suggests that these species are aggregated in sites which are rich of each types of species, whereas models predicts them more dispersed in the different sites. Such sites might be the less prone to hydrological perturbations and be subject to an accumulation of richness. Satellite species have lower dispersal abilities, and are more likely to install in stable biofilms. This is coherent with the relations between cooccurrences patterns and nestedness, because rare species occurs more likely in the most rich assemblages. From a more practical point of view, rare species might

provide insightful informations about the ecological conditions of assemblages, as suggested before (Lavoie et al., 2008; Gillett et al., 2011), and thus should not be suppressed in the bioassessment methodologies.

#### 3.3.4.5 The use of species-by-species predictive models

The use of predictive models allowed us to test whether coexistence between species were explainable by the only responses of species to environmental conditions. The use of pseudo-probabilities as output of these models and the consecutive drawing to form pseudo-communities allowed us to account for the variability of the predictions, and to test more efficiently the structure of assemblages. This methodology result in the modelling of pure gleasonian pseudo-communities (Gleason, 1926), where species assemble haphazardly and independently. Thus it provides a convenient null hypothesis for testing whether the community parameters can affect the species. These models can be used simultaneously to the classical null models to verify whether the patterns considered as significantly different from expected by chance would be explainable by environmental responses of species. In our study, the only use of classical random null models would have resulted in different and insufficient conclusions because of the environment-induced checkerboard distributions. Gotelli et Ulrich (2012) and Götzenberger et al. (2012) considered the accounting of environmental processes in null models as one of the challenge for further use. Our methodology might provide an answer for species which respond to gradients, while habitat constrained null models (e.g. Azeria et al., 2012) would be sufficient for species prone to more discrete habitat filtering. If such models were to be used in another study, we suggest to use several methods for predicting pseudo-probabilities of species because the different methods (Logistic regression and random forest) showed significantly different patterns of cooccurrences.

Many biomonitoring programs aim to assess the quality of ecological conditions of freshwater systems by comparing impacted communities to reference communities, i.e. communities present on the site if conditions were pristine. Some of the methods use near-pristine communities to compare directly with impacted communities, but others use virtual predictions of communities under pristine conditions, in order to avoid comparisons with "naturally different" communities. Most of these predictive methods use a classification-then-prediction design, where reference sites are separated among different groups and models predict which reference group impacted sites should be compared with (e.g Tison et al. 2007 ; Grenier et al. 2010 ; RIVPACS-like methods : Chessman 1999, Moss et al. 1999). However, Hallstan et al. (2012) compared classification-then-prediction and species-by-species predictive methods for bioassessment, and suggested that species-by-species models could outperform RIVPACS-type models, and could therefore provide a valuable alternative. However, none of the predictive species-by-species models used in our study resulted in an assemblage structure similar to the real one. These results demonstrate that the use of such models would result in biased comparison between reference communities and the predicted ones, at least for diatoms. We suggest that predictive biomonitoring methods should focus on general properties of diatom assemblages, e.g. proportion of traits, richness or diversity, but should not predict expected presences of species independently. Another solution might be the use of models

which predict species' presence simultaneously, such as the multiresponse artificial network methods, as suggested by Olden et al. (2006), or hierarchical modelling, as suggested by Ovaskainen et Soininen (2011).

### 3.3.4.6 Conclusions and perspectives

The environmental prediction of species method used in this study might be considered as a pure gleasonian (Gleason, 1926) null hypothesis of community structuring. The results revealed that in addition to the multiple independent responses of species to environment, diatom assemblages are partly structured by migration dynamics and complex clementsian factors, i.e. community properties shaping species responses to environment (Clements, 1916). Biomonitoring methodologies usually infer environmental conditions from individual species ecological preferences. Like our models, it underlies a pure gleasonian view of assemblages, which might lead to biased interpretation because of the additional structuring factors described in this study. However, accounting for these factors would require a more precise description of their effects, and further investigation is needed in that direction.

Future research should attempt to identify the determinants of the nestedness patterns at the appropriate scale. Determining both the factors which causes localities to accept more or less rich assemblages in the nested sequence, and the ones which determines the particular place of species in the nestedness patterns would allow to understand more precisely the determinants of diversity distribution of diatoms and their local coexistence. Researchers should pay a particular attention to the succession dynamics and dispersal abilities of species.

## Bibliography

- AFNOR (2000). Qualité de l'eau. détermination de l'indice biologique diatomées (ibd) - norme nf t90-354 -.
- Almeida-Neto, M., P. Guimarães, P. Guimarães Jr., R. Loyola, et W. Ulrich (2008). A consistent metric for nestedness analysis in ecological systems: Reconciling concept and measurement. *Oikos* 117(8), 1227–1239.
- Astorga, A., J. Oksanen, M. Luoto, J. Soininen, R. Virtanen, et T. Muotka (2011). Distance decay of similarity in freshwater communities: do macro- and microorganisms follow the same rules? *Global Ecology and Biogeography*.
- Azeria, E., J. Ibarzabal, et C. Hébert (2012). Effects of habitat characteristics and interspecific interactions on co-occurrence patterns of saproxylic beetles breeding in tree boles after forest fire: null model analyses. *Oecologia* 168, 1123–1135.
- Azovsky, A. I. (2002). Size-dependent species-area relationships in benthos: is the world more diverse for microbes? *Ecography* 25(3), 273–282.
- Breiman, L. (2001). Random forests. *Machine Learning* 45(1), 5–32.
- Bruno, J., J. Stachowicz, et M. Bertness (2003). Inclusion of facilitation into ecological theory. *Trends in Ecology and Evolution* 18(3), 119–125.

- Burkholder, J. M., R. G. Wetzel, et K. L. Klomparens (1990). Direct comparison of phosphate uptake by adnate and loosely attached microalgae within an intact biofilm matrix. *Applied and Environmental Microbiology* 56(9), 2882–2890.
- Chessman, B. C. (1999). Predicting the macroinvertebrate faunas of rivers by multiple regression of biological and environmental differences. *Freshwater Biology* 41(4), 747–757.
- Clements, F. E. (1916). *Plant Succession: An Analysis of the Development of Vegetation*. Number 242. Washington, DC: Carnegie institution of Washington.
- Connor, E. F. et D. Simberloff (1979). The assembly of species communities - chance or competition. *Ecology* 60(6), 1132–1140.
- Eulin, A. (1997). *Les communautés de diatomées épilithiques de la Garonne. Répartition naturelle et étude expérimentale in situ de la dynamique de colonisation sur substrat artificiel*. spécialité hydrobiologie, Université Paul Sabatier Toulouse III.
- Fenchel, T. et B. Finlay (2004). The ubiquity of small species: Patterns of local and global diversity. *BioScience* 54(8), 777–784.
- Finlay, B. (2002). Global dispersal of free-living microbial eukaryote species. *Science* 296(5570), 1061–1063.
- Gillett, N., Y. Pan, K. Manoylov, R. Stancheva, et C. Weilhoefer (2011). The potential indicator value of rare taxa richness in diatom-based stream bioassessment. *Journal of Phycology* 47(3), 471–482.
- Gillett, N., Y. Pan, K. Manoylov, et R. Stevenson (2011). The role of live diatoms in bioassessment: A large-scale study of western us streams. *Hydrobiologia* 665(1), 79–92.
- Gleason, H. A. (1926). The individualistic concept of the plant association. *Bulletin of the Torrey Botanical Club* 53(1), pp. 7–26.
- Gotelli, N. J. (2000). Null model analysis of species co-occurrence patterns. *Ecology* 81(9), 2606–2621.
- Gotelli, N. J. et G. L. Entsminger (2001). Swap and fill algorithms in null model analysis: Rethinking the knight’s tour. *Oecologia* 129(2), 281–291.
- Gotelli, N. J. et W. Ulrich (2012). Statistical challenges in null model analysis. *Oikos* 121(2), 171–180.
- Götzenberger, L., F. de Bello, K. Bråthen, J. Davison, A. Dubuis, A. Guisan, J. Lepš, R. Lindborg, M. Moora, M. Pärtel, L. Pellissier, J. Pottier, P. Vittoz, K. Zobel, et M. Zobel (2012). Ecological assembly rules in plant communities-approaches, patterns and prospects. *Biological Reviews* 87(1), 111–127.
- Grenier, M., S. Lek, M. Rodríguez, A. Rousseau, et S. Campeau (2010). Algae-based biomonitoring: Predicting diatom reference communities in unpolluted streams using classification trees, random forests, and artificial neural networks. *Water Quality Research Journal of Canada* 45(4), 413–425.

- Hallstan, S., R. K. Johnson, E. Willén, et U. Grandin (2012). Comparison of classification-then-modelling and species-by-species modelling for predicting lake phytoplankton assemblages. *Ecological Modelling* 231(0), 11 – 19.
- Hanski, I. (1982). Dynamics of regional distribution: the core and satellite species hypothesis. *Oikos* 38(2), 210–221.
- Heino, J. et J. Soininen (2005). Assembly rules and community models for unicellular organisms: Patterns in diatoms of boreal streams. *Freshwater Biology* 50(4), 567–577.
- Hoagland, K., S. Roemer, et J. Rosowski (1982). Colonization and community structure of 2 periphyton assemblages, with emphasis on the diatoms (bacillariophyceae). *American Journal of Botany* 69(2), 188–213.
- Hosmer, D. et S. Lemeshow (2000). *Applied Logistic Regression* (2nd edition ed.). John Wiley and Sons.
- Katoh, K. (1992). Correlation between cell density and dominant growth form of epilithic diatom assemblages. *Diatom Research* 7(1), 77–86.
- Kelly, M., H. Bennion, E. Cox, B. Goldsmith, J. Jamieson, S. Juggins, D. Mann, et R. Telford (2005). Common freshwater diatoms of Britain and Ireland: an interactive key.
- Krammer, K. et H. Lange-Bertalot (1986 - 1991). *Süßwasserflora von Mitteleuropa.*, pp. Bacillariophyceae 1. Teil: Naviculaceae. 876 p.; 2. Teil: Bacillariaceae, Epithemiaceae, Surirellaceae, 596 p.; 3. Teil: Centrales, Fragilariaceae, Eunotiaceae, 576 p.; 4. Teil: Achnanthaceae. Kritische Ergänzungen zu Navicula (Lineolatae) und Gomphonema. 437 p. G. Fischer Verlag.
- Lange, K., A. Liess, J. Piggott, C. Townsend, et C. Matthaei (2011). Light, nutrients and grazing interact to determine stream diatom community composition and functional group structure. *Freshwater Biology* 56(2), 264–278.
- Lavoie, I., P. J. Dillon, et S. Campeau (2008). The effect of excluding taxa and reducing taxonomic resolution on multivariate analyses and stream bioassessment. *Ecological Indicators* doi: 10.1016/j.ecolind.2008.04.003.
- Leflaive, J. et L. Ten-Hage (2009). Chemical interactions in diatoms: role of polyunsaturated aldehydes and precursors. *New Phytologist* 184(4), 794–805.
- Liaw, A. et M. Wiener (2002). Classification and regression by randomforest. *R News* 2(3), 18–22.
- Martiny, J., J. C. R. Bohannan, B.J.M. and Brown, J. Fuhrman, J. Green, M. Horner-Devine, M. Kane, J. Krumins, C. Kuske, P. Morin, S. Naeem, L. Ø vreås, A.-L. Reysenbach, V. Smith, et J. Staley (2006). Microbial biogeography: Putting microorganisms on the map. *Nature Reviews Microbiology* 4(2), 102–112.
- McCormick, P. V. et R. J. Stevenson (1991). Mechanisms of benthic algal succession in lotic environments. *Ecology* 72(5), pp. 1835–1848.

- McGill, B., B. Enquist, E. Weiher, et M. Westoby (2006). Rebuilding community ecology from functional traits. *Trends in Ecology and Evolution* 21(4), 178–185.
- Miklós, I. et J. Podani (2004). Randomization of presence-absence matrices: Comments and new algorithms. *Ecology* 85(1), 86–92.
- Morin, S., M. Bottin, N. Mazzella, F. Macary, F. Delmas, P. Winterton, et M. Coste (2009). Linking diatom community structure to pesticide input as evaluated through a spatial contamination potential (phytopixal): A case study in the neste river system (south-west france). *Aquatic Toxicology* 94(1), 28–39.
- Moss, D., J. F. Wright, M. T. Furse, et R. T. Clarke (1999). A comparison of alternative techniques for prediction of the fauna of running-water sites in great britain. *Freshwater Biology* 41(1), 167–181.
- Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, R. B. O’Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, et H. Wagner (2011). *vegan: Community Ecology Package*.
- Olden, J. D., M. K. Joy, et R. G. Death (2006). Rediscovering the species in community-wide predictive modeling. *Ecological Applications* 16(4), 1449–1460.
- Ovaskainen, O., J. b. Hottola, et J. Siitonen (2010). Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. *Ecology* 91(9), 2514–2521.
- Ovaskainen, O. et J. Soininen (2011). Making more out of sparse data: hierarchical modeling of species communities. *Ecology* 92(2), 289–295.
- Pan, Y. et R. L. Lowe (1994). Independent and interactive effects of nutrients and grazers on benthic algal community structure. *Hydrobiologia* 291(3), 201–209.
- Passy, S. et C. Larson (2011). Succession in stream biofilms is an environmentally driven gradient of stress tolerance. *Microbial Ecology* 62(2), 414–424.
- Passy, S. I. (2007). Diatom ecological guilds display distinct and predictable behavior along nutrient and disturbance gradients in running waters. *Aquatic Botany* 86(2), 171–178.
- Reynolds, C. (2006). *Ecology of phytoplankton (ecology, biodiversity and conservation)*. Cambridge University Press.
- Smol, J. et E. Stoermer (2010). *The diatoms: applications for the environmental and earth sciences* (2nd ed.). Cambridge: Cambridge University Press.
- Söderström, L. (1989). Regional distribution patterns of bryophyte species on spruce logs in northern sweden. *Bryologist* 92(3), 349–355.
- Soininen, J. (2008). The ecological characteristics of idiosyncratic and nested diatoms. *Protist* 159(1), 65–72.

- Soininen, J. et J. Heino (2007). Variation in niche parameters along the diversity gradient of unicellular eukaryote assemblages. *Protist* 158(2), 181–191.
- Stevenson, R. J. et R. Glover (1993). Effects of algal density and current on ion transport through periphyton communities. *Limnology and Oceanography* 38(6), 1276–1281.
- Tison, J., Y. S. Park, M. Coste, J. G. Wasson, F. Rimet, L. Ector, et F. Delmas (2007). Predicting diatom reference communities at the french hydrosystem scale: A first step towards the definition of the good ecological status. *Ecological Modelling* 203(1-2), 99–108.
- Tuji, A. (2000). The effect of irradiance on the growth of different forms of freshwater diatoms: Implications for succession in attached diatom communities. *Journal of Phycology* 36(4), 659–661.
- Ulrich, W., M. Almeida-Neto, et N. J. Gotelli (2009). A consumer's guide to nestedness analysis. *Oikos* 118(1), 3–17.
- Ulrich, W. et N. J. Gotelli (2007). Disentangling community patterns of nestedness and species co-occurrence. *Oikos* 116(12), 2053–2061.
- Vanelslander, B., A. De Wever, N. Van Oostende, P. Kaewnuratchadasorn, P. Vanormelingen, F. Hendrickx, K. Sabbe, et W. Vyverman (2009). Complementarity effects drive positive diversity effects on biomass production in experimental benthic diatom biofilms. *Journal of Ecology* 97(5), 1075–1082.
- Venables, W. N. et B. D. Ripley (2002). *Modern Applied Statistics with S* (Fourth ed.). New York: Springer.
- Weiher, E. et P. Keddy (1999). *Ecological assembly rules: perspectives, advances, retreats*. Cambridge, GBR: Cambridge university press.



## Deuxième partie

**Quelles méthodes pour  
analyser la structure établie  
des assemblages ?**



## Chapitre 4

# Les Cartes auto-organisatrices de Kohonen

### 4.1 Présentation générale

#### 4.1.1 Un réseau de neurones artificiels

Les cartes auto-organisatrices de Kohonen (SOM, pour Self-Organizing Maps) font partie de la famille de techniques issues de la recherche en intelligence artificielle appelée « réseaux de neurones artificiels ». Les chercheurs ayant développé ces techniques s'inspirent des facultés du cerveau pour élaborer des systèmes d'apprentissage au sein des ordinateurs grâce à des traitements de valeurs numériques (on parle de « machine learning » en anglais).

Ils s'inspirent d'une description simplifiée du fonctionnement du système nerveux. Les neurones sont des cellules élémentaires de ce système, capables de transmettre des signaux électriques et chimiques. Ils s'organisent en réseaux, dans lesquels ils sont liés les uns aux autres, de proche en proche, par des synapses. Ainsi, les informations circulent dans le cerveau en variant d'intensité selon la réponse de chaque neurone. Il en résulte une activation plus ou moins forte des différentes zones, qui traduit le traitement de l'information initiale à l'intérieur du cerveau. L'apprentissage va donc consister à ajuster les réponses de chaque neurone et la transmission des informations vers les autres neurones.

Dans les réseaux de neurones artificiels, chaque neurone est lié aux données d'entrée et aux autres neurones par des valeurs numériques. Ces valeurs numériques, appelées « poids » sont ajustées pendant une phase d'apprentissage, puis utilisées pour prédire des valeurs de sorties, ou simplement pour décrire les données. Il existe deux familles de réseaux de neurones : les réseaux de neurones supervisés et les réseaux de neurones non-supervisés. Les modèles supervisés sont des modèles prédictifs dans lesquels le jeu de données d'apprentissage est utilisé pour ajuster les réponses du modèle. La phase d'apprentissage va alors consister à modifier les poids de façon à ce que les valeurs de sortie ressemblent le plus possible aux valeurs de sortie du jeu d'apprentissage. Le modèle pourra ensuite être utilisé pour prédire des valeurs de sortie à partir d'un jeu ne conte-

nant que des valeurs d'entrées (prédicteurs). Dans le cas des SOM, qui font partie des réseaux non-supervisés, les poids seront modifiés de façon à décrire les différences entre les valeurs d'entrée, mais il n'y aura pas de valeur de sortie. Il s'agit dans ce cas de classer les données d'entrée de façon à ce que le réseau de neurones décrive au mieux les différences entre celles-ci.

## 4.1.2 Fonctionnement des SOM

### 4.1.2.1 Présentation

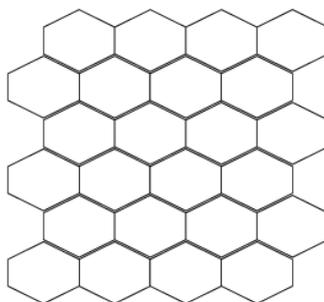


FIGURE 4.1 – Représentation d'une SOM, celle-ci comprend 24 cellules (ou neurones), six rangées horizontales, et quatre rangées verticales.

Les SOM se présentent sous la forme de cellules (neurones), le plus souvent hexagonales, disposées sur un plan (voir fig. 4.1). Le jeu de données est présenté sous la forme d'un tableau de données, comprenant  $m$  échantillons (lignes) et  $n$  espèces (colonnes). Chacun des neurones (ou cellules) est lié à chaque variable par un poids. Ce poids peut être considéré comme une valeur virtuelle de la variable, on parle donc de *vecteurs virtuels* pour les valeurs d'un neurone. Ainsi, la SOM représentée dans la figure 4.1 est définie par un tableau de valeurs de 24 lignes (une par neurone) et  $n$  variables (équivalentes aux  $n$  variables ou espèces du tableau de données).

### 4.1.2.2 Présentation de l'algorithme d'apprentissage

Les différentes étapes sont présentées dans la figure 4.2. Dans un premier temps, une analyse linéaire des gradients d'abondance des espèces, similaire à une Analyse en Composantes Principales est réalisée (voir paragraphe 1.4.3.2 p. 28). Cette analyse, grâce au calcul des deux vecteurs propres principaux, nous permet d'initialiser les valeurs virtuelles de la SOM sur deux dimensions, de façon linéaire.

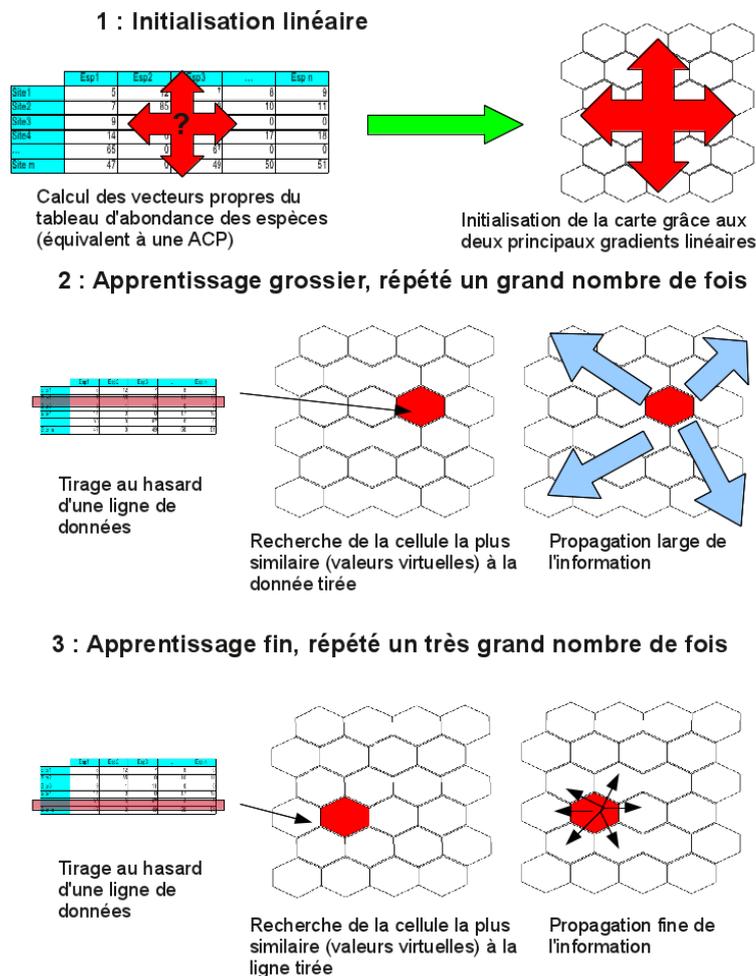


FIGURE 4.2 – Description de l’algorithme d’apprentissage des SOM

Puis ensuite va suivre un apprentissage récursif, répété un grand nombre de fois (figure 4.2).

- Un échantillon du tableau de données est tiré au hasard. Les valeurs de l'échantillon sont appelées « vecteur réel »,
- Une mesure de diversité  $\beta$  (voir page 25) est calculée entre le vecteur réel et chacun des vecteurs virtuels (valeurs des neurones),
- Le neurone dont les valeurs virtuelles sont les plus proches de celles du vecteur réel est désigné comme BMU (pour Best Matching Unit),
- Les valeurs du BMU sont modifiées pour ressembler encore davantage à celles de l'échantillon,
- Les valeurs de l'ensemble des vecteurs virtuels sont modifiées. Ces valeurs seront d'autant plus modifiées que la cellule est proche (spatialement, sur la carte) du BMU.

Grâce à ces étapes récursives, les cellules de la SOM proches spatialement vont avoir des valeurs ressemblantes pour chacune des espèces, alors que les cellules

éloignées vont avoir des valeurs très différentes.

Lors de cet apprentissage récursif, il y a deux phases majeures :

1. L'apprentissage grossier (rough phase) : durant cette phase, les valeurs des vecteurs virtuels vont être fortement modifiées à chaque étape, et cette modification sera significative sur une part considérable de la carte. Cette phase permet une prise en compte grossière des différences entre les échantillons afin de définir les principaux gradients de la carte, et les différentes grandes zones homogènes.
2. L'apprentissage fin (finetuning phase) : durant cette phase, seuls les vecteurs virtuels proches du BMU seront modifiés, et ces modifications seront très faibles. Cette phase, plus longue, permet d'affiner les valeurs virtuelles des cellules. Le but est que ces valeurs représentent au mieux les données présentes dans le tableau et les gradients linéaires ou non linéaires qui existent dans celui-ci.

#### 4.1.2.3 Les paramètres de l'algorithme d'apprentissage

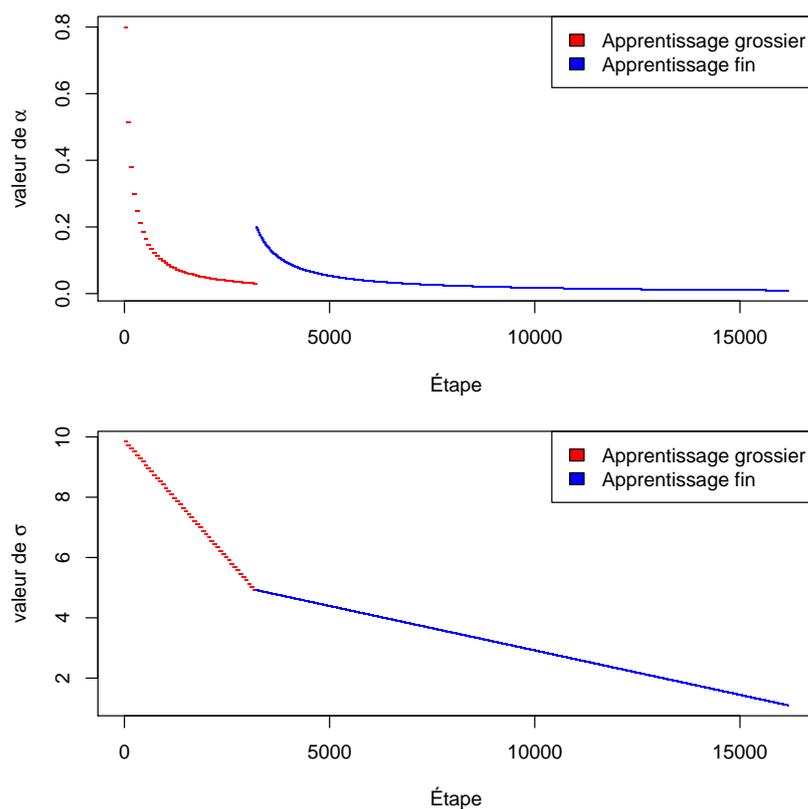


FIGURE 4.3 – Évolution des paramètres  $\alpha$  (paramètre d'apprentissage, en haut) et  $\sigma$  (paramètre de voisinage, en bas) durant l'apprentissage d'une SOM

La mise à jour des valeurs des vecteurs virtuels de la SOM se fait à l'aide de la fonction d'apprentissage, qui va déterminer la proportion des différences entre

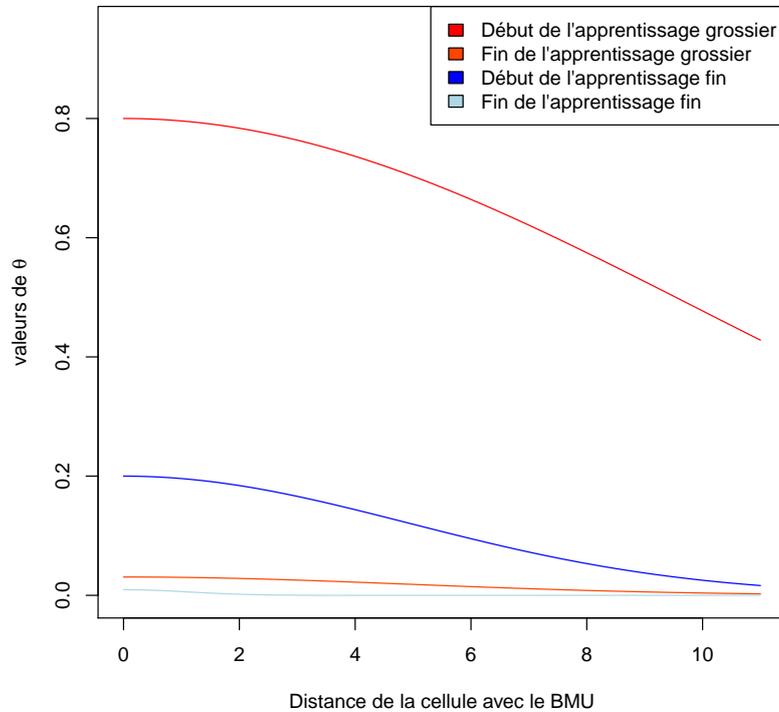


FIGURE 4.4 – Évolution du taux d'apprentissage  $\theta$  en fonction de l'étape de la SOM et de la distance de la cellule avec le BMU

les vecteurs réels et virtuels qui doit être réduite, selon l'étape, et la distance sur la carte entre la cellule et le BMU. Voici la formule de ces fonctions :

$$\theta = \alpha \times e^{\frac{-d^2}{2\sigma^2}}$$

où :

$\theta$  est la proportion de différence à réduire,

$\alpha$  est le paramètre d'apprentissage, qui va permettre de faire varier le taux global d'apprentissage,

$d$  est la distance sur la carte entre la cellule et le BMU et

$\sigma$  est le paramètre de voisinage, qui va permettre de faire diminuer  $\theta$  en fonction de la distance  $d$ .

Les paramètres  $\alpha$  et  $\sigma$  diminuent au fur et à mesure de l'apprentissage de la SOM (fig. 4.3), faisant varier la forme des fonctions d'apprentissage (fig. 4.4).

#### 4.1.2.4 Résultats d'une SOM

On utilise généralement deux mesures de la qualité de la représentation des données par la SOM finale :

**Erreur topographique** L'erreur topographique est la proportion de vecteurs réels pour lesquels le deuxième vecteur virtuel le plus similaire (deuxième BMU) ne correspond pas à une cellule voisine du BMU. Il s'agit de vérifier que ce sont bien les cellules les plus proches qui représentent les communautés les plus similaires.

**Erreur de quantification (quantization error)** L'erreur de quantification correspond à la somme des distances calculées entre chaque vecteur réel et son BMU.

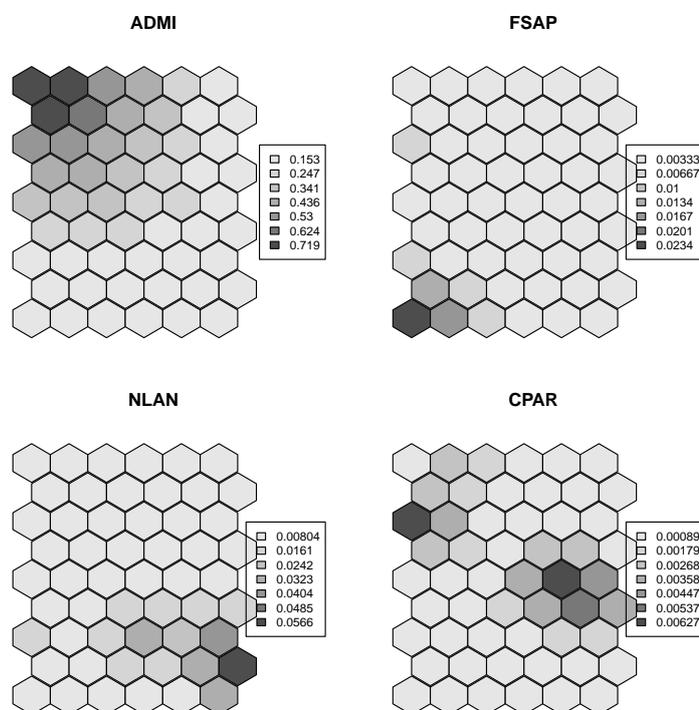


FIGURE 4.5 – Exemples de profils de quatre espèces sur une SOM réalisée sur un jeu de données de référence (bonne qualité de l'eau). Les valeurs représentées par l'échelle de gris sont les valeurs des vecteurs virtuels. ADM : *Achnanthydium minutissimum*, espèce pionnière principalement présente dans les eaux bien oxygénées. FSAP : *Fistulifera saprophila* espèce saprobe, typique des milieux riches en matières organiques. NLAN : *Navicula lanceolata*, espèce ubiquiste, montrant ses plus fortes abondance dans les sites de qualité d'eau moyenne. CPAR : *Cymbella parva*, espèce présente dans les sites de bonne qualité, elle montre ici une distribution bimodale.

À la fin de l'apprentissage, les valeurs des vecteurs virtuels sont représentatives des valeurs du tableau de données. En général, on choisit un nombre de

cellules très largement inférieur à celui du nombre  $m$  d'échantillons (une règle empirique veut que le nombre de neurones soit approximativement  $5 \times m^{0.543}$ ). Ainsi, les gradients contenus dans le tableau de données peuvent être analysés grâce à un nombre réduit de vecteurs virtuels, qui ont des valeurs pour chacune des espèces. Il est possible d'afficher les gradients d'abondance de chacune des espèces sur la carte SOM créée (fig. 4.5).

Il est aussi possible de repérer quel neurone est le BMU de chacun des vecteurs réels. Ainsi, on peut considérer que la SOM a réalisé un classement des différents échantillons en fonction des valeurs d'abondance des différentes espèces, et les a placés dans un nombre de groupes équivalent au nombre de neurones.

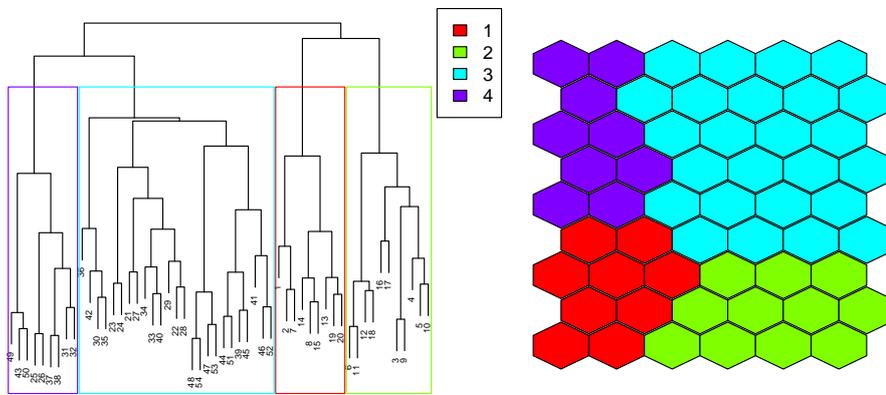


FIGURE 4.6 – Classification des neurones grâce à une classification hiérarchique ascendante, permettant une définition simultanée de zones sur la carte SOM.

On peut appliquer une classification hiérarchique ascendante aux cellules selon leurs valeurs virtuelles. Sur la carte SOM, les cellules des plus proches sont les plus similaires en terme de valeurs virtuelles, la classification va donc permettre la délimitation de zones (fig. 4.6), correspondant à des grands types d'assemblage, on parle de « *biotypes* ».

Lorsque des données environnementales sont disponibles, elles peuvent être projetées sur la carte SOM. Ainsi, on peut déterminer quels sont les gradients déterminant les différences entre les assemblages. Par exemple, dans la figure 4.7, on distingue un gradient de pH entre le coin inférieur droit et le coin supérieur gauche de la SOM, ce gradient étant inversé pour la conductivité. Cela permet de dire que la variation des assemblages de diatomées de Guadeloupe et de Martinique de cet exemple est dépendante de ces variables environnementales.

### 4.1.3 Avantages des SOM

Les SOM permettent de réaliser simultanément une ordination et une classification des données biologiques, tout en comparant la structure des communautés aux principaux déterminants environnementaux. Un de leurs avantages principaux est la prise en compte de gradients non-linéaires complexes, qui leur

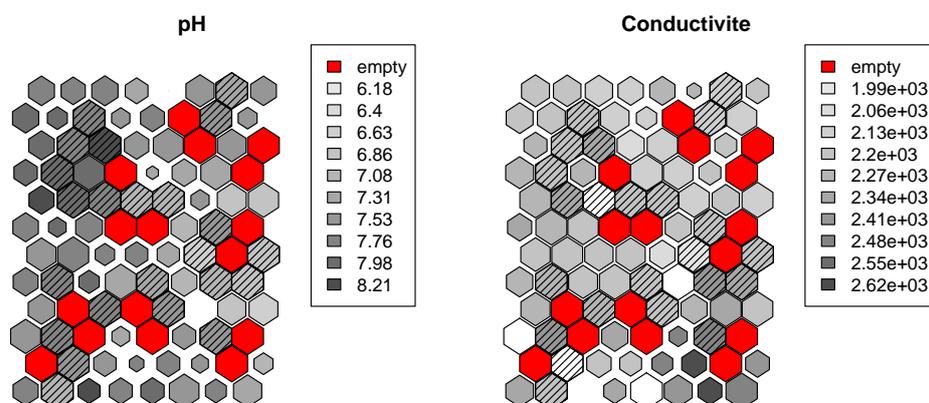


FIGURE 4.7 – Exemple de projection de variables environnementales (pH et conductivité) sur une SOM réalisée à partir d’abondances relatives de diatomées de territoires français d’Outre-Mer (Guadeloupe et Martinique). Chaque échantillon du tableau de données est attribué à son BMU lors de l’apprentissage de la SOM. On calcule les moyennes de variables supplémentaires pour chacune des cellules (la moyenne de chaque échantillon dont la cellule est le BMU), et les représenter sur la carte grâce à une échelle de gris. Les cellules rouges sont les cellules qui ne sont les BMU d’aucun échantillon réel (on ne peut donc pas calculer de moyenne sur les données environnementales réelles). La taille des cellules est d’autant plus petite que les données sont variables. Les cellules hachurées sont les BMU d’un seul site (on ne peut donc pas exprimer de variabilité des données réelles).

permettent de mieux décrire la distribution des assemblages biologiques que les techniques d’ordination classiques (Giraudel et Lek, 2001). Les SOM ont donc été régulièrement utilisées pour caractériser et décrire des communautés aquatiques variées (Chon, 2011), notamment des assemblages de poissons (Brosse et al., 2001), ou de macroinvertébrés benthiques (Cereghino et al., 2003; Jabiol et al., 2009). Les assemblages de diatomées ont aussi été analysés avec cette technique (Gevrey et al., 2004), afin de décrire leurs variations spatiales dans un but de bio-indication (Grenier et al., 2006; Tison et al., 2007; Coste et al., 2009), ou de décrire l’impact de substances toxiques sur leur composition (Morin et al., 2009).

En revanche les SOM demandent, comme les autres réseaux de neurones, un effort important de paramétrisation. Les paramètres que l’on peut faire varier sont nombreux : évolution des facteurs  $\alpha$  et  $\sigma$ , nombre de neurones, nombre d’étapes dans chacune des phases, mesure de distance utilisée (biodiversité  $\beta$ ). Ces paramètres doivent être adaptés à chaque type de jeu de données. La plupart des solutions logicielles proposent des options de paramétrage limitées, qui ne permettent pas d’adapter l’algorithme aux particularités des assemblages de diatomées (forts chevauchements de niches, nombreuses espèces rares, très forte diversité).

## 4.2 Adaptation des SOM aux assemblages de diatomées

Dans le cadre de cette thèse, nous avons développé un programme (sous la forme d'un paquetage R, disponible en annexe, avec un mode d'emploi sommaire) pour réaliser les SOM sur les assemblages de diatomées. En effet, les solutions existantes ne convenaient pas à nos besoins pour différentes raisons, que nous évoquons plus bas.

### 4.2.1 Problèmes posés

#### 4.2.1.1 Solutions logicielles existantes

Parmi les solutions logicielles proposées, la plus utilisée pour les communautés biologiques semble être la *somtoolbox* de Matlab<sup>TM</sup>. Malheureusement, Matlab contient peu d'outils supplémentaires d'analyse de données adaptés aux communautés, et le traitement des données oblige à utiliser des programmes complémentaires tels que R, beaucoup plus complet en termes d'analyses statistiques. Jean-Luc Giraudel a élaboré un paquetage sous R (*Emuann*), mais celui-ci, doté d'une interface graphique, ne fournissait pas la flexibilité nécessaire à l'adaptation de SOM spécialisées dans l'analyse des assemblages de diatomées. C'est donc à partir du code initial de J.-L. Giraudel (en langage C pour le logiciel R) que nous avons reprogrammé en totalité l'algorithme des SOM.

#### 4.2.1.2 Mesures de distances

La plupart des solutions logicielles utilisent la distance euclidienne pour mesurer les poids entre les neurones et les lignes des tableaux de données. Or les distances de Bray-Curtis (abondances absolues, voir paragraphe 1.4.2.2) et de Manhattan (abondances relatives) semblent plus adaptées aux communautés biologiques.

#### 4.2.1.3 Difficultés liées aux propriétés des diatomées

Les diatomées sont très diversifiées, et les matrices biologiques contiennent souvent un très grand nombre de colonnes (taxons). Du fait de cette diversité, du grand nombre d'espèces rares et des chevauchements de niches des espèces de diatomées, le traitement des données doit être adapté afin d'éviter l'instabilité et l'incohérence des résultats. Par exemple, l'utilisation de la distance de Bray-Curtis sur les assemblages de diatomées résultent parfois en des « groupes disjoints » (fig. 4.8) et des résultats très différents lors de l'application successive de SOM sur un même jeu de données. La distance de Bray-Curtis permet en effet une description plus complète des différences, en donnant le même poids à chacune des espèces dans son calcul. Des gradients d'abondance inverses entre plusieurs espèces peuvent alors perturber l'apprentissage, et il faut modifier l'algorithme pour pouvoir utiliser cette mesure (ou la distance de Manhattan, équivalente pour les abondances relatives)

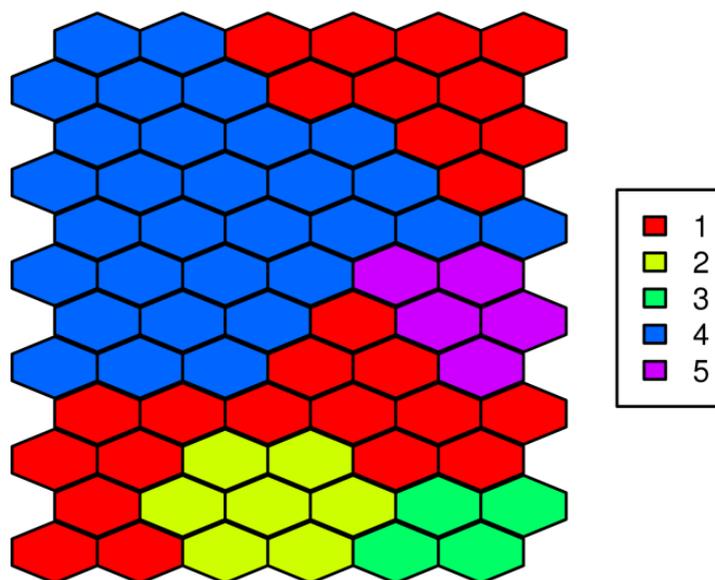


FIGURE 4.8 – Exemples de groupes disjoints, ce type d'incohérence, rare, résulte d'une mauvaise adaptation des paramètres d'apprentissage aux particularités des données (par exemple : nombre de colonnes, gradients inversés).

## 4.2.2 Solutions proposées

### 4.2.2.1 Initialisation linéaire

Habituellement, les algorithmes de SOM initialisent les vecteurs virtuels avec les valeurs d'échantillons tirés au hasard. Il en résulte une forte instabilité lors des premières phases de l'algorithme. Le paquetage que nous avons programmé propose une initialisation linéaire des vecteurs virtuels, basée sur les vecteurs propres d'une matrice de variance-covariance calculée sur le tableau de données (cette méthode correspond à celle d'une Analyse en Composantes Principales).

