



HAL
open science

Structuration spatiale des communautés aquatiques à large échelle : influence des facteurs environnementaux dans un contexte de changement climatique

Coline Picard

► **To cite this version:**

Coline Picard. Structuration spatiale des communautés aquatiques à large échelle : influence des facteurs environnementaux dans un contexte de changement climatique. Sciences de l'environnement. Université Claude Bernard - Lyon I, 2023. Français. NNT : 2023LYO10177 . tel-04744394

HAL Id: tel-04744394

<https://theses.hal.science/tel-04744394v1>

Submitted on 18 Oct 2024

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

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



**THÈSE de DOCTORAT DE
L'UNIVERSITÉ CLAUDE BERNARD LYON 1**

**École Doctorale n°341
Évolution, Écosystème, Microbiologie, Modélisation
(E2M2)**

Discipline : Hydroécologie

Soutenue publiquement le 2 octobre 2023, par :

Coline PICARD

**Structuration spatiale des
communautés aquatiques
Influence des facteurs environnementaux
à large échelle dans un contexte de
changement climatique**

Devant le jury composé de :

GRENOUILLET Gaël

Maître de Conférences, Université Toulouse 3 Paul Sabatier

Rapporteur

USSEGLIO-POLATERA Philippe

Professeur des Universités, Université de Lorraine, Metz

Rapporteur

DOLÉDEC Sylvain

Professeur des Universités, Université Lyon 1 Claude
Bernard

Président

JELIAZKOV Aliénor

Chargée de Recherche, INRAE - HYCAR, Antony

Examinatrice

Élisa

Chargée de Recherche, CNRS - iEES, Paris

Examinatrice

MOATAR Florentina

Directrice de recherche, INRAE - RiverLy, Villeurbanne

Directrice de thèse

FLOURY Mathieu

Grade/qualité, INRAE - HYCAR, Antony

Membre invité

MAIRE Anthony

Ingénieur-chercheur, EDF R&D - LNHE, Chatou

Membre invité

POULET Nicolas

Chargé de mission, Pôle écohydraulique OFB - IMFT,
Toulouse

Membre invité

Remerciements

Vous ne le croirez peut-être pas, mais ces pages de remerciements ont été les plus difficiles à écrire de ce manuscrit. Certainement parce que ce sont les derniers mots que j'ai eu à poser sur le papier (numériquement on s'entend), et qui marquent la fin d'une longue épopée. Réaliser une thèse c'est s'engager sur un chemin très bizarre, comme la traversée du pays des merveilles, où l'on ne sait jamais quel embranchement prendre et où l'on se sent bien souvent perdue. La route est semée de travail acharné, de déceptions qui se changent en frustrations, et de longs moments de solitude ; mais c'est aussi une occasion unique de nourrir l'avidité scientifique qui nous anime, et une formidable expérience humaine. Comme Alice, j'ai dialogué avec de drôles d'individus que je ne comprenais pas, suivi les conseils d'étranges personnages, et j'ai souvent rebroussé chemin. Et c'est ainsi que j'ai tracé mon propre chemin dans le monde de la recherche. De nombreuses personnes ont donc joué un rôle particulier dans cette aventure et je souhaite, en guise de remerciements, leur dédier ces quelques pages. L'exhaustivité est souvent difficile à atteindre, j'espère donc que vous vous reconnaîtrez au travers de cet échantillon que j'ai voulu représentatif.

Pour commencer je tiens à remercier mes encadrants, qui m'ont donné l'opportunité de travailler sur ce sujet de recherche et de comprendre un peu mieux le changement climatique et ses effets. Anthony, ce n'est pas un secret, cette première expérience d'encadrement de thèse n'a pas été des plus simples, et j'espère que tu en auras tiré autant de leçons que moi. Merci de m'avoir donné la chance de commencer à explorer le sujet par un stage, et d'avoir aussi bien préparé le lancement de cette thèse. Ta présence numérique en attendant de pouvoir regagner le bureau m'a permis de rester motivée durant les confinements, et même si, concrètement, un ordinateur me suffisait pour travailler, on sait tous les deux l'impact que cette période difficile et étrange a eu sur ce travail. Mathieu, ta présence discrète ne présageait pas des excellents conseils et des connaissances que tu m'as transmis. Merci de m'avoir offert une journée de terrain pour me sortir de mes codes et me faire découvrir tes autres sujets de recherche. Tu as toujours répondu présent à mes sollicitations, malgré le fait d'avoir suivi ce travail plutôt à distance, alors merci. Florentina, le défi de diriger l'encadrement d'une thèse en écologie n'était pas des plus simples et cela a permis d'apporter le regard d'une autre discipline sur nos approximations d'écologues, très précieux en sciences selon moi. Ton expérience dans l'encadrement de thèse a permis de me rassurer dans les moments où c'était nécessaire.