### 4.2.2.2 Mesures de distance

Le paquetage programmé contient pour l'instant trois mesures de distance : la distance euclidienne, la distance de Manhattan et la distance de Bray-Curtis. Par la suite, il faudra y ajouter des mesures de distances entre données binaires (présences/absences) d'espèces.

### 4.2.2.3 Calcul du nombre d'étape

Afin d'obtenir des résultats plus stables, et moins subjectifs, nous avons fortement augmenté le nombre d'étapes nécessaires à la réalisation d'une SOM, et ajouté un paramètre permettant de prendre en compte le nombre d'espèces dans le calcul automatique du nombre d'étapes. Cette augmentation est facilitée par l'utilisation du langage C, qui permet une grande rapidité de calcul.

Une autre différence essentielle de cet algorithme avec les autres est le tirage aléatoire des sites pendant les différentes phases d'apprentissage. Afin de donner la même importance à chacune des données durant les phases d'apprentissage grossier, seul leur ordre de prise en compte est tiré au hasard durant cette phase. Chacune des données est prise en compte avant un autre tirage. Les valeurs de  $\alpha$  et de  $\sigma$  ne diminuent pas avant que toutes les données aient été utilisées avec ces valeurs (fig. 4.3). Lors de l'apprentissage fin, les données sont simplement tirées au hasard, avec remise, comme dans les algorithmes habituels, mais la forte augmentation du nombre d'étapes durant cette phase nous assure la prise en compte de l'ensemble des données.

### 4.3 Mise à jour de la biotypologie des assemblages français

Ce travail a fait l'objet d'un partenariat avec l'ONEMA (Action 12-3 de la Convention 2011 : Développement et optimisation des méthodes de bioindication pour les cours d'eau). Le rapport officiel correspondant à ce travail est disponible en annexe (« *Typologie des flores diatomiques de référence et seuils d'état écologique à partir des notes d'IBD2007* ».)

#### 4.3.1 Buts de l'étude

Afin de pouvoir évaluer l'impact des pressions anthropiques sur les cours d'eau grâce aux assemblages de diatomées, il est important de décrire les variations « naturelles » de ces assemblages dans un premier temps (voir paragraphe 1.2.6). Tison et al. (2005) ont défini cinq régions sur le territoire français, permettant de décrire les principaux gradients naturels. Ces cinq régions ont été séparées grâce à l'application d'une SOM sur les assemblages de référence à l'échelle nationale. Les résultats de cette SOM ont révélé une forte composante spatiale dans la structuration des principaux groupes, et une bonne correspondance avec les Hydro-Éco-Régions (HER) de niveau 1 définies par Wasson et al. (2002).

Une nouvelle version de l'Indice Biologique Diatomée (voir paragraphe 1.2.5.1) a été validée et normalisée depuis la description de ces cinq biotypes. Il était donc important de redéfinir la variabilité de cet indice selon les régions et éventuellement de mettre à jour la typologie française des assemblages. De plus, l'application de la Directive Cadre européenne sur l'Eau (DCE) a permis la constitution d'une nouvelle base de données, mieux équilibrée géographiquement et contenant une quantité de données largement supérieure. En général, dans cette base de données, plusieurs échantillons sont disponibles par station de prélèvement, il est donc possible de calculer des données d'« assemblages moyens » dont les tests de Mantel ont montré qu'ils étaient mieux corrélés aux conditions environnementales (voir chapitre 2).

Le but de cette étude est de vérifier la validité des régions définies en 2005 pour décrire la variabilité des assemblages de références en utilisant (i) la nouvelle base de données, (ii) les SOM du paquetage décrit dans le paragraphe 4.2 (utilisation des distances de Manhattan, paramètres d'apprentissage adaptés aux diatomées) et (iii) des données d'assemblages moyens.

#### 4.3.2 Informations complémentaires sur les données et les méthodes

##### 4.3.2.1 Données utilisées

Les données utilisées proviennent de la « nouvelle base de données ». Seules les données issues des réseaux de référence (bonne qualité de l'eau) prélevées

entre 2005 et 2009 ont été prises en compte.

Un premier traitement des données, sous la forme d'une SOM sur l'ensemble des assemblages de diatomées de référence, a révélé deux problèmes importants :

- Une partie des assemblages a montré une forte proportion de taxons hétérotrophes (tolérant de fortes concentrations en nutriments) et polysaprobies (tolérant de fortes charges en matières organiques). Les données étaient donc fortement structurées par un gradient de qualité de l'eau, ajoutant du « bruit » à notre description des conditions biogéochimiques des sites de référence.
- Les assemblages des sites ayant été prélevés plusieurs années se sont souvent associés à des groupes différents de la SOM selon les années (17% des sites prélevés sur plusieurs années étaient présents dans au moins deux groupes différents selon les années).

Pour éviter un trop fort gradient de qualité des eaux dans les données, nous avons supprimé du jeu de référence les échantillons dans lesquels sept taxons repérés comme hétérotrophes ou polysaprobies comptaient ensemble pour plus de 4,5 % d'abondance relative. Ce seuil fut déterminé à partir de la distribution de la proportion de ces taxons dans les données de référence, et représente la valeur médiane, à laquelle on ajoute cinq fois l'écart inter-quartile (entre le premier et le troisième quartile). Les taxons concernés sont *Nitzschia capitellata*, *Fistulera saprophila*, *Nitzschia palea*, *Nitzschia paleacea*, *Luticola mutica*, *Gomphonema parvulum* et *Craticula accomoda*.

Afin de définir les régions de façon plus sûre et moins fluctuante, nous avons utilisé des assemblages moyens dans une deuxième SOM pour tous les sites qui comprenaient plusieurs prélèvements sur la période 2005-2009.

Le tableau de données final comprend l'abondance relative moyenne de 461 taxons sur 194 sites.

#### 4.3.2.2 Calcul d'entropie

Le but de cette étude étant de simplifier la régionalisation en HER (Wasson et al., 2002) pré-existante, il était important de pouvoir comparer la classification réalisée par la SOM à ces HER. Pour cela nous avons utilisé un calcul d'entropie issu de la théorie de l'information (correspondant à la mesure de diversité  $\alpha$  décrite page 25 ; Shannon, 1948)

$$E = - \sum_{a=1}^j \frac{n_{a\bullet}}{n_{\bullet\bullet}} \times \sum_{b=1}^5 \frac{n_{ab}}{n_{a\bullet}} \times \ln \frac{n_{ab}}{n_{a\bullet}}$$

où  $A$  est le classement des sites réalisé par la SOM,  $B$  l'appartenance des sites aux cinq régions,  $j$  est le nombre de groupes dans le classement par la SOM et  $n_{ab}$  est le nombre de sites appartenant au groupe  $a$  du classement et à la région  $b$ .

L'entropie est la mesure du « désordre » d'un classement par rapport à un autre. Quand chacun des groupes du premier classement ne contient que des individus d'un seul groupe du deuxième classement, l'entropie est nulle, on dit que le classement est « pur ». Quand les groupes du premier classement contiennent des individus de plusieurs groupes du deuxième classement, la valeur de l'entropie augmente.

L'entropie des groupes de la SOM quant aux HER a été calculée selon tous les niveaux de classification possibles (en prenant de 1 à  $n$  groupes de cellules,  $n$  étant le nombre de cellules). Ainsi une diminution rapide de l'entropie quand on

augmente le nombre de groupes suggère l'atteinte d'une correspondance entre les deux classements (le classement de la SOM devient très vite « pur »).

#### 4.3.2.3 Paramètres de la SOM

La SOM a été réalisée avec la distance de Manhattan. Après avoir testé plusieurs tailles, le nombre de neurones, 80, a été choisi afin de se rapprocher du critère  $5 \times n^{0.543}$  où  $n$  est le nombre de sites prélevés, et de valeurs minimales d'erreurs topographiques (0.0) et de quantization (82.30).

### 4.3.3 Résultats

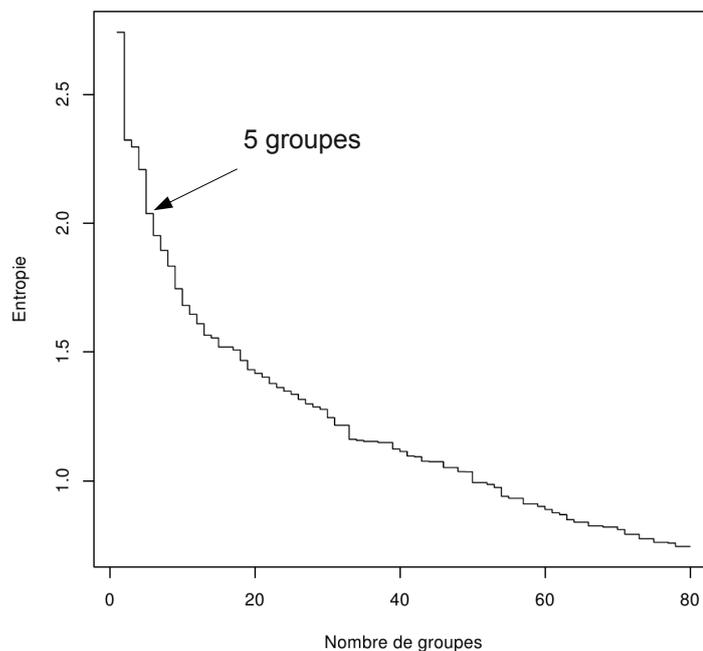


FIGURE 4.9 – Courbe d'entropie du classement des sites par la SOM pour les régions diatomiques pré-existantes

La courbe d'entropie montre une forte chute lors de la séparation des premiers groupes de la SOM (fig. 4.9), ce qui suggère que la SOM sépare en premier lieu les sites selon les mêmes facteurs que les HER. Malgré cela, la courbe d'entropie ne descend pas jusqu'à la valeur 0, signe qu'il y a des différences entre ces deux classements. Enfin la courbe d'entropie ne montre plus de fortes baisses après les cinq premières séparations, nous garderons donc cinq groupes, ce qui

permet de plus de rester homogène avec le nombre de régions diatomiques choisi en 2005.

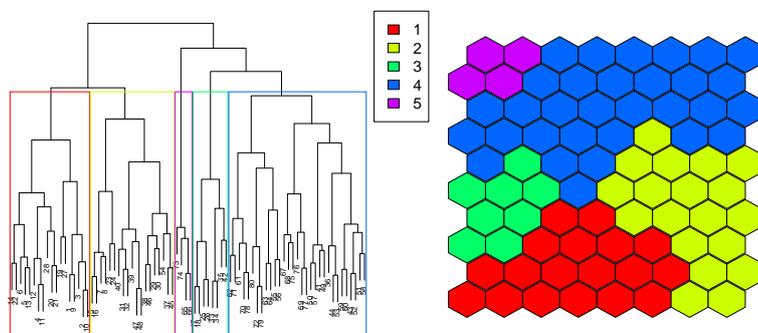


FIGURE 4.10 – Classification hiérarchique de la SOM et séparation des cinq groupes de la SOM.

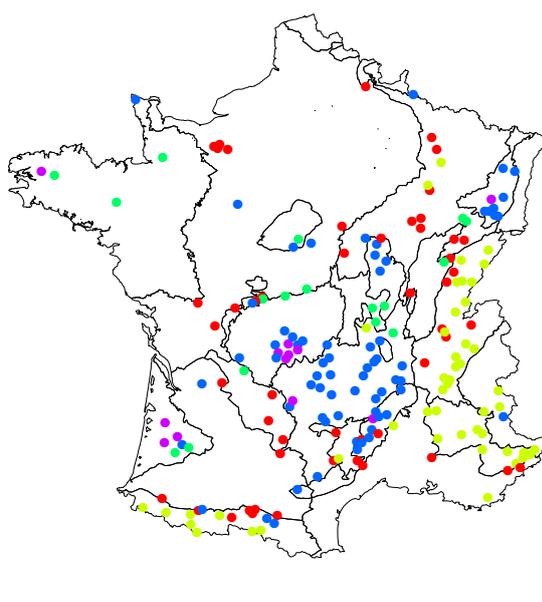


FIGURE 4.11 – Localisation des sites appartenant aux différents groupes de la SOM. Les couleurs correspondent à celles de la figure 4.10. Le fond de carte correspond aux Hydro-Éco-Régions de Wasson et al. (2002).

Les groupes formés par la SOM (fig. 4.10) se répartissent géographiquement, selon des délimitations qui correspondent plutôt bien à des frontières entre HER (fig. 4.11). Ainsi, comme l'ont montré Tison et al. (2005), les principales variations des communautés de diatomées sont en accord avec les facteurs qui ont

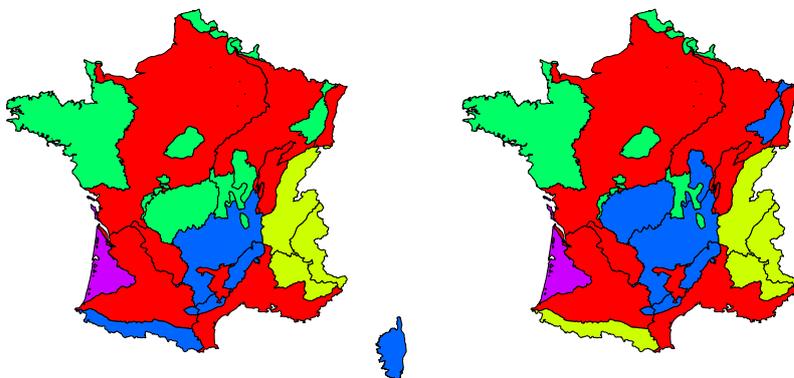


FIGURE 4.12 – Comparaison des régions pré-existantes (Tison et al., 2005) (à gauche) et de celles déterminées à l’aide de la SOM (à droite). Les couleurs correspondent à celles de la figure 4.10.

permis de définir les HER (climat, altitude et géologie).

On trouve des différences entre les régions définies ici et celle de Tison et al. (2005) (fig 4.12). Ces différences concernent principalement trois régions :

- Les Pyrénées, qui forment dorénavant un groupe de haute montagne avec les Alpes.
- Les Vosges et le nord du Massif-central qui rejoignent les régions de moyenne montagne peu carbonatées.

#### 4.3.4 Discussion

Les SOM développées dans le cadre de cette thèse ont permis de décrire la répartition régionale des assemblages de diatomées. Il est intéressant de noter que les différences de régions avec celles définies par Tison et al. (2005) concernent des régions de montagne. Dans le chapitre 3 de cette thèse, une étude de la structuration spatiale des assemblages avait montré que les assemblages de montagne étaient probablement soumis à des phénomènes de limites de dispersion des espèces. De plus, cette même étude avait montré que l’utilisation d’assemblages moyens sur plusieurs années (comme ici) permettaient d’obtenir des relations plus claires avec les conditions environnementales. Il nous semble donc que l’évolution de la stratification régionale française préconisée par cette étude est en accord avec la structure des assemblages décrite dans la première partie de cette thèse.

Aussi, ce traitement de données a révélé une importance plus importante des gradients altitudinaux français, en plaçant les régions d’altitude comparable dans les mêmes groupes. Ce résultat est en accord avec la biotypologie des États-Unis réalisée par Potapova et Charles (2002), dans laquelle l’altitude est un facteur majeur de séparation entre les régions diatomiques.

## Bibliographie

- Brosse, S., J. L. Giraudel, et S. Lek (2001). Utilisation of non-supervised neural networks and principal component analysis to study fish assemblages. *Ecological Modelling* 146(1-3), 159–166.
- Cereghino, R., Y.-S. Park, C. Arthur, et L. Sovan (2003). Predicting the species richness of aquatic insects in streams using a limited number of environmental variables. *Journal of the North American Benthological Society* 22(3), 442–456.
- Chon, T.-S. (2011). Self-organizing maps applied to ecological sciences. *Ecological Informatics* 6(1), 50–61.
- Coste, M., S. Boutry, J. Tison-Rosebery, et F. Delmas (2009). Improvements of the biological diatom index (BDI) : Description and efficiency of the new version (BDI-2006)mo. *Ecological Indicators* 9(4), 621–650.
- Gevrey, M., F. Rimet, Y. S. Park, J.-L. Giraudel, L. Ector, et S. Lek (2004). Water quality assessment using diatom assemblages and advanced modelling techniques. *Freshwater Biology* 49(2), 208–220.
- Giraudel, J. L. et S. Lek (2001). A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination. *Ecological Modelling* 146(1-3), 329–339.
- Grenier, M., S. Campeau, I. Lavoie, Y. S. Park, et S. Lek (2006). Diatom reference communities in quebec (canada) streams based on kohonen self-organizing maps and multivariate analyses. *Canadian Journal of Fisheries and Aquatic Sciences* 63(9), 2087–2106.
- Jabiol, J., B. Corbara, A. Dejean, et R. Céréghino (2009). Structure of aquatic insect communities in tank-bromeliads in a east-amazonian rainforest in french guiana. *Forest Ecology and Management* 257(1), 351–360.
- Morin, S., M. Bottin, N. Mazzella, F. Macary, F. Delmas, P. Winterton, et M. Coste (2009). Linking diatom community structure to pesticide input as evaluated through a spatial contamination potential (phytopixal) : A case study in the neste river system (south-west france). *Aquatic Toxicology* 94(1), 28–39.
- Potapova, M. G. et D. F. Charles (2002). Benthic diatoms in usa rivers : distributions along spatial and environmental gradients. *Journal of Biogeography* 29(2), 167–187.
- Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal* 27(3), 379–423.
- Tison, J., Y. S. Park, M. Coste, J. G. Wasson, L. Ector, F. Rimet, et F. Delmas (2005). Typology of diatom communities and the influence of hydroecoregions : A study on the french hydrosystem scale. *Water Research* 39(14), 3177–3188.

- Tison, J., Y. S. Park, M. Coste, J. G. Wasson, F. Rimet, L. Ector, et F. Delmas (2007). Predicting diatom reference communities at the french hydrosystem scale : A first step towards the definition of the good ecological status. *Ecological Modelling* 203(1-2), 99–108.
- Wasson, J. G., A. Chandesris, H. Pella, et L. Blanc (2002). Definition des hydroécorégions françaises métropolitaine. approche régionale de la typologie des eaux courantes et éléments pour la définition des peuplements de référence d'invertébrés - rapport final. Technical report, Ministère de l'Ecologie et du Développement Durable - Cemagref GMA. Biol. des Syst. Aquat. Lab. Hydroécol. Quantitative Lyon.

# Chapitre 5

## Les pattern trees

### 5.1 Présentation

#### 5.1.1 Généralités

La méthode des pattern trees est relativement récente, pour la présenter simplement, on peut dire qu'elle correspond à des arbres de décisions utilisant des règles de logique floue. Cette technique a déjà été utilisée pour déterminer la valeur indicatrice des diatomées du lac Prespa (Naumoski et Mitreski, 2011b).

##### 5.1.1.1 Les arbres de décisions

Les arbres de décisions représentent une famille de méthodes, parmi lesquelles la technique *CART* (Classification and Regression Trees ; Breiman et al., 1984) est souvent utilisée en Écologie (par exemple Garzón et al., 2006; Weilhöfer et Pan, 2008; Naumoski et Mitreski, 2009; Gillett et al., 2011). Le principe des arbres de décisions est de créer et de hiérarchiser un ensemble de règles pour classer un objet (prendre une décision). Par exemple, on peut imaginer de vouloir classer des sites dans les situations de référence ou impactées en fonction des variables physico-chimiques mesurées sur les sites (Fig. 5.1). Ces arbres de décisions utilisent des valeurs seuils pour les différents paramètres afin de partitionner les données de façon à obtenir des groupes les plus ressemblants possibles à ceux que l'on cherche à prédire (dans la figure 5.1 : référence ou impacté). Du fait de l'utilisation de ces valeurs seuils précises, les arbres de classification sont très sensibles aux données et peu robustes.

##### 5.1.1.2 La logique floue

La logique floue est une technique de plus en plus utilisée en intelligence artificielle, appliquée à des domaines très divers tels que la médecine, la robotique ou encore la météorologie (par exemple Mago et al., 2011; Rahoma et al., 2011). Le but de cette technique est de pouvoir intégrer dans des modèles des concepts imprécis, tels que ceux que nous utilisons lorsque l'on parle. En effet, il est difficile de traduire de manière mathématique ce que sont de « faibles valeurs de phosphate » ou des « des rivières plutôt acides ». Dans les exemples précédents, la logique floue peut permettre d'associer, pour chaque valeur mesurée

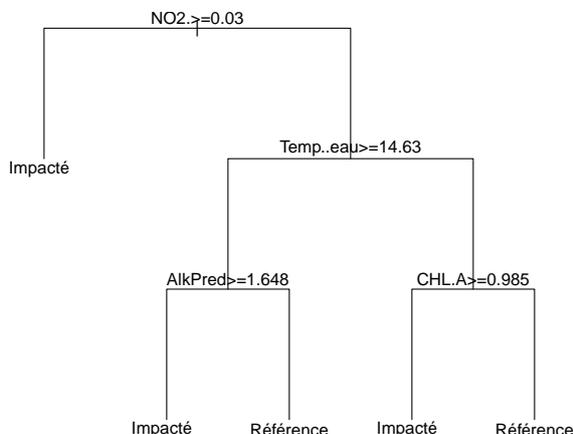


FIGURE 5.1 – Exemple d’arbre de classification (*CART*) : prédiction des sites de référence et impactés en fonction des variables physico-chimiques mesurées. Par exemple, une des règles de classification que l’on peut lire sur cet arbre est : *Si les nitrites ont une concentration supérieure à 0.03 mg.L<sup>-1</sup>, que la température de l’eau est supérieure à 14,53°C et que l’alcalinité est supérieure à 1.648, alors le site est impacté.*

de concentration en phosphate ou de pH, des valeurs permettant de déterminer à quel point les rivières sont « plutôt acides » et le phosphate « faible » (voir Fig 5.2). Du point de vue mathématique, la logique floue est souvent considérée comme une extension de la théorie des ensembles classique dans l’ « univers du discours ».

### 5.1.1.3 L’induction de pattern Trees

La technique des patterns trees va consister à construire des arbres tels que ceux issus de la méthode *CART*, mais en utilisant des règles établies en logique floue. L’induction des pattern trees est l’étape de l’algorithme qui va consister à choisir les règles floues à partir des données et à les hiérarchiser. Elle est précédée d’une étape dans laquelle les valeurs des différentes variables des modèles sont traduites en valeurs floues : la fuzzification (de l’anglais fuzzy : flou).

## 5.1.2 Paramétrage et algorithme des pattern trees

Les pattern trees étant une technique relativement récente, elle n’avait pas été implémentée dans le logiciel de statistique R. Afin de pouvoir utiliser efficacement cette méthode, nous l’avons donc programmée dans le langage C et R, sous forme d’un paquetage R appelé *patternTree*, inspiré principalement de Huang et al. (2008) et du programme Fuzzy-EBY, développé par Zhiheng

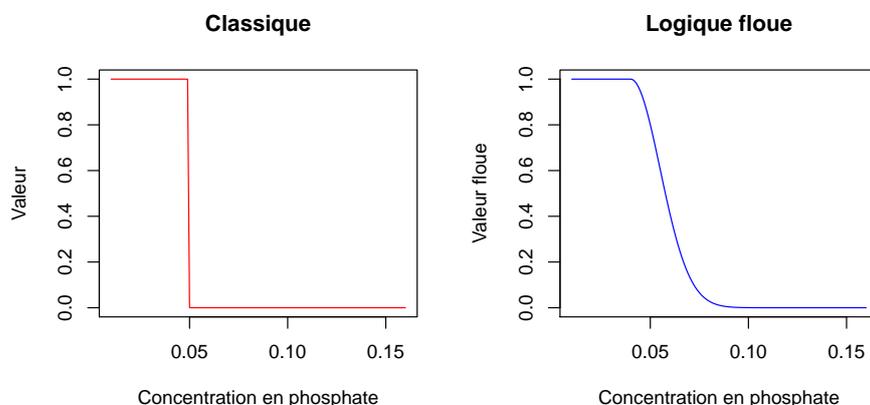


FIGURE 5.2 – Illustration de la logique Floue : « Faibles valeurs de phosphate ». Si la logique floue n'est pas utilisée, la traduction pour les modèles de l'expression « faibles valeurs de phosphate » utilise un seuil : en dessous de cette valeur seuil (ici 0.05) l'expression est vraie, au dessus de cette valeur l'expression est fausse. Lorsque l'on utilise la logique floue, l'évolution des valeurs de l'expression « faibles valeurs de phosphate » peut être plus progressive. Ces valeurs prennent donc en compte le fait que certaines conditions correspondent à un niveau « plutôt faible » ou « pas vraiment faible ». La logique floue est donc mieux adaptée aux concepts imprécis que l'on utilise dans le langage courant.

Huang (University of California, Berkeley). Ce travail de programmation a été réalisé en collaboration avec Andreja Naumoski, de l'université de Skopje et Nicolas Antunes, Irstea. Ce paragraphe a pour but de présenter plus précisément l'algorithme, les choix qui ont été faits lors de l'implémentation et les possibilités offertes par le paquetage patternTree.

### 5.1.2.1 La Fuzzification

La fuzzification permet d'obtenir des valeurs floues à partir de variables classiques. Il s'agit en fait de partager les valeurs de la variable en plusieurs classes, et ainsi d'attribuer des valeurs floues à toutes les unités d'échantillonnages selon ces classes.

Dans ce but différentes fonctions, de différentes formes, ont été créées (Huang et al., 2008; Naumoski et Mitreski, 2011b,a). Les exemples de fonctions disponibles dans le paquetage patternTree sont présentés dans la figure 5.3. Il est potentiellement possible d'utiliser un très grand nombre de fonctions différentes, l'important étant que la fonction mathématique représente bien la classe que l'on veut décrire. Aussi l'utilisation pourra varier selon les variables étudiées et les buts de l'analyse.

La première étape de la fuzzification des variables quantitatives consiste à trouver des intervalles de valeurs pour créer les différentes classes. Souvent, chaque intervalle est décrit par une valeur centrale, appelée  $\mu$  et une valeur de dispersion appelée  $\sigma$ . Par défaut, l'algorithme va diviser la variable quantitative en 5 classes (correspondant dans le discours à « très faible », « faible », « moyen »,

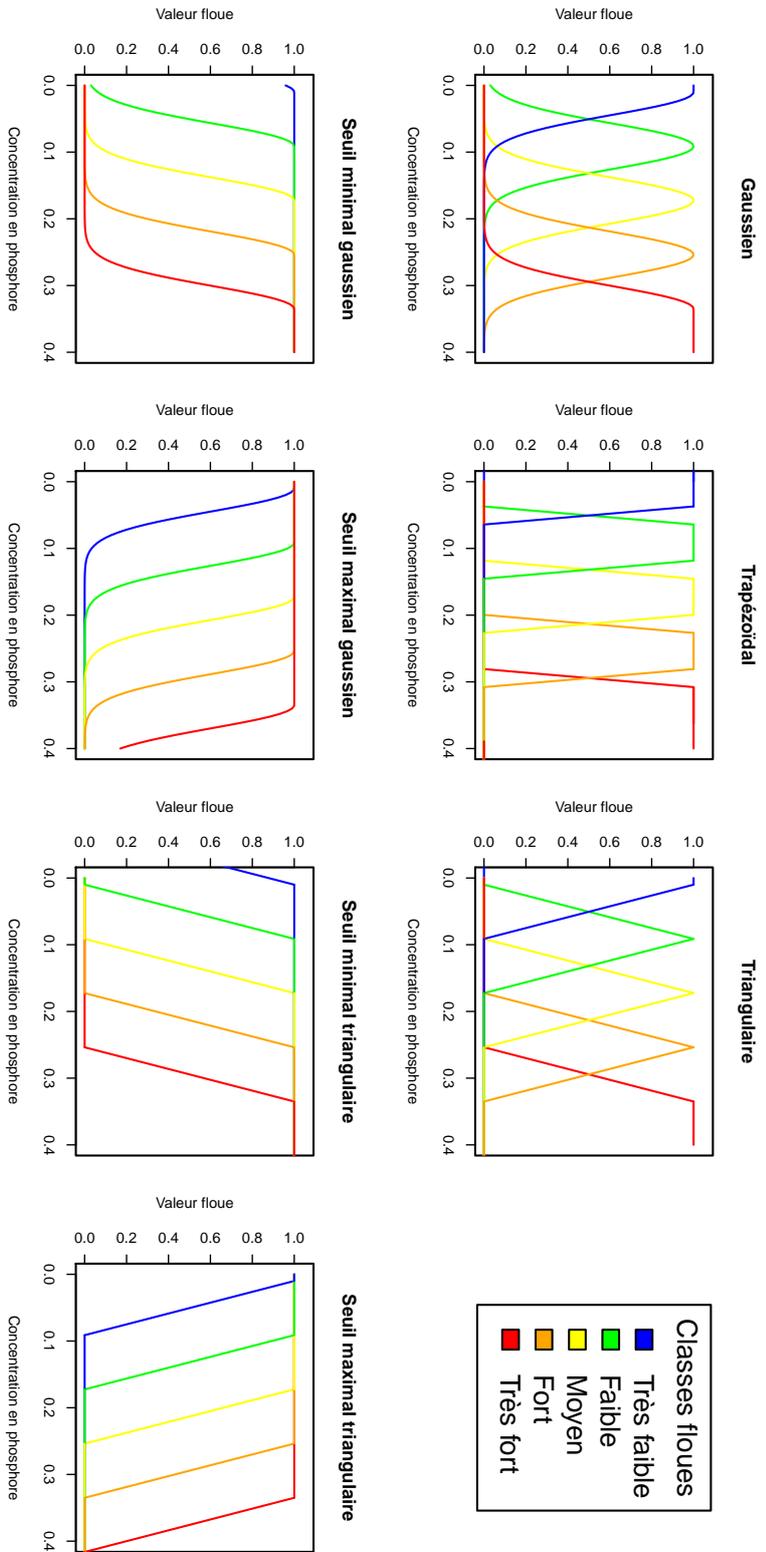


FIGURE 5.3 – Fonctions d'appartenance pour la fuzzification : exemple du phosphore. Ces graphiques montrent les différents types de fuzzification disponibles dans le paquetage patternTree. Chacune des fonctions est ici utilisée pour définir les valeurs de cinq classes floues (allant de « très faibles valeurs » à « très fortes valeurs de concentration en phosphore »), en fonction des valeurs initiales de phosphore. Les quatre dernières figures présentent des fonctions permettant d'utiliser des seuils, alors que les trois premières présentent des fonctions centrées autour d'une valeur  $\mu$ .

« fort », et « très fort »).

### 5.1.2.2 Les opérateurs logiques

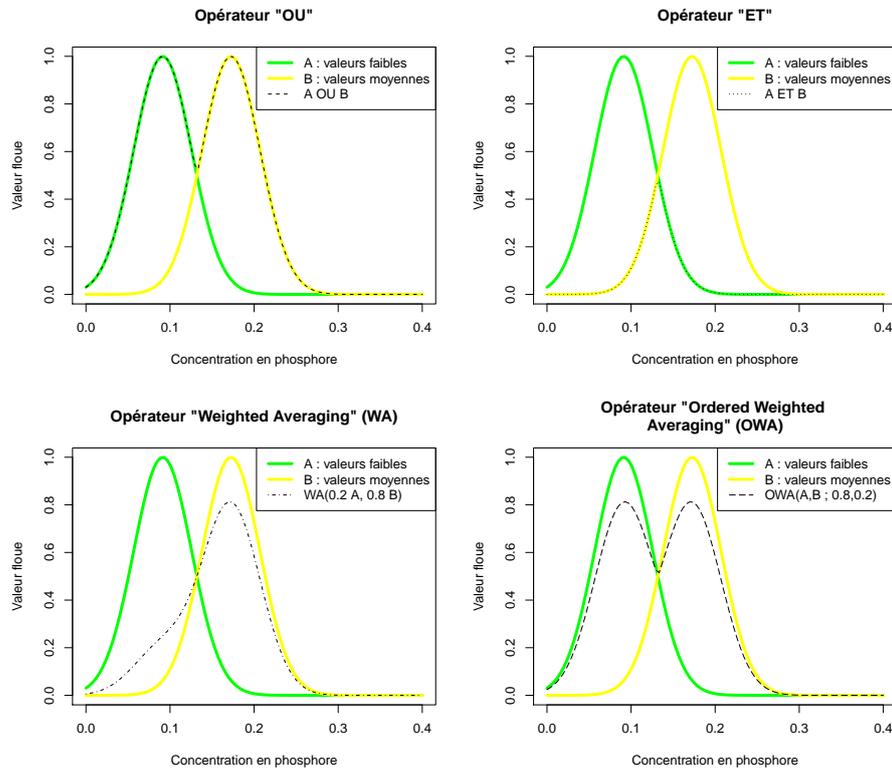


FIGURE 5.4 – Les opérateurs en logique floue. Quatre opérateurs sont présentés ici. Ils sont utilisés pour obtenir des valeurs floues pour des expressions liant deux classes floues (les classes "valeurs faibles" et "moyennes" de concentration en phosphore total). On voit que l'opérateur OU correspond à la fonction maximum, ET correspond à la fonction minimum, et les opérateurs WA et OWA à des situations intermédiaires. L'opérateur WA consiste à donner un poids différent à chacune des classes. L'opérateur OWA consiste à donner un poids différent à la classe au plus forte valeur et à la classe à plus faible valeur.

En logique classique, il existe trois opérateurs principaux : il s'agit de ET, OU et NON. Il est nécessaire de les traduire en logique floue, afin de créer des règles de classement (par exemple : *Si le niveau de phosphates est faible **ET** le niveau de nitrates est faible, alors le site est une référence*, ou encore *Si le niveau de phosphates est fort **OU** le niveau de nitrates est fort, alors le site est impacté*). La traduction mathématique des opérateurs ET et OU en logique floue peut être surprenante au premier abord, car il s'agit respectivement des fonctions « maximum » et « minimum » (voir Fig. 5.4).

Des opérateurs spécifiques ont aussi été développés spécialement pour la logique floue, et n'ont pas d'équivalents en logique classique. Les opérateurs WA

(Weighted Averaging) et OWA (Ordered Weighted Averaging) représentent des situations intermédiaires entre le ET et le OU dans le sens où ces opérateurs accordent des importances différentes aux propositions logiques qu'ils relient (voir Fig. 5.4). Ces opérateurs logiques ne sont pas encore utilisés dans le paquetage `patternTree` que nous avons développé, mais ils ont été implémentés et seront utilisés dans les versions ultérieures du paquetage.

### 5.1.2.3 Le paramètre d'optimisation

Optimiser les `pattern trees` revient à trouver les arbres dont les prédictions seront les plus proches de la variable à prédire. Nous devons donc utiliser une mesure de similarité entre les valeurs prédites et celles à prédire. Par défaut, le paquetage `patternTree` utilisera la statistique de dissimilarité *RMSE*, mais d'autres mesures ont été implémentées.

$$RMSE(A, B) = \sqrt{\frac{\sum_{j=1}^m (\mu_A(x_j) - \mu_B(x_j))^2}{m}}$$

où :

- $A$  est la variable prédite et  $B$  la variable à prédire.
- Les variables prédites et à prédire sont des suites de valeurs floues de longueur  $m$
- $\mu_A(x_j)$  représente la valeur floue n°  $j$  de la variable floue  $A$

La *RMSE* varie entre 0 (très mauvaise qualité de prédiction) et 1 (qualité parfaite de prédiction).

### 5.1.2.4 L'induction

L'induction des `pattern trees` est l'étape qui permet d'associer les variables floues afin de créer les règles floues finales, contenues dans des `pattern trees`. Sa fonction est comparable à l'apprentissage des SOM, mis à part que les SOM sont non-supervisées (pas de variable de sortie) et les `pattern trees` sont supervisés (on cherche à ajuster le modèle de façon à ce que ses valeurs de sortie ressemblent le plus possible à celles d'un jeu de données dit *d'apprentissage*).

Une description précise de l'algorithme d'induction utilisé dans le paquetage `patternTree` est disponible dans l'article de Huang et al. (2008).

### 5.1.2.5 « Simple pattern trees » et « general pattern trees »

A l'heure actuelle, seule l'induction des « simple pattern trees » (voir Huang et al., 2008) a été implémentée dans le paquetage `patternTree`, les « general pattern trees », feront partie des versions ultérieures du paquetage. Les « simple pattern trees » correspondent à des arbres dont toutes les branches correspondent à une seule classe logique, et qui sont reliées par les opérateurs logiques simples « ET » ou « OU ». Les « general pattern trees » peuvent permettre une architecture plus complexe des arbres, ressemblant par exemple à l'arbre de classification de la figure 5.1, et permettent également l'utilisation des opérateurs WA et OWA, présentés page 121.

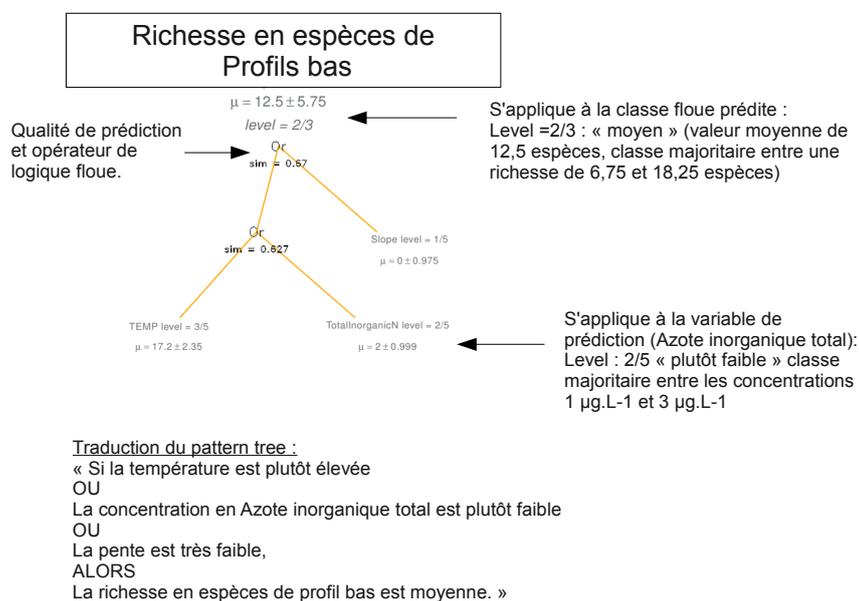


FIGURE 5.5 – Lecture d'un pattern tree, ici appliqué à la prédiction de la richesse en espèces de la guildes profil bas en fonction des variables environnementales.

### 5.1.2.6 Résultats des pattern trees

La figure 5.5 illustre l'exemple d'un pattern tree, et apporte des explications sur la façon de le lire.

## 5.2 Pattern trees et guildes des diatomées

(Naumoski et Mitreski, 2011b) ont utilisé les pattern trees dans le but de décrire des espèces peu connues du lac Prepa et d'en déduire leur capacité à caractériser les conditions du milieu. Nous les utiliserons ici pour former des règles floues concernant les guildes décrites par Passy (2007),

### 5.2.1 Matériel et méthode

Les données utilisées appartiennent à la base « historique » (voir page 30). Cette base de données comprend l'abondance relative de 847 espèces, prélevées dans 836 échantillons. La richesse spécifique des espèces appartenant aux trois guildes définies par Passy (2007) ont été calculées dans chacun des échantillons. Nous n'avons pas pu attribuer de guildes à 187 espèces (rares, ou dont la biologie n'a pas été décrite dans la littérature ; pour une description plus complète des références bibliographiques ayant conduit à l'attribution des guildes, voir page 77).

La base de données utilisée comprend aussi des données physico-chimiques, mesurées plusieurs fois (et moyennées) pendant une plage de temps de deux

mois comprenant le prélèvement de diatomées. Les variables mesurées sont :

- l' altitude (m, notée *ALT\_m*)
- la distance à la source (km, notée *Dist*),
- la pente (% , notée *Slope*)
- le pH (noté *PH*)
- la conductivité à 20 °C ( $\mu\text{S}/\text{cm}$ , notée *Cond20*)
- l'oxygène dissous (mg/l, noté *DissolvedOxygen* )
- l'azote total (mg/l, noté *TotalInorganicN*)
- les phosphates (mg/l, notés *PO<sub>4</sub>*)
- le calcium (mg/l, noté *Ca*)
- les carbonates (mg/l, notés *HCO<sub>3</sub>mg.l*)
- la température (°C, notée *TEMP*)
- les matières en suspension (mg/l, *SuspMatters*)
- l'ammonium (mg/l, noté *NH<sub>4</sub>*)
- les nitrates (mg/l, notés *NO<sub>3</sub>*)
- les nitrites (mg/l, notés *NO<sub>2</sub>* )
- l'alcalinité (meq/l, notée *Alkmeq*)

Les patterns trees ont permis de déterminer des règles floues expliquant la richesse des différentes guildes à partir des données environnementales. Ils ont été appliqués à partir du paquetage *patternTree*, selon l'algorithme et les calculs décrits dans le paragraphe 5.1. Nous les avons paramétrés de façon à ce qu'il ne puissent pas comprendre plus de 3 variables floues après induction, ce qui permet de former des règles plus simples et plus faciles à énoncer. Les variables environnementales ont été « fuzzifiées » en cinq classes floues, en utilisant une fuzzification « normale » (fig 5.3). Les proportions de guildes et leur richesses ont été fuzzifiées en trois classes floues, en utilisant aussi les fonctions normales. Seuls les résultats relatifs aux les classes floues décrivant les richesses spécifiques les plus fortes (level=3/3) seront présentés ici, car ce sont celles qui permettent de décrire au mieux les préférences environnementales des guildes.

## 5.2.2 Résultats et discussion

Les résultats de l'extraction des règles floues sont présentés dans la figure 5.6 Globalement, les trois pattern trees ont montré de fortes RMSE (entre 0.7 et 0.8, sur une échelle allant de 0 à 1). Ce résultat tend à montrer que les conditions environnementales agissent tout aussi fortement sur la diversité intra-gilde que sur la dominance d'une guildes sur les autres en termes d'abondance.

Les sites affichant de fortes richesses spécifiques en espèces appartenant à la guildes de profil bas présentent de faibles niveaux de nitrates et de fortes concentrations de calcium. Il est possible que cette affinité pour les fortes concentrations de calcium soit dû à un artefact de la base de données, où les sites de régions montagneuses à très forte hydrologie sont aussi les plus carbonatés. Les faibles niveaux de phosphate sont en accord avec la description que Passy (2007) a donné de cette guildes, tolérant les très faibles niveaux de nutriments.