Je tiens à remercier tous les membres qui composent le jury de thèse pour avoir accepté d'évaluer ce travail et de m'avoir offert des échanges très stimulants lors de la soutenance, en présentiel malgré la distance à parcourir pour certains. En premier lieu, merci à mes deux rapporteurs, Gaël Grenouillet et Philippe Usseglio-Polatera, d'avoir accepté d'évaluer ce manuscrit et pour leurs questions avisées lors de la soutenance. J'ai également apprécié le regard de chacun de mes examinateurs, pas toujours aussi pessimiste que le mien sur certains résultats. Merci donc à Éliisa Thebault, Aliénor Jeliaskov et Sylvain Dolédec. Merci enfin à Nicolas Poulet, d'avoir fermé le bal des questions lors de la soutenance, avec

humour et sans ménagement.

Je tiens à remercier l'implication de chacun des membres du comité de suivi, Jean-Michel Olivier, Laëtita Buisson, Cédric Mondy et Pablo Tedesco, aussi bien dans les réflexions scientifiques qu'organisationnelles et leur mise en garde sur la quantité de travail à accomplir dans le temps imparti. La diversité des profils de chacun a permis un comité riche en échanges et en bons conseils.

La particularité d'une thèse CIFRE c'est de faire partie de deux équipes et, bien que cela ne soit pas toujours confortable d'être entre deux, cela m'a donné accès à un réseau de scientifiques et d'humains hors norme. P76 vous m'avez très bien accueillie et vos portes sont toujours restées ouvertes, quelle que soit la teneur des discussions. Je pense notamment aux pauses thé/chocolats dans le bureau de Laure et Véronique ; à Sylvie, qui met un point d'honneur à inclure les stagiaires et doctorants dans la vie du projet Biodiv' ; à l'équipe des sportifs (Sabine, Flo, Émilie, Pascaline, Carole, Clémentine) pour les moments de course à pied, gym et autres défouloirs ; à Julie pour ton encouragement à secouer le cocotier vestimentaire et le partage de tes recettes ; à Christine pour ton humour. Côté Lyon, mes venues à ecoflows ont été marquées par de nombreux échanges sur le développement de mes modèles, merci donc à Maxime, Jérémy, Nicolas et David pour votre temps. Mes venues ont aussi été l'occasion de conseils avisés, de témoignage de soutien, et d'invitation aux moments conviviaux.

Au-delà de mes propres équipes, de nombreux autres collègues ont été présents. À la team des *machine learners*, Fabrice, Sami, Marouanne, Pablo, Antoine et Théo, pour leurs conseils même s'ils ne connaissent pas les poissons. À toutes les personnes qui m'ont aidé, scientifiquement ou moralement : Valentin et Olivier du MFEE ; Mélanie pour ses encouragements perpétuels et sa sagesse sur la vie ; Thibaut et tes irruptions pour discuter longuement des nouveautés de l'intelligence artificielle, en me laissant de nombreux gribouillis sur mon tableau tout propre.

Je tiens à saluer l'ensemble des collègues qui ont oeuvré à la bancarisation et à la production et sans qui ce travail n'aurait pas pu voir le jour. Je pense en premier lieu aux collègues de l'onema/AFB/OFB, qui réalisent les campagnes de terrain alimentant Nāïades. Ces données uniques sont très précieuses et sont le fruit d'un énorme boulot, j'en ai bien conscience pour y avoir moi-même un peu participé. Merci aussi à ceux qui ont accepté de partager leurs jeux de données, en étant ravis que ça puisse servir : Pierre Steinbach, l'équipe de SYRAH-CE, etc. La science a besoin de ces partages !

Un travail de thèse ne pourrait avoir lieu sans le travail et les conseils avisés de confrères scientifiques. Je tiens donc à saluer les auteurs de (merveilleux) packages R, qui sortent parfois pile quand on en a besoin, et qui sont ravis de répondre à quelques (nombreux) mails même si les demandes sont bien trop particulières : K.Ganz, R.Hornung, etc. Merci aussi à ceux qui m'ont accrodé un peu de leur temps et de leur expérience pour essayer de m'aider à avancer, comme Wilfried Thuiller, François Munoz, Jean-Nicolas Beisel.

Plusieurs personnes m'ont aidé à croire en moi et à repousser mes limites professionnelles en m'incitant à poursuivre mes études et débiter ma carrière scientifique par un doctorat. Je pense notamment à Sabrina Gaba, mon encadrante de stage de M2, qui était étonnée que je considère ne pas avoir les compétences et le niveau pour m'engager sur la voie du doctorat. Je pense à Loïc Bollache, qui m'a envoyé sans relâche une multitude de sujets de thèse jusqu'à ce que je finisse par en accepter un. Et enfin, je pense également à Valérie Verneaux, qui m'avait noté le nom de Philippe Usseglio-Polatera à la naissance du projet, parce qu'une thèse sur les invertébrés doit avoir Philippe dans son jury.

Il est temps de passer au remerciements un peu plus personnels, aux personnes qui m'ont accompagnée dans cette aventure et l'ont marquée d'une pierre blanche. La team des stagiaires, promo 2019, pour les découvertes de nos sujets parfois si différents (microbio nous voilà !) : Corentin, Arthi, Franz-Arnold, et les autres. Aux amis qui m'ont convaincu que vivre à Paris ce n'est pas si terrible, et qui m'ont fait découvrir les bons côtés de la capitale : notamment Florian, Laurie, Thomas W. À ceux qui ont toujours été là, et qui ont fait le chemin pour venir le jour J : Thibault, Romain, Franz-Arnold. À tous ceux qui auraient voulu venir mais ont quand même été présents, en visio ou même en pensée. Aux jeunes collègues que je n'oublierai jamais, Mathilde, Sophie, Charles et Sarah. À ma camarade de thèse, Hanieh, qui aurait trouvé sa place dans les autres paragraphes, en tant que collègue, productrice de données, camarade de thèse, mais surtout en tant qu'amie.

Un remerciement tout particulier à Azénor, petit titi, mon pompon breton rose flashy, ma collègue de galères doctorales. Ta force de caractère et ton entêtement ont été des inspirations pour moi. Le soutien mutuel qu'on a réussi à s'apporter, même quand on était au bout du rouleau toutes les deux, a été presque magique. Te voir soutenir quelques mois avant moi m'a donné de la force pour atteindre le but ultime, et je suis ravie d'avoir pu partager cette aventure avec toi, et devenir docteurs quasiment en même temps. Je souhaite de continuer à te voir évoluer dans la vie, et grandir un peu plus chaque jour.

Sévane, tu m'as soutenue dès le début dans la poursuite de mes études. Merci d'avoir accepté de sociabiliser, et de m'avoir suivie partout, avec le sourire et sérénité (apparente on le sait bien !) quoiqu'il arrive le jour de ma soutenance. Je te dois les illustrations détournées pour les invertébrés et des conseils avisés pour la mise en page de ce manuscrit. Et en réalité je te dois bien plus : une amitié sans faille, une présence indéfectible et une relation qui m'est très précieuse. Ton art et surtout ta capacité d'adaptation malgré les coups durs m'impressionnent un peu plus chaque jour.

J'ai appris de mon père à cultiver la curiosité et l'émerveillement, ceux qui nous animent enfants pour les choses qui nous entourent ; et me concernant ceux pour les êtres dont la vie plus secrète se déroule sous la surface de nos rivières. Il m'a aussi transmis sa passion pour les conférences inspirantes d'Hubert Reeves, astrophysicien et vulgarisateur de renom, à qui je tiens à rendre hommage pour son éternel optimisme. Mon ambition et la persévérance dans le travail ont été nourris par ma mère. Ange, je n'ai pas compté les heures que tu as passées à me conseiller en programmation et à me dépanner en ins-

tallations sous Linux, en visio, l'horreur (vive l'absence de droits admin et les versions obsolètes livrées par défaut !). Notre curiosité et notre envie de d'apprendre sont des moteurs incroyables qui ont nourri nos échanges si précieux. Romane, tu as assisté d'un point de vue dijonnais à ce doctorat, et pourtant tu as été avec moi à chaque instant. Je sais que vous êtes fiers de moi et je suis encore plus fière de vous. On s'en est sortis ensemble, et nous continuerons de grandir ensemble, la fratrie c'est éternel. Votre présence, votre amour et votre fierté m'ont portée, particulièrement le jour de ma soutenance. Voir les personnes que vous devenez chaque jour m'inspire. Alors merci !