Les guildes mobiles et profil haut présentent de fortes richesses spécifiques dans les sites caractérisés par de fortes températures. De telles conditions correspondent en général aux cours d'eau de plaine, dans lesquels les fortes températures sont accompagnées de faibles courants et de niveaux élevés de nutriments. Passy (2007), Passy et Larson (2011), et Lange et al. (2011) ont montré que les espèces de la guildes de profil haut s'installaient préférentiellement dans les

## Fortes richesses spécifiques

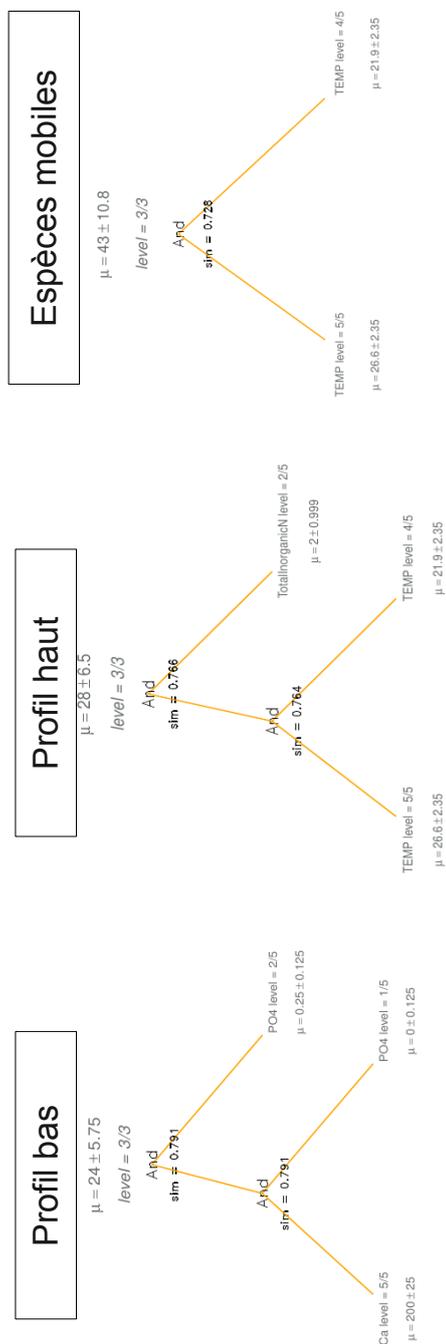


FIGURE 5.6 – Pattern trees appliqués à la prédiction des fortes richesses spécifiques des espèces des différentes, à partir des variables environnementales

sites riches en nutriments et subissant peu de perturbations hydrologiques. Ces espèces, vivant au dessus des espèces de profil bas, ont alors tendance à créer un gradient vertical de ressource (Burkholder et al., 1990), en consommant celles-ci avant qu'elles atteignent les niveaux inférieurs du biofilm. Ces gradients de ressources sont profitables aux espèces mobiles, qui peuvent adapter leur position dans le biofilm selon leur besoins en nutriments. Ce mécanisme tend à expliquer les résultats obtenus ici.

D'autres résultats, plus nombreux, sont disponibles en annexe, dans le document intitulé *Utilisation des Pattern Trees pour l'études des relations entre traits écologiques et biologiques et paramètres environnementaux. (Partenariat avec l'ONEMA-Action 12-3)*. Ils montrent eux aussi la possibilité d'utiliser les pattern trees afin de décrire les relations entre l'environnement et les traits des espèces.

### 5.2.3 Conclusions

La richesses spécifique en espèces de profil bas montre une relation intéressante avec les niveau de nutriments. Cette relation permet de penser que sa mesure pourrait s'avérer utile dans le cadre du calcul d'indices multimétriques tels que ceux utilisés aux États-Unis (Hill et al., 2000, voir page 11).

Cette étude a aussi montré la capacité des pattern trees à décrire de façon simple les gradients environnementaux agissant sur les assemblages de diatomées. Cette technique efficace pourrait se révéler intéressante dans plusieurs domaines :

**Règles d'assemblages** (Weiher et Keddy, 1999, voir page 1999). Le principe des règles d'assemblages est d'énoncer des règles simples, telles que celles que les pattern trees sont capables de déduire d'un jeu de données, concernant la façon dont les différentes espèces s'assemblent pour former des communautés. A ce sujet, Keddy et Weiher (1999) ont décrit les différentes étapes menant à l'élaboration de règles d'assemblages, et l'une d'entre elle consiste à tester à l'aide de tests statistiques ou de modèles nuls si ces règles affectent significativement les communautés. Il paraît donc nécessaire de développer une méthodologie associée aux pattern trees permettant de tester si les prédictions de ces modèles sont supérieures à celles obtenues aléatoirement.

**Recherche de seuils d'effets des polluants** Afin d'évaluer l'état écologique des cours d'eau, il est régulièrement demandé aux chercheurs de déterminer les seuils de concentration d'un polluant à partir desquels on observe un effet significatif sur les communautés biologiques. Or un des problèmes majeurs des méthodologies actuelles est lié au fait qu'elles sont fortement soumises aux erreurs de mesures, rendant difficile la détermination de seuils précis et fixes. Basés sur le principe de la logique floue, les pattern trees sont peu sensibles à ces fluctuations des mesure : l'utilisation d'un grand nombre de classes floues pour la prédiction du bon état écologique, évalué au préalable, pourrait s'avérer une solution efficace afin s'affranchir de ce type d'erreur lors de la détermination des seuils.

## Bibliographie

- Breiman, L., J. Friedman, R. Olshen, et C. Stone (1984). *Classification and regression tree*. Wadsworth and Brooks.
- Burkholder, J. M., R. G. Wetzel, et K. L. Klomparens (1990). Direct comparison of phosphate uptake by adnate and loosely attached microalgae within an intact biofilm matrix. *Applied and Environmental Microbiology* 56(9), 2882–2890.
- Garzón, M. B., R. Blazek, M. Neteler, R. S. d. Dios, H. S. Ollero, et C. Furlanello (2006). Predicting habitat suitability with machine learning models : The potential area of pinus sylvestris l. in the iberian peninsula. *Ecological Modelling* 197(3-4), 383–393.
- Gillett, N., Y. Pan, K. Manoylov, R. Stancheva, et C. Weillhoefer (2011). The potential indicator value of rare taxa richness in diatom-based stream bioassessment. *Journal of Phycology* 47(3), 471–482.
- Hill, B., A. Herlihy, P. Kaufmann, R. Stevenson, F. McCormick, et C. Burch Johnson (2000). Use of periphyton assemblage data as an index of biotic integrity. *Journal of the North American Benthological Society* 19(1), 50–67.
- Huang, Z., T. D. Gedeon, et M. Nikravesh (2008). Pattern trees induction : A new machine learning method. *Fuzzy Systems, IEEE Transactions on Fuzzy Systems* 16(4), 958–970.
- Keddy, P. et E. Weiher (1999). Introduction : The scope and goals of research on assembly rules. In E. Weiher et P. Keddy (Eds.), *Ecological assembly rules : perspectives, advances, retreats*, Chapter Introduction, pp. 1–20. Cambridge, GBR : Cambridge university press.
- Lange, K., A. Liess, J. Piggott, C. Townsend, et C. Matthaei (2011). Light, nutrients and grazing interact to determine stream diatom community composition and functional group structure. *Freshwater Biology* 56(2), 264–278.
- Mago, V., A. Mago, P. Sharma, et J. Mago (2011). Fuzzy logic based expert system for the treatment of mobile tooth. *Advances in Experimental Medicine and Biology* 696, 607–614.
- Naumoski, A. et K. Mitreski (2009). Climate change influence on diatoms biodiversity in lake prespa. pp. 25–30.
- Naumoski, A. et K. Mitreski (2011a). Diatom classification with novel bell based classification algorithm. *Communications in Computer and Information Science* 83 CCIS, 245–254.
- Naumoski, A. et K. Mitreski (2011b). Novel algorithm for diatom classification in lake prespa using log-normal distribution. *Ecology & Hydrobiology* 11(1), 23–34.
- Passy, S. et C. Larson (2011). Succession in stream biofilms is an environmentally driven gradient of stress tolerance. *Microbial Ecology* 62(2), 414–424.

- Passy, S. I. (2007). Diatom ecological guilds display distinct and predictable behavior along nutrient and disturbance gradients in running waters. *Aquatic Botany* 86(2), 171–178.
- Rahoma, W., U. Rahoma, et A. Hassan (2011). Application of neuro-fuzzy techniques for solar radiation. *Journal of Computer Science* 7(10), 1605–1611.
- Weiher, E. et P. Keddy (1999). *Ecological assembly rules : perspectives, advances, retreats*. Cambridge, GBR : Cambridge university press.
- Weilhoefer, C. et Y. Pan (2008). Using change-point analysis and weighted averaging approaches to explore the relationships between common benthic diatoms and in-stream environmental variables in mid-atlantic highlands streams, usa. *Hydrobiologia* 614(1), 259–274.





## Chapitre 6

# Conclusions et perspectives

### 6.1 Structure des assemblages et dispersion des espèces de diatomées

Les travaux de la première partie de cette thèse ont consisté à déterminer si des processus écologiques négligés chez les diatomées (limites de dispersion et relations biotiques) pouvaient avoir un impact sur la distribution et la structure de leurs assemblages. Dans le chapitre deux, nous avons démontré que des structures spatiales, indépendantes de l'environnement, influençaient partiellement ces assemblages. Dans le chapitre trois, les patrons de coexistence locale des diatomées n'ont pas abouti à une démonstration claire de l'influence des relations biotiques, mais les analyses ont permis de mettre à jour l'influence de paramètres généraux de la communauté sur la distribution des espèces. Dans les deux cas, il semblerait que les dynamiques de dispersion des espèces jouent un rôle majeur. Pourtant, les micro-organismes sont rarement considérés comme soumis à de tels phénomènes du fait des très grandes tailles de populations et des fortes capacités de dispersion passive liées à leur petite taille (Finlay, 2002; Fenchel et Finlay, 2004).

Cette vision des communautés de diatomées est en partie héritée d'expériences menées sur les micro-organismes marins. Persson (2002) a récolté 100 ml de sédiments marins sur la côte ouest de la Norvège. En mettant en culture les organismes récoltés dans des milieux de nature physico-chimique différente, il fut capable de retrouver près de 25% des espèces de dinoflagellés et de diatomées décrites le long de la côte norvégienne. Ceci lui a permis de développer l'idée que pour les micro-organismes « Tous sont partout, et l'environnement sélectionne ». Pourtant, la démonstration n'est pas complète : il est possible que ce cosmopolitisme ne soit propre qu'à une partie des micro-organismes, et qu'il concerne particulièrement les micro-organismes marins pour lesquels la dispersion est facilitée par la quasi-absence de barrières de dispersion (Soininen et al., 2007). Or la dispersion des diatomées benthiques en rivières est certainement limitée par l'hostilité du milieu environnant leur habitat, les résultats du chapitre deux ont d'ailleurs montré que les structures spatiales des assemblages étaient liées à de potentielles barrières géographiques. Aussi de nombreux chercheurs se sont opposés à l'idée du cosmopolitisme des micro-organismes et ont démontré que ceux-ci présentaient des patrons biogéographiques similaires

à ceux des macro-organismes (Martiny et al., 2006; Vanormelingen et al., 2008; Verleyen et al., 2009; Astorga et al., 2011). Ces travaux de thèse sont en accord avec les études citées, et ont de plus permis d'orienter la description des facteurs à l'origine de ces patrons géographiques vers des phénomènes de barrières géographiques (chapitre 2), de dynamiques de colonisation (chapitre 3) et de nestedness (chapitre 3).

## 6.2 Nestedness

Les résultats du chapitre deux ont révélé une perspective intéressante : l'étude de la nestedness. Il apparaît que ce patron est fortement lié aux structures de coexistence locale des espèces. Or, il existe des méthodes simples permettant d'analyser ses causes et ses déterminants (Ulrich et al., 2009), en répondant aux questions suivantes : quels sont les paramètres environnementaux locaux ayant un effet sur la distribution de la diversité, et quelles sont les caractéristiques des espèces déterminant leur valence écologique et la capacité de certaines d'entre elles à coloniser un grand nombre de milieux différents ? Ce dernier point pourrait s'avérer intéressant en bioindication, car il pourrait permettre de déterminer si des espèces sont rares car elles sont liées à des conditions environnementales rares (bon bioindicateurs), ou si elles sont rares pour des raisons indépendantes de leurs préférences environnementales telles que leurs capacités limitées de dispersion (mauvais bioindicateurs).

De plus, il est possible que la nestedness soit liée à la stratification régionale des assemblages. On peut faire l'hypothèse que les assemblages de chaque région homogène environnementalement se déclinent selon une séquence de communautés « emboîtées », du fait de relations de colonisation/extinction entre les communautés vivant dans des habitats équivalents et interconnectés. On peut aussi penser que ces séquences de nestedness au sein des régions sont dépendantes de la quantité de nutriments, à l'origine de différences de productivités, et donc de densités de populations et de capacités de recolonisation différentes. La nestedness deviendrait de cette manière un moyen inédit de comprendre l'effet de l'enrichissement des cours d'eau sur les communautés aquatiques. Afin de tester cette hypothèse de séquences régionales de nestedness, une solution utilisant les SOM présentée dans le chapitre quatre pourrait être envisagée. Les SOM décrivant la stratification régionale des assemblages, le calcul d'une statistique de nestedness pour tous les niveaux de classification, comme nous l'avons fait pour l'entropie (voir explications page 111 et fig. 4.9), permettrait de déterminer si la nestedness est maximale pour une échelle géographique particulière.

## 6.3 Présence et abondances des espèces

Il est intéressant de remarquer que les chapitres deux et trois, basé sur l'abondance relative des espèces d'une part, et leur simple présence de l'autre ont fourni des résultats similaires, montrant l'importance particulière des dynamiques de dispersion des espèces.

Les relations entre l'occurrence des espèces et les relations de dominance dans les communautés aquatiques sont plutôt méconnues. Pour les diatomées, on sait qu'il y a une relation forte entre des présences fréquentes régionales et la

capacité d'atteindre de fortes abondances à l'échelle des sites (voir paragraphe sur les espèces core-satellite, page 77). Les notions d'occurrence et d'abondance relative se traduisent en termes de diversité par les notions de richesse spécifique et d'evenness (voir page 25).

Soininen et al. (sous presse) ont étudié la relation entre richesse spécifique et evenness dans une méta-analyse reprenant les résultats de 229 études sur les communautés aquatiques. Cette étude n'a pas montré de patrons clairs et a permis de souligner le besoin de développement de méthodologies permettant de modéliser des hypothèses nulles de répartition aléatoire des abondances parmi les différentes espèces d'un assemblage. Aussi, en collaboration avec Sophia Passy, de l'université d'Arlington (Texas, USA), nous avons développé un tel modèle nul. Ce modèle nul nous permettra de tester l'influence des phénomènes de relation biotique, de niche et de dispersion aléatoire des espèces sur les relations de dominance entre les espèces (alors que l'étude présentée dans le chapitre trois se focalisait sur l'occurrence des espèces). Ce travail permettra donc de pallier une limite des travaux de thèse présentés ici, dans lesquels l'effet des relations biotiques a été décrit uniquement à partir de l'occurrence des espèces.

## 6.4 Le modèle de la bioindication

Ces travaux de thèse ont permis de souligner une structuration complexe des assemblages diatomiques, dans laquelle la forte dépendance des espèces aux conditions environnementales s'accompagne de processus complémentaires, comme l'influence des dynamiques de dispersion et l'influence des communautés sur les espèces. Ces phénomènes ont pour effet de créer des réponses complexes, non-linéaires et en partie aléatoires, des assemblages aux conditions environnementales. Or le modèle implicite utilisé par la bioindication ne prend pas suffisamment en compte cette complexité, et utilise un lien mécanistique entre l'environnement et la réponse de chacune des espèces.

La seconde partie de ce mémoire a souligné la possibilité de développer des méthodes adaptées à la structure établie des assemblages, qui pourront permettre de décrire de façon plus rigoureuse la répartition et les réponses des espèces de diatomées et d'affiner les pratiques de bioindication. Afin d'améliorer l'évaluation de la qualité écologique des cours d'eau par les diatomées, il semble important :

- d'affiner la stratification régionale dans les régions de montagne ;
- de tester la possibilité de calculer les valeurs spécifiques ; concernant les préférences des espèces sur une base régionale
- de tester la possibilité de déterminer pour chaque espèce des valeurs de sensibilité aux facteurs de dispersion et d'agrégation dans le calcul des indices, comme on le fait déjà pour la prise en compte de la sténocécie de l'espèce (voir page 10) ;
- de développer des métriques telles que la richesse en espèces des différentes guildes des assemblages permettant par la suite le développement d'indices multi-métriques.

## Bibliographie

- Astorga, A., J. Oksanen, M. Luoto, J. Soininen, R. Virtanen, et T. Muotka (2011). Distance decay of similarity in freshwater communities : do macro- and microorganisms follow the same rules? *Global Ecology and Biogeography*.
- Fenchel, T. et B. Finlay (2004). The ubiquity of small species : Patterns of local and global diversity. *BioScience* 54(8), 777–784.
- Finlay, B. (2002). Global dispersal of free-living microbial eukaryote species. *Science* 296(5570), 1061–1063.
- Martiny, J., J. C. R. Bohannan, B.J.M. and Brown, J. Fuhrman, J. Green, M. Horner-Devine, M. Kane, J. Krumins, C. Kuske, P. Morin, S. Naeem, L. Øvreås, A.-L. Reysenbach, V. Smith, et J. Staley (2006). Microbial biogeography : Putting microorganisms on the map. *Nature Reviews Microbiology* 4(2), 102–112.
- Persson, A. (2002). Proliferation of cryptic protists and germination of resting stages from untreated sediment samples with emphasis on dinoflagellates. *Ophelia* 55(3), 151–166.
- Soininen, J., J. J. Lennon, et H. Hillebrand (2007). A multivariate analysis of beta diversity across organisms and environments. *Ecology* 88(11), 2830–2838.
- Soininen, J., S. Passy, et H. Hillebrand (in press). The relationship between species richness and evenness : a meta-analysis of studies across aquatic ecosystems. *Oecologia*.
- Ulrich, W., M. Almeida-Neto, et N. J. Gotelli (2009). A consumer's guide to nestedness analysis. *Oikos* 118(1), 3–17.
- Vanormelingen, P., E. Verleyen, et W. Vyverman (2008). The diversity and distribution of diatoms : From cosmopolitanism to narrow endemism. *Biodiversity and Conservation* 17(2), 393–405.
- Verleyen, E., W. Vyverman, M. Sterken, D. A. Hodgson, A. De Wever, S. Juggins, B. Van de Vijver, V. J. Jones, P. Vanormelingen, D. Roberts, R. Flower, C. Kilroy, C. Souffreau, et K. Sabbe (2009). The importance of dispersal related and local factors in shaping the taxonomic structure of diatom meta-communities. *Oikos* 118(8), 1239–1249.

# Annexes



## Annexe A

# Rapport ONEMA sur la biotypologie utilisant les SOM



**Partenariat 2011** – Qualité des masses d'eau (cours d'eau, plans d'eau, masses d'eau de transition)  
*Action 12-3 : Développement et optimisation des méthodes de bioindication pour les cours d'eau*



# Typologie des flores diatomiques de référence et seuils d'état écologique à partir des notes d'IBD2007

*Révision sur la base des données issues du  
RRef et du RCS*

*Rapport final*

*Juliette Rosebery, Marius Bottin, Michel Coste, François Delmas*

*Septembre 2011*

## ***SOMMAIRE***

CONTEXTE, OBJECTIF ET DEMARCHE	1
METHODOLOGIE	2
1-BASES DE DONNEES	2
Jeu de données total	2
Jeu de données de référence	2
2- OUTILS D'ANALYSE	2
3-FIXATION DES SEUILS D'ETAT ECOLOGIQUE	3
RESULTATS	4
1-TYPOLOGIE	4
2-CALCUL DES SEUILS	6
3-COMPARAISON AVEC LA CLASSIFICATION SELON LA GRILLE INDICIELLE DE 2009	7
CONCLUSIONS	9
Références	10

## ***ANNEXES***

Annexe 1A : Répartition géographique des 5 groupes floristiques obtenus par la carte auto-organisante sur la base de l'ensemble des relevés	11
Annexe 1B : Répartition géographique des 5 groupes floristiques obtenus par la carte auto-organisante sur la base des relevés moyennés	12
Annexe 2 : Liste des espèces caractéristiques de chacun des 5 groupes floristiques	13
Annexe 3A : Valeurs brutes pour la comparaison des niveaux d'état écologique entre la grille de 2009 et celle de 2011, notes arrondies au demi-point le plus proche	16
Annexe 3B : Valeurs brutes pour la comparaison des niveaux d'état écologique 2011, selon que les notes seuils sont arrondies au demi ou au dixième de point le plus proche	16
Annexe 4 : Nouvelle grille indicielle proposée	17

## ***LISTE DES FIGURES***

Figure 1 : Résultat de la classification des prélèvements par la carte auto-organisante	4
Figure 2 : Courbe d'entropie calculée sur la variable « région diatomée obtenue en 2005 »	5
Figure 3 : Carte révisée des régions naturelles diatomées (fond HER1)	6
Figure 4 : Classifications en niveaux d'état écologique selon la grille indicienne de 2009 (EE2009) ou la grille proposée ici (EE2011)	8

## ***LISTE DES TABLEAUX***

Tableau 1 : Seuils physico-chimiques (en mg/l) pour la validation des relevés de référence utilisés durant le second exercice d'intercalibration	4
Tableau 2 : Limites d'état écologique obtenues par région diatomées	7
Tableau 3 : Limites d'état écologique, notes arrondies au demi-point le plus proche	7
Tableau 4 : Classifications en niveaux d'état écologique selon la grille indicienne de 2009 (EE2009) ou la grille proposée ici (EE2011) pour l'HER1	8
Tableau 5 : Classifications en niveaux d'état écologique selon la grille indicienne de 2009 (EE2009) ou la grille proposée ici (EE2011) pour l'HER4	8
Tableau 6 : Classifications en niveaux d'état écologique selon la grille indicienne de 2009 (EE2009) ou la grille proposée ici (EE2011) pour l'HER21	9

**Typologie des flores diatomiques de référence  
et seuils d'état écologique à partir des notes d'IBD2007 :  
Révision sur la base des données issues du Réseau de Référence  
et du Réseau de Contrôle et de Surveillance**

## **CONTEXTE, OBJECTIF ET DEMARCHE**

Le niveau des notations indicielles de qualité des écosystèmes « cours d'eau » est soumis à une variation naturelle non négligeable selon l'entité hydro-écorégionale à laquelle appartient l'hydrosystème considéré, et ceci quelle que soit la métrique utilisée. Afin de pouvoir déterminer une limite de Bon Etat Ecologique comparable d'une Hydro-Ecorégion (HER) à une autre, en 2005 et ce à partir d'une base de données historique constituée par le Cemagref de Bordeaux, une typologie des communautés diatomiques de référence avait été établie. Après avoir mis en correspondance ces communautés avec les différents types de masses d'eau, une grille de notes de référence et limites d'état écologique avait été proposée pour l'IBD2000. Les ratios de qualité écologique ou EQR dérivés des notes seuils pour les limites d'état écologique Très Bon/Bon et Bon/Moyen avaient été validés lors du premier exercice d'intercalibration européen.

En 2009, suivant l'entrée en vigueur de l'IBD2007 (AFNOR, 2007), cette grille de notes et EQR associés avait été revue puis intégrée à l'arrêté du 25 janvier 2010 (arrêté relatif aux méthodes et critères d'évaluation de l'état écologique, de l'état chimique et du potentiel écologique des eaux de surface). Le très récent second exercice d'intercalibration européen a validé ces nouvelles limites d'état écologique.

Aujourd'hui, l'exercice consiste à mettre à l'épreuve ces nouvelles limites exprimées en EQR, en analysant les données disponibles issues de la prospection du Réseau de Contrôle et de Surveillance de 2007 à 2009, et du réseau de Référence de 2005 à 2007.

La même démarche que celle mise au point en 2005 a été suivie. Tout d'abord une nouvelle typologie des communautés diatomiques de référence a été établie à l'aide d'un réseau de neurones non supervisé (carte auto-organisante). La correspondance entre ces communautés et les HER de niveau 1 a été étudiée, afin de donner une assise géographique à ces flores naturelles, et d'attribuer à chaque masse d'eau sa communauté de référence. 5 régions diatomiques ont ainsi été mises en évidence, régions très similaires à celles obtenues en 2005. Pour chacune de ces régions et donc pour chacune des différentes masses d'eau la composant, une note de référence a été calculée. Les notes seuils d'état écologique sont dérivées des EQR validés lors du second exercice d'intercalibration.

La nouvelle grille indicielle a été fixée après expertise de ce scénario. Le diagnostic d'état écologique des masses d'eau bassin par bassin est proposé sur la base de cette grille, et comparé au diagnostic de 2009.

## METHODOLOGIE

### 1-BASES DE DONNEES

#### *Jeu de données total*

Les données sont issues de la prospection du Réseau de Contrôle et de Surveillance de 2007 à 2009, et de la prospection des stations validées du réseau de référence de 2005 à 2007. En tout, la base comporte 3263 échantillons et 1152 taxons.

L'harmonisation taxonomique a été réalisée selon le référentiel du logiciel Omnidia 5.3, base 2009.

Nous avons émis certains critères de sélection des relevés à conserver, afin de limiter les principales sources de variabilité des communautés pouvant interférer avec l'effet écorégion et/ou altération. Ces critères sont les suivants :

-Substrats : galets uniquement (échantillons sur substrats végétaux éliminés) ;

-Saison : seuls les relevés de saison chaude (mois de mai, juin, juillet et août) ont été conservés, à raison d'un seul relevé par an.

-Prélèvements selon la norme NF T 90-354 (AFNOR, 2007)

Cette pratique a permis de mieux maîtriser la variabilité naturelle due à la saison et aux supports échantillonnés.

De plus, seuls les prélèvements pour lesquels les données physico-chimiques compatibles avec la date du prélèvement biologique (entre 45 jours avant et 15 jours après) étaient disponibles ont été conservés.

#### *Jeu de données de référence*

La typologie des communautés de référence pour le re-calcul des valeurs seuils d'état écologique concerne uniquement les prélèvements issus de la prospection 2005-2007 des stations validées du Réseau de Référence, et après application des critères de sélection ci-dessus.

Un tri supplémentaire a été réalisé afin de supprimer les prélèvements comportant un pourcentage d'espèces hétérotrophes et/ou polysaprobies non négligeable et caractérisant de ce fait une pollution suffisante pour biaiser l'analyse typologique.

Ainsi, 26 prélèvements présentant des abondances cumulées >4,5% des espèces suivantes ont été supprimés : *Nitzschia capitellata*, *Fistulifera saprophila*, *Nitzschia palea*, *Nitzschia paleacea*, *Luticola mutica*, *Gomphonema parvulum var. parvulum f. saprophilum*, *Craticula accomoda*.

4,5% représente l'écart maximal acceptable entre le % représenté par l'abondance cumulée de ces taxons dans un relevé et le 3<sup>ème</sup> quartile de cette abondance dans le jeu de données de références validées. 4,5% vaut 5 fois la distance inter-quartile (entre le 1<sup>er</sup> et le 3<sup>ème</sup> quartile).

La base de référence pour la typologie comporte à ce stade 389 prélèvements, pour 462 taxons dénombrés.

### 2- OUTILS D'ANALYSE

En vue de repérer les principaux types de communautés diatomiques présents sur le territoire national et en conditions de référence, les relevés floristiques issus des stations validées du réseau de référence sur la période 2005-2007 ont été analysés à l'aide d'un réseau artificiel de

neurones appelé Carte Auto-Organisante (ou SOM pour Self Organizing Map, Kohonen 1995), dont l'intérêt en écologie a déjà largement été démontré (Giraudel & Lek, 2001).

La SOM est une méthode d'ordination permettant de classer les individus dans un espace à 2 dimensions, par projection non linéaire et en fonction de leur similarité. Nos échantillons ont donc été classés selon la similarité des communautés diatomiques qu'ils hébergent. Les échantillons très similaires seront classés dans la même cellule, des échantillons proches dans 2 cellules proches, des échantillons très différents dans 2 cellules très éloignées etc. La qualité de la classification peut être appréhendée par 2 paramètres :

-l'erreur topographique : c'est le pourcentage de sites pour lesquels la cellule la plus proche n'est pas à côté de la 2<sup>ème</sup> cellule la plus proche de leur composition taxonomique. Ce paramètre évalue la qualité du gradient mis en place sur la carte.

-l'erreur de quantisation : c'est la somme des distances entre les sites et leurs cellules les plus proches, toujours du point de vue de la composition taxonomique.

Une analyse ascendante hiérarchique appliquée sur ces résultats a permis de dégager 5 groupes, chacun d'entre eux représentant une communauté particulière liée à des conditions environnementales spécifiques, et pouvant donc être considéré comme un écotype naturel.

Une analyse d'entropie (Shannon, 1948) a permis de comparer cette classification des communautés à la classification obtenue en 2005 (Tison et al, 2005). L'entropie permet d'apprécier si deux classements de variables sont similaires (ici la classification en 5 groupes de 2005 ou l'actuelle), en calculant un indice de diversité des groupes d'un classement au sein des groupes de l'autre classement. Lors de la classification hiérarchique ascendante appliquée aux résultats de la SOM, des nouveaux groupes sont créés en regroupant successivement d'autres groupes deux à deux. Il en résulte une augmentation du niveau d'entropie au sein des nouveaux groupes du fait que la probabilité de créer des groupes purs diminue.

Pour chacun de ces 5 groupes, nous avons pu déterminer la liste des espèces caractéristiques grâce au calcul de l'Indval (Annexe 2), indice mis au point par Dufrêne et Legendre (1997). Cet indice tient compte de l'abondance et de la probabilité d'occurrence de chaque espèce dans chacun des groupes, et attribue ainsi à chaque espèce une valeur indicatrice dont la valeur la plus forte correspond au groupe qu'elle caractérise le mieux. Un test de permutation (test de Monte-Carlo) a ensuite été appliqué à ces données d'IndVal afin de s'assurer de leur significativité (ce qui revient à tester que l'affectation par l'IndVal d'une espèce dans un groupe n'est pas due au hasard). L'existence de 5 assemblages typiques et comportant des différences marquées a ainsi pu être vérifiée.

Une dernière SOM a été réalisée sur des communautés moyennes, obtenues à partir des différents prélèvements disponibles sur un même site (plusieurs campagnes). Ceci permettant 1) de vérifier que les groupes obtenus ne sont pas représentatifs d'une variabilité annuelle des communautés mais bien d'une variabilité régionale et 2) de statuer définitivement sur l'appartenance d'une HER 1 à une région diatomées.

Les logiciels R et PcOrd4 ont été utilisés pour les analyses de données.

### *3-FIXATION DES SEUILS D'ETAT ECOLOGIQUE*

Pour le calcul des seuils, une sélection supplémentaire a été réalisée : tout prélèvement dont les caractéristiques physico-chimiques dépassaient les seuils maximum admis lors du 2<sup>nd</sup> round d'intercalibration (Tableau 1) ont été exclus de la base de données.

type GIG	N-NH4	N-NO3	P-PO4	DBO5
M1, 2, 4, 5	<0,09	<1,15	<0,06	<2
A1, 2	<0,1	<4	<0,04	<2,4
C2	<0,05	<4	<0,03	<2,4
C3	<0,05	<2	<0,02	<2
C1, 4, 5, 6	<0,1	<4	<0,04	<2,4

**Tableau 1** : Seuils physico-chimiques (en mg/l) pour la validation des relevés de référence utilisés durant le second exercice d'intercalibration (Type GIG : nomenclature européenne des types de masses d'eau)

La base finale comporte 291 échantillons.

Selon la même procédure qu'en 2005 et 2009, nous avons fixé la valeur de référence de l'IBD2007 à la médiane de la distribution des notes d'indice des stations de référence.

Le calcul des Ratios de Qualité Ecologique (EQR) repose sur la formule de base suivante :

$EQR = (\text{indice diatomique observé} - \text{note minimale}) / (\text{indice diatomique de référence} - \text{note minimale})$

Les seuils suivants exprimés en EQR ont été validés par le second round d'intercalibration et correspondent aux seuils adoptés dans la grille indicielle de 2009 pour l'IBD2007 :

Très-Bon/Bon : 0,94

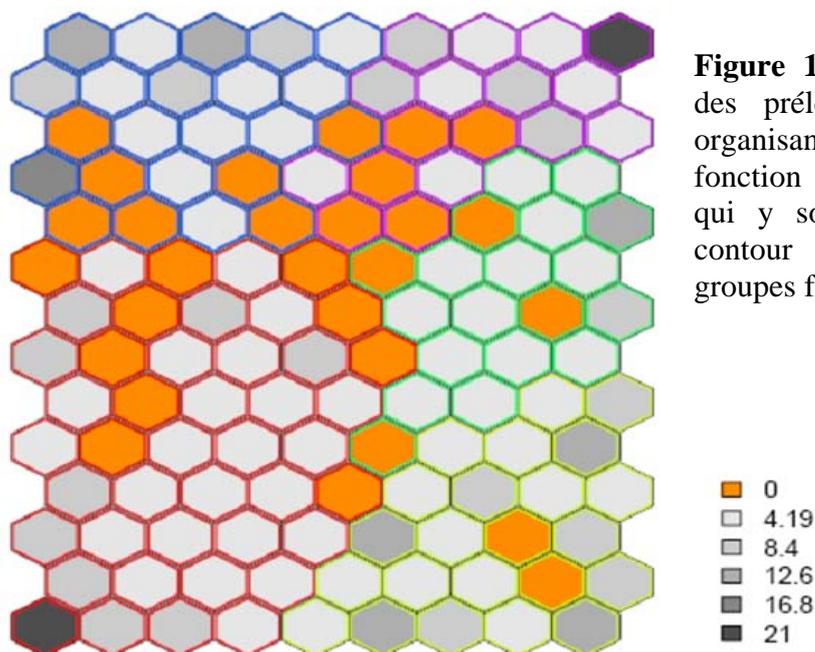
Bon/Moyen : 0,78

Ces seuils sont appliqués à la nouvelle note de référence afin d'en dériver les notes seuils d'état écologique.

## RESULTATS

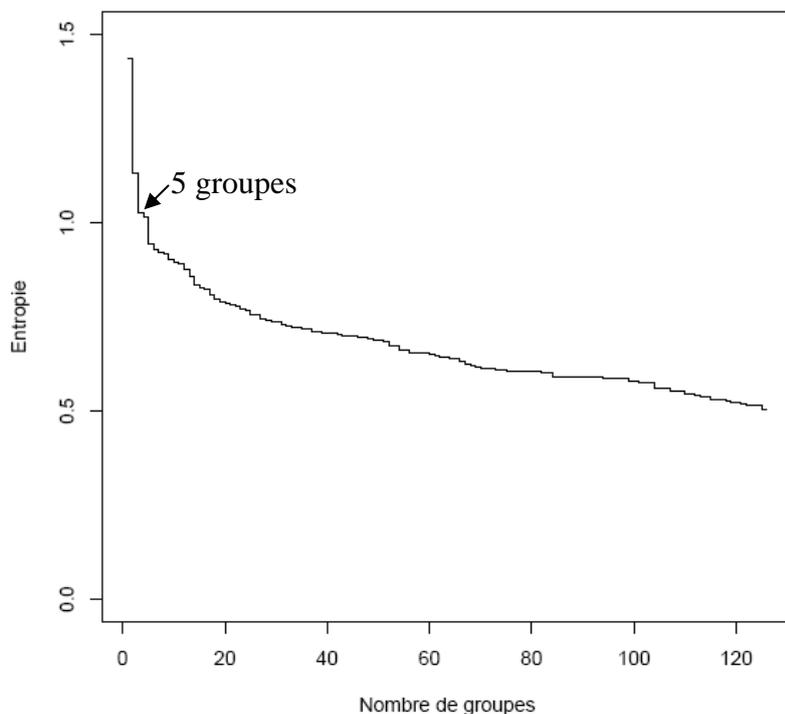
### 1-TYPOLOGIE

La classification des prélèvements par carte auto-organisante est très satisfaisante d'après les critères de validation classiques (erreur topographique : 0,02290076 ; erreur de quantisation : 178,4755) (Figure 1).



**Figure 1** : Résultat de la classification des prélèvements par la carte auto-organisante : la couleur de la cellule est fonction du nombre minimal de relevés qui y sont groupés. Les couleurs de contour des cellules délimitent les 5 groupes floristiques naturels.

Le choix d'un niveau de classification à 5 groupes se justifie écologiquement, et permet de retrouver quasiment les 5 mêmes régions naturelles qu'en 2005 (et 2009). La courbe d'entropie calculée sur la variable « région diatomée obtenue en 2005 » montre en effet qu'on arrive rapidement à des niveaux très bas (Figure 2), et qu'à 5 groupes on obtient donc une très bonne correspondance avec les résultats de 2005.

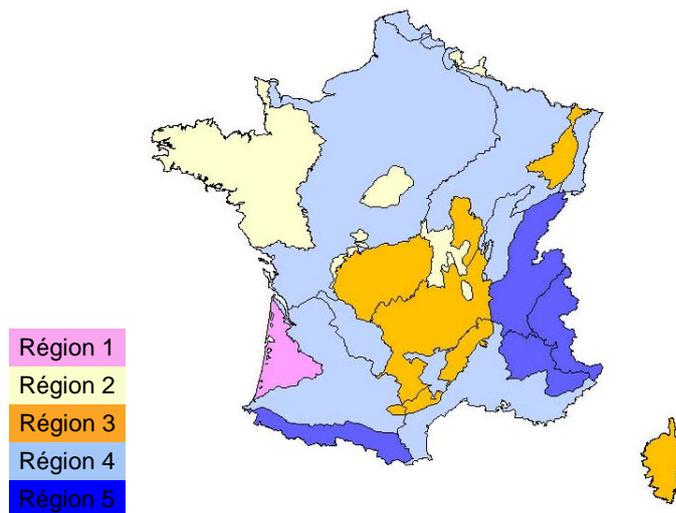


**Figure 2** : Courbe d'entropie calculée sur la variable « région diatomée obtenue en 2005 »

Les figures en annexe 1A montrent la projection géographique de ces 5 nouveaux groupes sur fond de HER1.

On observe toutefois qu'au sein des Cévennes et du Massif Central Nord différents groupes sont simultanément représentés, sous l'effet d'une variabilité inter-annuelle des flores peut-être, et qu'il est nécessaire de statuer définitivement sur la classification de ces HER en région diatomées. Nous avons donc conduit une nouvelle SOM sur les communautés moyennées (non représentée) (194 relevés moyennés ; erreur topographique : 0; erreur de quantisation : 82,30). Ainsi l'annexe 1B montre la projection géographique des 5 groupes obtenus à partir des communautés moyennées, intégrant la variabilité temporelle des communautés, lorsqu'un même site est prospecté au cours de plusieurs campagnes annuelles.

La figure 3 ci-dessous fait la synthèse de ces observations et propose la carte des régions naturelles révisée.



**Figure 3** : Carte révisée des régions naturelles diatomées (fond HER1).

Région 1 : contexte acide des Landes (note minimale : 5/20)

Région 2 : plaines et faibles altitudes peu carbonatées (note minimale : 1/20)

Région 3 : régions moyennement élevées à élevées, peu carbonatées : Massif Central Nord et Sud, Vosges, Cévennes (note minimale : 5/20)

Région 4 : plaines et faibles altitudes carbonatées (note minimale : 1/20)

Région 5 : montagnes, plutôt carbonatées à franchement carbonatées (note minimale : 5/20)

On observe par rapport à la carte de 2005 (utilisée également en 2009) que les Pyrénées rejoignent les Alpes pour former un groupe de hautes montagnes, et que les Vosges et le Massif Central Nord se classent avec la région 3 (région 2 auparavant).

Ce nouveau classement paraît plus logique que celui de 2005 d'après les caractéristiques hydro-géologiques et altitudinales des HER1 concernées.

Certains sites présentent systématiquement des flores acidophiles et se classent avec la région 1 malgré le fait qu'ils ne se situent pas dans les Landes: Elorn à Commana (Massif Armoricaïn), le ruisseau de Menorrupt à Sapoïs (Vosges), le Cayla à Sousceyrac (Massif Central Sud), l'Allier à Chasserades (Cévennes), et plusieurs points du Massif Central Nord : la Maulde à Gentioux-Pigerolles, la Combade à Doms, la Diège à Sornac, la Vézère au Bugeat, la Corrèze des Pradines à Gourdon-Murat.

L'Annexe 2 présente après calcul de l'IndVal les espèces caractéristiques de ces 5 types de flores naturelles.

## *2-CALCUL DES SEUILS*

La note de référence de chacun des 5 nouveaux groupes floristiques obtenus correspond à la médiane des notes de référence au sein de ce groupe.

Les limites Très Bon Etat/Bon Etat (TB/B) et Bon Etat/Etat Moyen (B/Moy) sont dérivées des limites exprimées en EQR validées par le second exercice d'intercalibration (EQR valant respectivement 0,94 et 0,78). Les limites suivantes sont fixées aux EQR 0,55 et 0,30 conformément à la grille proposée en 2009.

Le tableau 2 présente les notes seuils ainsi obtenues :

	note de référence	très bon/ bon	bon/moyen	moyen/ médiocre	médiocre/ mauvais	note minimale
région 1	18,4	17,6	15,5	12,4	9	5
région 2	17,4	16,4	13,8	10	5,9	1
région 3	19	18,2	15,9	12,7	9,2	5
région 4	18,1	17,1	14,3	10,4	6,1	1
région 5	20	19,1	16,7	13,3	9,5	5

**Tableau 2** : Limites d'état écologique obtenues par région diatomée

Le Tableau 3 ci-dessus propose d'arrondir les notes précédemment obtenues au demi-point le plus proche, afin d'obtenir une grille plus « opérationnelle », compte-tenu de l'imprécision attachée à une note d'indice (cet arrondi avait été utilisé de la même façon en 2005 et 2009) :

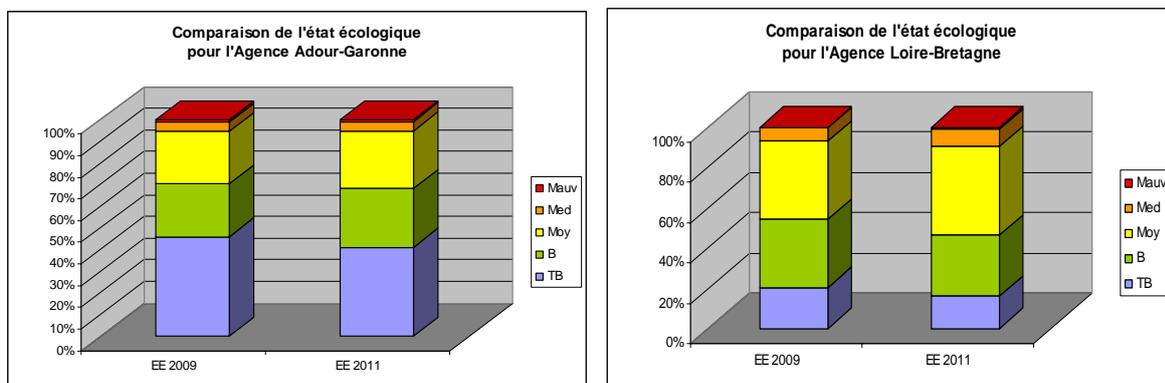
	note de référence	très bon/ bon	bon/moyen	moyen/ médiocre	médiocre/ mauvais	note minimale
région 1	18,5	17,5	15,5	12,5	9	5
région 2	17,5	16,5	14	10	6	1
région 3	19	18	16	13	9,5	5
région 4	18	17	14,5	10,5	6	1
région 5	20	19	16,5	13	9,5	5

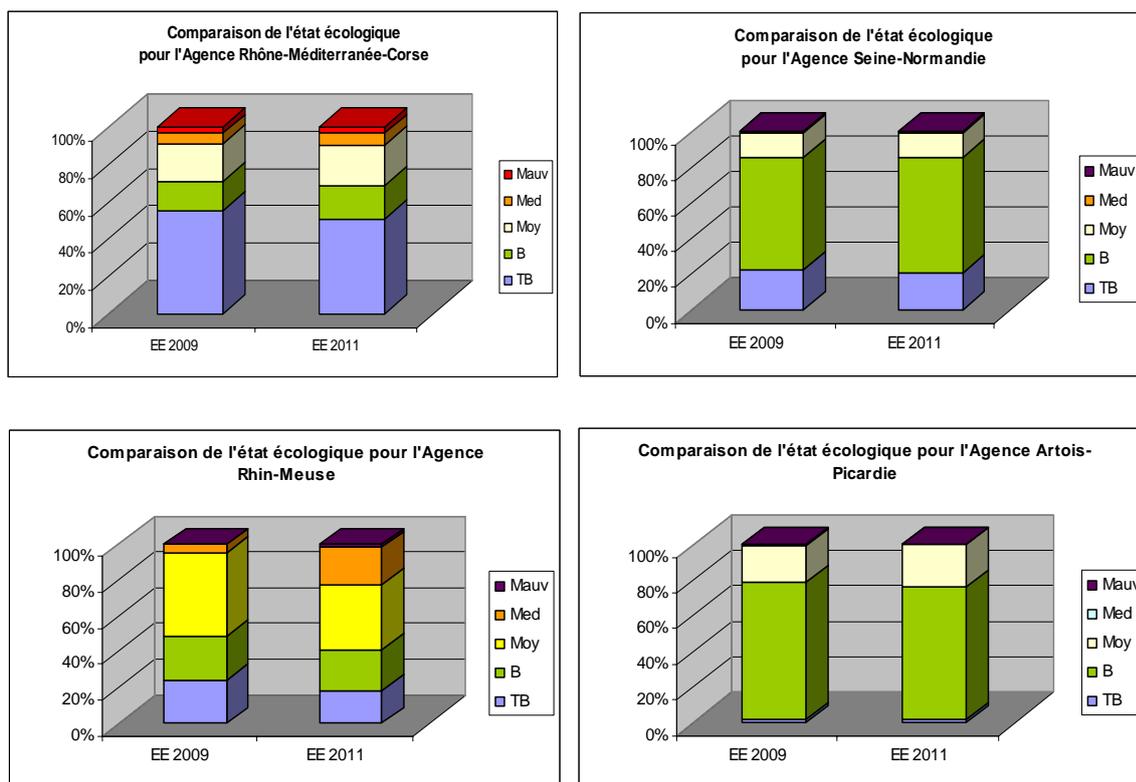
**Tableau 3** : Limites d'état écologique, notes arrondies au demi-point le plus proche

La grille indicielle dérivée du Tableau 3 est proposée en Annexe 4. Par rapport à 2009 il n'est plus proposé de prendre en compte un basculement sur les notes seuils de la région 1 en présence d'un contexte particulièrement acide (hors Landes).

### 3-COMPARAISON AVEC LA CLASSIFICATION SELON LA GRILLE INDICIELLE DE 2009

Les données utilisées ici sont issues de la prospection du Réseau de Contrôle et de Surveillance de 2007 à 2009, et de la prospection des stations validées du réseau de référence de 2005 à 2007 (export complet de la base Pandore d'avril 2011). Si l'on compare la classification en niveaux d'état écologique, bassin par bassin, d'après la grille indicielle de 2009 ou la grille revue ici (notes arrondies au demi-point le plus proche), ces classifications sont identiques selon un test classique du  $\chi^2$  au seuil 0,05 (Figure 4 et Annexe 3A pour les valeurs brutes), sauf pour Loire-Bretagne et Rhin-Meuse. L'attribution d'un niveau d'état écologique est faite d'après la note d'IBD2007.





**Figure 4 :** Classifications en niveaux d'état écologique selon la grille indicielle de 2009 (EE2009) ou la grille proposée ici (EE2011) (TB : Très Bon ; B : Bon ; Moy : Moyen ; Med : Médiocre ; Mauv : Mauvais).

Si l'on regarde plus précisément la répartition en niveaux d'état écologique concernant les 3 HER ayant changé de région, on observe des différences significatives ( $p < 0.05$ ).

Pour l'HER1 « Pyrénées », les différences semblent *a priori* modestes :

HER1	Très Bon	Bon	Moyen
EE 2009	135	12	3
EE 2011	119	25	6

**Tableau 4 :** Classifications en niveaux d'état écologique selon la grille indicielle de 2009 (EE2009) ou la grille proposée ici (EE2011) pour l'HER1

L'état écologique pour l'HER4 « Vosges » pourrait en revanche avoir été davantage surestimé en 2009, ce qui impacte directement l'évaluation actuelle de l'état écologique pour le bassin Rhin-Meuse :

HER4	Très Bon	Bon	Moyen	Médiocre
EE 2009	56	24	29	0
EE 2011	40	20	38	11

**Tableau 5 :** Classifications en niveaux d'état écologique selon la grille indicielle de 2009 (EE2009) ou la grille proposée ici (EE2011) pour l'HER4

Il en est de même pour l'HER21 « Massif Central Nord », ce qui explique par ailleurs que le bassin Loire-Bretagne obtient également en 2011 un état écologique moyen inférieur à celui de 2009.

HER21	TB	B	Moy	Med	Mauv
EE 2005	152	111	46	3	0
EE 2011	89	82	116	24	1

**Tableau 6** : Classifications en niveaux d'état écologique selon la grille indicielle de 2009 (EE2009) ou la grille proposée ici (EE2011) pour l'HER21

Outre le fait que ce nouveau classement est plus logique du point de vue des caractéristiques géochimiques et altitudinales de ces HER, si l'on regarde de plus près les notes de référence de chacune d'elles, celles-ci sont tout à fait en adéquation avec la note de référence de leur région d'attribution.

Ainsi l'HER21 affiche une note de référence de 18,7/20 et l'HER 4 une note de 19,5/20, ces 2 HER appartenant à la région 3 dont la note de référence est fixée à 19/20 (en 2009, les HER 4 et 21 étaient classées dans la région 2 dont la note de référence était fixée à 17,5/20).

Egalement, la note de référence de l'HER1 est de 20/20, identique à celle de sa région de rattachement (région 5).

Concernant les Landes, la baisse de la note de référence par rapport à 2009 ne semble pas affecter significativement l'évaluation de l'état écologique (Annexe 3A).

Enfin, il est à noter que la classification en niveaux d'état écologique obtenue pour chaque bassin à partir des notes arrondies au demi-point supérieur (Tableau 3), ne diffère pas significativement de la classification obtenue à partir des notes brutes issues du calcul (Tableau 2) (Annexe 3B).

## CONCLUSIONS

Selon la même démarche que celle adoptée en 2005 et 2009, et à partir des nouvelles données issues de la prospection des stations de référence validées, une typologie des communautés naturelles de diatomées benthiques a été réalisée. Les 5 mêmes grands types de communautés ont été retrouvés, mais le tracé des régions de répartition de celles-ci, sur la base des HER1, a quelque peu évolué.

De nouvelles notes de référence ont été calculées, assez comparables aux notes fixées en 2009, et les notes seuils d'état écologiques ont été dérivées des EQR validés lors du 2<sup>nd</sup> round d'intercalibration. La grille indicielle obtenue diffère peu de celle de 2009. Les notes seuils sont fixées au demi-point, cet arrondissement jugé plus pratique du point de vue opérationnel n'entraînant pas de modification significative du classement en niveaux d'état écologique au sein des différents bassins.

En revanche, la nouvelle grille indicielle impacte significativement les HER dont la « région diatomée » d'appartenance a été revue depuis 2009. Le classement actuel aboutit ainsi pour les Vosges et le Massif Central Nord à un état écologique moyen plus bas qu'avec le classement de 2009, ceci se répercutant au niveau de diagnostic global des bassins Rhin-Meuse et Loire-Bretagne. Ce nouvel état des lieux est néanmoins jugé plus logique, du point de vue des caractéristiques géochimiques et altitudinales de ces HER, ainsi que du point de vue de leur note de référence.

Enfin, le principe de basculement sur la grille de l'HER Landes en cas d'acidité contextuelle n'est pas conservé.

## Références

AFNOR – Norme NF T90-354 Décembre 2007. Qualité de l'eau - Détermination de l'Indice Biologique Diatomées (IBD)

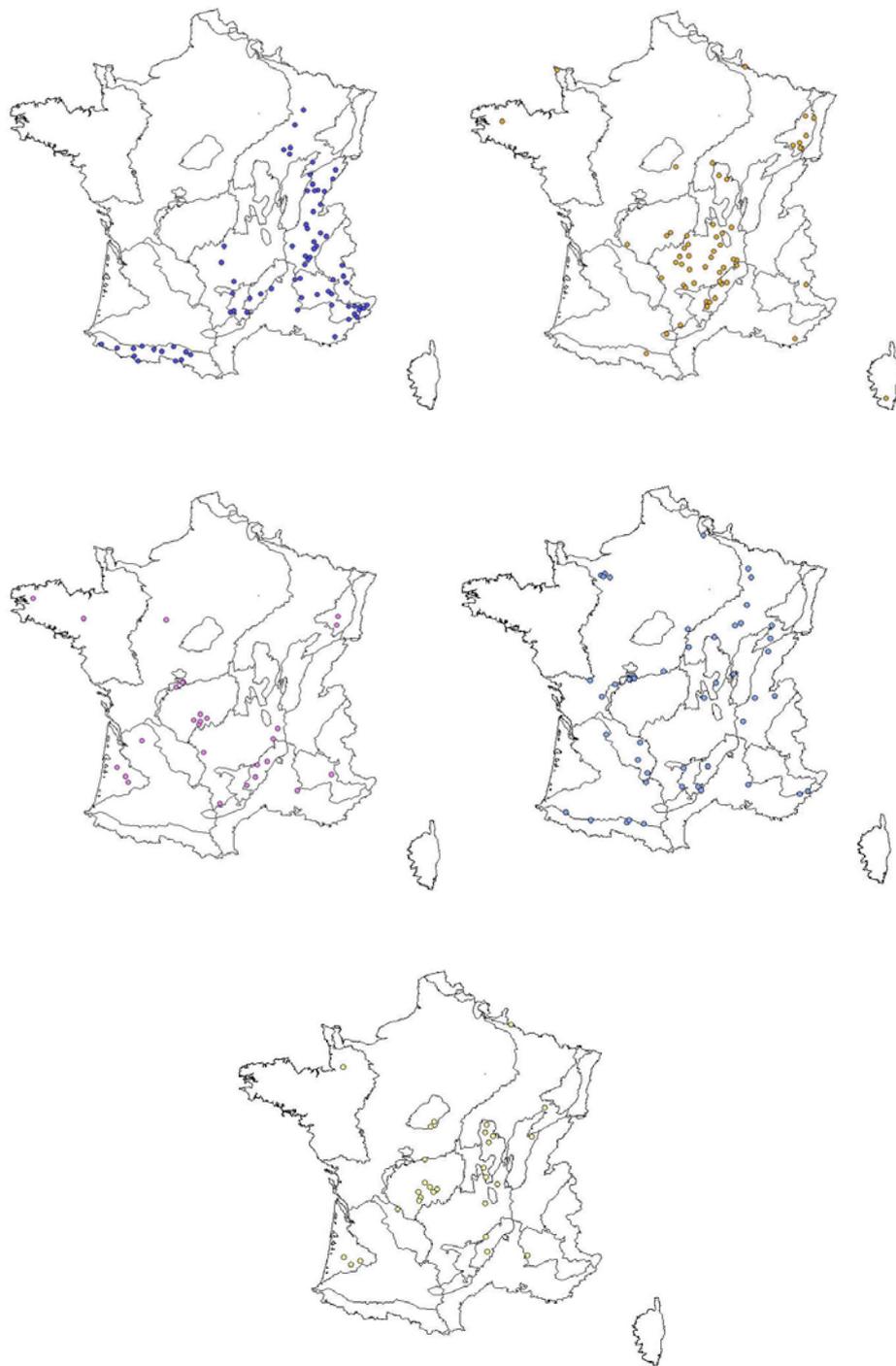
Dufrêne M. and Legendre, P., 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecological Monographs*, 67: 345-66

Giraudel, J.L. and Lek, S., 2001. A comparison of Self-Organizing Map algorithm and some conventional statistical methods for ecological community ordination. *Ecological Modelling*, 141: 329-339.

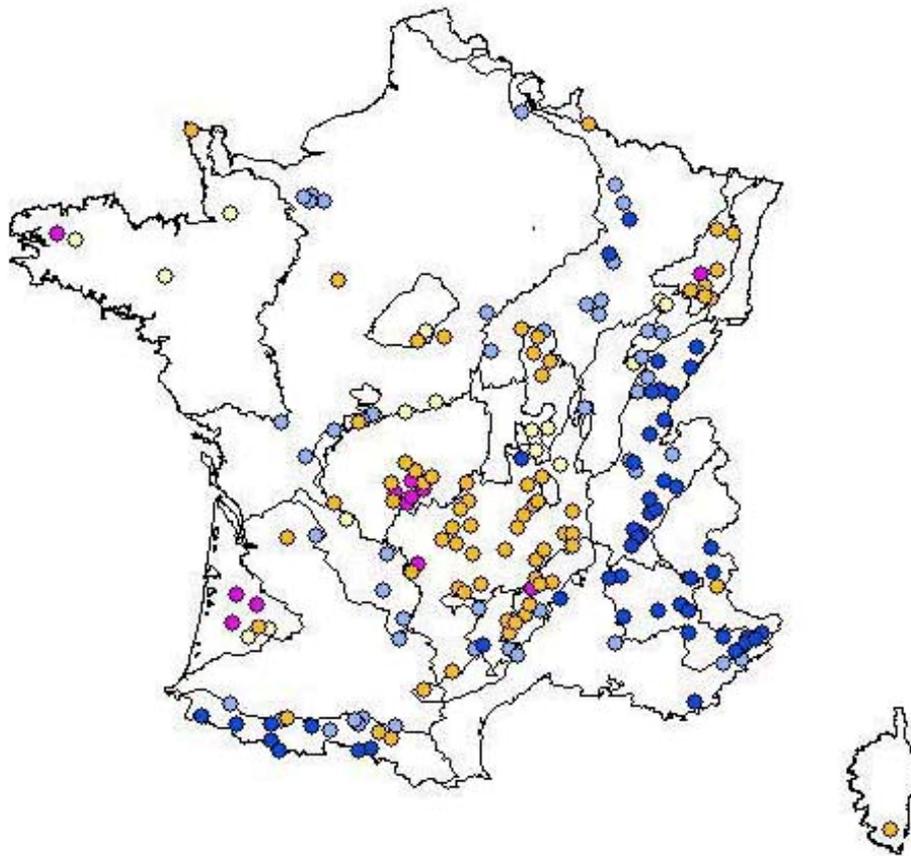
Kohonen, T., 1995. *Self-Organizing Maps*, volume 30 of Springer Series in Information Sciences}. Springer, Berlin, Heidelberg. (Second Extended Edition 1997).

Shannon, C. E., 1948. A mathematical theory of communication. *Bell System Technical Journal*, vol. 27, pp. 379-423 and 623-656.

Tison, J., Coste, M., Delmas, F., 2005.- Flores diatomiques des cours d'eau : proposition de valeurs limites du « Bon Etat » pour l'IPS et l'IBD. Ministère de l'Environnement et du Développement Durable, Direction de l'Eau : 20 p.



**Annexe 1A** : Répartition géographique des 5 groupes floristiques obtenus par la carte auto-organisante sur la base de l'ensemble des relevés



**Annexe 1B** : Répartition géographique des 5 groupes floristiques obtenus par la carte auto-organisante sur la base des relevés moyennés.

## Annexe 2 : Liste des espèces caractéristiques de chacun des 5 groupes floristiques

Code	Dénomination de l'espèce	Région	IndVal	p value
KOBG	Karayevia oblongella (Oestrup) M. Aboal	1	70,7	0,001
EMIN	Eunotia minor (Kützing) Grunow in Van Heurck	1	40,5	0,001
ADHE	Achnanthes helveticum (Hustedt) Monnier Lange-Bertalot & Ector	1	39,9	0,001
FCAP	Fragilaria capucina Desmazieres var. capucina	1	38,8	0,001
NAAN	Navicula angusta Grunow	1	35,3	0,001
FVIR	Fragilaria virescens Ralfs	1	35,2	0,001
EEXI	Eunotia exigua (Brebisson ex Kützing) Rabenhorst	1	33,5	0,001
TFLO	Tabellaria flocculosa (Roth) Kützing	1	32,6	0,001
ETEN	Eunotia tenella (Grunow) Hustedt	1	31,8	0,001
SRBA	Surirella roba Leclercq	1	30,2	0,001
EINC	Eunotia incisa Gregory var. incisa	1	28,8	0,001
ADMI	Achnanthes minutissimum (Kützing) Czarnecki	1	26,9	0,005
GEXL	Gomphonema exilissimum (Grunow) Lange-Bertalot & Reichardt	1	26,4	0,001
ADSO	Achnanthes subatomoides (Hustedt) Monnier, Lange-Bertalot et Ector	1	25,1	0,001
FRUM	Fragilaria rumpens (Kütz.) G.W.F. Carlson	1	20,6	0,002
EMUC	Eunotia mucophila (Lange-Bert. & Norpel Schempp) Lange-Bertalot	1	19,9	0,001
EIMP	Eunotia implicata Nörpel. Lange-Bertalot & Alles	1	18,2	0,001
ABRT	Achnanthes bioretii (Germain) Edlund	1	17,8	0,001
FGRA	Fragilaria gracilis Østrup	1	16,4	0,003
PSEL	Pinnularia subcapitata Gregory var. elongata Krammer	1	16,3	0,001
FKRA	Frustulia krammeri Lange-Bertalot & Metzeltin	1	14,9	0,001
GPVL	Gomphonema parvulus Lange-Bertalot & Reichardt	1	14,4	0,001
ESUB	Eunotia subarcuatoidea Alles Nörpel & Lange-Bertalot	1	14,2	0,001
ENNG	Encyonema neogracile Krammer	1	13	0,001
EUIN	Eunotia intermedia (Krasske ex Hustedt) Nörpel & Lange-Bertalot	1	12,7	0,001
FBID	Fragilaria bidens Heiberg	1	7	0,018
TVEN	Tabellaria ventricosa Kützing	1	6,6	0,009
EARC	Eunotia arcus Ehrenberg var. arcus	1	6,4	0,011
NNOT	Navicula notha Wallace	1	6,4	0,01
ADMS	Adafia minuscula (Grunow) Lange-Bertalot	1	5,7	0,027
GCLA	Gomphonema clavatum Ehr.	1	5,3	0,048
EPEC	Eunotia pectinalis (Dyallwyn) Rabenhorst var. pectinalis	1	4,6	0,017
BVIT	Brachysira vitrea (Grunow) Ross in Hartley	1	3,9	0,033
NLAN	Navicula lanceolata (Agardh) Ehrenberg	2	63,7	0,001
PLFR	Planothidium frequentissimum (Lange-Bertalot) Lange-Bertalot	2	57,6	0,001
NCRY	Navicula cryptocephala Kützing	2	40,8	0,001
NGRE	Navicula gregaria Donkin	2	39,4	0,001
NPAL	Nitzschia palea (Kützing) W. Smith	2	38,3	0,001
ADSH	Achnanthes subhudsonis (Hustedt) H. Kobayasi	2	35,9	0,001
NRHY	Navicula rhynchocephala Kützing	2	35	0,001
EOMI	Eolimna minima (Grunow) Lange-Bertalot	2	33,1	0,001
ESLE	Encyonema silesiacum (Bleisch in Rabh.) D.G. Mann	2	31,4	0,003
MVAR	Melosira varians Agardh	2	27,3	0,001
GPAR	Gomphonema parvulum (Kützing) Kützing var. parvulum f. parvulum	2	26,8	0,001
FCVA	Fragilaria capucina Desmazieres var. vaucheriae (Kützing) Lange-Bertalot	2	26,1	0,001
ENMI	Encyonema minutum (Hilse in Rabh.) D.G. Mann	2	25,1	0,002
AAMB	Aulacoseira ambigua (Grunow) Simonsen	2	23,9	0,001
SSVE	Staurosira venter (Ehr.) Cleve & Moeller	2	23,6	0,001
NGER	Navicula germainii Wallace	2	23,4	0,001
ADEU	Achnanthes eutrophilum (Lange-Bertalot) Lange-Bertalot	2	23,2	0,001
PDAU	Planothidium dauyi (Foged) Lange-Bertalot	2	21	0,001
LGOE	Luticola goeppertiana (Bleisch in Rabenhorst) D.G. Mann	2	19	0,001
NINC	Nitzschia inconspicua Grunow	2	18,2	0,006
NANT	Navicula antonii Lange-Bertalot	2	18,1	0,002
DSTE	Discostella stelligera (Cleve et Grunow) Houk & Klee	2	17,9	0,001

Code	Dénomination de l'espèce	Région	IndVal	p value
PTPE	Planothidium peragallii (Brun & Heribaud)Round & Bukhtiyarova	2	16	0,001
ADRI	Achnanthydium rivulare Potapova & Ponader	2	14,7	0,002
NACD	Nitzschia acidoclinata Lange-Bertalot	2	14,4	0,001
KALA	Karayevia laterostrata (Hustedt) Bukhtiyarova	2	14,1	0,001
NPAD	Nitzschia palea (Kützing) W.Smith var.debilis(Kützing)Grunow in Cl. & Grun	2	12,7	0,001
SSEM	Sellaphora seminulum (Grunow) D.G. Mann	2	12,6	0,005
ESBM	Eolimna subminuscula (Manguin) Moser Lange-Bertalot & Metzeltin	2	12,3	0,009
NROS	Navicula rostellata Kützing	2	12,3	0,001
SPUP	Sellaphora pupula (Kützing) Mereschkowsky	2	11,8	0,005
GMIC	Gomphonema micropus Kützing var. micropus	2	11,7	0,008
SANG	Surirella angusta Kützing	2	11,3	0,004
SBRV	Staurosira brevistriata (Grunow) Grunow	2	10,2	0,003
HCAP	Hippodonta capitata (Ehr.)Lange-Bert.Metzeltin & Witkowski	2	9,4	0,003
PRST	Planothidium rostratum (Oestrup) Lange-Bertalot	2	9,1	0,005
PPRO	Parlibellus protracta (Grunow) Witkowski Lange-Bertalot & Metzeltin	2	8,1	0,004
ADCT	Achnanthydium catenatum (Bily & Marvan) Lange-Bertalot	2	7,8	0,014
CMLF	Craticula molestiformis (Hustedt) Lange-Bertalot	2	7,8	0,011
AUSB	Aulacoseira subborealis (Nygaard) Denys, Muylaert & Krammer	2	7,2	0,005
PCLT	Placoneis clementis (Grun.) Cox	2	6,9	0,007
FVUL	Frustulia vulgaris (Thwaites) De Toni	2	6,8	0,005
NSUA	Nitzschia subacicularis Hustedt in A.Schmidt et al.	2	6,7	0,011
NSHR	Navicula schroeteri Meister var. schroeteri	2	6,6	0,011
PGRN	Planothidium granum (Hohn & Hellerman) Lange-Bertalot	2	6,6	0,004
GITA	Gomphonema italicum Kützing	2	5,4	0,009
SEXG	Stauroforma exiguiformis (Lange-Bertalot) Flower Jones et Round	2	4,8	0,027
CPLI	Cocconeis placentula Ehrenberg var.lineata (Ehr.)Van Heurck	3	73,3	0,001
ADSU	Achnanthydium subatomus (Hustedt) Lange-Bertalot	3	52,1	0,001
PTLA	Planothidium lanceolatum(Brebisson ex Kützing) Lange-Bertalot	3	36,5	0,001
RSIN	Reimeria sinuata (Gregory) Kociolek & Stoermer	3	34,7	0,001
COPL	Cocconeis pseudolineata (Geitler) Lange-Bertalot	3	28,1	0,001
ADDA	Achnanthydium daonense (Lange-Bertalot) Lange-Bertalot Monnier & Ector	3	21,8	0,001
GRHB	Gomphonema rhombicum M. Schmidt	3	21,8	0,001
GACC	Geissleria acceptata (Hust.) Lange-Bertalot & Metzeltin	3	20,9	0,001
GCFU	Gomphoneis caldifuga (Lange-Bertalot & Reichardt)Tuij	3	20,7	0,001
DPER	Diadesmis perpusilla (Grunow) D.G. Mann in Round & al.	3	18,5	0,001
ADSP	Adlafia sp.	3	16,9	0,001
ADLS	Adlafia suchlandtii (Hustedt) Moser Lange-Bertalot & Metzeltin	3	11,7	0,003
DMES	Diatoma mesodon (Ehrenberg) Kützing	3	11,5	0,027
NULA	Nupela lapidosa (Lange-Bertalot) Lange-Bertalot var.lapidosa	3	10,4	0,003
NEXI	Navicula exilis Kützing	3	9,9	0,012
PSRE	Psammothidium rechtensis (Leclercq) Lange-Bertalot	3	9,4	0,002
NHAN	Nitzschia hantzschiana Rabenhorst	3	8,5	0,006
FARC	Fragilaria arcus (Ehrenberg) Cleve var. arcus	3	7,9	0,04
ADMM	Adlafia minuscula var. muralis (Grunow) Lange-Bertalot	3	5,3	0,034
NLUN	Navicula lundii Reichardt	3	4,4	0,036
NTPT	Navicula tripunctata (O.F.Müller) Bory	4	82,7	0,001
APED	Amphora pediculus (Kützing) Grunow	4	79,5	0,001
NCTE	Navicula cryptotenella Lange-Bertalot	4	69,4	0,001
CEUG	Cocconeis euglypta Ehrenberg	4	50,3	0,001
CPED	Cocconeis pediculus Ehrenberg	4	44,3	0,001
GMIN	Gomphonema minutum(Ag.)Agardh f. minutum	4	40,6	0,001
NDIS	Nitzschia dissipata(Kützing)Grunow var.dissipata	4	39	0,001
NRCH	Navicula reichardtiana Lange-Bertalot var. reichardtiana	4	37,7	0,001
FSBH	Fallacia subhamulata (Grunow in V. Heurck) D.G. Mann	4	34,5	0,001
CBAC	Caloneis bacillum (Grunow) Cleve	4	33,2	0,001

Code	Dénomination de l'espèce	Région	IndVal	p value
NCPR	Navicula capitatoradiata Germain	4	33	0,001
GSCI	Gyrosigma sciotense (Sullivan et Wormley) Cleve	4	27,9	0,001
RABB	Rhoicosphenia abbreviata (C.Agardh) Lange-Bertalot	4	27,6	0,001
AINA	Amphora inariensis Krammer	4	27,3	0,001
NFON	Nitzschia fonticola Grunow in Cleve et Möller	4	26,6	0,002
ACOP	Amphora copulata (Kütz) Schoeman & Archibald	4	24,4	0,001
NSOC	Nitzschia sociabilis Hustedt	4	22,4	0,001
GOLI	Gomphonema olivaceum (Hornemann) Brébisson var. olivaceum	4	21,9	0,001
CAEX	Cymbella excisa Kützing var. excisa	4	19	0,002
DVUL	Diatoma vulgare Bory	4	18,2	0,001
RUNI	Reimeria uniseriata Sala Guerrero & Ferrario	4	18,2	0,001
SBRE	Surirella brebissonii Krammer & Lange-Bertalot var.brebissonii	4	17,3	0,001
DTEN	Denticula tenuis Kützing	4	16,7	0,038
NAMP	Nitzschia amphibia Grunow f.amphibia	4	15,9	0,001
ADLB	Achnantheridium lauenburgianum (Hustedt) Monnier Lange-Bertalot & Ector	4	15,3	0,001
GPRI	Gomphonema pumilum var. rigidum Reichardt & Lange-Bertalot	4	15,2	0,001
GYAT	Gyrosigma attenuatum (Kützing) Rabenhorst	4	13,2	0,001
KGES	Kolbesia gessneri (Hustedt) Aboal	4	13	0,001
DPST	Discostella pseudostelligera (Hustedt) Houk et Klee	4	12,2	0,008
FLEN	Fallacia lenzi (Hustedt) Lange-Bertalot	4	11,6	0,001
CEUO	Cocconeis euglyptoides (Geitler) Lange-Bertalot	4	11,3	0,001
NTRV	Navicula trivialis Lange-Bertalot var. trivialis	4	10,7	0,002
ENLB	Encyonema lange-bertalotii Krammer morphotype 1	4	10	0,008
SIDE	Simonsenia delognei Lange-Bertalot	4	9,9	0,006
DOCU	Diploneis oculata (Brebisson) Cleve	4	9,8	0,002
NREC	Nitzschia recta Hantzsch in Rabenhorst	4	9,2	0,013
DOBL	Diploneis oblongella (Naegeli) Cleve-Euler	4	8,8	0,008
CINV	Cyclostephanos invisitatus (Hohn & Hellerman) Theriot Stoermer & Hakansson	4	8	0,005
DSEP	Diploneis separanda Lange-Bertalot	4	7,3	0,008
ACAF	Achnantheridium affine (Grun) Czarnecki	4	6,7	0,045
NCTO	Navicula cryptotenelloides Lange-Bertalot	4	6,6	0,033
CMEN	Cyclotella meneghiniana Kützing	4	6,2	0,017
DMAR	Diploneis marginestriata Hustedt	4	4,7	0,03
ADPY	Achnantheridium pyrenaicum (Hustedt) Kobayasi	5	78,2	0,001
GELG	Gomphonema elegans (Reichardt & Lange-Bertalot) Monnier & Ector	5	28,9	0,001
GPUM	Gomphonema pumilum (Grunow) Reichardt & Lange-Bertalot	5	27,2	0,001
ACLI	Achnantheridium lineare W. Smith	5	22,3	0,001
DEHR	Diatoma ehrenbergii Kützing	5	19,6	0,001
ECPM	Encyonopsis minuta Krammer & Reichardt	5	17,9	0,001
GMPU	Gomphonema micropumilum Reichardt	5	17,8	0,001
CPAR	Cymbella parva (W.Sm.) Kirchner in Cohn	5	16	0,001
ENCM	Encyonopsis microcephala (Grunow) Krammer	5	16	0,005
CAFF	Cymbella affinis Kützing var. affinis	5	14,6	0,001
DTMO	Diatoma tenue Agardh var. moniliformis Kützing	5	14,4	0,002
GLAT	Gomphonema lateripunctatum Reichardt & Lange-Bertalot	5	14	0,001
GTER	Gomphonema tergestinum Fricke	5	13,9	0,002
SSTM	Sellaphora stroemii (Hustedt) Mann	5	13,5	0,001
ADAM	Achnantheridium atomoides Monnier, Lange-Bertalot & Ector	5	12,8	0,013
GOCU	Gomphonema occultum Reichardt & Lange-Bertalot	5	11,1	0,002
ESUM	Encyonopsis subminuta Krammer & Reichardt	5	10,3	0,001
ECKR	Encyonopsis krammeri Reichardt	5	9,5	0,006
DDEL	Delicata delicatula (Kützing) Krammer var. delicatula	5	7,8	0,007
GROS	Gomphonema rosenstockianum Lange-Bertalot & Reichardt	5	7,8	0,037
BMIC	Brachysira microcephala (Grunow) Compère	5	7,7	0,015
CEXF	Cymbella excisiformis Krammer var. excisiformis	5	7,3	0,008
CVUL	Cymbella vulgata Krammer var. vulgata Krammer	5	7,1	0,014
ADKR	Achnantheridium kranzii (Lange-Bertalot) Round & Bukhtiyarova	5	6,6	0,023
GAGV	Gomphonema angustivalva E. Reichardt	5	6,6	0,018

**Annexe 3A** : Valeurs brutes pour la comparaison des niveaux d'état écologique entre la grille de 2009 et celle de 2011, notes arrondies au demi-point le plus proche (\* : p value <0,05 désignant une différence significative entre les classements)

		Très-Bon	Bon	Moyen	Médiocre	Mauvais	p value kh <sup>2</sup>
Adour-Garonne	EE 2009	315	168	167	29	7	0,32
	EE 2011	290	181	194	34	7	
Loire-Bretagne	EE 2009	208	343	400	62	6	0,002*
	EE 2011	166	305	448	87	13	
Rhône-Med-Corse	EE 2009	638	180	235	66	38	0,27
	EE 2011	588	208	246	74	41	
Seine-Normandie	EE 2009	76	219	47	2	0	0,87
	EE 2011	70	222	50	1	1	
Rhin-Meuse	EE 2009	55	54	105	11	0	7,87E-6*
	EE 2011	41	52	82	46	4	
Artois-Picardie	EE 2009	2	103	28	1	0	0,52
	EE 2011	2	103	29	0	0	
HER4 Vosges	EE 2009	56	24	29	0	0	0,00001*
	EE 2011	40	20	31	18	0	
HER21 Massif Central N	EE 2009	152	111	46	3	0	4,9E-14*
	EE 2011	89	82	116	24	1	
HER1 Pyrénées	EE 2009	135	12	3	0	0	0,03*
	EE 2011	119	25	6	0	0	
HER13 Landes	EE 2009	11	9	12	6	2	0,8
	EE 2011	16	7	11	5	1	

**Annexe 3B** : Valeurs brutes pour la comparaison des niveaux d'état écologique 2011, selon que les notes seuils sont arrondies au demi ou au dixième de point le plus proche (toutes les p value sont >0,05, aucune différence significative entre les classements n'a donc été répertoriée).

		Très-Bon	Bon	Moyen	Médiocre	Mauvais	p value kh <sup>2</sup>
Adour-Garonne	demi-point	290	181	194	34	7	0,48
	10ème de point	279	210	183	27	7	
Loire-Bretagne	demi-point	166	305	448	87	13	0,52
	10ème de point	165	341	423	78	12	
Rhône-Med-Corse	demi-point	588	208	246	74	41	0,85
	10ème de point	577	230	238	72	40	
Seine-Normandie	demi-point	70	222	50	1	1	0,81
	10ème de point	66	232	45	0	1	
Rhin-Meuse	demi-point	41	52	82	46	4	0,67
	10ème de point	40	60	87	35	3	
Artois-Picardie	demi-point	2	100	32	0	0	0,83
	10ème de point	3	101	31	0	0	
HER4 Vosges	demi-point	40	20	31	18	0	0,85
	10ème de point	39	21	35	14	0	
HER21 Massif Central	demi-point	89	82	116	24	1	0,54
	10ème de point	80	95	120	16	1	
HER1 Pyrénées	demi-point	119	25	6	0	0	0,99
	10ème de point	118	26	6	0	0	
HER13 Landes	demi-point	16	7	11	5	1	1
	10ème de point	16	7	11	5	1	

## Annexe 4 : Nouvelle grille indicielle proposée

		Valeurs inférieures des limites de Classes d'Etat Ecologique par type (IBD 2007)					
		Rangs (bassin Loir e-Bretagne)	8, 7	6	5	4	3, 2, 1
<b>IBD 2007</b>		Rangs (autres bassins)	8, 7, 6	5	4	3	2, 1
Hydroécorégions de niveau 1		Cas général, cours d'eau exogène de l'HER de niveau 1 indiquée ou HER de niveau 2	Très Grands	Grands	Moyens	Petits	Très Petits
20	DEPOTS ARGILE SABLEUX	Cas général		17-14,5-10,5-6		17-14,5-10,5-6	17-14,5-10,5-6
		HER2 n°31, 33, 39, 119 (Artois-Picardie)		17-14,5-10,5-6		17-14,5-10,5-6	
		Exogène de l'HER 9		17-14,5-10,5-6			
		Exogène de l'HER 21		18-16-13-9,5	18-16-13-9,5		
21	MASSIF CENTRAL NORD	Cas général		18-16-13-9,5		18-16-13-9,5	18-16-13-9,5
		Exogène de l'HER 19		18-16-13-9,5		18-16-13-9,5	18-16-13-9,5
		Exogène de l'HER 8		#			
		Exogène de l'HER 19 ou 8		17-14,5-10,5-6			
3	MASSIF CENTRAL SUD	Cas général					
		Exogène de l'HER 19					
		Exogène de l'HER 8					
		Exogène de l'HER 19 ou 8		17-14,5-10,5-6			
17	DEPRESSIONS SEDIMENTAIRES	Cas général			16,5-14-10-6	16,5-14-10-6	16,5-14-10-6
		Exogène de l'HER 3 ou 21	#	18-16-13-9,5	18-16-13-9,5	18-16-13-9,5	18-16-13-9,5
		Exogène de l'HER 3 ou 21					
		Exogène de l'HER 5		17-14,5-10,5-6			
15	PLAINE SAONE	Cas général		19-16,5-13-9,5		19-16,5-13-9,5	
		Exogène de l'HER 10		17-14,5-10,5-6		17-14,5-10,5-6	17-14,5-10,5-6
5	JURA / PRE-ALPES DU NORD	Cas général		19-16,5-13-9,5	19-16,5-13-9,5	19-16,5-13-9,5	19-16,5-13-9,5
		Exogène de l'HER 2		19-16,5-13-9,5			
TTGA	FLEUVES ALPINS	Cas général	#				
2	ALPES INTERNES	Cas général		19-16,5-13-9,5		19-16,5-13-9,5	19-16,5-13-9,5
7	PRE-ALPES DU SUD	Cas général			19-16,5-13-9,5		19-16,5-13-9,5
		Exogène de l'HER 2		19-16,5-13-9,5			
		Exogène de l'HER 2 ou 7	17-14,5-10,5-6	19-16,5-13-9,5			
		Exogène de l'HER 7		19-16,5-13-9,5			
		Exogène de l'HER 8		18-16-13-9,5			
		Exogène de l'HER 1	17-14,5-10,5-6	19-16,5-13-9,5			
6	MEDITERRANEE	Cas général		17-14,5-10,5-6		17-14,5-10,5-6	17-14,5-10,5-6
		Exogène de l'HER 2		18-16-13-9,5		18-16-13-9,5	18-16-13-9,5
		Exogène de l'HER 2 ou 7		18-16-13-9,5		18-16-13-9,5	18-16-13-9,5
		Exogène de l'HER 7		18-16-13-9,5		18-16-13-9,5	18-16-13-9,5
		Exogène de l'HER 8		18-16-13-9,5		18-16-13-9,5	18-16-13-9,5
		Exogène de l'HER 1		18-16-13-9,5		18-16-13-9,5	18-16-13-9,5
8	CEVENNES	Cas général		18-16-13-9,5		18-16-13-9,5	18-16-13-9,5
		A-her2 n°70		18-16-13-9,5		18-16-13-9,5	18-16-13-9,5
		A-her2 n°22		18-16-13-9,5		18-16-13-9,5	18-16-13-9,5
16	CORSE	B-her2 n°88		18-16-13-9,5		18-16-13-9,5	18-16-13-9,5
19	GRANDS CAUSSES	Cas général				17-14,5-10,5-6	
		Exogène de l'HER 8		18-16-13-9,5			
11	CAUSSES AQUITAINS	Cas général				17-14,5-10,5-6	17-14,5-10,5-6
		Exogène de l'HER 3 et/ou 21	17-14,5-10,5-6	18-16-13-9,5	18-16-13-9,5	18-16-13-9,5	
		Exogène des HER 3, 8, 11 ou 19	17-14,5-10,5-6		17-14,5-10,5-6		
		Exogène de l'HER 3 ou 8		17-14,5-10,5-6		17-14,5-10,5-6	
		Exogène de l'HER 1		17-14,5-10,5-6		17-14,5-10,5-6	17-14,5-10,5-6
13	LANDES	Cas général			17,5-15,5-12,5-9	17,5-15,5-12,5-9	17,5-15,5-12,5-9
1	PYRENEES	Cas général		19-16,5-13-9,5	19-16,5-13-9,5	19-16,5-13-9,5	19-16,5-13-9,5
12	ARMORICAIN	A-Centre-Sud		16,5-14-10-6	16,5-14-10-6	16,5-14-10-6	16,5-14-10-6
		B-Ouest-Nord Est		16,5-14-10-6	16,5-14-10-6	16,5-14-10-6	16,5-14-10-6
TTGL	LA LOIRE	Cas général	17-14,5-10,5-6				
		A-her2 n°57			17-14,5-10,5-6	17-14,5-10,5-6	
9	TABLES CALCAIRES	Cas général	17-14,5-10,5-6	17-14,5-10,5-6	17-14,5-10,5-6	17-14,5-10,5-6	17-14,5-10,5-6
		Exogène de l'HER 10		17-14,5-10,5-6	17-14,5-10,5-6	17-14,5-10,5-6	
		Exogène de l'HER 21		18-16-13-9,5	18-16-13-9,5		
		Exogène de l'HER 21					
10	COTES CALCAIRES EST	Cas général	17-14,5-10,5-6	17-14,5-10,5-6	17-14,5-10,5-6	17-14,5-10,5-6	17-14,5-10,5-6
		Exogène de l'HER 4		18-16-13-9,5	18-16-13-9,5		
4	VOSGES	Cas général		18-16-13-9,5		18-16-13-9,5	18-16-13-9,5
22	ARDENNES	Exogène de l'HER 10	16,5-14-10-6			16,5-14-10-6	16,5-14-10-6
		Cas général			16,5-14-10-6	16,5-14-10-6	16,5-14-10-6
18	ALSACE	Cas général			17-14,5-10,5-6		17-14,5-10,5-6
		Exogène de l'HER 4		18-16-13-9,5	18-16-13-9,5	18-16-13-9,5	

a-b-c-d : a = limite inférieure du très bon état, b = limite inférieure du bon état, c = limite inférieure de l'état moyen, d = limite inférieure de l'état médiocre  
# : absence de référence. En gris : type inexistant

Région diatomées	Note de référence
1	18,5
2	17,5
3	19
4	18
5	20

Annexe B

# Rapport ONEMA sur les Pattern trees





*Utilisation des Pattern Trees pour  
l'étude des relations entre traits  
écologiques et biologiques et  
paramètres environnementaux*

*Rapport final*

*Marius Bottin, Julie Guéguen, Nicolas Antunes, Juliette Rosebery*

*Janvier 2012*

## Contexte de programmation et de réalisation

---

La mise en place de la DCE nécessite une mise en conformité des méthodes de bioindication pour prendre en compte les définitions normatives, répondre à toute la gamme des pressions anthropiques, intégrer des métriques fonctionnelles plus fiables à large échelle, améliorer les capacités de diagnostic et harmoniser les méthodes au niveau européen. Ce travail de fond prépare et accompagne la définition des indicateurs à intégrer dans le SEEE.