Didier, Christine, vous avez su m'apporter un soutien très précieux et me donner du réconfort quand j'en avais besoin. Je ne m'attendais pas à ce que mes beaux-parents soient aussi consciencieux dans la relecture de chacune des pages de ce manuscrit (photo à l'appui !). C'est certainement grâce à vous et votre travail que ces pages sont plus compréhensibles et dans un français correct (pour un manuscrit scientifique j'entends). L'enthousiasme dont vous avez fait preuve avec Sophie, Valentin, Odile et Alain, ainsi que tous vos proches, m'a gonflé de joie et de confiance pour la soutenance. Christine, je tiens à souligner ton organisation sans faille pour la préparation du pot de ma soutenance, pot que tout le monde a beaucoup apprécié, particulièrement le gâteau à la broche.

Et le meilleur pour la fin. Nicolas, nous avons survécu à cette thèse, ce qui me laisse persuader que nous pourrions survivre ensemble à beaucoup de choses dans la vie. Tu as reçu le premier la consécration, le 9 novembre 2020, bien tristement derrière une caméra (fichu COVID) mais avec brio comme dans tout ce que tu entreprends. Tu as ainsi pu me guider sur la voie du doctorat, en sachant que peu importe comment j'y arriverai, je terminerai cette thèse. Ton soutien, même lorsque je voulais tout arrêter, lorsque mon syndrome de l'imposteur se faisait de plus en plus menaçant, lorsque la thèse a pris plus de place que tout le reste, m'ont prouvé combien tu tenais à moi. Ton regard de physicien a nourri mes réflexions et je souhaite que l'on poursuive notre quête du savoir à deux, en continuant à dissenter sur le sens de l'univers.

Je dédie ce travail à Nini !

Résumé

Le changement climatique engendre une modification profonde et durable des conditions de vie au sein des écosystèmes, entraînant le bouleversement des communautés qui y vivent. La faune abritée au sein des écosystèmes dulçaquicoles représente plus d'un tiers des vertébrés à l'échelle mondiale, et son déclin est plus alarmant que dans les autres milieux (terrestres et marins). Malgré cela, les connaissances nécessaires à la mise en place de mesures de gestion en vue de protéger ces milieux et d'y adapter les usages restent incomplètes. L'objectif de cette thèse a été d'apporter des éléments sur la structuration spatiale des communautés aquatiques à large échelle, et ce dans un contexte de changement climatique, en vue de répondre à des besoins *in fine* très appliqués.

Le choix entre des variables environnementales directes (*i.e.* température de l'eau, variables hydrauliques) ou proximales (température de l'air, débits) des conditions de vie des organismes a d'abord été discuté. Puis, la structuration spatiale des communautés de poissons et de macroinvertébrés du bassin versant de la Loire ont été décrites par analyses multivariées. L'intérêt de l'approche multi-compartiments biologiques, c'est-à-dire la combinaison des données des poissons et des macroinvertébrés, a été montré au regard du manque de concordance spatiale entre les deux compartiments étudiés. Les communautés ligériennes sont structurées le long du gradient environnemental amont-aval, par la température de l'eau, les conditions hydrauliques et la topographie.

Le recours à la modélisation de la distribution spatiale des espèces (SDM) est une méthode de choix pour identifier les déterminants environnementaux majeurs des espèces et anticiper les potentiels changement de distribution futurs. La plupart des SDM ne sont basés que sur des données de présence-absence, alors que l'utilisation de données d'abondance, bien que plus complexe à modéliser, apporte des informations précieuses sur l'état de santé des populations et leur capacité de résilience face à un environnement changeant. Une méthode de modélisation par *machine learning* des classes d'abondance a été proposée, avec l'application de solutions techniques pour résoudre les difficultés inhérentes aux données d'abondance. Ces SDM ont ensuite permis de projeter les distributions potentielles futures des espèces aux horizons temporels 2050 et 2080 et d'évaluer l'impact du changement climatique sur les abondances des communautés de poissons et de macroinvertébrés.

Enfin, une comparaison de la similarité des patrons de structuration spatiaux des communautés aquatiques des grands bassins versants français, Loire, Seine, Rhône et Garonne, a été réalisée afin d'évaluer la portée du travail réalisé sur le bassin versant de la Loire.

Mots-clés : modèles de distribution d'espèces ; SDM ; abondance ; poisson ; macroinvertébrés ; communautés aquatiques ; bassin versant de la Loire ; changement climatique ; transférabilité.