## Les auteurs

---

Marius Bottin  
maris.bottin@cemagref.fr  
Cemagref Bordeaux

Julie Guéguen  
Julie.gueguen@cemagref.fr  
Cemagref Bordeaux

Nicolas Antunes  
Anciennement Cemagref Bordeaux

Juliette Rosebery  
Juliette.rosebery@cemagref.fr  
Cemagref Bordeaux

## Les correspondants

---

Onema : Yorick Reyjol, ONEMA-DAST, [yorick.reyjol@onema.fr](mailto:yorick.reyjol@onema.fr)  
Référence du document :

Cemagref : Yves Souchon, Cemagref, [yves.souchon@cemagref.fr](mailto:yves.souchon@cemagref.fr)  
Référence du document :

[Autres renseignements nécessaires à la mise sur le Portail « les documents techniques sur l'eau », à renseigner si possible]

<b>Droits d'usage :</b>	<i>Par ex, accès libre</i>
<b>Couverture géographique :</b>	<i>Par ex, dépt du Loiret, commune de Limoges</i>
<b>Niveau géographique</b> [un seul choix] :	<i>Mondial, national, régional, départemental, communal</i>
<b>Niveau de lecture</b> [plusieurs choix possibles] :	<i>Professionnels, experts</i>
<b>Nature de la ressource :</b>	<i>Document</i>

Utilisation des Pattern Trees pour l'étude des relations entre traits écologiques et  
biologiques et paramètres environnementaux

Rapport final

Marius Bottin, Julie Guéguen, Nicolas Antunes, Juliette Rosebery

**Sommaire**

<b>Résumé .....</b>	<b>4</b>
<b>Abstracts .....</b>	<b>5</b>
<b>Synthèse pour l'action opérationnelle (<a href="#">rapports finaux</a>) .....</b>	<b>6</b>
<b>Corps du document .....</b>	<b>7</b>

**Utilisation des Pattern Trees pour l'étude des relations entre traits écologiques et biologiques et paramètres environnementaux**

Marius Bottin, Julie Guéguen, Nicolas Antunes, Juliette Rosebery

**Résumé**

RESUME

Le travail réalisé dans le cadre de la convention Onema-Cemagref 2010 sur la pertinence des traits écologiques et biologiques comme descripteurs synthétiques des communautés de diatomées, avait permis d'après étude bibliographique de déterminer une liste de traits candidats dans la perspective de la mise en œuvre d'un indice multimétrique. A l'issue d'un rapide et simple test de ces métriques s'était imposée la nécessité d'entamer un travail d'investigation sur les relations entre ces traits et les conditions du milieu, à partir :

- d'une base de données conséquente,
- et avec l'aide de techniques biomathématiques performantes permettant de relier paramètres environnementaux et traits dans une logique de type pression/impact.

Ce sont les résultats de ce travail que nous reportons dans le présent document. Le modèle mathématique utilisé est une technique récente appelée « pattern trees ». Nous attirons l'attention sur le fait qu'un important travail a été réalisé pour mener à bien cette étude (6 mois de CDD Ingénieur d'Etudes en biomathématiques ont été nécessaires) : bibliographie, codage, mises au point et création d'un package « R » spécifique.

Ce document doit être considéré étant principalement d'ordre méthodologique, la méthode proposée et mise en œuvre pouvant être transposée à tout maillon biologique.

MOTS CLES (THEMATIQUE ET GEOGRAPHIQUE)

**Diatomées – pattern trees - traits écologiques – traits biologiques – indice multimétrique**

**Using Pattern Trees for the study of relationships between ecological and biological traits of diatoms and environmental parameters.**

Marius Bottin, Julie Guéguen, Nicolas Antunes, Juliette Rosebery

**Abstracts**

*[obligatoire pour un document final,  
de préférence en une page y compris les mots-clés]*

**ABSTRACT**

---

The work done within the framework of the Convention Onema-Cemagref 2010 on the relevance of ecological and biological traits as synthetic descriptors of diatom communities, allowed after a review of the literature to determine a list of candidates traits for the implementation of a multimetric index. After a quick and simple test of these metrics was imposed the need to begin a work on the relationships between these traits and environmental conditions, from:

- A consistent database,
- And with the help of powerful biomathematical techniques to link environmental parameters and communities following a pressure / impact model.

We report in this paper the results of this work. The mathematical model used is a recent technique called "pattern trees." We draw attention to the fact that important work was done to complete the study (6 months of an engineer in biomathematics were necessary): bibliography, coding, development and creation of a package "R" specific.

This document should be considered as methodological, knowing that the proposed method and its implementation can be transposed to any biological element.

**Key words (thematic and geographical area)**

---

Diatoms, pattern trees, ecological traits, biological traits, multimetric index

## **Utilisation des Pattern Trees pour l'étude des relations entre traits écologiques et biologiques et paramètres environnementaux**

Marius Bottin, Julie Guéguen, Nicolas Antunes, Juliette Rosebery

### **Synthèse pour l'action opérationnelle**

*[4 pages maximum]*

*[Cette synthèse ne concerne que les rapports finaux. Il est très important que le rapport final des résultats des recherches, soit accompagné d'une synthèse présentant les **principaux acquis transférables**.*

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

#### **[La synthèse devra comporter :**

- *Contexte général (quelle situation, quels enjeux), objectifs généraux du projet (à quelles questions veut-on répondre) et quelques éléments de méthodologie*
- *Principaux acquis transférables obtenus et implications pratiques, recommandations, réalisations pratiques, limites de l'utilisation*
- *Pour en savoir plus : quelques références (dont les livrables, articles prévus...); adresse électronique et/ou site internet de(s) auteur(s)]*

*Titre court, auteur(s) [sur les pages sans en-tête]*

**Utilisation des Pattern Trees pour l'étude des relations entre traits écologiques et biologiques et paramètres environnementaux**

Marius Bottin, Julie Guéguen, Nicolas Antunes, Juliette Rosebery

Introduction	8
1-Méthodologie	8
1-1 Données exploitées	8
1-2 Les Pattern Trees	10
1-2-1 Généralités	10
1-2-2 Paramétrage et algorithme des pattern trees	11
1-2-3 Application à nos objectifs	14
2-Résultats	14
2-1 Traits biologiques et écologiques	14
2-1-1 Traits biologiques : Guildes de Passy	14
2-1-2 Traits écologiques de van Dam	15
2-2 Test concernant l'état écologique	17
3 Synthèse et perspectives	19
3-1 Traits	19
3-2 Etat écologique	20
Littérature	21
Annexe 1 : Tableaux des classes de paramètres	22
Annexe 2 : Pattern trees	

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

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

*Titre court, auteur(s) [sur les pages sans en-tête]*

## **Introduction**

Le travail réalisé dans le cadre de la convention Onema-Cemagref 2010 sur la pertinence des traits écologiques et biologiques comme descripteurs synthétiques des communautés de diatomées (Rosebery et al., 2010), avait permis d'après étude bibliographique de déterminer une liste de traits candidats dans la perspective de la mise en œuvre d'un indice multimétrique. A l'issue d'un rapide et simple test de ces métriques s'était imposée la nécessité d'entamer un travail d'investigation sur les relations entre ces traits et les conditions du milieu, à partir :

- d'une base de données conséquente,
- et avec l'aide de techniques biomathématiques performantes permettant de relier paramètres environnementaux et traits dans une logique de type pression/impact.

Ce sont les résultats de ce travail que nous reportons dans le présent document. Le modèle mathématique utilisé est une technique récente appelée « pattern trees ». Nous attirons l'attention sur le fait qu'un important travail d'ordre bio-mathématique et informatique a été réalisé pour mener à bien cette étude : bibliographie, codage, mises au point et création d'un package « R » spécifique.

Ce document doit être considéré comme étant principalement d'ordre méthodologique, la méthode proposée et mise en œuvre pouvant être transposée à tout maillon biologique.

Notre objectif était de valider la pertinence des traits présélectionnés en 2010, soit les traits de van Dam et des guildes de Passy, en caractérisant qualitativement et quantitativement les paramètres environnementaux induisant ces traits. La démarche ne consiste pourtant pas à définir à proprement parler des seuils à partir desquels les traits sont retrouvés dans le milieu, mais plutôt de tenter de décrire quels paramètres déterminent leur présence et leur intensité. Dans la dernière partie du document, nous tentons d'appliquer la démarche à la prédiction de l'état écologique par région diatomées. Encore une fois, les résultats obtenus ne peuvent être considérés comme des seuils mais plutôt comme des ambiances générales du milieu induisant un déclassement ou non de l'état écologique par l'IBD2007.

## **1- METHODOLOGIE**

### **1-1 Données exploitées**

#### *La base de données*

La base de données comprend 3458 échantillons prélevés sur 1689 stations de 2005 à 2009 (prospections du Réseau de Référence et du Réseau de Contrôle et Surveillance) (Figure 1). Au total 1272 taxons différents ont été dénombrés, et uniquement ceux présents dans plus d'un échantillon soit 1164 taxons ont été conservés pour les analyses. Chaque taxon, lorsque l'information existe, est codé selon le panel de traits écologiques et biologiques retenu (voir paragraphe suivant).

Afin d'éviter toute source de variabilité pouvant interférer dans les relations communautés-environnement, tous les échantillons retenus ont été prélevés en période estivale et sur galets.

Chaque échantillon de la base est caractérisé par sa classe d'état écologique. Ces classes d'état ont été obtenues à partir des notes seuils d'état écologique proposées dans le rapport « Typologie des flores diatomiques de référence et seuils d'état écologique à partir des notes d'IBD2007 : Révision sur la base des données issues du Réseau de Référence et du Réseau de Contrôle et de Surveillance » (livrable Onema-Cemagref 2011).



**Figure 1 :** Carte des sites étudiés

#### *Les traits écologiques testés*

Différentes classifications des espèces de diatomées en fonction de leur sensibilité à certains paramètres environnementaux ont été réalisées. La plus utilisée de ces classifications reste celle de van Dam et al. (1994), on y trouve par exemple le trait écologique « sensibilité à la matière organique », présentant les modalités suivantes : oligosaprobe, bêtamésosaprobe, alphamésosaprobe, alphamésosaprobe à polysaprobe, polysaprobe. Nous appellerons ces traits les traits van Dam, et nous en testerons diverses modalités (saprobie, trophie, N-hétérotrophie).

#### *Les traits biologiques testés*

Passy (2007) écrit que les tendances adaptatives les plus fortes au sein de communautés diatomiques semblent être : (1) la transition vers des formes plus longues et érigées lorsque la pression hydrologique ou de broutage diminue, ou encore lorsque le milieu s'appauvrit en nutriments et en lumière ; (2) la transition vers des formes mobiles face à une augmentation de la pression naturelle ou anthropique. L'auteur propose ainsi 3 « guildes », que nous testerons ici :

- 1 : espèces adhérentes, attachées par l'apex de façon parallèle au substrat, érigées, centriques solitaires et espèces à lente mobilité (*Achnanthes*, *Achnanthidium*, *Amphora*, *Cocconeis*, *Cyclotella*, *Cymbella*, *Hannaea*, *Meridion*, *Opephora*, *Reimeria*).
- 2 : grandes espèces incluant des formes érigées, filamenteuses, ou formant de longues chaînes, coloniales (*Diatoma*, *Ellerbeckia*, *Fragilaria*, *Gomphonema*, *Gomphoneis*, *Melosira*, *Synedra*).
- 3 : espèces à grande mobilité (*Navicula*, *Nitzschia*, *Sellaphora*, *Surirella*)

## 1-2 Les pattern trees

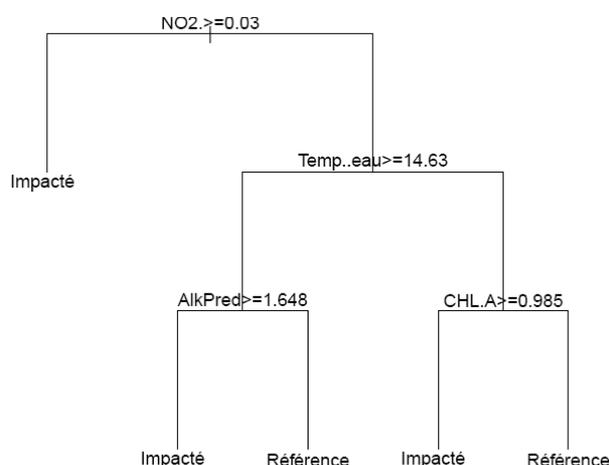
### 1-2-1 Généralités

La méthode des pattern trees est relativement récente, et, pour la présenter grossièrement, correspond à des arbres de décisions utilisant des règles de logique floue. Cette technique a déjà été utilisée pour déterminer la valeur indicatrice de diatomées du lac Prespa (Naumoski and Mitreski, 2011b).

#### *Les arbres de décisions*

Les arbres de décisions constituent une famille de méthodes, parmi lesquelles la technique CART (Classification and Regression Trees ; Breiman et al., 1984) souvent utilisée en écologie (par exemple Garzón et al., 2006 ; Weilhoefer and Pan, 2008 ; Naumoski and Mitreski, 2009 ; Gillett et al., 2011).

Le principe des arbres de décisions est de créer et de hiérarchiser un ensemble de règles pour classer un objet (prendre une décision). Par exemple, on peut imaginer de vouloir classer des sites en « référence » ou « impacté » en fonction des variables physico-chimiques mesurées sur les sites (Figure 2). Ces arbres de décisions utilisent des valeurs seuils pour les différents paramètres afin de partitionner les données de façon à obtenir des groupes les plus ressemblants possibles à ceux que l'on cherche à prédire (dans la figure 2 : référence ou impacté). Du fait de l'utilisation de ces valeurs seuils précises, les arbres de classifications sont très sensibles aux données et peu robustes.

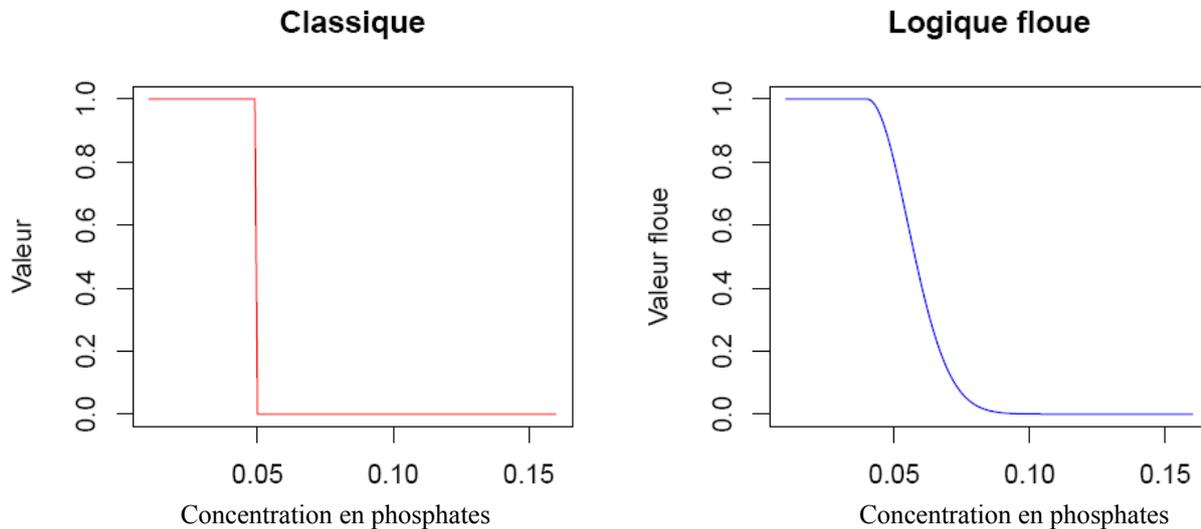


**Figure 2** : Exemple d'arbre de classification (CART) : prédiction des sites de références et impactés en fonction des variables physico-chimiques mesurées. Par exemple, une des règles de classification que l'on peut lire sur cet arbre est : si les nitrites ont une concentration supérieure à 0,03, que la température de l'eau est supérieure à 14,53°C et que l'alcalinité est supérieure à 1,648, alors le site est impacté.

#### *La logique floue*

La logique floue est une technique de plus en plus utilisée en intelligence artificielle, appliquée à des domaines très divers tels que la médecine, la robotique ou encore la météorologie (par exemple Mago et al., 2011 ; Rahoma et al., 2011). Le but de cette technique est de pouvoir intégrer dans des modèles des concepts imprécis, tels que ceux que nous utilisons lorsque l'on parle. En effet, il est difficile de traduire de manière mathématique ce que sont de «faibles valeurs de phosphates» ou des «des rivières plutôt acides».

exemples précédents, la logique floue peut permettre d'associer, pour chaque valeur mesurée de concentration en phosphates ou de pH, des valeurs permettant de déterminer à quel point les rivières sont «plutôt acides » et le phosphate « faible » (Figure 3). Du point de vue mathématique, la logique floue est souvent considérée comme une extension de la théorie des ensembles classiques dans l'«univers du discours».



**Figure 3** : Illustration de la logique floue : «Faibles valeurs de phosphates». Quand on n'utilise pas la logique floue, la traduction de l'expression «faibles valeurs de phosphates» pour les modèles utilise un seuil : en dessous de cette valeur seuil (ici 0,05) l'expression est vraie, au-dessus de cette valeur l'expression est fausse. Lorsque l'on utilise la logique floue, la progression des valeurs de l'expression «faibles valeurs de phosphates» peut être plus progressive. Ces valeurs prennent donc en compte le fait que certaines d'entre elles correspondent à un niveau «plutôt faible» ou «pas vraiment faible». La logique floue est donc mieux adaptée aux concepts imprécis que l'on utilise dans le langage courant.

#### *L'induction de pattern trees*

La technique des patterns trees va consister à construire des arbres tels que ceux construits par la méthode CART, mais en utilisant des règles établies en logique floue. L'induction des pattern trees est l'étape de l'algorithme qui va consister à choisir les règles floues à partir des données et à les hiérarchiser. Elle est précédée d'une étape dans laquelle les valeurs des différentes variables des modèles sont traduites en valeurs floues : la fuzzification (de l'anglais fuzzy : flou).

#### **1-2-2 Paramétrage et algorithme des pattern trees**

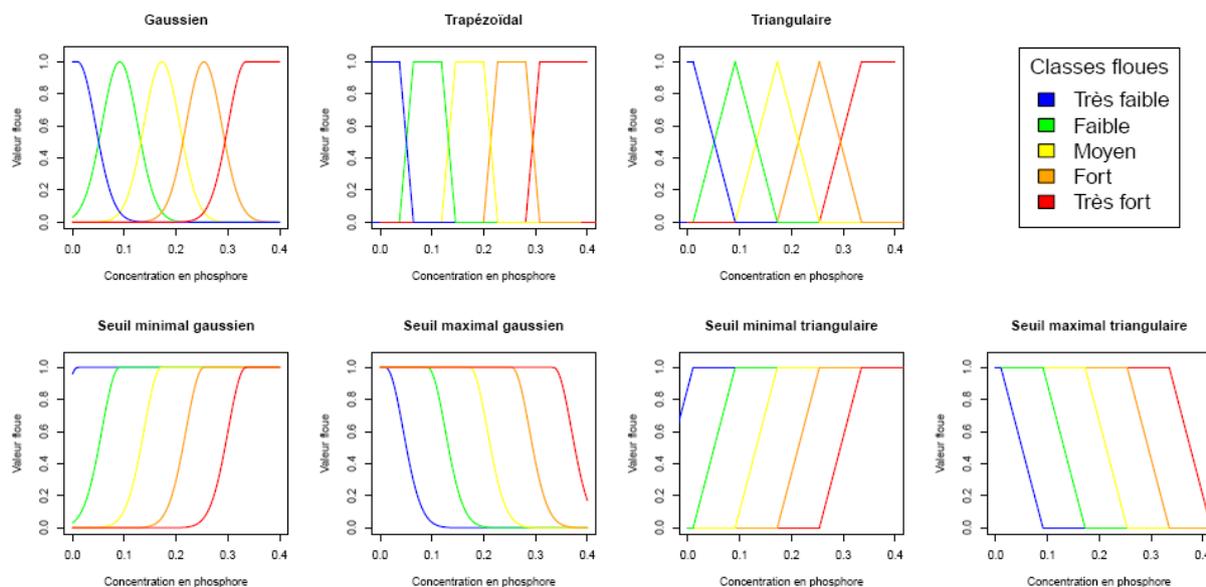
Les pattern trees étant une technique relativement récente, celle-ci n'avait jamais été implémentée dans le logiciel de statistique R. Afin de pouvoir utiliser efficacement cette méthode, nous l'avons donc programmée dans les langages C et R, sous forme d'un paquet R appelé patternTree, inspiré principalement de Huang et al. (2008) et du programme Fuzzy-EBY, développé par Zhiheng Huang (University of California, Berkeley). Ce travail de programmation a été réalisé Nicolas Antunes, recruté au Cemagref dans ce but, en collaboration avec Andreja Naumoski, de l'université de Skopje. Ce paragraphe a pour but de présenter plus précisément l'algorithme, les choix qui ont été faits lors de l'implémentation et les possibilités offertes par le paquet patternTree.

### La fuzzification

La fuzzification est l'étape des pattern trees qui permet de donner des valeurs floues aux données entrées dans le modèle, à partir de variables classiques. Il s'agit en fait de partager les valeurs de la variable en plusieurs classes, et de donner des valeurs floues à toutes les unités d'échantillonnage pour chacune de ces classes.

Pour cela, différentes fonctions ayant différentes allures ont été créées (Huang et al., 2008 ; Naumoski and Mitreski, 2011b, 2011a). Les exemples de fonctions disponibles dans le paquet patternTree sont présentés dans la Figure 4. Il est potentiellement possible d'utiliser un très grand nombre de fonctions différentes, l'important étant que la fonction mathématique représente bien la classe que l'on veut décrire. Aussi l'utilisation pourra varier selon les variables étudiées et les buts des analyses.

La première étape de la fuzzification des variables quantitatives consiste à trouver des intervalles de valeurs pour créer les différentes classes. Souvent, chaque intervalle est décrit par une valeur centrale, appelée  $\mu$  et une valeur de dispersion appelée  $\sigma$ . Par défaut, l'algorithme va diviser la variable quantitative en 5 classes (correspondant dans le discours à «très faible», «faible», «moyen», «fort», et «très fort»).

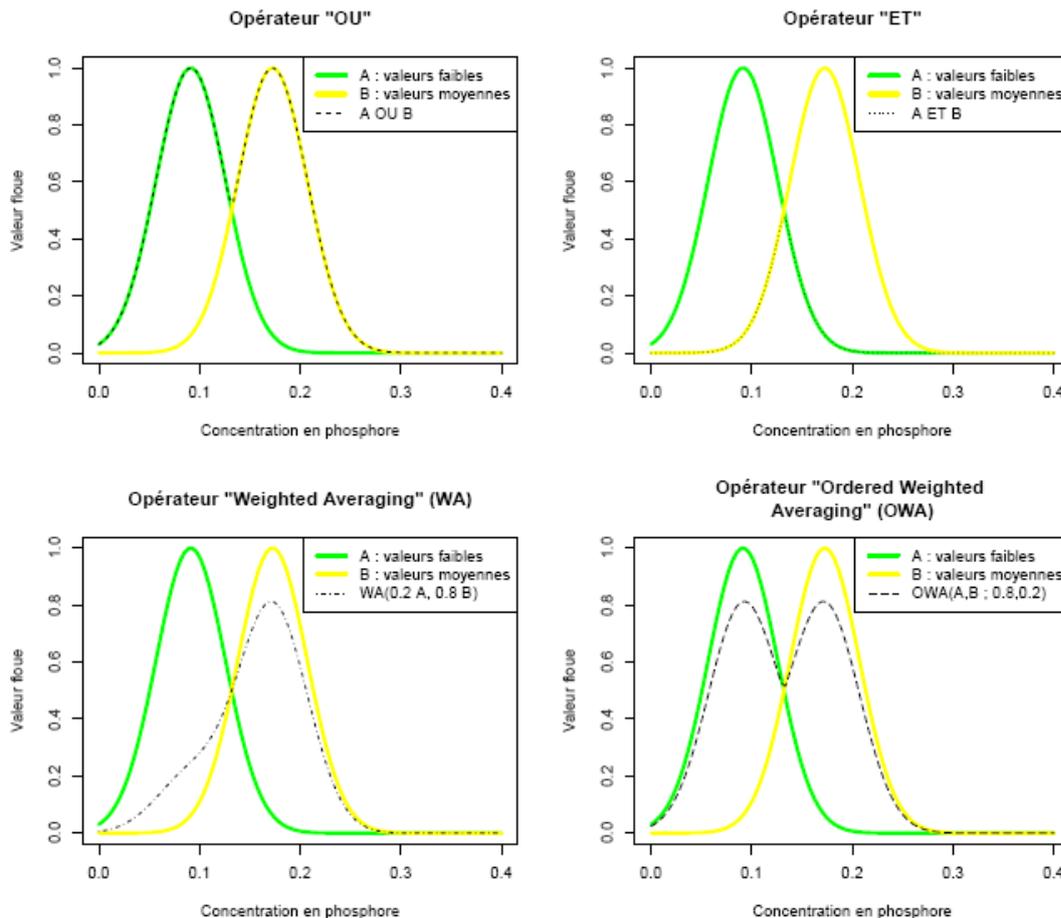


**Figure 4** : Fonctions d'appartenance pour la fuzzification : exemple du phosphore. Ces différents graphiques montrent les différents types de fuzzification disponibles dans le paquet patternTree. Chacune des fonctions est ici utilisée pour définir les valeurs de cinq différentes classes floues (allant de « très faibles valeurs » à « très fortes valeurs » de concentration en phosphore), en fonction des valeurs initiales de phosphore. Les quatre dernières figures présentent des fonctions permettant d'utiliser des seuils, alors que les trois premières présentent des fonctions centrées autour d'une valeur  $\mu$ .

### Les opérateurs logiques

En logique classique, il existe trois opérateurs logiques principaux : il s'agit de ET, OU et NON. Il est important de pouvoir traduire ces opérateurs en logique floue, afin créer des règles de classement (par exemple : si le niveau de phosphates est faible ET le niveau de nitrates est faible, alors le site est une référence, ou encore si le niveau de phosphates est fort OU le niveau en nitrates est fort, alors le site est impacté). La traduction mathématique des opérateurs ET et OU en logique floue peut être surprenante au premier abord, car il s'agit tout simplement respectivement des fonctions «maximum» et «minimum» (Figure 5).

Dans le cas de la logique floue et des pattern trees des opérateurs logiques spécifiques ont été aussi développés. Les opérateurs WA (Weighted Averaging) et OWA (Ordered Weighted Averaging) représentent des situations intermédiaires entre le ET et le OU dans le sens où ces opérateurs accordent des importances différentes aux propositions logiques qu'ils relient. Ces opérateurs logiques ne sont pas encore utilisés dans le paquet patternTree que nous avons développé, mais ils ont été implémentés et seront utilisés dans les versions ultérieures du paquet.



**Figure 5** : Les opérateurs en logique floue. Quatre opérateurs sont présentés ici. Ils sont utilisés pour obtenir des valeurs floues pour des expressions liant deux classes floues (les classes "valeurs faibles" et "moyennes" de concentration en phosphore total). On voit que l'opérateur OU correspond à la fonction maximum, ET correspond à la fonction minimum, et les opérateurs WA et OWA à des situations intermédiaires. L'opérateur WA consiste à donner un poids différent à chacune des classes. L'opérateur OWA consiste à donner un poids différent à la classe à plus fortes valeurs et à la classe à plus faibles valeurs.

#### *Le paramètre d'optimisation*

Optimiser les pattern trees revient à trouver les arbres dont les prédictions seront les plus proches de la variable à prédire. Nous devons donc utiliser une mesure de similarité entre les valeurs prédites et celles à prédire. Par défaut, le paquet patternTree utilisera la statistique de dissimilarité RMSE, mais d'autres mesures ont été implémentées.

$$RMSE(A, B) = \sqrt{\frac{\sum_{j=1}^m (\mu_A(x_j) - \mu_B(x_j))^2}{m}}$$

où :

- A est la variable prédite et B la variable à prédire.
- Les variables prédites et à prédire sont des suites de valeurs floues de longueur m
- $\mu_A(x_j)$  représente la valeur floue n° j de la variable floue A

### **1-2-3 Application à nos objectifs**

Pour chaque trait testé seront présentés 3 pattern trees, car chaque trait sera subdivisé en 3 classes, traduisant :

- classe 1 : trait peu présent
- classe 2 : trait moyennement présent
- classe 3 : trait fortement présent dans la communauté considérée

Une classe représente un intervalle décrit par une valeur centrale appelée  $\mu$ , et une valeur de dispersion. Ces 3 classes pour chaque trait seront prédites à partir des paramètres environnementaux disponibles.

Concernant le test sur les classes d'état, les 5 classes habituelles seront considérées : très bon, bon, moyen, médiocre, mauvais.

Pour chaque trait testé, ou chaque classe d'état testée, les paramètres environnementaux seront également divisés en classes, mais cette fois en 5 classes, disposant également d'une valeur centrale et d'une valeur de dispersion. Ainsi par exemple la conductivité avant analyse sera divisée en conductivité très faible, faible, moyenne, élevée, très élevée. Ces tableaux de classes sont présentés en Annexe 1.

## **2- RESULTATS**

Les pattern trees obtenus sont présentés en Annexe 2.

### **2-1 Traits biologiques et écologiques**

#### **2-1-1 Traits biologiques : Guildes de Passy**

##### *Guilde 1 – espèces adhérentes -*

Ce groupe d'espèce se retrouve en abondance (classe 3) dans des milieux à forte hydrodynamique, de type montagne, avec une altitude élevée et une température de l'eau basse (8,4°C en moyenne). Ces résultats sont en accord avec la définition théorique de Passy : représentant le premier stade de colonisation du substrat, cette guilda caractérise une communauté en voie de construction d'un biofilm mais qui par la forte hydrodynamique ne parviendra pas au stade mature. On retrouve également de fortes proportions d'espèces appartenant à cette guilda dans les cas de pollution toxique aigue (non démontrable à partir de notre base de données).

On remarque ainsi que les paramètres prédisant les différentes classes d'abondance de cette guilda ne sont pas en relation avec la pression anthropique, mais plutôt liés à l'altitude.

Nous recommandons que cette guilda soit utilisée comme alerte dans l'interprétation des indices : une bonne note d'indice concernant une communauté composée à plus de 50% de telles espèces pionnières peut cacher une forte pollution toxique, ou un prélèvement non représentatif car réalisé après un évènement hydrologique particulier.

### *Guilde 2 – grandes espèces -*

Toujours en accord avec la définition de Passy, ces espèces de formes longues et érigées sont abondantes dans des milieux pauvres en nutriments (nitrates dans l'arbre). D'après leurs arbres respectifs, on remarque qu'elles sont peu présentes lorsque les espèces mobiles (guilde 3) sont abondantes (graphes très similaires pour les classes d'abondance opposées).

### *Guilde 3 – espèces mobiles -*

Ces espèces à grande mobilité sont présentes en abondance dans les parties aval de cours d'eau, riches en phosphore total (classe 5 : moyenne de 0,215 mg/l Pt). Ce résultat conforte la corrélation significative obtenue en 2010 entre cette guilde et la métrique d'anthropisation ( $\text{NH}_4+\text{PO}_4$ ), sur la base des données historiques du Cemagref.

## **2-1-2 Traits écologiques de van Dam**

### **N-hétérotrophie (capacité à croître hétérotrophiquement)**

#### *N-autotrophes sensibles*

Selon nos résultats ces formes montrent une abondance maximale lorsque les nutriments sont en faible concentration (en moyenne 0,1 mg/l de  $\text{NO}_3$ , 0,01 mg/l de  $\text{PO}_4$ ), et l'alcalinité faible. Une augmentation de la conductivité tend à les faire disparaître. Alcalinité et conductivité ayant une origine anthropique mais également géochimique, ce trait semble être affecté par la typologie naturelle des cours d'eau, et doit donc être interprété avec précaution.

#### *N-autotrophes tolérantes*

On retrouve ces formes en tête de cours d'eau, en contexte montagneux (oxygène dissous élevé et température basse). Ces espèces participent à l'auto-épuration notamment du carbone organique dissous. *Achnanthydium minutissimum* est une espèce N-hétérotrophe tolérante typique : bien que généralement considérée polluo-sensible par les indices, on peut la retrouver en tête de cours d'eau en présence d'une charge modérée de carbone organique, tant que la concentration en oxygène dissous est élevée.

#### *N-hétérotrophes facultatives et obligatoires*

Ces deux formes semblent partager globalement le même profil : peu représentées lorsque les niveaux de  $\text{PO}_4$  et  $\text{NH}_4$  sont bas (tête de cours d'eau avec température basse), et abondantes à l'aval (oxygène dissous bas et température haute) plus riche en nitrates et matières organiques (DBO5 de 3,65 mg/l en moyenne).

Ces deux formes gagneraient certainement à être regroupées lors d'interprétation de l'état écologique des masses d'eau, les séparer ne semble pas apporter d'information supplémentaire pertinente.

## **Saprobie**

### *Oligosaprobies*

De façon très logique, on retrouve ces formes en abondance en tête de cours d'eau, mais peu dans les secteurs aval riches en phosphore total.

On remarque cependant, de façon similaire au cas des N-hétérotrophes sensibles, que les principaux paramètres prédictifs des formes oligosaprobies sont liés à la géochimie (pH, alcalinité, conductivité), et donc à la typologie des cours d'eau. Ainsi les oligosaprobies affectionneraient particulièrement selon nos résultats les eaux naturellement peu minéralisées.

### *Beta-mesosaprobies et alpha-mesosaprobies*

Les arbres de ces deux formes sensées être assez proches du point de vue écologique auraient dus être comparables, mais il est intéressant de noter qu'ils sont bien différents. Les beta-mesosaprobies sont prédites en forte abondance lorsque les niveaux de pollution organique sont bas (0,2 mg/l de carbone organique en moyenne), et c'est plutôt l'inverse concernant les alpha-mesosaprobies (DBO5 de 3,65 mg/l en moyenne prédit la classe 3, et 0,2 mg/l de carbone organique prédit la classe 1). On continue à observer parmi les paramètres de prédiction des facteurs liés à la géochimie (faibles abondances de beta-mesosaprobies). D'après nos résultats les formes oligo- et beta-mesosaprobies pourraient gagner à être regroupées.

### *Alpha à polysaprobies et polysaprobies*

Ces arbres sont assez comparables entre eux et à celui des alpha-mesosaprobies. Ils diffèrent en revanche de ce dernier par des niveaux de conductivité haute (1070  $\mu\text{s}/\text{cm}$  en moyenne) parmi les paramètres semblant prédire de fortes abondances de formes alpha à polysaprobies et polysaprobies. Ces conductivités hautes sont par ailleurs accompagnées de matières organiques et d'orthophosphates (classes respectives maximales).

On ne retrouve pas d'effet typologique potentiel dans les arbres de ces formes tolérantes, à l'inverse de ceux des formes sensibles.

## **Trophie**

### *Oligotrophes et oligo-mesotrophes*

Les profils sont comparables, ces deux formes se retrouvent en abondance en tête de cours d'eau naturellement peu minéralisés. On retrouve l'effet typologique repéré systématiquement pour les formes sensibles.

### *Mesotrophes*

L'alcalinité en tant que telle n'entre plus dans la prédiction des abondances extrêmes, mais on retrouve malgré tout le pH. Un niveau faible de nitrates et de carbone organique entraîne une forte abondance des taxons mésotrophes, qui semblent inféodés aux parties amont des cours d'eau.

### *Meso-eutrophes*

Ces formes sont faiblement présentes dans les zones marquées par un contexte altitudinal élevé et une alcalinité faible. Elles sont abondantes dans les zones assez minéralisées et au pH alcalin, mais malgré tout de bonne qualité (carbone organique faible et trop fortes conductivités –classe 5- exclues). On peut penser qu'elles abondent dans des petits cours d'eau lents et peu anthropisés de plaine.

### *Eutrophes*

On retrouve les taxons eutrophes franchement à l'aval des cours d'eau (altitude faible, température haute). Aucun paramètre d'anthropisation n'apparaît parmi les paramètres prédictifs de l'arbre, aussi il semblerait que ces formes ne soient pas systématiquement liées à d'importants niveaux de pollution.

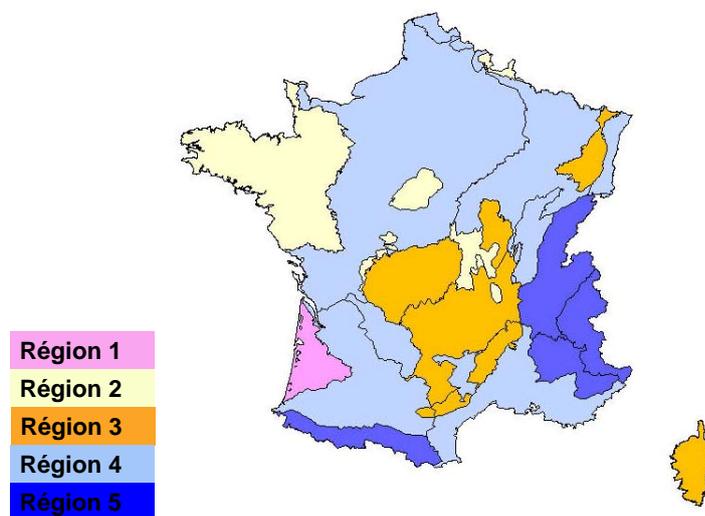
### *Hypereutrophes*

Cette fois les paramètres d'anthropisation sont bien corrélés avec la présence de ces formes hypereutrophes. On les retrouve à l'aval de cours d'eau, dans des conditions chargées en matière organique notamment (classe maximale de DBO5).

Si les taxons eutrophes ne semblent pas diagnostiquer sans équivoque un état écologique très médiocre (difficulté à différencier une eutrophisation d'origine naturelle ou anthropique), les taxons hypereutrophes en revanche portent un message clair.

## 2-2 Test concernant l'état écologique

Nous proposons dans cette partie d'observer quels paramètres prédisent le mieux les différentes classes d'état écologique par région diatomées (Figures 6 et 7).



**Figure 6** : Carte révisée des régions naturelles diatomées (fond HER1).

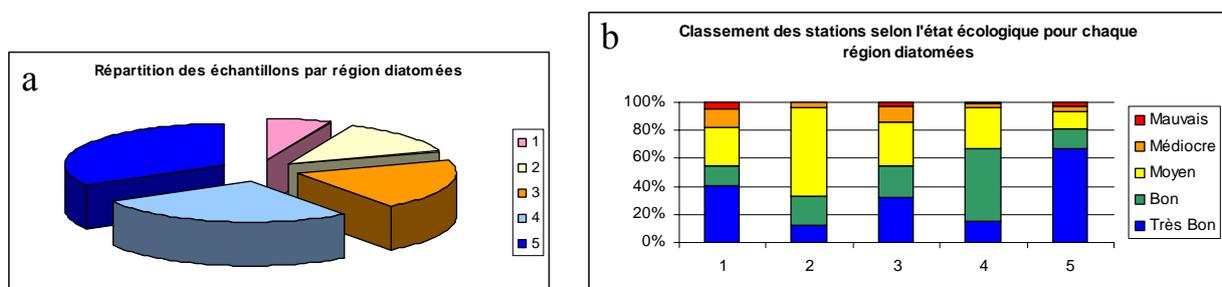
Région 1 : contexte acide des Landes (note minimale : 5/20)

Région 2 : plaines et faibles altitudes peu carbonatées (note minimale : 1/20)

Région 3 : régions moyennement élevées à élevées, peu carbonatées : Massif Central Nord et Sud, Vosges, Cévennes (note minimale : 5/20)

Région 4 : plaines et faibles altitudes carbonatées (note minimale : 1/20)

Région 5 : montagnes, plutôt carbonatées à franchement carbonatées (note minimale : 5/20)



**Figure 7** : a) Répartition du nombre d'échantillons de la base selon la région diatomées, b) Classement des stations selon leur état écologique, pour chaque région diatomée.

Chaque paramètre environnemental est décrit selon la logique floue en 5 niveaux d'intensité, et ces niveaux afin de tenir compte des spécificités typologiques sont redéfinis pour chaque région (annexe 1).

Les classes prédites de limite de bon état concernant les paramètres physico-chimiques sont comparées à titre indicatif aux seuils officiels de l'arrêté du 25/01/10 relatif aux méthodes et critères d'évaluation de l'état écologique, de l'état chimique et du potentiel écologique des eaux de surface.

### *Région 1*

A partir d'un niveau d'orthophosphates de classe 3, soit 0,09 mg/l en moyenne dans cette région, un site n'est plus en bon état écologique selon l'IBD2007. Cette valeur est bien inférieure à celle de l'arrêté d'évaluation, fixant à 0,5 mg/l la limite du bon état écologique pour ce paramètre. L'état moyen est caractérisé par les forts taux d'orthophosphates et de phosphore total, et l'ajout de  $\text{NH}_4$  et/ou de carbone organique dégrade encore l'état écologique.

### *Région 2*

Comme précédemment, à partir d'un niveau d'orthophosphates de classe 3 un site n'est plus en bon état écologique selon l'IBD2007, mais cette fois ce niveau correspond à 0,31 mg/l en moyenne dans cette région, valeur toujours en-deçà de celle de l'arrêté.

Même si on n'observe pas de relation claire entre les niveaux d'état écologique et la classe de pH, il faut noter que ce paramètre est important pour la prédiction de l'état écologique. Le pH étant lié à la géochimie, on peut se poser la question d'un effet typologique même au sein de cette région supposée assez homogène.

Enfin il est difficile d'expliquer de façon logique le rôle des nitrates dans la prédiction du bon état écologique : les nitrates en classe 4 (en moyenne 37,8 mg/l) participent à prédire un bon état écologique, et en classe 1 (en moyenne 0,6 mg/l) participent à prédire un état écologique médiocre.

### *Région 3*

Dans cette région montagneuse le bon et le très bon état écologique concernent particulièrement les plus hautes altitudes, associées à un pH neutre et une conductivité inférieure ou égale à environ 100 $\mu\text{s}/\text{cm}$ .

Ensuite classiquement une élévation de la température associée à une augmentation de la concentration en  $\text{NH}_4$  (0,08 mg/l en moyenne) fait basculer l'opération de contrôle hors du bon état. Le seuil de  $\text{NH}_4$  marquant la limite du bon état écologique selon l'arrêté d'évaluation se situe à 0,5 mg/l.

Il est intéressant de noter que dans cette région, les nitrites semblent jouer un rôle significatif concernant la dégradation de l'état écologique, l'état médiocre puis mauvais étant associé aux classes 3 à 5 de ce paramètre (respectivement 0,05 à 0,09 mg/ en moyenne).

### *Région 4*

Dans cette région de plaines carbonatées, un site n'est plus en bon état écologique à partir d'un niveau d'orthophosphates de classe 4 (0,35 mg/l en moyenne), associé à un niveau de carbone organique de classe 4 également (4,82 mg/l en moyenne). La valeur d'orthophosphates est plus proche cette fois du seuil de l'arrêté. En revanche la valeur de carbone organique est plus restrictive que celle de l'arrêté (7mg/l). Les nitrates semblent peu informatifs de l'état écologique de cette région selon l'IBD2007.

### Région 5

Il est assez étonnant de remarquer que dans ce contexte alpin un site est toujours classé en bon état écologique malgré des taux de carbone organique dissous et de PO<sub>4</sub> déjà élevés pour la région (respectivement 3,25 mg/l et 0,1 mg/l en moyenne), bien que toujours en-deçà des seuils officiels. Le caractère torrentiel de ces cours d'eau (forte oxygénation, température basse, renouvellement rapide des eaux) doit favoriser une bonne auto-épuration et les taxons sensibles se maintiennent malgré l'apparition de taxons notamment N-autotrophes tolérants (*Achnanthydium minutissimum*) ou saprobes (*Nitzschia palea*), ceux-ci restant malgré tout minoritaires.

Ainsi le basculement en état médiocre ou mauvais est davantage dû à un abaissement du taux d'oxygène dissous ou à une élévation de la température, inversant le ratio formes sensibles/formes tolérantes. Ces conditions semblent associées à des teneurs en NH<sub>4</sub> majorées.

## 3 SYNTHÈSE ET PERSPECTIVES

### 3-1 Traits

Les proportions d'espèces appartenant aux guildes de Passy 1 et 3 paraissent clairement informatives, respectivement quant à la survenue d'un événement particulier (pollution toxique aigue, crue), ou quant au niveau trophique du site étudié. Le pourcentage d'espèces mobiles est d'ailleurs une métrique adoptée aux Etats unis par l'Indice Périphytique d'Intégrité Biologique (Hill et al., 2000).

La synthèse des résultats relatifs aux traits est présentée sous la forme du tableau ci-dessous. Les métriques jugées significatives des contextes rencontrés y sont reportées.

Zones amont de référence, contexte montagneux	Zones amont, contexte montagneux, charge organique modérée	Cours lents de plaine, peu anthropisés	Zones aval très anthropisées	Communautés en phase de recolonisation
-Gilde 1 -(Oligo- à mésotrophes) -(N-autotrophes sensibles) -(Oligo- à betamesosaprobes)	N-autotrophes tolérantes (auto-épuration)	Méso-eutrophes	-Gilde 3 -N-hétérotrophes facultative + obligatoires -Alphameso à polysaprobes + polysaprobes -Hypereutrophes	Gilde 1 (message d'alerte : stress toxique ou hydrologique)

**Tableau 1** : Les différents traits et les conditions environnementales qu'ils caractérisent.

Entre parenthèses : formes polluo-sensibles semblant affectionner particulièrement les zones peu minéralisées. Le message écologique porté doit donc être considéré avec précaution dans les zones naturellement alcalines.

Ces résultats permettent de bien valider et sélectionner les informations écologiques portées par les diatomées, et de préfigurer un futur indice multimétrique auquel les traits retenus sont tous potentiellement candidats. Le pourcentage de taxons appartenant à la guildes 1 devrait être systématiquement calculé avant d'interpréter les résultats d'indices classiques.

Nos résultats permettent également d'orienter le travail proposé dans le cadre de la convention Onema-Irstea 2012. Ce travail visera à exploiter la méthode des pattern trees afin de dresser un profil de sensibilité des espèces clés listées par l'IBD2007, ce vis-à-vis des paramètres majeurs et à partir de la base de données utilisée ici.

Dans la présente étude, 18% des espèces ne disposaient pas de code van Dam (dont 7% des espèces clés de l'IBD2007), nous espérons en 2012 obtenir de telles informations pour 100% des espèces.

A la lumière des résultats actuels, et si cela s'avère possible, il sera primordial de veiller à bien décorrélérer eutrophisation naturelle et pression anthropique, afin de rendre non ambiguë la disparition de taxons sensibles.

### **3-2 Etat écologique**

Bien que le présent travail ne vise en aucun cas et ne peut délivrer des seuils d'éléments chimiques déclassants selon l'IBD2007, il est intéressant d'observer que les conclusions sont largement cohérentes avec le travail mené par l'équipe de Lyon sur l'étude des relations entre chimie et biologie dans les cours d'eau (Bougon et Ferréol, 2012).

Concernant le  $PO_4$  :

-La région diatomées 4 correspond globalement aux régions H2 et H3 du document cité ci-dessus (eaux dures, altitudes intermédiaires et basses). Si on moyenne pour ces 2 dernières régions la valeur du seuil de  $PO_4$  au-delà duquel la station n'est plus en bon état (BE), on obtient 0,35 mg/l, très proche de la classe de  $PO_4$  obtenue pour notre région diatomées 4 (0,30 mg/l en moyenne).

-Notre région 1 appartient à la région S3 (eaux douces, altitudes basses). La classe limite de BE pour la région 1 se situe à 0,09 mg/l en moyenne de  $PO_4$ , très proche du seuil de 0,10 mg/l de la région S3.

-Notre région diatomée 5 correspond à la région H1 (eaux dures, altitudes hautes), et la classe limite du BE pour le  $PO_4$  est de 0,1 mg/l en moyenne, identique au seuil de la région H1.

Concernant le carbone organique dissous :

La région diatomées 4 correspond aux régions H2 et H3 (eaux dures, altitudes intermédiaires et basses). Si on moyenne pour ces 2 dernières régions la valeur du seuil de  $PO_4$  au-delà duquel la station n'est plus en bon état, on obtient 4,5 mg/l, très proche de la classe limite de BE de  $PO_4$  de notre région diatomées 4 (4,82 mg/l en moyenne).

Concernant le  $NH_4$  :

Globalement S2 correspond à notre région 3. Dans ce cas notre classe limite de BE de  $NH_4$  est plus bas (0,08 mg/l en moyenne au lieu de 0,2 mg/l).

Ces fortes convergences de résultats, même si les objectifs respectifs sont bien distincts, permettent d'avancer que les seuils notamment de  $PO_4$  utilisés dans l'arrêté d'évaluation ne permettent pas de garantir un bon état écologique selon l'élément de qualité biologique « diatomée ».

## LITTERATURE

Bougon N. et Ferréol, M. (2012). Physico-Chimie soutenant la biologie - Identification des seuils de Bon Etat Physico-chimique des cours d'eau - Rapport d'étape. Pôle hydroécologie des cours d'eau Cemagref/Onema.

Breiman, L., J. Friedman, R. Olshen, and C. Stone (1984). Classification and regression tree. Wadsworth and Brooks.

Garzón, M. B., R. Blazek, M. Neteler, R. S. d. Dios, H. S. Ollero, and C. Furlanello (2006). Predicting habitat suitability with machine learning models : The potential area of *pinus sylvestris* l. in the iberian peninsula. *Ecological Modelling* 197(3-4), 383–393.

Gillett, N., Y. Pan, K. Manoylov, R. Stancheva, and C. Weilhoefer (2011). The potential indicator value of rare taxa richness in diatom-based stream bioassessment. *Journal of Phycology* 47(3), 471–482.

Hill BH, Herlihy AT, Kaufmann PR, Stevenson RJ, McCormick FH, Johnson CB (2000). Use of periphyton assemblage data as an index of biotic integrity. *Journal of the North American Benthological Society* 19: 50-67.

Huang, Z., T. D. Gedeon, and M. Nikravesh (2008). Pattern trees induction : A new machine learning method. *Fuzzy Systems, IEEE Transactions on Fuzzy Systems* 16(4), 958–970.

Mago, V., A. Mago, P. Sharma, and J. Mago (2011). Fuzzy logic based expert system for the treatment of mobile tooth. *Advances in Experimental Medicine and Biology* 696, 607–614.

Naumoski, A. and K. Mitreski (2009). Climate change influence on diatoms biodiversity in lake prespa. pp. 25–30.

Naumoski, A. and K. Mitreski (2011a). Diatom classification with novel bell based classification algorithm. *Communications in Computer and Information Science* 83 CCIS, 245–254.

Naumoski, A. and K. Mitreski (2011b). Novel algorithm for diatom classification in lake prespa using log-normal distribution. *Ecohydrology & Hydrobiology* 11(1), 23–34.

Passy SI. 2007. Diatom ecological guilds display distinct and predictable behavior along nutrient and disturbance gradients in running waters. *Aquatic Botany* 86: 171-178.

Rahoma, W., U. Rahoma, and A. Hassan (2011). Application of neuro-fuzzy techniques for solar radiation. *Journal of Computer Science* 7(10), 1605–1611.

Tison-Rosebery, J., Morin, S., Coste, M. (2010). Traits biologiques et écologiques comme descripteurs synthétiques des communautés diatomiques Etat de l'art et premiers tests de métriques. *Convention Onema-Cemagref / Action 14*, 24p.

van Dam H, Mertens A, Sinkeldam J. (1994). A coded checklist and ecological indicator values of freshwater diatoms from the Netherlands. *Netherlands Journal of Aquatic Ecology* 28: 117-133.

Weilhoefer, C. and Y. Pan (2008). Using change-point analysis and weighted averaging approaches to explore the relationships between common benthic diatoms and in-stream environmental variables in mid-atlantic highlands streams, usa. *Hydrobiologia* 614(1), 259–274.

# ANNEXE 1

## Tableau des paramètres

code	paramètre	unité
Temp.eau	Température	T°C
pH	pH	
altitude	altitude	m
O2.dissous	O2.dissous	mg/l
MES	Matières en suspension	mg/l
NO3-	nitrates	mg/l
DBO5	demande biologique en O2 5 jours	mg/l
NO2-	nitrites	mg/l
Orthophosp	orthophosphates	mg/l
NH4+	ammonium	mg/l
P.total	phosphore total	mg/l
AlkPred	alcalinité	meq/l
CondPred	conductivité	µs/cm
C.Orga	carbonioque organique dissous	mg/l

## Région 1

parametres	Classe_1	Classe_2	Classe_3	Classe_4	Classe_5
Temp.eau	12.9 +/- 0.7	14.3 +/- 0.7	15.7 +/- 0.7	17.1 +/- 0.7	18.5 +/- 0.7
pH	6.6 +/- 0.139	6.88 +/- 0.139	7.15 +/- 0.139	7.43 +/- 0.139	7.71 +/- 0.139
altitude	5 +/- 11.8	28.5 +/- 11.8	52 +/- 11.8	75.5 +/- 11.8	99 +/- 11.8
O2.dissous	8.05 +/- 0.162	8.37 +/- 0.162	8.7 +/- 0.162	9.02 +/- 0.162	9.35 +/- 0.162
MES	2 +/- 1.62	5.25 +/- 1.62	8.5 +/- 1.62	11.8 +/- 1.62	15 +/- 1.62
NO3-	0.1 +/- 1.44	2.98 +/- 1.44	5.85 +/- 1.44	8.73 +/- 1.44	11.6 +/- 1.44
DBO5	0.6 +/- 0.237	1.08 +/- 0.237	1.55 +/- 0.237	2.03 +/- 0.237	2.5 +/- 0.237
NO2-	0.01 +/- 0.01	0.03 +/- 0.01	0.05 +/- 0.01	0.07 +/- 0.01	0.09 +/- 0.01
Orthophosp	0.02 +/- 0.0169	0.0537 +/- 0.0169	0.0875 +/- 0.0169	0.121 +/- 0.0169	0.155 +/- 0.0169
NH4+	0.04 +/- 0.00687	0.0537 +/- 0.00687	0.0675 +/- 0.00687	0.0812 +/- 0.00687	0.095 +/- 0.00687
P.total	0.02 +/- 0.0125	0.045 +/- 0.0125	0.07 +/- 0.0125	0.095 +/- 0.0125	0.12 +/- 0.0125
AlkPred	0 +/- 1.58	3.16 +/- 1.58	6.32 +/- 1.58	9.48 +/- 1.58	12.6 +/- 1.58
CondPred	78.3 +/- 25.5	129 +/- 25.5	180 +/- 25.5	231 +/- 25.5	282 +/- 25.5
C.Orga	1.84 +/- 0.633	3.11 +/- 0.633	4.37 +/- 0.633	5.64 +/- 0.633	6.9 +/- 0.633

## Région 2

parametres	Classe_1	Classe_2	Classe_3	Classe_4	Classe_5
Temp.eau	11.5 +/- 1.48	14.5 +/- 1.48	17.4 +/- 1.48	20.4 +/- 1.48	23.3 +/- 1.48
pH	6.44 +/- 0.268	6.97 +/- 0.268	7.51 +/- 0.268	8.04 +/- 0.268	8.58 +/- 0.268
altitude	2 +/- 26.6	55.2 +/- 26.6	108 +/- 26.6	162 +/- 26.6	215 +/- 26.6
O2.dissous	5.15 +/- 0.81	6.78 +/- 0.81	8.4 +/- 0.81	10 +/- 0.81	11.6 +/- 0.81
MES	1 +/- 4	9 +/- 4	17 +/- 4	25 +/- 4	33 +/- 4
NO3-	0.6 +/- 6.21	13 +/- 6.21	25.4 +/- 6.21	37.8 +/- 6.21	50.2 +/- 6.21
DBO5	1.05 +/- 0.319	1.69 +/- 0.319	2.32 +/- 0.319	2.96 +/- 0.319	3.6 +/- 0.319
NO2-	0.01 +/- 0.0275	0.065 +/- 0.0275	0.12 +/- 0.0275	0.175 +/- 0.0275	0.23 +/- 0.0275
Orthophosp	0.01 +/- 0.075	0.16 +/- 0.075	0.31 +/- 0.075	0.46 +/- 0.075	0.61 +/- 0.075
NH4+	0.01 +/- 0.0256	0.0612 +/- 0.0256	0.112 +/- 0.0256	0.164 +/- 0.0256	0.215 +/- 0.0256
P.total	0.015 +/- 0.0469	0.109 +/- 0.0469	0.202 +/- 0.0469	0.296 +/- 0.0469	0.39 +/- 0.0469
AlkPred	0 +/- 3.27	6.53 +/- 3.27	13.1 +/- 3.27	19.6 +/- 3.27	26.1 +/- 3.27
CondPred	55.5 +/- 79.8	215 +/- 79.8	375 +/- 79.8	534 +/- 79.8	694 +/- 79.8
C.Orga	1.12 +/- 1.57	4.26 +/- 1.57	7.41 +/- 1.57	10.6 +/- 1.57	13.7 +/- 1.57

## Région 3

parametres	Classe_1	Classe_2	Classe_3	Classe_4	Classe_5
Temp.eau	8.55 +/- 1.96	12.5 +/- 1.96	16.4 +/- 1.96	20.3 +/- 1.96	24.2 +/- 1.96
pH	6.45 +/- 0.263	6.97 +/- 0.263	7.5 +/- 0.263	8.02 +/- 0.263	8.55 +/- 0.263
altitude	0 +/- 143	287 +/- 143	574 +/- 143	860 +/- 143	1150 +/- 143
O2.dissous	6.9 +/- 0.567	8.03 +/- 0.567	9.17 +/- 0.567	10.3 +/- 0.567	11.4 +/- 0.567
MES	1 +/- 2.74	6.48 +/- 2.74	12 +/- 2.74	17.4 +/- 2.74	22.9 +/- 2.74
NO3-	0.1 +/- 1.56	3.23 +/- 1.56	6.35 +/- 1.56	9.48 +/- 1.56	12.6 +/- 1.56
DBO5	0.5 +/- 0.387	1.27 +/- 0.387	2.05 +/- 0.387	2.82 +/- 0.387	3.6 +/- 0.387
NO2-	0.01 +/- 0.0106	0.0312 +/- 0.0106	0.0525 +/- 0.0106	0.0738 +/- 0.0106	0.095 +/- 0.0106
Orthophosp	0.01 +/- 0.0313	0.0725 +/- 0.0313	0.135 +/- 0.0313	0.198 +/- 0.0313	0.26 +/- 0.0313
NH4+	0.02 +/- 0.01	0.04 +/- 0.01	0.06 +/- 0.01	0.08 +/- 0.01	0.1 +/- 0.01
P.total	0.01 +/- 0.0238	0.0575 +/- 0.0238	0.105 +/- 0.0238	0.153 +/- 0.0238	0.2 +/- 0.0238
AlkPred	0 +/- 2.24	4.49 +/- 2.24	8.98 +/- 2.24	13.5 +/- 2.24	18 +/- 2.24
CondPred	14 +/- 50.8	116 +/- 50.8	217 +/- 50.8	318 +/- 50.8	420 +/- 50.8
C.Orga	0.4 +/- 1.14	2.69 +/- 1.14	4.98 +/- 1.14	7.26 +/- 1.14	9.55 +/- 1.14

## Région 4

parametres	Classe_1	Classe_2	Classe_3	Classe_4	Classe_5
Temp.eau	9.8 +/- 1.93	13.7 +/- 1.93	17.5 +/- 1.93	21.4 +/- 1.93	25.2 +/- 1.93
pH	7.27 +/- 0.164	7.6 +/- 0.164	7.93 +/- 0.164	8.26 +/- 0.164	8.58 +/- 0.164
altitude	1 +/- 39	79 +/- 39	157 +/- 39	235 +/- 39	313 +/- 39
O2.dissous	6 +/- 0.731	7.46 +/- 0.731	8.92 +/- 0.731	10.4 +/- 0.731	11.8 +/- 0.731
MES	1 +/- 4.12	9.25 +/- 4.12	17.5 +/- 4.12	25.8 +/- 4.12	34 +/- 4.12
NO3-	0.1 +/- 6.2	12.5 +/- 6.2	24.9 +/- 6.2	37.3 +/- 6.2	49.7 +/- 6.2
DBO5	0.5 +/- 0.419	1.34 +/- 0.419	2.18 +/- 0.419	3.01 +/- 0.419	3.85 +/- 0.419
NO2-	0.01 +/- 0.0281	0.0663 +/- 0.0281	0.123 +/- 0.0281	0.179 +/- 0.0281	0.235 +/- 0.0281
Orthophosp	0.01 +/- 0.0569	0.124 +/- 0.0569	0.238 +/- 0.0569	0.351 +/- 0.0569	0.465 +/- 0.0569
NH4+	0.01 +/- 0.0169	0.0437 +/- 0.0169	0.0775 +/- 0.0169	0.111 +/- 0.0169	0.145 +/- 0.0169
P.total	0.0125 +/- 0.0297	0.0719 +/- 0.0297	0.131 +/- 0.0297	0.191 +/- 0.0297	0.25 +/- 0.0297
AlkPred	5.1 +/- 3.73	12.6 +/- 3.73	20 +/- 3.73	27.5 +/- 3.73	35 +/- 3.73
CondPred	186 +/- 94.1	374 +/- 94.1	562 +/- 94.1	750 +/- 94.1	938 +/- 94.1
C.Orga	0.2 +/- 0.771	1.74 +/- 0.771	3.28 +/- 0.771	4.82 +/- 0.771	6.37 +/- 0.771

## Région 5

parametres	Classe_1	Classe_2	Classe_3	Classe_4	Classe_5
Temp.eau	3.6 +/- 2.56	8.73 +/- 2.56	13.9 +/- 2.56	19 +/- 2.56	24.1 +/- 2.56
pH	7.55 +/- 0.125	7.8 +/- 0.125	8.05 +/- 0.125	8.3 +/- 0.125	8.55 +/- 0.125
altitude	1 +/- 184	369 +/- 184	738 +/- 184	1110 +/- 184	1470 +/- 184
O2.dissous	6.86 +/- 0.78	8.42 +/- 0.78	9.98 +/- 0.78	11.5 +/- 0.78	13.1 +/- 0.78
MES	1 +/- 2.38	5.75 +/- 2.38	10.5 +/- 2.38	15.2 +/- 2.38	20 +/- 2.38
NO3-	0.1 +/- 1.05	2.2 +/- 1.05	4.3 +/- 1.05	6.4 +/- 1.05	8.5 +/- 1.05
DBO5	0.5 +/- 0.275	1.05 +/- 0.275	1.6 +/- 0.275	2.15 +/- 0.275	2.7 +/- 0.275
NO2-	0.02 +/- 0.00292	0.0258 +/- 0.00292	0.0317 +/- 0.00292	0.0375 +/- 0.00292	0.0433 +/- 0.00292
Orthophosp	0.01 +/- 0.015	0.04 +/- 0.015	0.07 +/- 0.015	0.1 +/- 0.015	0.13 +/- 0.015
NH4+	0.05 +/- 0.00125	0.0525 +/- 0.00125	0.055 +/- 0.00125	0.0575 +/- 0.00125	0.06 +/- 0.00125
P.total	0.01 +/- 0.00625	0.0225 +/- 0.00625	0.035 +/- 0.00625	0.0475 +/- 0.00625	0.06 +/- 0.00625
AlkPred	3.75 +/- 2.63	9.01 +/- 2.63	14.3 +/- 2.63	19.5 +/- 2.63	24.8 +/- 2.63
CondPred	29 +/- 93.8	216 +/- 93.8	404 +/- 93.8	592 +/- 93.8	779 +/- 93.8
C.Orga	0.2 +/- 0.381	0.963 +/- 0.381	1.73 +/- 0.381	2.49 +/- 0.381	3.25 +/- 0.381

## Guilde 1

parametres	Classe_1	Classe_2	Classe_3	Classe_4	Classe_5
Temp.eau	8.4 +/- 2.14	12.7 +/- 2.14	17 +/- 2.14	21.3 +/- 2.14	25.5 +/- 2.14
pH	6.63 +/- 0.283	7.2 +/- 0.283	7.77 +/- 0.283	8.33 +/- 0.283	8.9 +/- 0.283
altitude	0 +/- 87.8	176 +/- 87.8	351 +/- 87.8	526 +/- 87.8	702 +/- 87.8
O2.dissous	6.05 +/- 0.744	7.54 +/- 0.744	9.02 +/- 0.744	10.5 +/- 0.744	12 +/- 0.744
MES	1 +/- 3.88	8.75 +/- 3.88	16.5 +/- 3.88	24.2 +/- 3.88	32 +/- 3.88
NO3-	0.1 +/- 4.26	8.63 +/- 4.26	17.2 +/- 4.26	25.7 +/- 4.26	34.2 +/- 4.26
DBO5	0.5 +/- 0.394	1.29 +/- 0.394	2.07 +/- 0.394	2.86 +/- 0.394	3.65 +/- 0.394
NO2-	0.01 +/- 0.02	0.05 +/- 0.02	0.09 +/- 0.02	0.13 +/- 0.02	0.17 +/- 0.02
Orthophosp	0.01 +/- 0.0444	0.0987 +/- 0.0444	0.187 +/- 0.0444	0.276 +/- 0.0444	0.365 +/- 0.0444
NH4+	0.01 +/- 0.0144	0.0388 +/- 0.0144	0.0675 +/- 0.0144	0.0963 +/- 0.0144	0.125 +/- 0.0144
P.total	0.01 +/- 0.0256	0.0613 +/- 0.0256	0.113 +/- 0.0256	0.164 +/- 0.0256	0.215 +/- 0.0256
NK	0.51 +/- 0.0987	0.707 +/- 0.0987	0.905 +/- 0.0987	1.1 +/- 0.0987	1.3 +/- 0.0987
AlkPred	0 +/- 4.15	8.31 +/- 4.15	16.6 +/- 4.15	24.9 +/- 4.15	33.2 +/- 4.15
CondPred	14 +/- 132	278 +/- 132	542 +/- 132	806 +/- 132	1070 +/- 132
C.Orga	0.2 +/- 1.05	2.3 +/- 1.05	4.4 +/- 1.05	6.5 +/- 1.05	8.6 +/- 1.05
racine	7.69 +/- 13.9	35.4 +/- 13.9	63.2 +/- 13.9	NA	NA

## Guilde 2

parametres	Classe_1	Classe_2	Classe_3	Classe_4	Classe_5
Temp.eau	8.4 +/- 2.14	12.7 +/- 2.14	17 +/- 2.14	21.3 +/- 2.14	25.5 +/- 2.14
pH	6.63 +/- 0.283	7.2 +/- 0.283	7.77 +/- 0.283	8.33 +/- 0.283	8.9 +/- 0.283
altitude	0 +/- 87.8	176 +/- 87.8	351 +/- 87.8	526 +/- 87.8	702 +/- 87.8
O2.dissous	6.05 +/- 0.744	7.54 +/- 0.744	9.02 +/- 0.744	10.5 +/- 0.744	12 +/- 0.744
MES	1 +/- 3.88	8.75 +/- 3.88	16.5 +/- 3.88	24.2 +/- 3.88	32 +/- 3.88
NO3-	0.1 +/- 4.26	8.63 +/- 4.26	17.2 +/- 4.26	25.7 +/- 4.26	34.2 +/- 4.26
DBO5	0.5 +/- 0.394	1.29 +/- 0.394	2.07 +/- 0.394	2.86 +/- 0.394	3.65 +/- 0.394
NO2-	0.01 +/- 0.02	0.05 +/- 0.02	0.09 +/- 0.02	0.13 +/- 0.02	0.17 +/- 0.02
Orthophosp	0.01 +/- 0.0444	0.0987 +/- 0.0444	0.187 +/- 0.0444	0.276 +/- 0.0444	0.365 +/- 0.0444
NH4+	0.01 +/- 0.0144	0.0388 +/- 0.0144	0.0675 +/- 0.0144	0.0963 +/- 0.0144	0.125 +/- 0.0144
P.total	0.01 +/- 0.0256	0.0613 +/- 0.0256	0.113 +/- 0.0256	0.164 +/- 0.0256	0.215 +/- 0.0256
NK	0.51 +/- 0.0987	0.707 +/- 0.0987	0.905 +/- 0.0987	1.1 +/- 0.0987	1.3 +/- 0.0987
AlkPred	0 +/- 4.15	8.31 +/- 4.15	16.6 +/- 4.15	24.9 +/- 4.15	33.2 +/- 4.15
CondPred	14 +/- 132	278 +/- 132	542 +/- 132	806 +/- 132	1070 +/- 132
C.Orga	0.2 +/- 1.05	2.3 +/- 1.05	4.4 +/- 1.05	6.5 +/- 1.05	8.6 +/- 1.05
racine	0 +/- 10	20 +/- 10	40 +/- 10	NA	NA

## Guilde 3

parametres	Classe_1	Classe_2	Classe_3	Classe_4	Classe_5
Temp.eau	8.4 +/- 2.14	12.7 +/- 2.14	17 +/- 2.14	21.3 +/- 2.14	25.5 +/- 2.14
pH	6.63 +/- 0.283	7.2 +/- 0.283	7.77 +/- 0.283	8.33 +/- 0.283	8.9 +/- 0.283
altitude	0 +/- 87.8	176 +/- 87.8	351 +/- 87.8	526 +/- 87.8	702 +/- 87.8
O2.dissous	6.05 +/- 0.744	7.54 +/- 0.744	9.02 +/- 0.744	10.5 +/- 0.744	12 +/- 0.744
MES	1 +/- 3.88	8.75 +/- 3.88	16.5 +/- 3.88	24.2 +/- 3.88	32 +/- 3.88
NO3-	0.1 +/- 4.26	8.63 +/- 4.26	17.2 +/- 4.26	25.7 +/- 4.26	34.2 +/- 4.26
DBO5	0.5 +/- 0.394	1.29 +/- 0.394	2.07 +/- 0.394	2.86 +/- 0.394	3.65 +/- 0.394
NO2-	0.01 +/- 0.02	0.05 +/- 0.02	0.09 +/- 0.02	0.13 +/- 0.02	0.17 +/- 0.02
Orthophosp	0.01 +/- 0.0444	0.0987 +/- 0.0444	0.187 +/- 0.0444	0.276 +/- 0.0444	0.365 +/- 0.0444
NH4+	0.01 +/- 0.0144	0.0388 +/- 0.0144	0.0675 +/- 0.0144	0.0963 +/- 0.0144	0.125 +/- 0.0144
P.total	0.01 +/- 0.0256	0.0613 +/- 0.0256	0.113 +/- 0.0256	0.164 +/- 0.0256	0.215 +/- 0.0256
NK	0.51 +/- 0.0987	0.707 +/- 0.0987	0.905 +/- 0.0987	1.1 +/- 0.0987	1.3 +/- 0.0987
AlkPred	0 +/- 4.15	8.31 +/- 4.15	16.6 +/- 4.15	24.9 +/- 4.15	33.2 +/- 4.15
CondPred	14 +/- 132	278 +/- 132	542 +/- 132	806 +/- 132	1070 +/- 132
C.Orga	0.2 +/- 1.05	2.3 +/- 1.05	4.4 +/- 1.05	6.5 +/- 1.05	8.6 +/- 1.05
racine	6.67 +/- 18.7	44.1 +/- 18.7	81.6 +/- 18.7	NA	NA

## N-autotrophes sensibles

parametres	Classe_1	Classe_2	Classe_3	Classe_4	Classe_5
Temp.eau	8.4 +/- 2.14	12.7 +/- 2.14	17 +/- 2.14	21.3 +/- 2.14	25.5 +/- 2.14
pH	6.63 +/- 0.283	7.2 +/- 0.283	7.77 +/- 0.283	8.33 +/- 0.283	8.9 +/- 0.283
altitude	0 +/- 87.8	176 +/- 87.8	351 +/- 87.8	526 +/- 87.8	702 +/- 87.8
O2.dissous	6.05 +/- 0.744	7.54 +/- 0.744	9.02 +/- 0.744	10.5 +/- 0.744	12 +/- 0.744
MES	1 +/- 3.88	8.75 +/- 3.88	16.5 +/- 3.88	24.2 +/- 3.88	32 +/- 3.88
NO3-	0.1 +/- 4.26	8.63 +/- 4.26	17.2 +/- 4.26	25.7 +/- 4.26	34.2 +/- 4.26
DBO5	0.5 +/- 0.394	1.29 +/- 0.394	2.07 +/- 0.394	2.86 +/- 0.394	3.65 +/- 0.394
NO2-	0.01 +/- 0.02	0.05 +/- 0.02	0.09 +/- 0.02	0.13 +/- 0.02	0.17 +/- 0.02
Orthophosp	0.01 +/- 0.0444	0.0987 +/- 0.0444	0.187 +/- 0.0444	0.276 +/- 0.0444	0.365 +/- 0.0444
NH4+	0.01 +/- 0.0144	0.0388 +/- 0.0144	0.0675 +/- 0.0144	0.0963 +/- 0.0144	0.125 +/- 0.0144
P.total	0.01 +/- 0.0256	0.0613 +/- 0.0256	0.113 +/- 0.0256	0.164 +/- 0.0256	0.215 +/- 0.0256
NK	0.51 +/- 0.0987	0.707 +/- 0.0987	0.905 +/- 0.0987	1.1 +/- 0.0987	1.3 +/- 0.0987
AlkPred	0 +/- 4.15	8.31 +/- 4.15	16.6 +/- 4.15	24.9 +/- 4.15	33.2 +/- 4.15
CondPred	14 +/- 132	278 +/- 132	542 +/- 132	806 +/- 132	1070 +/- 132
C.Orga	0.2 +/- 1.05	2.3 +/- 1.05	4.4 +/- 1.05	6.5 +/- 1.05	8.6 +/- 1.05
racine	0 +/- 11.5	23 +/- 11.5	46 +/- 11.5	NA	NA

## N-autotrophes tolérantes

parametres	Classe_1	Classe_2	Classe_3	Classe_4	Classe_5
Temp.eau	8.4 +/- 2.14	12.7 +/- 2.14	17 +/- 2.14	21.3 +/- 2.14	25.5 +/- 2.14
pH	6.63 +/- 0.283	7.2 +/- 0.283	7.77 +/- 0.283	8.33 +/- 0.283	8.9 +/- 0.283
altitude	0 +/- 87.8	176 +/- 87.8	351 +/- 87.8	526 +/- 87.8	702 +/- 87.8
O2.dissous	6.05 +/- 0.744	7.54 +/- 0.744	9.02 +/- 0.744	10.5 +/- 0.744	12 +/- 0.744
MES	1 +/- 3.88	8.75 +/- 3.88	16.5 +/- 3.88	24.2 +/- 3.88	32 +/- 3.88
NO3-	0.1 +/- 4.26	8.63 +/- 4.26	17.2 +/- 4.26	25.7 +/- 4.26	34.2 +/- 4.26
DBO5	0.5 +/- 0.394	1.29 +/- 0.394	2.07 +/- 0.394	2.86 +/- 0.394	3.65 +/- 0.394
NO2-	0.01 +/- 0.02	0.05 +/- 0.02	0.09 +/- 0.02	0.13 +/- 0.02	0.17 +/- 0.02
Orthophosp	0.01 +/- 0.0444	0.0987 +/- 0.0444	0.187 +/- 0.0444	0.276 +/- 0.0444	0.365 +/- 0.0444
NH4+	0.01 +/- 0.0144	0.0388 +/- 0.0144	0.0675 +/- 0.0144	0.0963 +/- 0.0144	0.125 +/- 0.0144
P.total	0.01 +/- 0.0256	0.0613 +/- 0.0256	0.113 +/- 0.0256	0.164 +/- 0.0256	0.215 +/- 0.0256
NK	0.51 +/- 0.0987	0.707 +/- 0.0987	0.905 +/- 0.0987	1.1 +/- 0.0987	1.3 +/- 0.0987
AlkPred	0 +/- 4.15	8.31 +/- 4.15	16.6 +/- 4.15	24.9 +/- 4.15	33.2 +/- 4.15
CondPred	14 +/- 132	278 +/- 132	542 +/- 132	806 +/- 132	1070 +/- 132
C.Orga	0.2 +/- 1.05	2.3 +/- 1.05	4.4 +/- 1.05	6.5 +/- 1.05	8.6 +/- 1.05
racine	33.3 +/- 12.7	58.8 +/- 12.7	84.2 +/- 12.7	NA	NA

## N-hétérotrophes facultatifs

parametres	Classe_1	Classe_2	Classe_3	Classe_4	Classe_5
Temp.eau	8.4 +/- 2.14	12.7 +/- 2.14	17 +/- 2.14	21.3 +/- 2.14	25.5 +/- 2.14
pH	6.63 +/- 0.283	7.2 +/- 0.283	7.77 +/- 0.283	8.33 +/- 0.283	8.9 +/- 0.283
altitude	0 +/- 87.8	176 +/- 87.8	351 +/- 87.8	526 +/- 87.8	702 +/- 87.8
O2.dissous	6.05 +/- 0.744	7.54 +/- 0.744	9.02 +/- 0.744	10.5 +/- 0.744	12 +/- 0.744
MES	1 +/- 3.88	8.75 +/- 3.88	16.5 +/- 3.88	24.2 +/- 3.88	32 +/- 3.88
NO3-	0.1 +/- 4.26	8.63 +/- 4.26	17.2 +/- 4.26	25.7 +/- 4.26	34.2 +/- 4.26
DBO5	0.5 +/- 0.394	1.29 +/- 0.394	2.07 +/- 0.394	2.86 +/- 0.394	3.65 +/- 0.394
NO2-	0.01 +/- 0.02	0.05 +/- 0.02	0.09 +/- 0.02	0.13 +/- 0.02	0.17 +/- 0.02
Orthophosp	0.01 +/- 0.0444	0.0987 +/- 0.0444	0.187 +/- 0.0444	0.276 +/- 0.0444	0.365 +/- 0.0444
NH4+	0.01 +/- 0.0144	0.0388 +/- 0.0144	0.0675 +/- 0.0144	0.0963 +/- 0.0144	0.125 +/- 0.0144
P.total	0.01 +/- 0.0256	0.0613 +/- 0.0256	0.113 +/- 0.0256	0.164 +/- 0.0256	0.215 +/- 0.0256
NK	0.51 +/- 0.0987	0.707 +/- 0.0987	0.905 +/- 0.0987	1.1 +/- 0.0987	1.3 +/- 0.0987
AlkPred	0 +/- 4.15	8.31 +/- 4.15	16.6 +/- 4.15	24.9 +/- 4.15	33.2 +/- 4.15
CondPred	14 +/- 132	278 +/- 132	542 +/- 132	806 +/- 132	1070 +/- 132
C.Orga	0.2 +/- 1.05	2.3 +/- 1.05	4.4 +/- 1.05	6.5 +/- 1.05	8.6 +/- 1.05
racine	0 +/- 9.78	19.6 +/- 9.78	39.1 +/- 9.78	NA	NA

## N-hétérotrophes obligatoires

parametre	Classe_1	Classe_2	Classe_3	Classe_4	Classe_5
Temp.eau	8.4 +/- 2.14	12.7 +/- 2.14	17 +/- 2.14	21.3 +/- 2.14	25.5 +/- 2.14
pH	6.63 +/- 0.283	7.2 +/- 0.283	7.77 +/- 0.283	8.33 +/- 0.283	8.9 +/- 0.283
altitude	0 +/- 87.8	176 +/- 87.8	351 +/- 87.8	526 +/- 87.8	702 +/- 87.8
O2.dissous	6.05 +/- 0.744	7.54 +/- 0.744	9.02 +/- 0.744	10.5 +/- 0.744	12 +/- 0.744
MES	1 +/- 3.88	8.75 +/- 3.88	16.5 +/- 3.88	24.2 +/- 3.88	32 +/- 3.88
NO3-	0.1 +/- 4.26	8.63 +/- 4.26	17.2 +/- 4.26	25.7 +/- 4.26	34.2 +/- 4.26
DBO5	0.5 +/- 0.394	1.29 +/- 0.394	2.07 +/- 0.394	2.86 +/- 0.394	3.65 +/- 0.394
NO2-	0.01 +/- 0.02	0.05 +/- 0.02	0.09 +/- 0.02	0.13 +/- 0.02	0.17 +/- 0.02
Orthophosp	0.01 +/- 0.0444	0.0987 +/- 0.0444	0.187 +/- 0.0444	0.276 +/- 0.0444	0.365 +/- 0.0444
NH4+	0.01 +/- 0.0144	0.0388 +/- 0.0144	0.0675 +/- 0.0144	0.0963 +/- 0.0144	0.125 +/- 0.0144
P.total	0.01 +/- 0.0256	0.0613 +/- 0.0256	0.113 +/- 0.0256	0.164 +/- 0.0256	0.215 +/- 0.0256
NK	0.51 +/- 0.0987	0.707 +/- 0.0987	0.905 +/- 0.0987	1.1 +/- 0.0987	1.3 +/- 0.0987
AlkPred	0 +/- 4.15	8.31 +/- 4.15	16.6 +/- 4.15	24.9 +/- 4.15	33.2 +/- 4.15
CondPred	14 +/- 132	278 +/- 132	542 +/- 132	806 +/- 132	1070 +/- 132
C.Orga	0.2 +/- 1.05	2.3 +/- 1.05	4.4 +/- 1.05	6.5 +/- 1.05	8.6 +/- 1.05
racine	0 +/- 5.95	11.9 +/- 5.95	23.8 +/- 5.95	NA	NA

## Oligosaprobes

parametres	Classe_1	Classe_2	Classe_3	Classe_4	Classe_5
Temp.eau	8.4 +/- 2.14	12.7 +/- 2.14	17 +/- 2.14	21.3 +/- 2.14	25.5 +/- 2.14
pH	6.63 +/- 0.283	7.2 +/- 0.283	7.77 +/- 0.283	8.33 +/- 0.283	8.9 +/- 0.283
altitude	0 +/- 87.8	176 +/- 87.8	351 +/- 87.8	526 +/- 87.8	702 +/- 87.8
O2.dissous	6.05 +/- 0.744	7.54 +/- 0.744	9.02 +/- 0.744	10.5 +/- 0.744	12 +/- 0.744
MES	1 +/- 3.88	8.75 +/- 3.88	16.5 +/- 3.88	24.2 +/- 3.88	32 +/- 3.88
NO3-	0.1 +/- 4.26	8.63 +/- 4.26	17.2 +/- 4.26	25.7 +/- 4.26	34.2 +/- 4.26
DBO5	0.5 +/- 0.394	1.29 +/- 0.394	2.07 +/- 0.394	2.86 +/- 0.394	3.65 +/- 0.394
NO2-	0.01 +/- 0.02	0.05 +/- 0.02	0.09 +/- 0.02	0.13 +/- 0.02	0.17 +/- 0.02
Orthophosp	0.01 +/- 0.0444	0.0987 +/- 0.0444	0.187 +/- 0.0444	0.276 +/- 0.0444	0.365 +/- 0.0444
NH4+	0.01 +/- 0.0144	0.0388 +/- 0.0144	0.0675 +/- 0.0144	0.0963 +/- 0.0144	0.125 +/- 0.0144
P.total	0.01 +/- 0.0256	0.0613 +/- 0.0256	0.113 +/- 0.0256	0.164 +/- 0.0256	0.215 +/- 0.0256
NK	0.51 +/- 0.0987	0.707 +/- 0.0987	0.905 +/- 0.0987	1.1 +/- 0.0987	1.3 +/- 0.0987
AlkPred	0 +/- 4.15	8.31 +/- 4.15	16.6 +/- 4.15	24.9 +/- 4.15	33.2 +/- 4.15
CondPred	14 +/- 132	278 +/- 132	542 +/- 132	806 +/- 132	1070 +/- 132
C.Orga	0.2 +/- 1.05	2.3 +/- 1.05	4.4 +/- 1.05	6.5 +/- 1.05	8.6 +/- 1.05
racine	0 +/- 8.33	16.7 +/- 8.33	33.3 +/- 8.33	NA	NA

## Beta-mesosaprobies

parametres	Classe_1	Classe_2	Classe_3	Classe_4	Classe_5
Temp.eau	8.4 +/- 2.14	12.7 +/- 2.14	17 +/- 2.14	21.3 +/- 2.14	25.5 +/- 2.14
pH	6.63 +/- 0.283	7.2 +/- 0.283	7.77 +/- 0.283	8.33 +/- 0.283	8.9 +/- 0.283
altitude	0 +/- 87.8	176 +/- 87.8	351 +/- 87.8	526 +/- 87.8	702 +/- 87.8
O2.dissous	6.05 +/- 0.744	7.54 +/- 0.744	9.02 +/- 0.744	10.5 +/- 0.744	12 +/- 0.744
MES	1 +/- 3.88	8.75 +/- 3.88	16.5 +/- 3.88	24.2 +/- 3.88	32 +/- 3.88
NO3-	0.1 +/- 4.26	8.63 +/- 4.26	17.2 +/- 4.26	25.7 +/- 4.26	34.2 +/- 4.26
DBO5	0.5 +/- 0.394	1.29 +/- 0.394	2.07 +/- 0.394	2.86 +/- 0.394	3.65 +/- 0.394
NO2-	0.01 +/- 0.02	0.05 +/- 0.02	0.09 +/- 0.02	0.13 +/- 0.02	0.17 +/- 0.02
Orthophosp	0.01 +/- 0.0444	0.0987 +/- 0.0444	0.187 +/- 0.0444	0.276 +/- 0.0444	0.365 +/- 0.0444
NH4+	0.01 +/- 0.0144	0.0388 +/- 0.0144	0.0675 +/- 0.0144	0.0963 +/- 0.0144	0.125 +/- 0.0144
P.total	0.01 +/- 0.0256	0.0613 +/- 0.0256	0.113 +/- 0.0256	0.164 +/- 0.0256	0.215 +/- 0.0256
NK	0.51 +/- 0.0987	0.707 +/- 0.0987	0.905 +/- 0.0987	1.1 +/- 0.0987	1.3 +/- 0.0987
AlkPred	0 +/- 4.15	8.31 +/- 4.15	16.6 +/- 4.15	24.9 +/- 4.15	33.2 +/- 4.15
CondPred	14 +/- 132	278 +/- 132	542 +/- 132	806 +/- 132	1070 +/- 132
C.Orga	0.2 +/- 1.05	2.3 +/- 1.05	4.4 +/- 1.05	6.5 +/- 1.05	8.6 +/- 1.05
racine	18.8 +/- 14.4	47.5 +/- 14.4	76.2 +/- 14.4	NA	NA