Abstract

Climate change is causing profound and lasting changes to living conditions within ecosystems, leading to major changes in the communities that inhabit them. More than a third of the world's vertebrates live in freshwater ecosystems, and their decline is more alarming than in other ecosystems (terrestrial and marine ones). Despite this, the knowledge needed to implement management measures to protect these environments and adapt uses to them remains incomplete. The aim of this PhD thesis was to provide information on the spatial structuration of aquatic communities on a large scale, in a context of climate change, with a view to meeting highly applied needs.

First, we discuss, the choice between direct environmental variables (e.g. water temperature, hydraulic variables) or proximal variables (air temperature, discharge) for the living conditions of organisms. The spatial structuration of fish and macroinvertebrate communities in the Loire catchment was then described using multivariate analyses. The value of the cross-taxa approach, i.e. the combination of fish and macroinvertebrates data, was demonstrated in view of the lack of spatial concordance between the two biological compartments studied. Thus, Loire communities are structured along the upstream-downstream environmental gradient, by water temperature, hydraulic conditions and topography.

Species distribution modelling (SDM) is the method of choice for identifying the major environmental determinants of species structuration and anticipating potential future changes in distribution. Most SDM are based solely on presence-absence data, whereas the use of abundance data, although more complex to model, provides valuable information on the health of populations and their resilience capacity in a changing environment. A method for modelling abundance classes using *machine learning* was proposed, with the application of technical solutions to resolve the difficulties inherent in abundance data. These SDM were then used to project potential future species distributions for the 2050 and 2080 time horizons and to assess the impact of climate change on the abundances of fish and macroinvertebrate communities.

Finally, a comparison of the similarity of the spatial structuration patterns of the aquatic communities of the major French river basins, the Loire, Seine, Rhône and Garonne, was carried out in order to assess the scope of the work carried out on the Loire river basin.

Keywords : species distribution models ; SDM ; abundance ; fish ; macroinvertebrates ; aquatic communities ; Loire River basin ; climate change ; transferability.

Avant-propos ---

Ce travail de thèse a bénéficié du dispositif de financement CIFRE (Convention Industrielle de Formation par la REcherche) de l'ANRt (Association Nationale de la Recherche et de la technologie). Il a été réalisé en accueil exclusif au sein du LNHE (Laboratoire National d'Hydraulique et Environnement) de la R&D d'EDF (Électricité de France ; adresse complète : EDF R&D - Département LNHE - 6 quai Watier, 78400 CHATOU), en partenariat avec le laboratoire RiverLy de INRAE (Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement - adresse complète : Unité RiverLy 5 rue de la Doua - CS 20244 69625 Villeurbanne Cedex).



Sommaire

Remerciements	I
Résumé	V
Abstract	VI
Avant-propos	VII
Liste des figures	XII
Liste des tableaux	XVII
Liste des acronymes	XIX
Chapitre I. Introduction générale	1
A La théorie de la niche écologique	2
A.1 Des définitions multiples	2
A.2 Les hypothèses liées à la niche écologique	5
A.3 Comment faire le lien entre la niche écologique théorique et l'espace géographique observable ?	6
B Processus à l'origine de la distribution spatiale des abondances des espèces aquatiques	9
B.1 Influence des processus biogéographiques sur les patrons de structuration spatiaux	9
B.2 Quelles sont les principales variables environnementales qui structurent les communautés aquatiques dulçaquicoles ?	12
B.3 Le rôle des interactions biotiques	17
B.4 Les modifications induites par les changements globaux	19
C Contexte et enjeux de la modélisation de la distribution spatiale des abondances	22
C.1 Quelle méthodologie pour étudier la distribution spatiale des abondances ?	22
C.2 Quelles sont les méthodes de modélisation à disposition ?	24
C.3 Hypothèses en modélisation de la distribution des espèces	29
C.4 Modélisation de la distribution des abondances des espèces	31
D Cadrage de l'intérêt et de l'échelle d'étude de la thèse	34
D.1 Objectifs de compréhension des processus structurants les communautés aquatiques dans un contexte de changements globaux	34

D.2	Quelle échelle d'étude idéale ? Le bassin versant	35
D.3	Objectifs de la thèse et organisation du manuscrit	36
D.4	Contexte et enjeux industriels	38
Chapitre II.	Présentation de la zone d'étude et des données utilisées	39
A	Présentation de la zone d'étude principale : le bassin versant de la Loire	40
A.1	Régimes hydrologique et thermique	41
A.2	La faune aquatique ligérienne et ses particularités	47
A.3	Tendance d'évolution sous l'effet du changement climatique	48
B	Constitution des jeux de données biologiques et environnementaux	50
B.1	Les données biologiques	50
B.2	Le réseau H comme unité géographique	61
B.3	Les données environnementales du bassin versant de la Loire	65
B.4	Représentativité des conditions environnementales du bassin dans les jeux de données biologiques	78
	Résumé du chapitre	80
Chapitre III.	Apport des variables directes et de l'approche multi-compartiments pour l'analyse de la structuration des communautés aquatiques	81
A	Introduction	82
B	Données et méthodologie	83
B.1	Variables environnementales directes et proxies	85
B.2	Données poissons et macroinvertébrés	85
B.3	Analyse de la structuration des communautés par ordination mul- tivariable	86
B.4	Évaluation de la concordance des patrons spatiaux entre les com- munautés aquatiques	86
B.5	Comparaison de l'apport des variables environnementales directes par rapport aux variables proxies	87
C	Patrons de structuration obtenus à partir des ordinations multivariées	88
D	Intérêt des variables environnementales directes par rapport aux proxies dans la structuration des communautés aquatiques	93
D.1	Procédure de validation croisée de Monte-Carlo et du pouvoir pré- dictif de la RDA	94
D.2	Analyse de sensibilité des résultats de la RDA aux incertitudes des variables environnementales dérivées de modèles physiques	97
E	Discussion	99
E.1	Quel type de variables pour expliquer la structuration spatiale des communautés aquatiques ?	99
E.2	Limites à l'explication de la structuration spatiale des communautés	101
E.3	Concordance entre les compartiments poissons et macroinvertébrés	102
	Résumé du chapitre	106

Chapitre IV. Modélisation de la distribution des abondances de multiples taxons	107
A Introduction	108
B Données et méthodologie de modélisation des abondances des espèces aquatiques	110
B.1 Sélection des données biologiques	110
B.2 Méthodologie déployée pour modéliser la distribution spatiale des abondances	117
C Intérêt de l'application des <i>ordinal forest</i> pour la modélisation de la distribution spatiale des classes d'abondance	127
C.1 Évaluation des performances de deux méthodologies de modélisation des classes d'abondance par <i>ordinal forest</i>	127
C.2 Interprétation de la relation entre les classes d'abondance et l'environnement	136
D Intégration des interactions biotiques dans les modèles	143
D.1 Méthode d'intégration des relations trophiques dans les modèles de distribution	144
D.2 Évaluation de l'apport des relations trophiques à la modélisation de la distribution des classes d'abondance	147
Résumé du chapitre	154
Chapitre V. Projections futures des communautés aquatiques ligériennes sous scénario de changement climatique et conséquences sur leur structuration spatiale	155
A Introduction	156
B Données de scénarios de changement climatique	159
C Comparaison des conditions environnementales entre les horizons temporels	166
D Modifications de la distribution des espèces	169
D.1 Les différents types de modification de la distribution des espèces	169
D.2 Mesure de la modification de la distribution des espèces	169
E Modifications à l'échelle des communautés	172
E.1 Modification de la diversité des communautés	173
E.2 Taux de changement des communautés	180
F Conclusion et perspectives	184
Résumé du chapitre	187
Chapitre VI. Comparaison des patrons de structuration des communautés entre grands bassins versants français	189
A Introduction	190
B Données biologiques et environnementales à l'échelle des quatre bassins versants	192
B.1 Données biologiques	193
B.2 Données environnementales à l'échelle nationale	195

C	Similarité de la structuration des communautés et de ses déterminants environnementaux entre bassins versants	196
C.1	Comparaison de la structuration des communautés entre bassins versants	196
C.2	Déterminants environnementaux de la structuration des communautés	200
D	Influence du groupe d'espèces sur la transférabilité entre bassins versants	202
E	Conclusion et perspectives	204
	Résumé du chapitre	206
	Chapitre VII. Conclusion et perspectives de recherche	207
A	Bilan des travaux	208
B	Perspectives de recherche	212
	Références bibliographiques	223
	Annexes	237
	Annexe A. Liste des communications réalisées pendant la thèse	239
	Annexe B. Article publié dans Plos One	239
	Annexe C. Procédure de jointure des réseaux T-NET et RHT	284
	Annexe D. Boxplot des performances des modèles de la distribution des classes d'abondance pour les espèces de poissons et les genres de macroinvertébrés	285
	Annexe E. Exemples supplémentaires de <i>Partial Dependance Plot</i>	286
	Annexe F. Rang des prédicteurs avant intégration des indicateurs de ressources macroinvertébrés dans les modèles basés sur le jeu de données communs	287
	Annexe G. Cartographies des projections futures pour les poissons (n = 26 espèces) et les macroinvertébrés (n = 31 genres)	288