## Alpha-mesosaprobies

parametres	Classe_1	Classe_2	Classe_3	Classe_4	Classe_5
Temp.eau	8.4 +/- 2.14	12.7 +/- 2.14	17 +/- 2.14	21.3 +/- 2.14	25.5 +/- 2.14
pH	6.63 +/- 0.283	7.2 +/- 0.283	7.77 +/- 0.283	8.33 +/- 0.283	8.9 +/- 0.283
altitude	0 +/- 87.8	176 +/- 87.8	351 +/- 87.8	526 +/- 87.8	702 +/- 87.8
O2.dissous	6.05 +/- 0.744	7.54 +/- 0.744	9.02 +/- 0.744	10.5 +/- 0.744	12 +/- 0.744
MES	1 +/- 3.88	8.75 +/- 3.88	16.5 +/- 3.88	24.2 +/- 3.88	32 +/- 3.88
NO3-	0.1 +/- 4.26	8.63 +/- 4.26	17.2 +/- 4.26	25.7 +/- 4.26	34.2 +/- 4.26
DBO5	0.5 +/- 0.394	1.29 +/- 0.394	2.07 +/- 0.394	2.86 +/- 0.394	3.65 +/- 0.394
NO2-	0.01 +/- 0.02	0.05 +/- 0.02	0.09 +/- 0.02	0.13 +/- 0.02	0.17 +/- 0.02
Orthophosp	0.01 +/- 0.0444	0.0987 +/- 0.0444	0.187 +/- 0.0444	0.276 +/- 0.0444	0.365 +/- 0.0444
NH4+	0.01 +/- 0.0144	0.0388 +/- 0.0144	0.0675 +/- 0.0144	0.0963 +/- 0.0144	0.125 +/- 0.0144
P.total	0.01 +/- 0.0256	0.0613 +/- 0.0256	0.113 +/- 0.0256	0.164 +/- 0.0256	0.215 +/- 0.0256
NK	0.51 +/- 0.0987	0.707 +/- 0.0987	0.905 +/- 0.0987	1.1 +/- 0.0987	1.3 +/- 0.0987
AlkPred	0 +/- 4.15	8.31 +/- 4.15	16.6 +/- 4.15	24.9 +/- 4.15	33.2 +/- 4.15
CondPred	14 +/- 132	278 +/- 132	542 +/- 132	806 +/- 132	1070 +/- 132
C.Orga	0.2 +/- 1.05	2.3 +/- 1.05	4.4 +/- 1.05	6.5 +/- 1.05	8.6 +/- 1.05
racine	3.57 +/- 9.82	23.2 +/- 9.82	42.9 +/- 9.82	NA	NA

## Alphameso-polysaprobe

parametres	Classe_1	Classe_2	Classe_3	Classe_4	Classe_5
Temp.eau	8.4 +/- 2.14	12.7 +/- 2.14	17 +/- 2.14	21.3 +/- 2.14	25.5 +/- 2.14
pH	6.63 +/- 0.283	7.2 +/- 0.283	7.77 +/- 0.283	8.33 +/- 0.283	8.9 +/- 0.283
altitude	0 +/- 87.8	176 +/- 87.8	351 +/- 87.8	526 +/- 87.8	702 +/- 87.8
O2.dissous	6.05 +/- 0.744	7.54 +/- 0.744	9.02 +/- 0.744	10.5 +/- 0.744	12 +/- 0.744
MES	1 +/- 3.88	8.75 +/- 3.88	16.5 +/- 3.88	24.2 +/- 3.88	32 +/- 3.88
NO3-	0.1 +/- 4.26	8.63 +/- 4.26	17.2 +/- 4.26	25.7 +/- 4.26	34.2 +/- 4.26
DBO5	0.5 +/- 0.394	1.29 +/- 0.394	2.07 +/- 0.394	2.86 +/- 0.394	3.65 +/- 0.394
NO2-	0.01 +/- 0.02	0.05 +/- 0.02	0.09 +/- 0.02	0.13 +/- 0.02	0.17 +/- 0.02
Orthophosp	0.01 +/- 0.0444	0.0987 +/- 0.0444	0.187 +/- 0.0444	0.276 +/- 0.0444	0.365 +/- 0.0444
NH4+	0.01 +/- 0.0144	0.0388 +/- 0.0144	0.0675 +/- 0.0144	0.0963 +/- 0.0144	0.125 +/- 0.0144
P.total	0.01 +/- 0.0256	0.0613 +/- 0.0256	0.113 +/- 0.0256	0.164 +/- 0.0256	0.215 +/- 0.0256
NK	0.51 +/- 0.0987	0.707 +/- 0.0987	0.905 +/- 0.0987	1.1 +/- 0.0987	1.3 +/- 0.0987
AlkPred	0 +/- 4.15	8.31 +/- 4.15	16.6 +/- 4.15	24.9 +/- 4.15	33.2 +/- 4.15
CondPred	14 +/- 132	278 +/- 132	542 +/- 132	806 +/- 132	1070 +/- 132
C.Orga	0.2 +/- 1.05	2.3 +/- 1.05	4.4 +/- 1.05	6.5 +/- 1.05	8.6 +/- 1.05
racine	0 +/- 9.38	18.8 +/- 9.38	37.5 +/- 9.38	NA	NA

## Polysaprobe

parametres	Classe_1	Classe_2	Classe_3	Classe_4	Classe_5
Temp.eau	8.4 +/- 2.14	12.7 +/- 2.14	17 +/- 2.14	21.3 +/- 2.14	25.5 +/- 2.14
pH	6.63 +/- 0.283	7.2 +/- 0.283	7.77 +/- 0.283	8.33 +/- 0.283	8.9 +/- 0.283
altitude	0 +/- 87.8	176 +/- 87.8	351 +/- 87.8	526 +/- 87.8	702 +/- 87.8
O2.dissous	6.05 +/- 0.744	7.54 +/- 0.744	9.02 +/- 0.744	10.5 +/- 0.744	12 +/- 0.744
MES	1 +/- 3.88	8.75 +/- 3.88	16.5 +/- 3.88	24.2 +/- 3.88	32 +/- 3.88
NO3-	0.1 +/- 4.26	8.63 +/- 4.26	17.2 +/- 4.26	25.7 +/- 4.26	34.2 +/- 4.26
DBO5	0.5 +/- 0.394	1.29 +/- 0.394	2.07 +/- 0.394	2.86 +/- 0.394	3.65 +/- 0.394
NO2-	0.01 +/- 0.02	0.05 +/- 0.02	0.09 +/- 0.02	0.13 +/- 0.02	0.17 +/- 0.02
Orthophosp	0.01 +/- 0.0444	0.0987 +/- 0.0444	0.187 +/- 0.0444	0.276 +/- 0.0444	0.365 +/- 0.0444
NH4+	0.01 +/- 0.0144	0.0388 +/- 0.0144	0.0675 +/- 0.0144	0.0963 +/- 0.0144	0.125 +/- 0.0144
P.total	0.01 +/- 0.0256	0.0613 +/- 0.0256	0.113 +/- 0.0256	0.164 +/- 0.0256	0.215 +/- 0.0256
NK	0.51 +/- 0.0987	0.707 +/- 0.0987	0.905 +/- 0.0987	1.1 +/- 0.0987	1.3 +/- 0.0987
AlkPred	0 +/- 4.15	8.31 +/- 4.15	16.6 +/- 4.15	24.9 +/- 4.15	33.2 +/- 4.15
CondPred	14 +/- 132	278 +/- 132	542 +/- 132	806 +/- 132	1070 +/- 132
C.Orga	0.2 +/- 1.05	2.3 +/- 1.05	4.4 +/- 1.05	6.5 +/- 1.05	8.6 +/- 1.05
racine	0 +/- 2.42	4.84 +/- 2.42	9.68 +/- 2.42	NA	NA

## Oligotrophes

parametres	Classe_1	Classe_2	Classe_3	Classe_4	Classe_5
Temp.eau	8.4 +/- 2.14	12.7 +/- 2.14	17 +/- 2.14	21.3 +/- 2.14	25.5 +/- 2.14
pH	6.63 +/- 0.283	7.2 +/- 0.283	7.77 +/- 0.283	8.33 +/- 0.283	8.9 +/- 0.283
altitude	0 +/- 87.8	176 +/- 87.8	351 +/- 87.8	526 +/- 87.8	702 +/- 87.8
O2.dissous	6.05 +/- 0.744	7.54 +/- 0.744	9.02 +/- 0.744	10.5 +/- 0.744	12 +/- 0.744
MES	1 +/- 3.88	8.75 +/- 3.88	16.5 +/- 3.88	24.2 +/- 3.88	32 +/- 3.88
NO3-	0.1 +/- 4.26	8.63 +/- 4.26	17.2 +/- 4.26	25.7 +/- 4.26	34.2 +/- 4.26
DBO5	0.5 +/- 0.394	1.29 +/- 0.394	2.07 +/- 0.394	2.86 +/- 0.394	3.65 +/- 0.394
NO2-	0.01 +/- 0.02	0.05 +/- 0.02	0.09 +/- 0.02	0.13 +/- 0.02	0.17 +/- 0.02
Orthophosp	0.01 +/- 0.0444	0.0987 +/- 0.0444	0.187 +/- 0.0444	0.276 +/- 0.0444	0.365 +/- 0.0444
NH4+	0.01 +/- 0.0144	0.0388 +/- 0.0144	0.0675 +/- 0.0144	0.0963 +/- 0.0144	0.125 +/- 0.0144
P.total	0.01 +/- 0.0256	0.0613 +/- 0.0256	0.113 +/- 0.0256	0.164 +/- 0.0256	0.215 +/- 0.0256
NK	0.51 +/- 0.0987	0.707 +/- 0.0987	0.905 +/- 0.0987	1.1 +/- 0.0987	1.3 +/- 0.0987
AlkPred	0 +/- 4.15	8.31 +/- 4.15	16.6 +/- 4.15	24.9 +/- 4.15	33.2 +/- 4.15
CondPred	14 +/- 132	278 +/- 132	542 +/- 132	806 +/- 132	1070 +/- 132
C.Orga	0.2 +/- 1.05	2.3 +/- 1.05	4.4 +/- 1.05	6.5 +/- 1.05	8.6 +/- 1.05
racine	0 +/- 3.45	6.9 +/- 3.45	13.8 +/- 3.45	NA	NA

## Oligo-mesotrophes

parametres	Classe_1	Classe_2	Classe_3	Classe_4	Classe_5
Temp.eau	8.4 +/- 2.14	12.7 +/- 2.14	17 +/- 2.14	21.3 +/- 2.14	25.5 +/- 2.14
pH	6.63 +/- 0.283	7.2 +/- 0.283	7.77 +/- 0.283	8.33 +/- 0.283	8.9 +/- 0.283
altitude	0 +/- 87.8	176 +/- 87.8	351 +/- 87.8	526 +/- 87.8	702 +/- 87.8
O2.dissous	6.05 +/- 0.744	7.54 +/- 0.744	9.02 +/- 0.744	10.5 +/- 0.744	12 +/- 0.744
MES	1 +/- 3.88	8.75 +/- 3.88	16.5 +/- 3.88	24.2 +/- 3.88	32 +/- 3.88
NO3-	0.1 +/- 4.26	8.63 +/- 4.26	17.2 +/- 4.26	25.7 +/- 4.26	34.2 +/- 4.26
DBO5	0.5 +/- 0.394	1.29 +/- 0.394	2.07 +/- 0.394	2.86 +/- 0.394	3.65 +/- 0.394
NO2-	0.01 +/- 0.02	0.05 +/- 0.02	0.09 +/- 0.02	0.13 +/- 0.02	0.17 +/- 0.02
Orthophosp	0.01 +/- 0.0444	0.0987 +/- 0.0444	0.187 +/- 0.0444	0.276 +/- 0.0444	0.365 +/- 0.0444
NH4+	0.01 +/- 0.0144	0.0388 +/- 0.0144	0.0675 +/- 0.0144	0.0963 +/- 0.0144	0.125 +/- 0.0144
P.total	0.01 +/- 0.0256	0.0613 +/- 0.0256	0.113 +/- 0.0256	0.164 +/- 0.0256	0.215 +/- 0.0256
NK	0.51 +/- 0.0987	0.707 +/- 0.0987	0.905 +/- 0.0987	1.1 +/- 0.0987	1.3 +/- 0.0987
AlkPred	0 +/- 4.15	8.31 +/- 4.15	16.6 +/- 4.15	24.9 +/- 4.15	33.2 +/- 4.15
CondPred	14 +/- 132	278 +/- 132	542 +/- 132	806 +/- 132	1070 +/- 132
C.Orga	0.2 +/- 1.05	2.3 +/- 1.05	4.4 +/- 1.05	6.5 +/- 1.05	8.6 +/- 1.05
racine	0 +/- 3.45	6.9 +/- 3.45	13.8 +/- 3.45	NA	NA

## Mesotrophes

parametres	Classe_1	Classe_2	Classe_3	Classe_4	Classe_5
Temp.eau	8.4 +/- 2.14	12.7 +/- 2.14	17 +/- 2.14	21.3 +/- 2.14	25.5 +/- 2.14
pH	6.63 +/- 0.283	7.2 +/- 0.283	7.77 +/- 0.283	8.33 +/- 0.283	8.9 +/- 0.283
altitude	0 +/- 87.8	176 +/- 87.8	351 +/- 87.8	526 +/- 87.8	702 +/- 87.8
O2.dissous	6.05 +/- 0.744	7.54 +/- 0.744	9.02 +/- 0.744	10.5 +/- 0.744	12 +/- 0.744
MES	1 +/- 3.88	8.75 +/- 3.88	16.5 +/- 3.88	24.2 +/- 3.88	32 +/- 3.88
NO3-	0.1 +/- 4.26	8.63 +/- 4.26	17.2 +/- 4.26	25.7 +/- 4.26	34.2 +/- 4.26
DBO5	0.5 +/- 0.394	1.29 +/- 0.394	2.07 +/- 0.394	2.86 +/- 0.394	3.65 +/- 0.394
NO2-	0.01 +/- 0.02	0.05 +/- 0.02	0.09 +/- 0.02	0.13 +/- 0.02	0.17 +/- 0.02
Orthophosp	0.01 +/- 0.0444	0.0987 +/- 0.0444	0.187 +/- 0.0444	0.276 +/- 0.0444	0.365 +/- 0.0444
NH4+	0.01 +/- 0.0144	0.0388 +/- 0.0144	0.0675 +/- 0.0144	0.0963 +/- 0.0144	0.125 +/- 0.0144
P.total	0.01 +/- 0.0256	0.0613 +/- 0.0256	0.113 +/- 0.0256	0.164 +/- 0.0256	0.215 +/- 0.0256
NK	0.51 +/- 0.0987	0.707 +/- 0.0987	0.905 +/- 0.0987	1.1 +/- 0.0987	1.3 +/- 0.0987
AlkPred	0 +/- 4.15	8.31 +/- 4.15	16.6 +/- 4.15	24.9 +/- 4.15	33.2 +/- 4.15
CondPred	14 +/- 132	278 +/- 132	542 +/- 132	806 +/- 132	1070 +/- 132
C.Orga	0.2 +/- 1.05	2.3 +/- 1.05	4.4 +/- 1.05	6.5 +/- 1.05	8.6 +/- 1.05
racine	0 +/- 5.47	10.9 +/- 5.47	21.9 +/- 5.47	NA	NA

## Meso-eutrophes

parametres	Classe_1	Classe_2	Classe_3	Classe_4	Classe_5
Temp.eau	8.4 +/- 2.14	12.7 +/- 2.14	17 +/- 2.14	21.3 +/- 2.14	25.5 +/- 2.14
pH	6.63 +/- 0.283	7.2 +/- 0.283	7.77 +/- 0.283	8.33 +/- 0.283	8.9 +/- 0.283
altitude	0 +/- 87.8	176 +/- 87.8	351 +/- 87.8	526 +/- 87.8	702 +/- 87.8
O2.dissous	6.05 +/- 0.744	7.54 +/- 0.744	9.02 +/- 0.744	10.5 +/- 0.744	12 +/- 0.744
MES	1 +/- 3.88	8.75 +/- 3.88	16.5 +/- 3.88	24.2 +/- 3.88	32 +/- 3.88
NO3-	0.1 +/- 4.26	8.63 +/- 4.26	17.2 +/- 4.26	25.7 +/- 4.26	34.2 +/- 4.26
DBO5	0.5 +/- 0.394	1.29 +/- 0.394	2.07 +/- 0.394	2.86 +/- 0.394	3.65 +/- 0.394
NO2-	0.01 +/- 0.02	0.05 +/- 0.02	0.09 +/- 0.02	0.13 +/- 0.02	0.17 +/- 0.02
Orthophosp	0.01 +/- 0.0444	0.0987 +/- 0.0444	0.187 +/- 0.0444	0.276 +/- 0.0444	0.365 +/- 0.0444
NH4+	0.01 +/- 0.0144	0.0388 +/- 0.0144	0.0675 +/- 0.0144	0.0963 +/- 0.0144	0.125 +/- 0.0144
P.total	0.01 +/- 0.0256	0.0613 +/- 0.0256	0.113 +/- 0.0256	0.164 +/- 0.0256	0.215 +/- 0.0256
NK	0.51 +/- 0.0987	0.707 +/- 0.0987	0.905 +/- 0.0987	1.1 +/- 0.0987	1.3 +/- 0.0987
AlkPred	0 +/- 4.15	8.31 +/- 4.15	16.6 +/- 4.15	24.9 +/- 4.15	33.2 +/- 4.15
CondPred	14 +/- 132	278 +/- 132	542 +/- 132	806 +/- 132	1070 +/- 132
C.Orga	0.2 +/- 1.05	2.3 +/- 1.05	4.4 +/- 1.05	6.5 +/- 1.05	8.6 +/- 1.05
racine	0 +/- 7.29	14.6 +/- 7.29	29.2 +/- 7.29	NA	NA

## Eutrophes

parametres	Classe_1	Classe_2	Classe_3	Classe_4	Classe_5
Temp.eau	8.4 +/- 2.14	12.7 +/- 2.14	17 +/- 2.14	21.3 +/- 2.14	25.5 +/- 2.14
pH	6.63 +/- 0.283	7.2 +/- 0.283	7.77 +/- 0.283	8.33 +/- 0.283	8.9 +/- 0.283
altitude	0 +/- 87.8	176 +/- 87.8	351 +/- 87.8	526 +/- 87.8	702 +/- 87.8
O2.dissous	6.05 +/- 0.744	7.54 +/- 0.744	9.02 +/- 0.744	10.5 +/- 0.744	12 +/- 0.744
MES	1 +/- 3.88	8.75 +/- 3.88	16.5 +/- 3.88	24.2 +/- 3.88	32 +/- 3.88
NO3-	0.1 +/- 4.26	8.63 +/- 4.26	17.2 +/- 4.26	25.7 +/- 4.26	34.2 +/- 4.26
DBO5	0.5 +/- 0.394	1.29 +/- 0.394	2.07 +/- 0.394	2.86 +/- 0.394	3.65 +/- 0.394
NO2-	0.01 +/- 0.02	0.05 +/- 0.02	0.09 +/- 0.02	0.13 +/- 0.02	0.17 +/- 0.02
Orthophosp	0.01 +/- 0.0444	0.0987 +/- 0.0444	0.187 +/- 0.0444	0.276 +/- 0.0444	0.365 +/- 0.0444
NH4+	0.01 +/- 0.0144	0.0388 +/- 0.0144	0.0675 +/- 0.0144	0.0963 +/- 0.0144	0.125 +/- 0.0144
P.total	0.01 +/- 0.0256	0.0613 +/- 0.0256	0.113 +/- 0.0256	0.164 +/- 0.0256	0.215 +/- 0.0256
NK	0.51 +/- 0.0987	0.707 +/- 0.0987	0.905 +/- 0.0987	1.1 +/- 0.0987	1.3 +/- 0.0987
AlkPred	0 +/- 4.15	8.31 +/- 4.15	16.6 +/- 4.15	24.9 +/- 4.15	33.2 +/- 4.15
CondPred	14 +/- 132	278 +/- 132	542 +/- 132	806 +/- 132	1070 +/- 132
C.Orga	0.2 +/- 1.05	2.3 +/- 1.05	4.4 +/- 1.05	6.5 +/- 1.05	8.6 +/- 1.05
racine	20 +/- 17.1	54.3 +/- 17.1	88.6 +/- 17.1	NA	NA

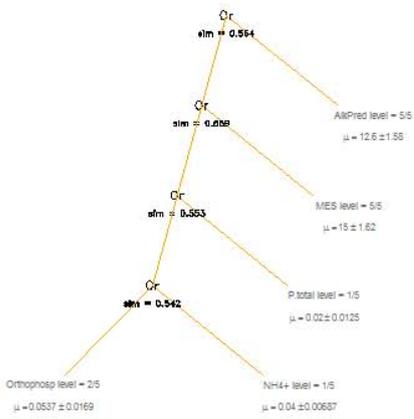
## Hypereutrophes

parametres	Classe_1	Classe_2	Classe_3	Classe_4	Classe_5
Temp.eau	8.4 +/- 2.14	12.7 +/- 2.14	17 +/- 2.14	21.3 +/- 2.14	25.5 +/- 2.14
pH	6.63 +/- 0.283	7.2 +/- 0.283	7.77 +/- 0.283	8.33 +/- 0.283	8.9 +/- 0.283
altitude	0 +/- 87.8	176 +/- 87.8	351 +/- 87.8	526 +/- 87.8	702 +/- 87.8
O2.dissous	6.05 +/- 0.744	7.54 +/- 0.744	9.02 +/- 0.744	10.5 +/- 0.744	12 +/- 0.744
MES	1 +/- 3.88	8.75 +/- 3.88	16.5 +/- 3.88	24.2 +/- 3.88	32 +/- 3.88
NO3-	0.1 +/- 4.26	8.63 +/- 4.26	17.2 +/- 4.26	25.7 +/- 4.26	34.2 +/- 4.26
DBO5	0.5 +/- 0.394	1.29 +/- 0.394	2.07 +/- 0.394	2.86 +/- 0.394	3.65 +/- 0.394
NO2-	0.01 +/- 0.02	0.05 +/- 0.02	0.09 +/- 0.02	0.13 +/- 0.02	0.17 +/- 0.02
Orthophosp	0.01 +/- 0.0444	0.0987 +/- 0.0444	0.187 +/- 0.0444	0.276 +/- 0.0444	0.365 +/- 0.0444
NH4+	0.01 +/- 0.0144	0.0388 +/- 0.0144	0.0675 +/- 0.0144	0.0963 +/- 0.0144	0.125 +/- 0.0144
P.total	0.01 +/- 0.0256	0.0613 +/- 0.0256	0.113 +/- 0.0256	0.164 +/- 0.0256	0.215 +/- 0.0256
NK	0.51 +/- 0.0987	0.707 +/- 0.0987	0.905 +/- 0.0987	1.1 +/- 0.0987	1.3 +/- 0.0987
AlkPred	0 +/- 4.15	8.31 +/- 4.15	16.6 +/- 4.15	24.9 +/- 4.15	33.2 +/- 4.15
CondPred	14 +/- 132	278 +/- 132	542 +/- 132	806 +/- 132	1070 +/- 132
C.Orga	0.2 +/- 1.05	2.3 +/- 1.05	4.4 +/- 1.05	6.5 +/- 1.05	8.6 +/- 1.05
racine	0 +/- 3.26	6.52 +/- 3.26	13 +/- 3.26	NA	NA

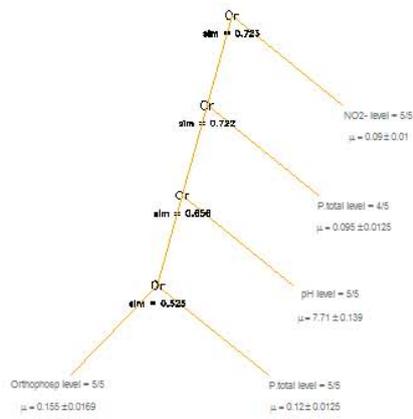
## **ANNEXE 2**

# REGION 1

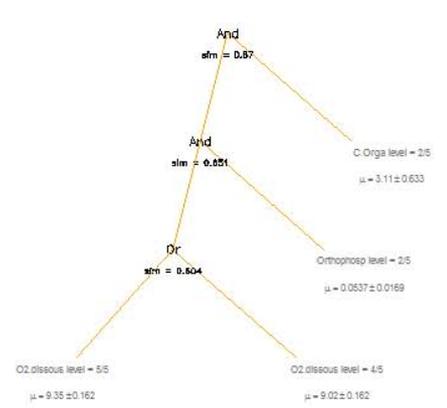
root = TB.r1.



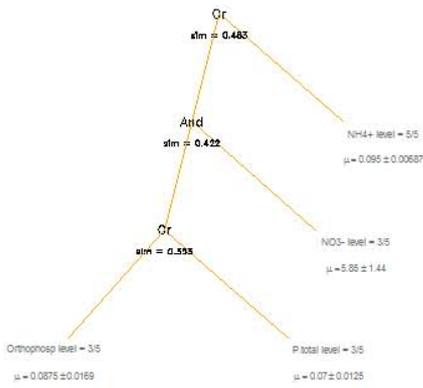
root = Moy.r2.



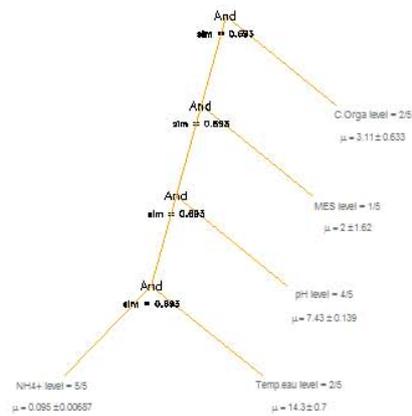
root = B.r3.



root = Med.r4.

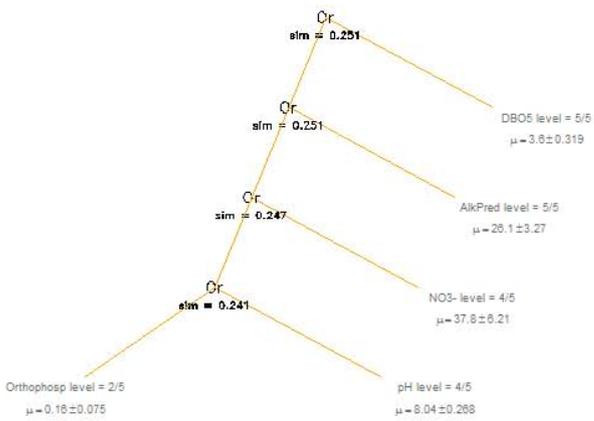


root = Mauv.r5.

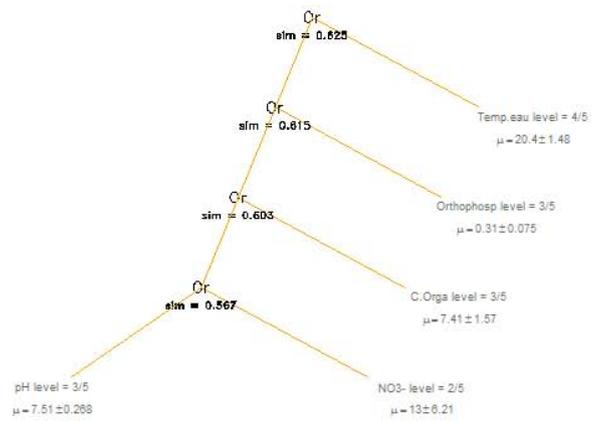


## REGION 2

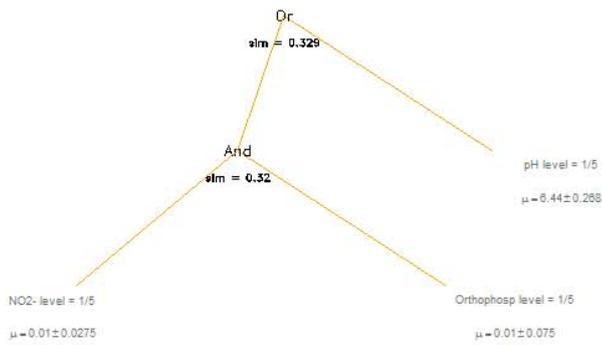
*root = B.r1.*



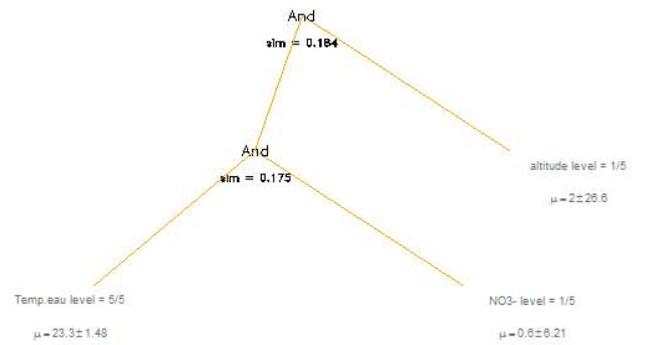
*root = Moy.r2.*



*root = TB.r3.*

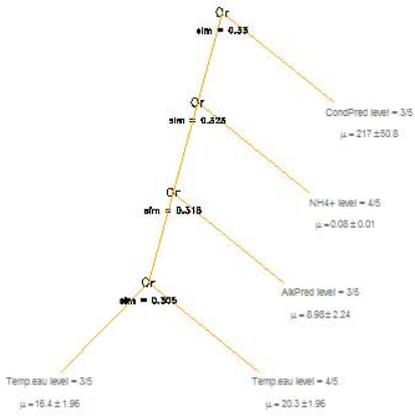


*root = Med.r4.*

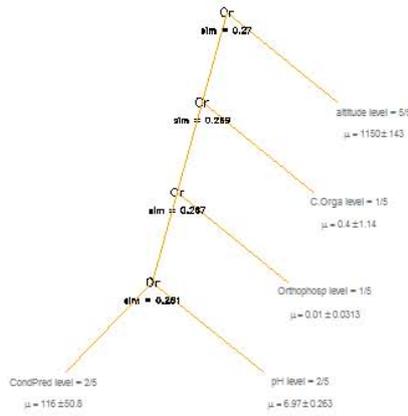


# REGION 3

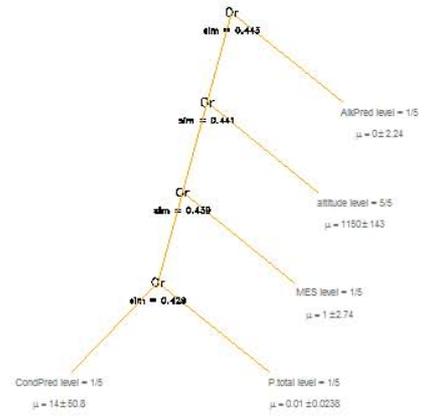
root = Moy.r1.



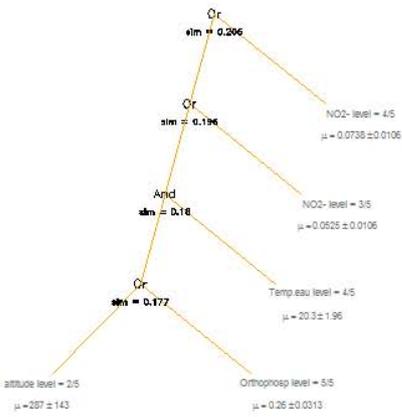
root = B.r2.



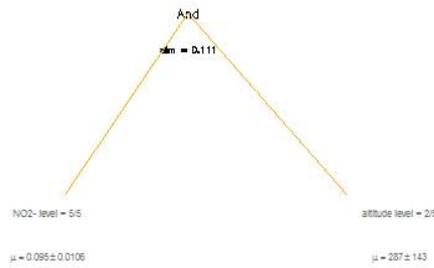
root = TB.r3.



root = Med.r4.

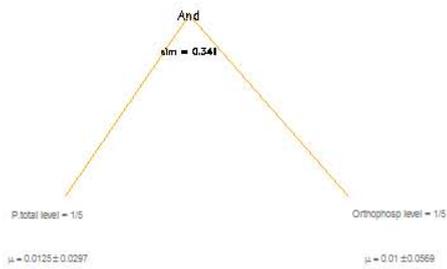


root = Mauv.r5.

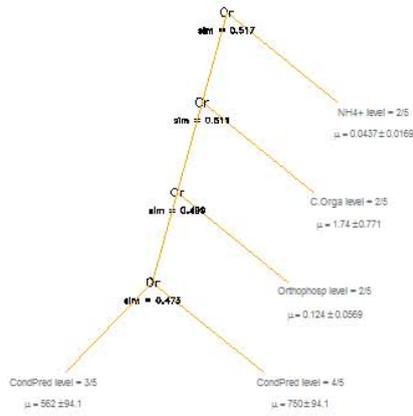


# REGION 4

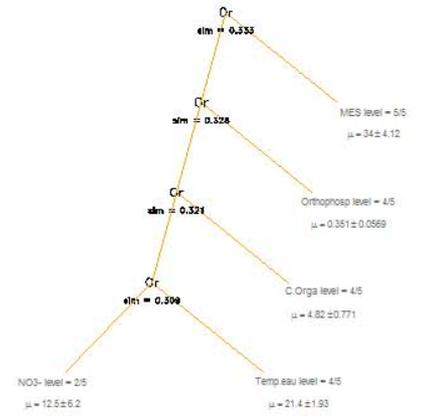
root = TB.r1.



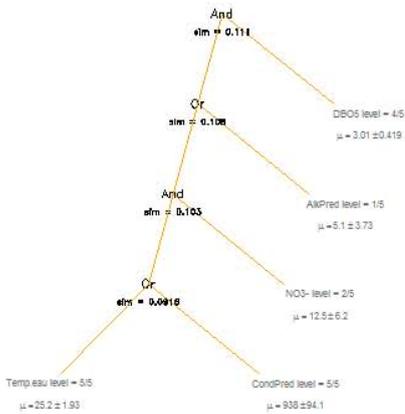
root = B.r2.



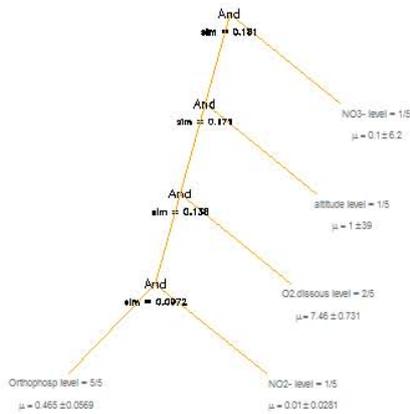
root = Moy.r3.



root = Med.r4.

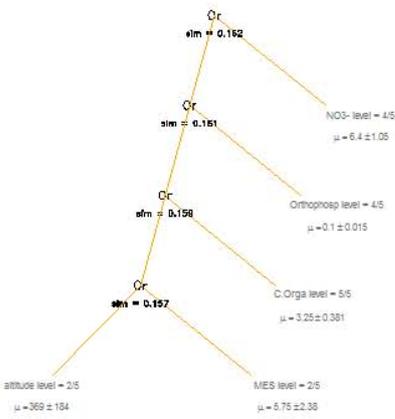


root = Mauv.r5.

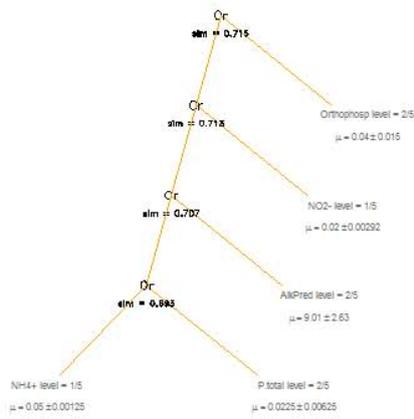


# REGION 5

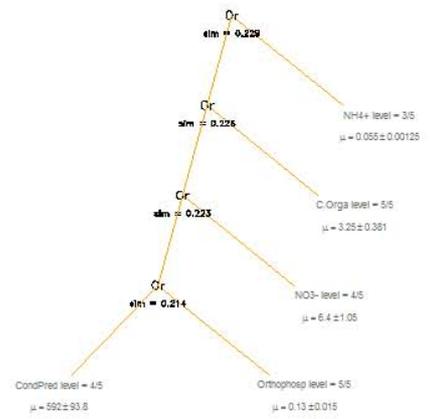
root = B.r1.



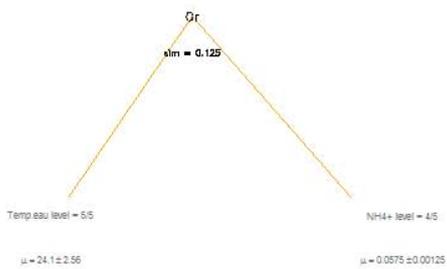
root = TB.r2.



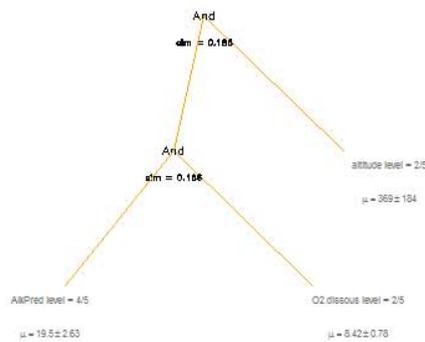
root = Moy.r3.



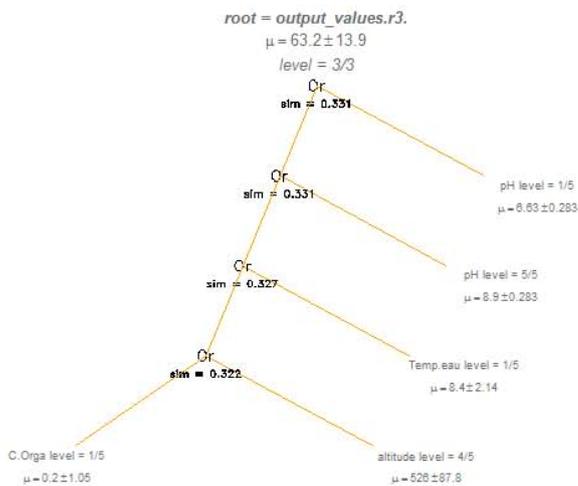
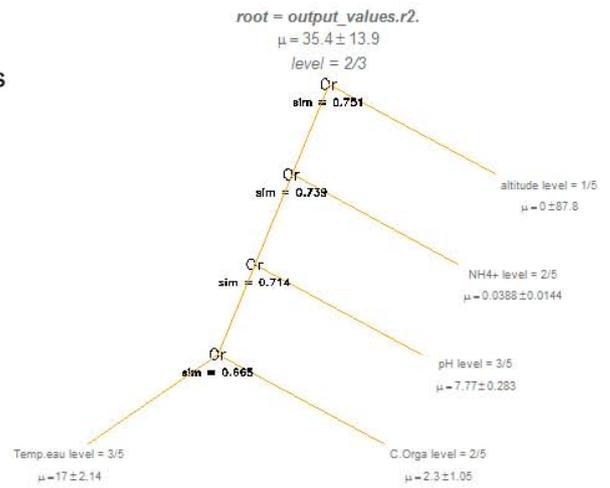
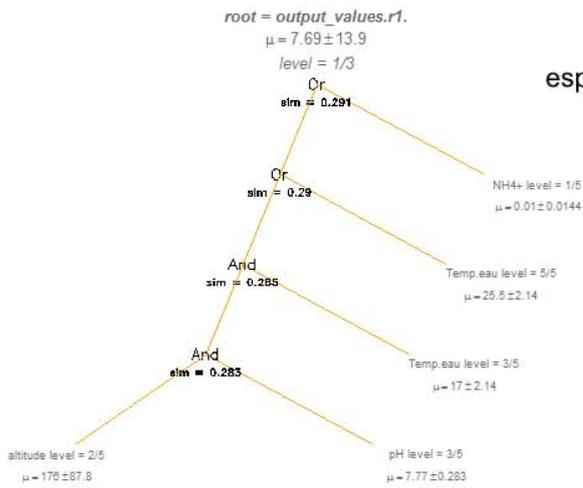
root = Mauv.r4.



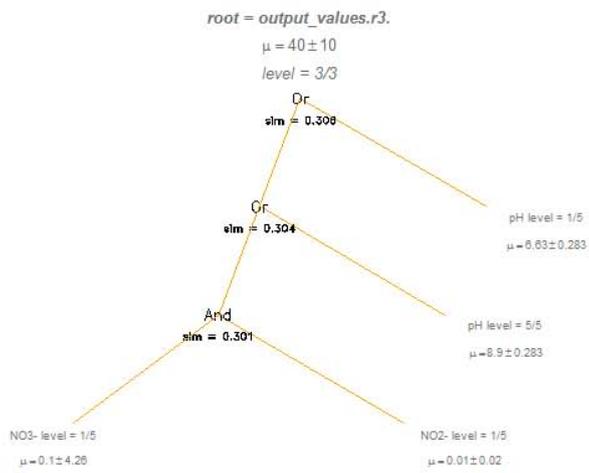
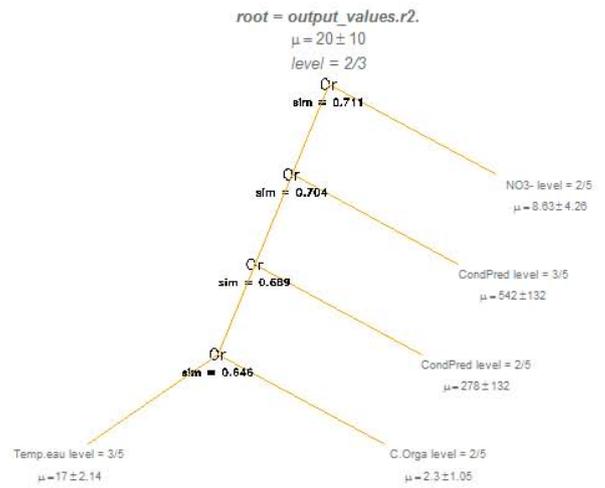
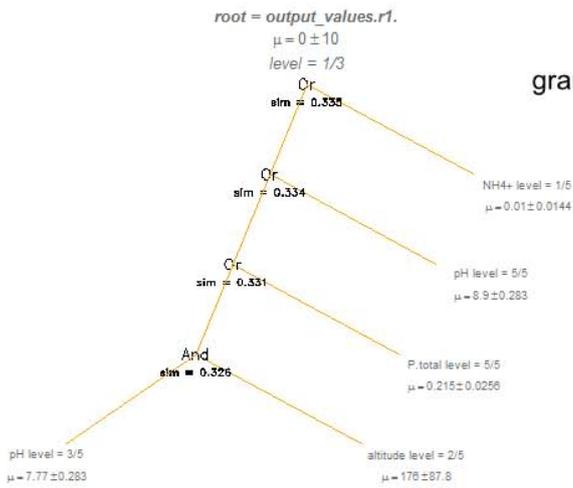
root = Med.r5.