Liste des figures

Figure 1	Chronologie des conceptualisations majeures de la niche écologique . . .	3
Figure 2	Illustrations du concept de niche écologique et du partitionnement de niche en présence d'un compétiteur.	4
Figure 3	Diagramme BAM	7
Figure 4	Schématisation de la structuration dendritique d'un réseau hydrographique	10
Figure 5	Schématisation de la théorie des filtres environnementaux	12
Figure 6	La règle des pentes définie par Huet (1949)	14
Figure 7	Les dix biocénotypes définis dans la biotypologie de Verneaux (1973). .	15
Figure 8	Structuration amont-aval des différentes caractéristiques hydromorphologiques d'un cours d'eau	18
Figure 9	Évolution de l'écart de température moyenne annuelle en fonction des trois scénarios de RCPs en France.	20
Figure 10	Principe des modèles statistiques de distribution d'espèces	26
Figure 11	Exemple de distribution avec une sur-inflation en zéros.	33
Figure 12	Résumé graphique du plan du manuscrit.	37
Figure 13	Carte altitudinale et hydro-écorégions du bassin versant de la Loire . .	42
Figure 14	Régimes hydrologiques observés dans le bassin versant de la Loire . . .	44
Figure 15	Régimes thermiques observés dans le bassin versant de la Loire	46
Figure 16	Constitution du jeu de données	50
Figure 17	Répartition du nombre de stations en fonction de leur nombre d'opérations d'échantillonnage de la faune piscicole en fonction des périodes 1994-2017 et 2007-2017	60
Figure 18	Représentation du réseau H sur le bassin versant de la Loire	62
Figure 19	Schéma de l'extraction et du traitement des données biologiques	63
Figure 20	Localisation des tronçons portant une station de suivi biologique. . . .	64
Figure 21	Données d'entrée du modèle T-NET.	68
Figure 22	Cartographie des moyennes interannuelles de la hauteur en eau et de la température de l'eau	69
Figure 23	Cartographie de la probabilité d'intermittence des cours d'eau	71

Figure 24	Cartographie de la variable synthétique d'occupation du sol	73
Figure 25	Cartographie de la densité de population aux abords directs des cours d'eau	74
Figure 26	Ouvrages référencés dans le référentiel des obstacles à l'écoulement . .	75
Figure 27	Indicateurs de pression sur la continuité écologique	77
Figure 28	Représentativité des ordres de Strahler dans chacun des jeux de don- nées	78
Figure 29	Ordination biologique basée sur les données poissons	89
Figure 30	Ordinations biologiques basées sur les données macroinvertébrés	90
Figure 31	Ordination biologique basée sur les données poissons et macroinver- tébrés	91
Figure 32	Richesse taxonomique illustrée dans les ordinations biologiques	92
Figure 33	Importance de chaque variable environnementale dans la structuration des ordinations biologiques	93
Figure 34	Distribution des biais annuels entre les valeurs observées et simulées pour la température de l'eau et le débit	97
Figure 35	Histogrammes des CPUE > 0 observées sur les tronçons échantillonnés (N = 484) pour les espèces de poissons	115
Figure 36	Histogrammes des CPUE > 0 observées sur les tronçons échantillonnés (N = 1077) pour les genres de macroinvertébrés	116
Figure 37	Transformation des valeurs de CPUE en classes d'abondance	119
Figure 38	Corrélogramme entre les 10 prédicteurs environnementaux utilisés dans la modélisation de la distribution des abondances	121
Figure 39	Exemple de sur-inflation en zéros	123
Figure 40	Principe de la méthode SMOTE	125
Figure 41	Schéma des différentes étapes de la modélisation de la distribution des abondances d'une espèce	127
Figure 42	Exemples de comparaisons cartographiques des modèles non-SCUT vs SCUT sur les tronçons pour les données observées	128
Figure 43	Comparaison des performances des modèles sur données non-SCUT vs. SCUT	131
Figure 44	Exemples de matrice de confusion de deux modèles calibrés sur des données non-SCUT vs SCUT	133
Figure 45	Relation entre le gain de performance entre les modèles SCUT et non- SCUT et les caractéristiques des espèces	135
Figure 46	Rang des prédicteurs dans les modèles d'abondances des espèces de poissons	138
Figure 47	Rang des prédicteurs dans les modèles d'abondances des genres de macroinvertébrés	139