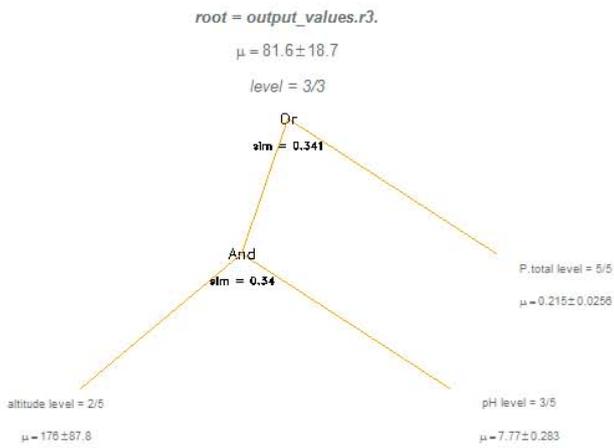
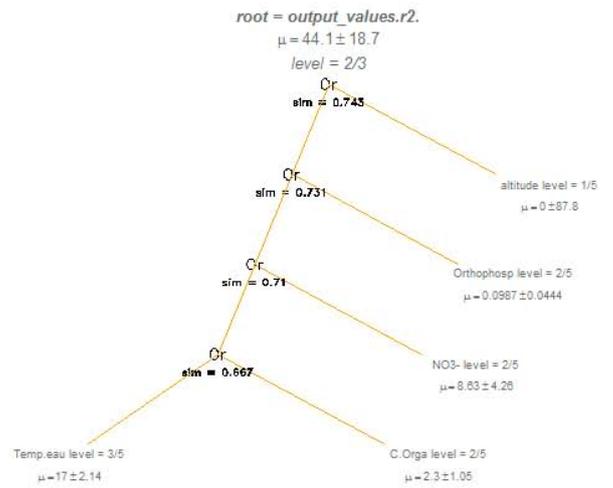
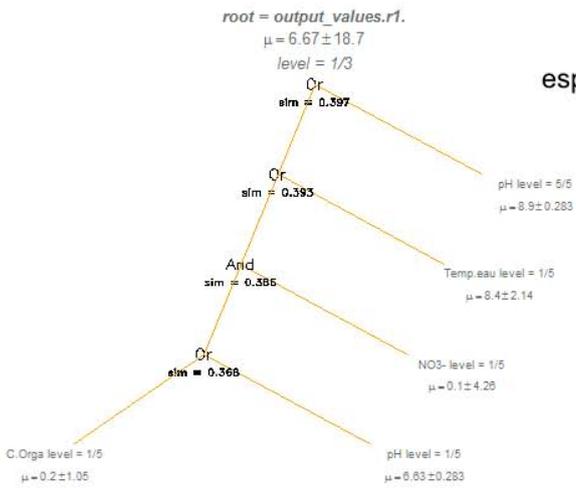
## Gilde, espèces adhérentes



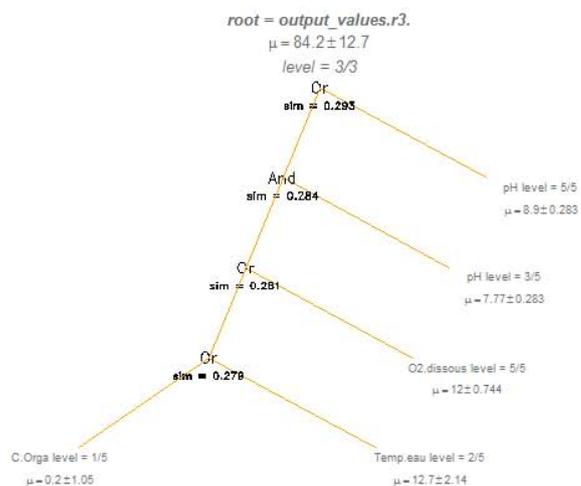
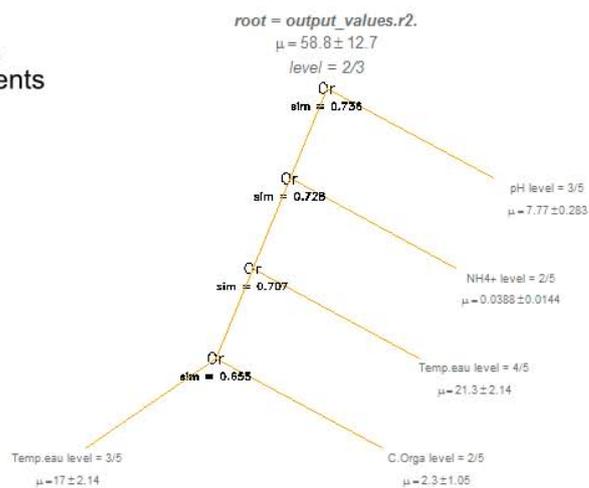
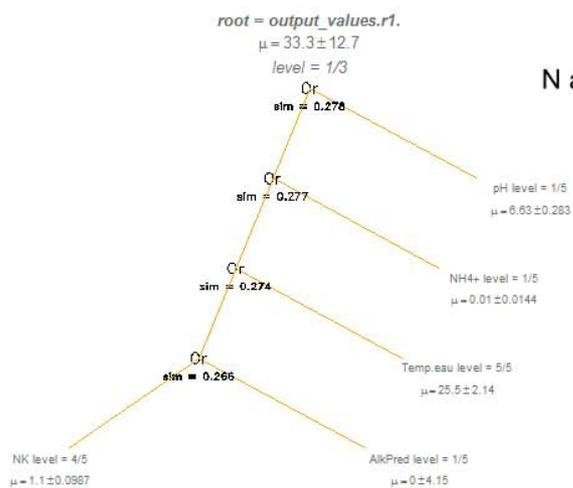
## Gilde, grandes espèces



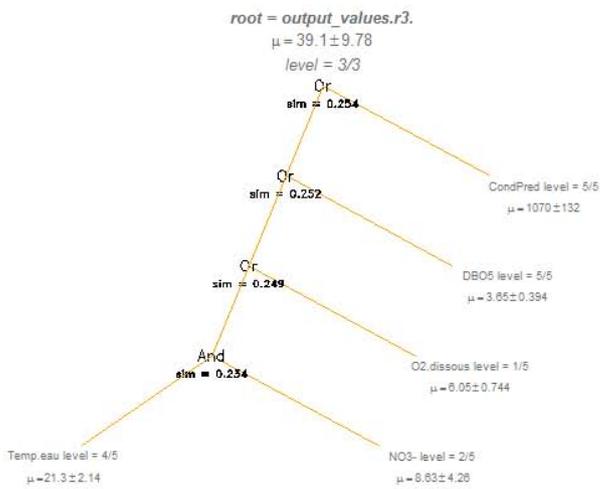
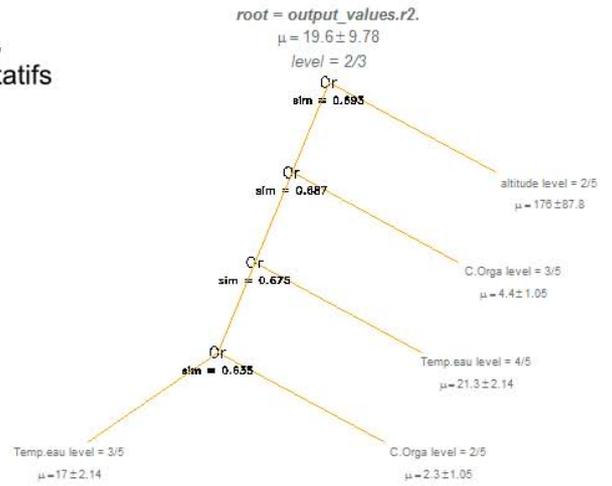
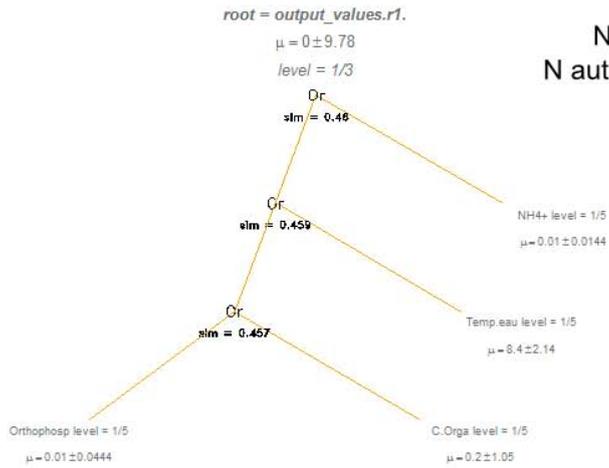
## Gilde, espèces mobiles



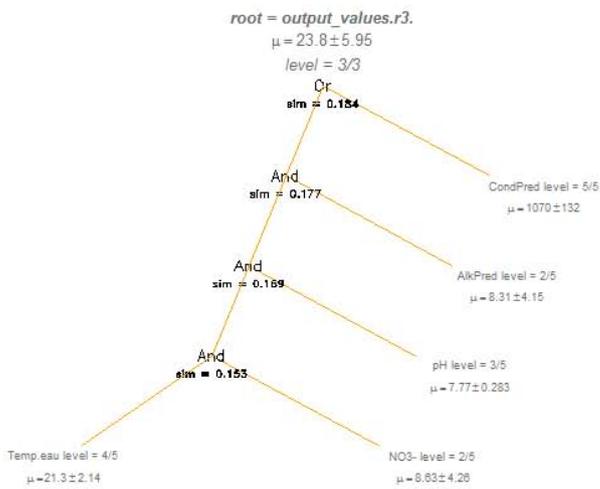
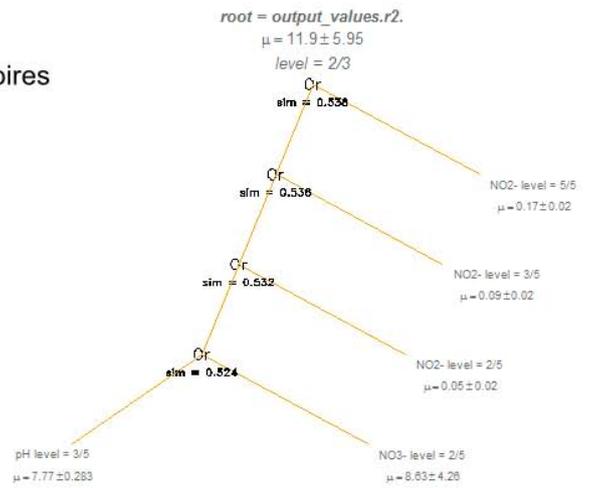
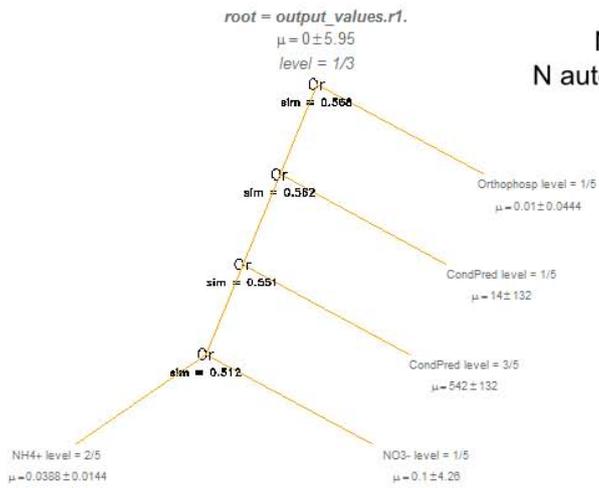
## N hétérotrophie, N autotrophes tolérants



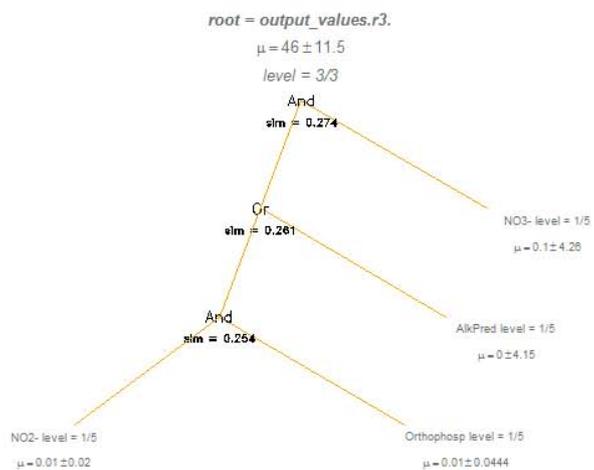
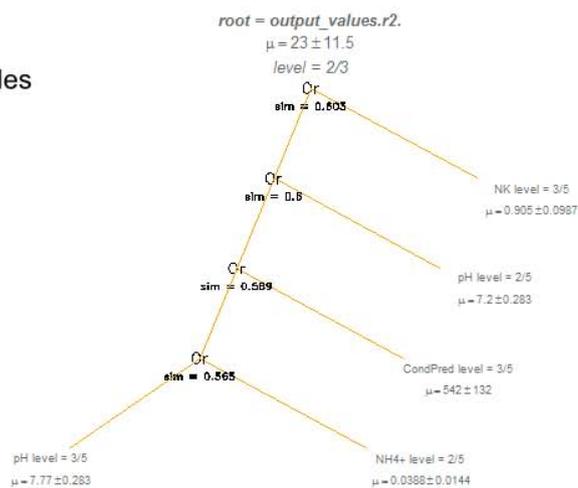
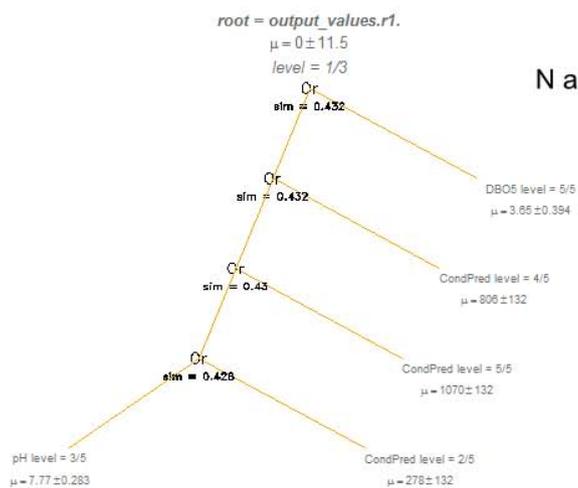
## N hétérotrophie, N autotrophes facultatifs



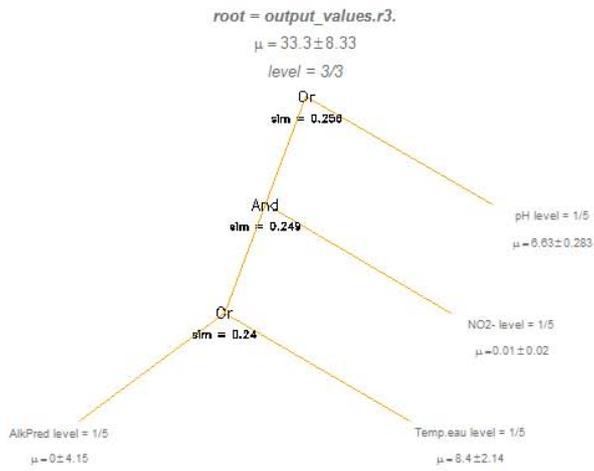
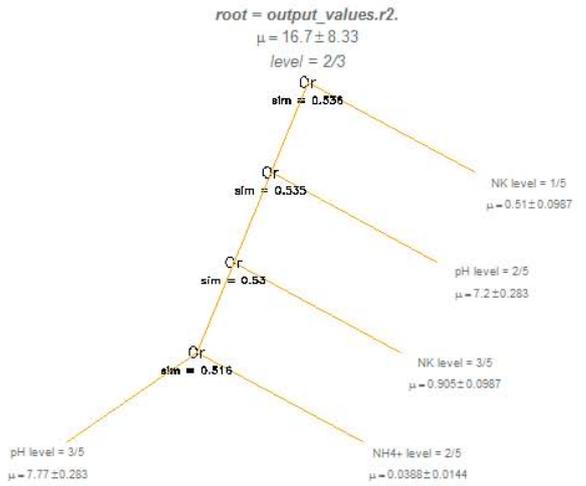
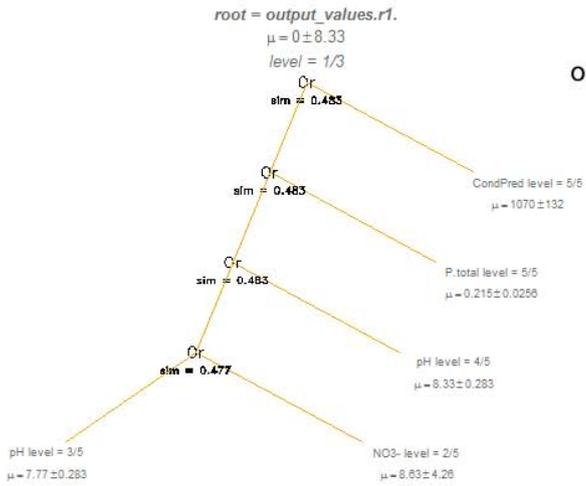
## N hétérotrophie, N autotrophes obligatoires



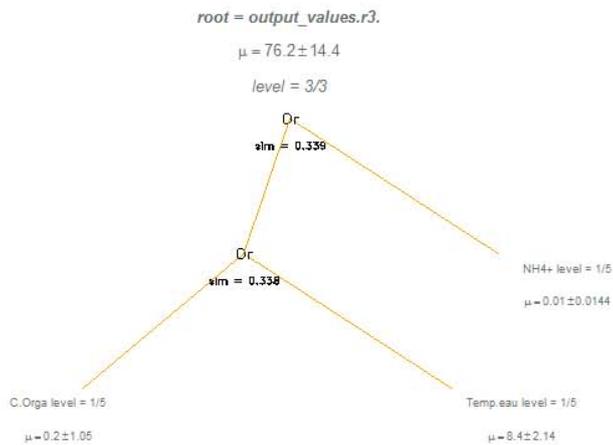
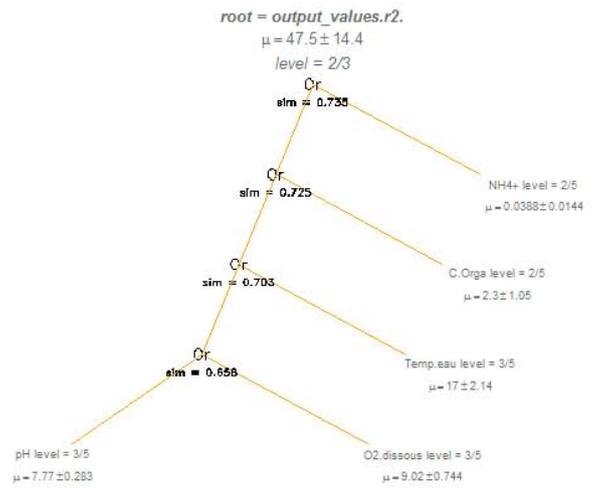
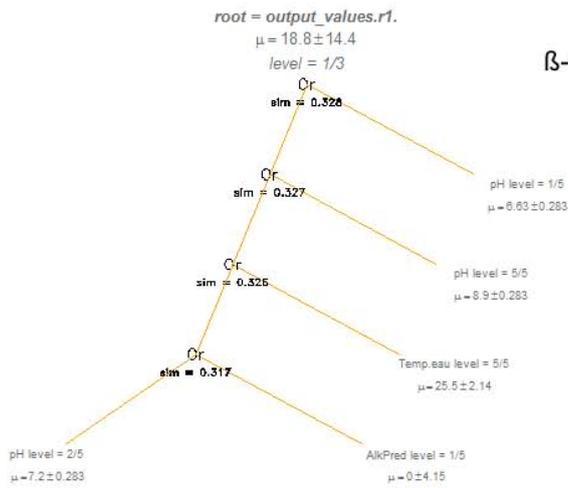
N hétérotrophie,  
N autotrophes sensibles



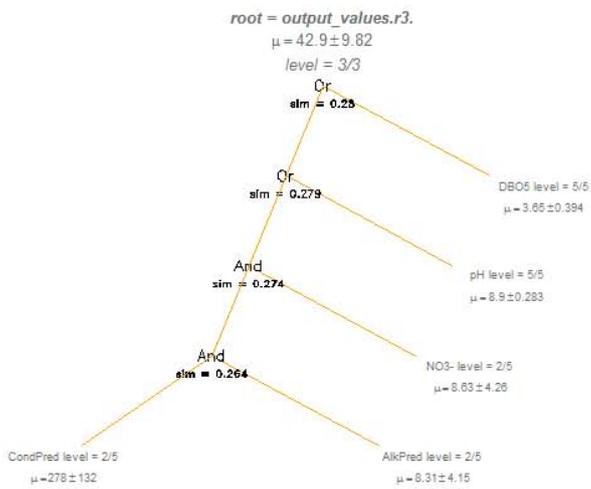
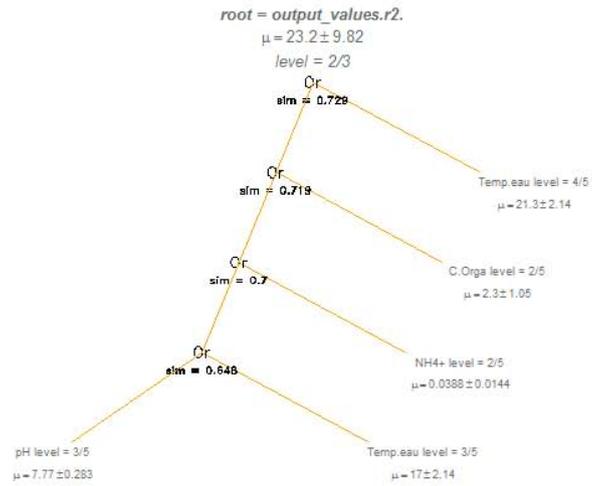
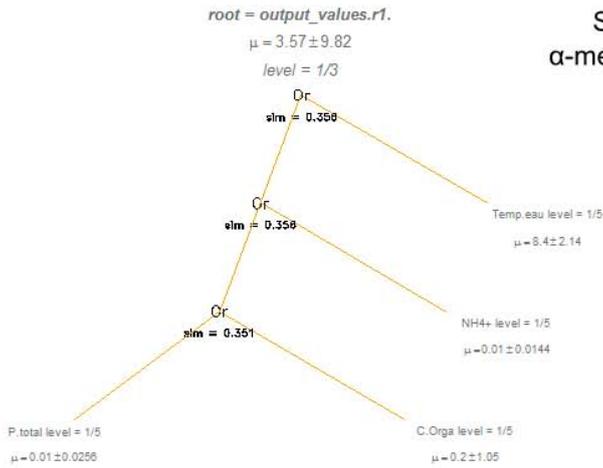
Saprobie, oligosaprobies



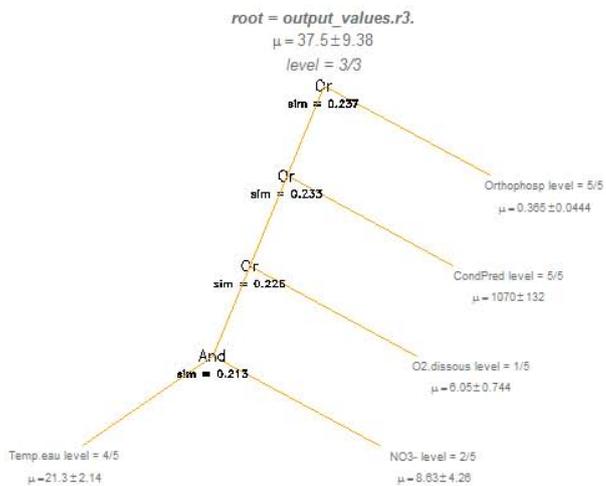
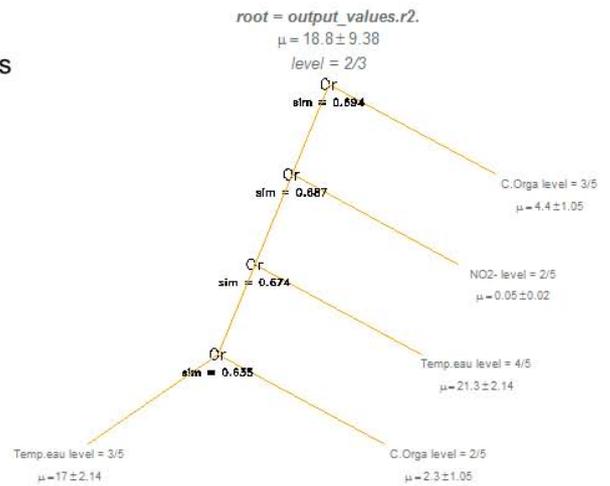
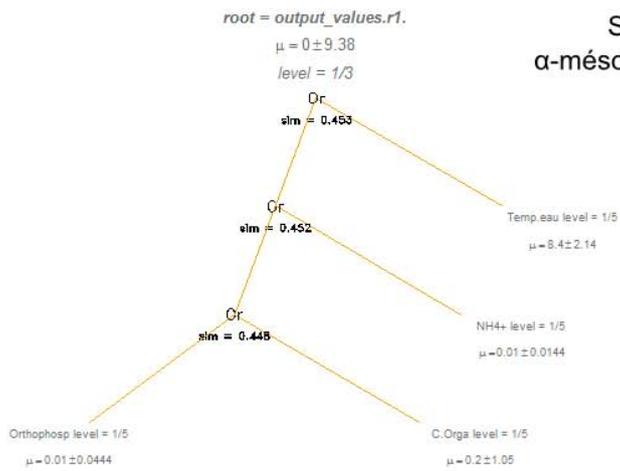
## Saprobie, β-mésosaprobies



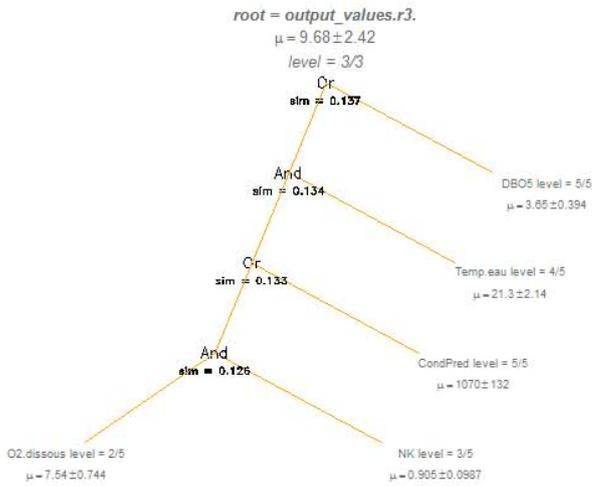
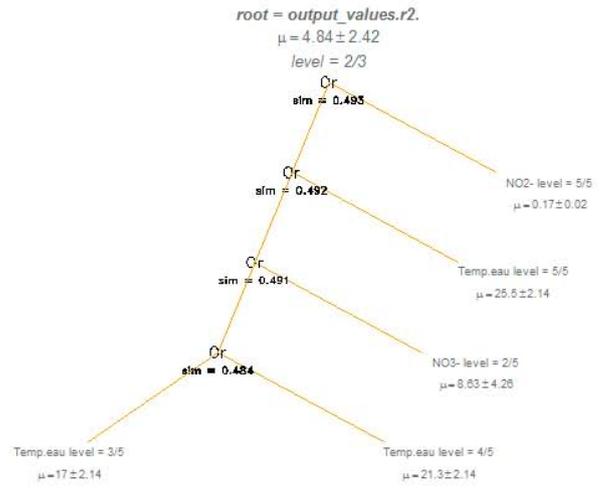
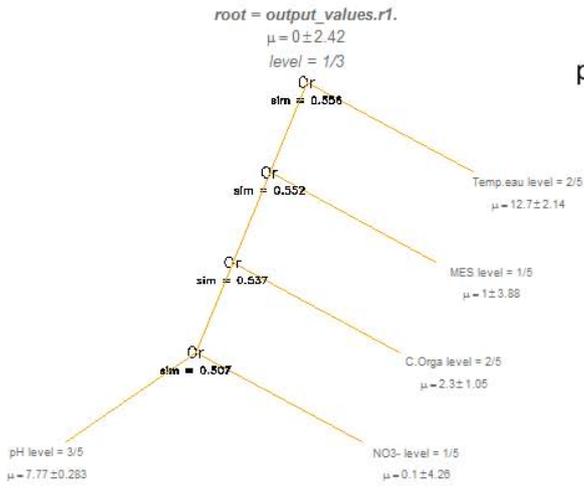
## Saprobie, α-mésosaprobies



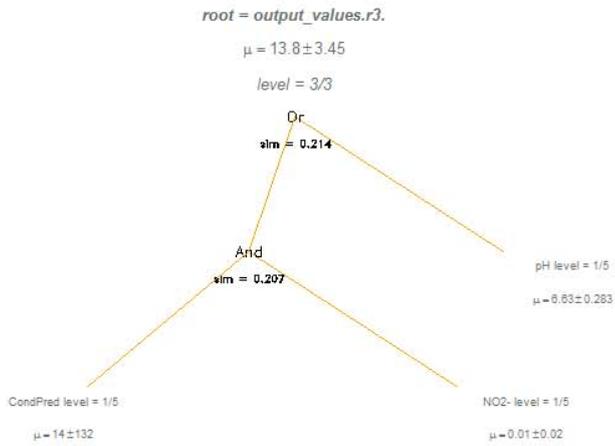
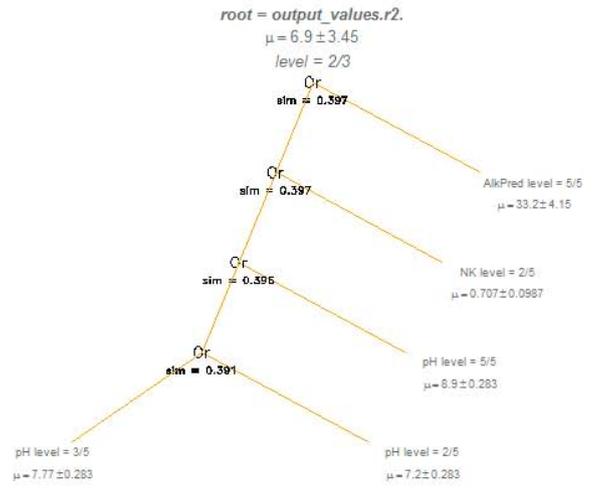
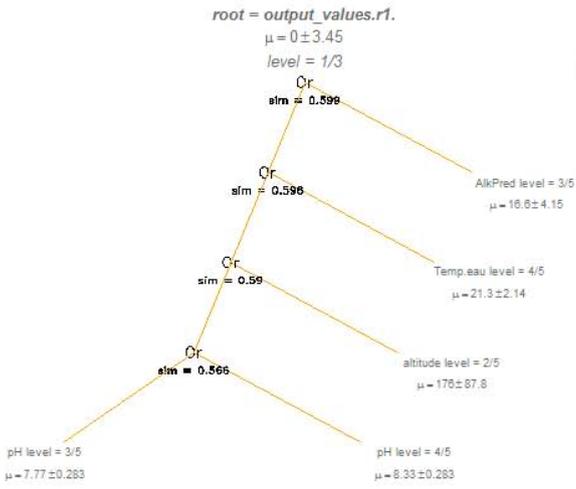
## Saprobie, α-méso-polysaprobies



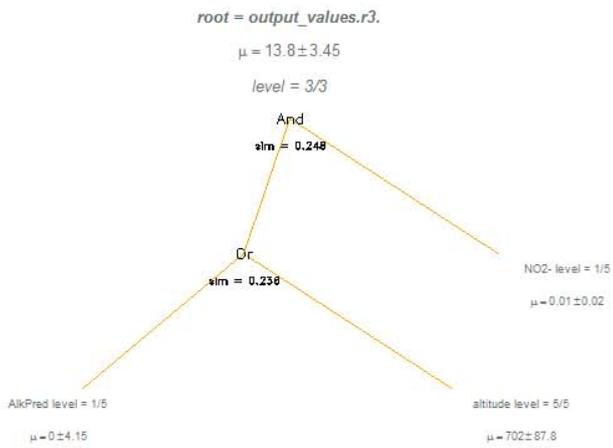
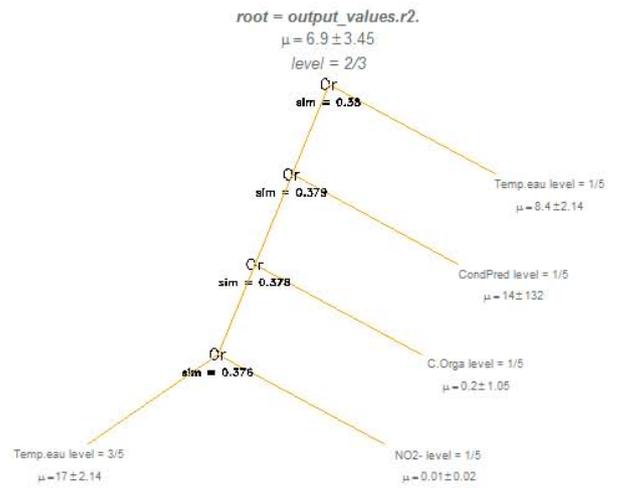
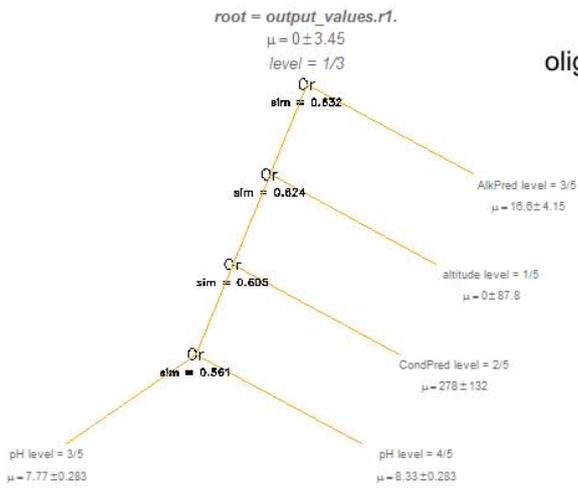
## Saprobie, polysaprobies



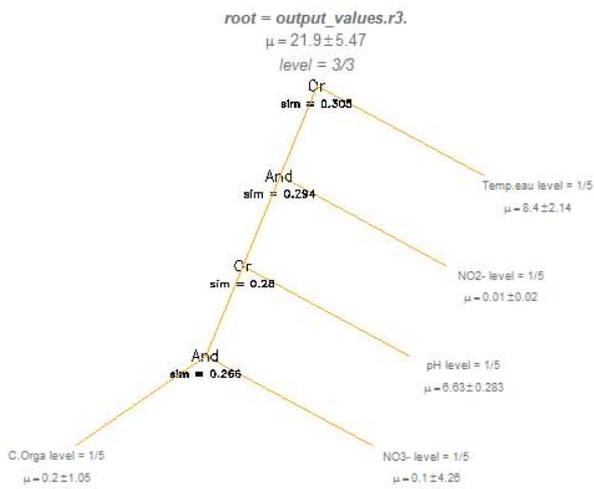
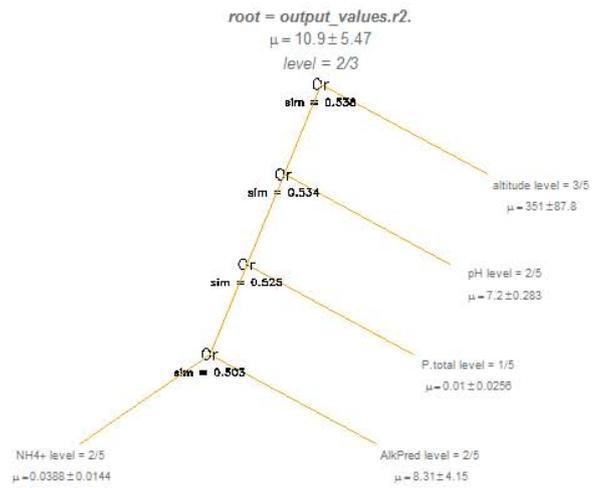
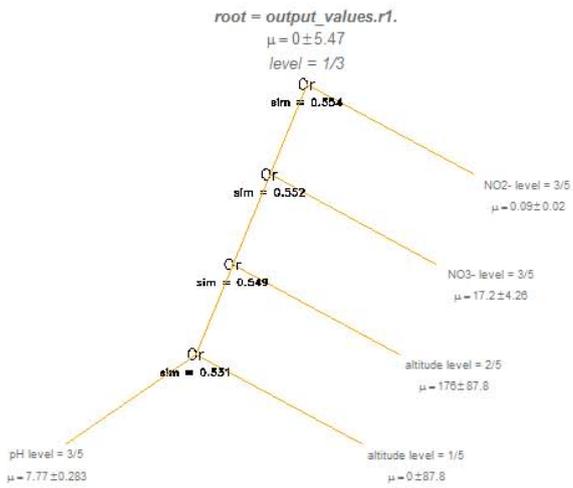
## Trophie, oligotrophes



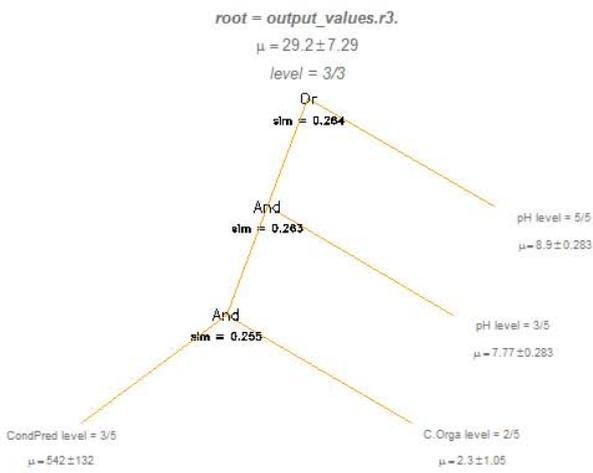
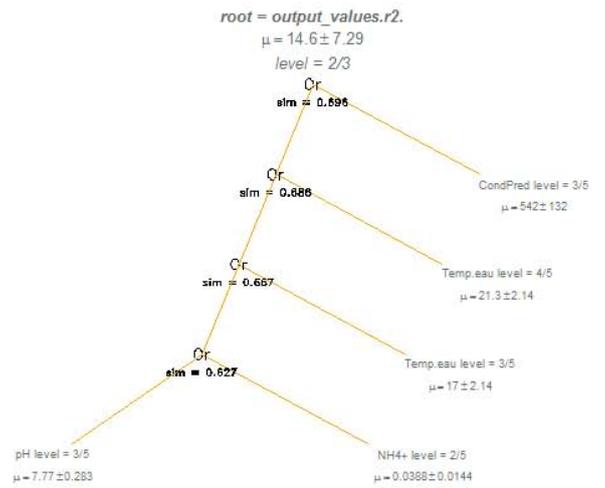
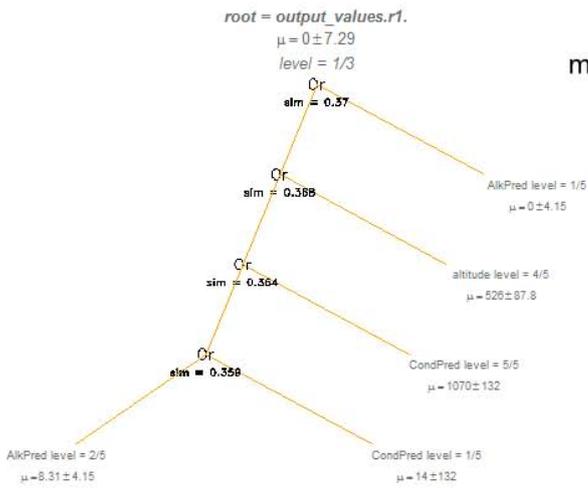
## Trophie, oligo-mésotrophes



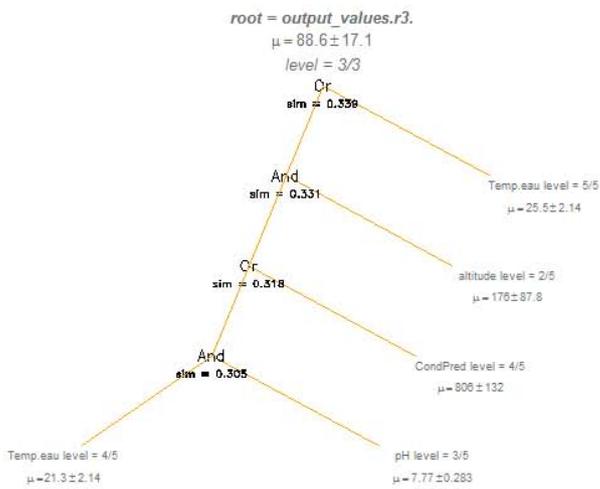
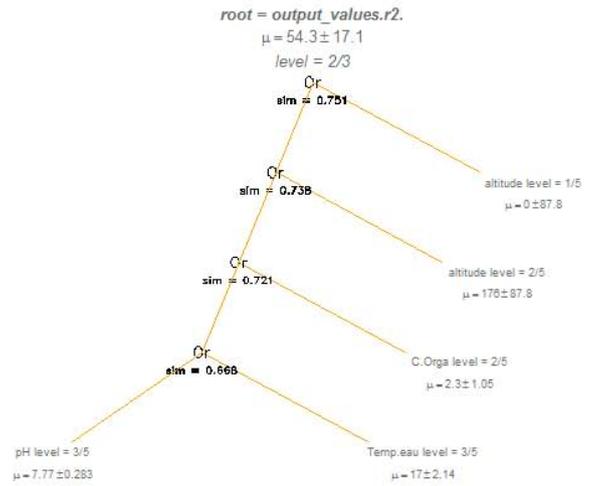
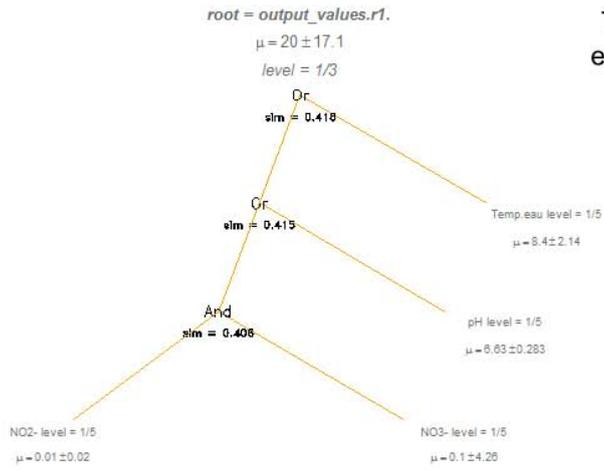
## Trophie, mésotrophes



## Trophie, méso-eutrophes



## Trophie, eutrophes



## Trophie, hyper-eutrophes