Figure 48	<i>Partial Dependence Plot</i> des quatre prédicteurs majeurs pour deux modèles poissons (truite commune et perche fluviatile) et les deux modèles macroinvertébrés (Potamanthus et Neureclipsis)	141
Figure 49	Exemple de <i>Partial Dependence Plot</i> de P_{inter}	142
Figure 50	<i>Fuzzy Coding Analysis</i> (FCA) à 6 groupes sur les 7 traits considérés pour définir les groupes de macroinvertébrés en fonction de leur susceptibilité à la prédation. Les modalités sont représentées sur le graphe de gauche et les taxons sur le graphe de droite.	146
Figure 51	Corrélogramme des prédicteurs utilisés dans les SDM avec intégration de la ressource macroinvertébrés. Les valeurs présentées sont les coefficients de corrélation de Spearman (ρ). Les prédicteurs de la ressource macroinvertébrés sont annotés comme suit : abondance du groupe vulnérable lentique = « ab vuln. len. », abondance du groupe vulnérable lotique = « ab vuln. lo. », indice de Piélou = « pielou » ; le reste des correspondances avec les codes prédicteurs peut être consultée au tableau 9.	147
Figure 52	Comparaison des performances des modèles avant et après intégration des indicateurs de ressource macroinvertébrés	148
Figure 53	Rang des prédicteurs après intégration des indicateurs de ressource macroinvertébrés	150
Figure 54	<i>Partial Dependence Plot</i> des indicateurs de ressource macroinvertébrés dans les modèles poissons	151
Figure 55	Modifications théoriques attendues des abondances d'une espèce en fonction de l'adéquation entre la niche d'une espèce et les conditions environnementales disponibles entre deux horizons temporels	158
Figure 56	Changements de précipitations et de température de l'air en France d'ici à la fin du siècle par rapport à la période historique 1976-2005 sous le scénario RCP 8.5.	160
Figure 57	Écarts de la température moyenne interannuelle de l'eau entre la période de référence (2007-2017) et les horizons temporels H2050 et H2080 sous le scénario RCP 8.5	162
Figure 58	Écarts du ratio percentile 90 sur percentile 10 de la température de l'eau entre la période de référence (2007-2017) et les horizons temporels H2050 et H2080 sous le scénario RCP 8.5	163
Figure 59	Écarts de la vitesse moyenne de courant interannuelle entre la période de référence (2007-2017) et les horizons temporels H2050 et H2080 sous le scénario RCP 8.5	164
Figure 60	Écarts du ratio percentile 90 sur percentile 10 de la vitesse de courant entre la période de référence (2007-2017) et les horizons temporels H2050 et H2080 sous le scénario RCP 8.5	165

Figure 61	Similarité des conditions environnementales rencontrées dans le jeu d'entraînement par rapport aux conditions de projections à l'échelle du réseau H aux différents horizons temporels	168
Figure 62	Distributions potentielles projetées pour l'ablette, la truite, Choroterpes et Trichostegia aux trois horizons temporels	170
Figure 63	<i>Species Range Change</i> pour chaque espèce de poissons et genres de macroinvertébrés	172
Figure 64	Occurrence des classes d'abondance projetées pour les 26 espèces de poissons	173
Figure 65	Occurrence des classes d'abondance projetées pour les 31 genres de macroinvertébrés	174
Figure 66	Modification de la richesse spécifique entre les trois horizons temporels.	175
Figure 67	Modification de la richesse spécifique pour les poissons entre les trois horizons temporels.	176
Figure 68	Modification de la richesse spécifique pour les macroinvertébrés entre les trois horizons temporels.	177
Figure 69	Changements des patrons de richesse spécifique et de β -diversité des communautés formées par les espèces de poissons et de genres de macroinvertébrés par ordre de Strahler	179
Figure 70	<i>Temporal Beta Diversity Index</i> (TBI) moyen entre l'horizon H2010 et les horizons H2050 (barres noires) et H2080 (barres grises) par ordre de Strahler	181
Figure 71	Composants B et C du <i>Temporal Beta Diversity Index</i> (TBI) entre l'horizon H2010 et les horizons H2050 (barres noires) et H2080 (barres grises) par ordre de Strahler	181
Figure 72	<i>Temporal Beta Diversity Index</i> (TBI) calculés entre les horizons H2050 et H2010, et entre H2080 et H2010	182
Figure 73	Partitionnement du <i>Temporal Beta Diversity Index</i> (TBI) en gains et pertes pour les deux horizons temporels H2050 et H2080	183
Figure 74	Hydroécorégions de niveau 1 à l'échelle de la France	192
Figure 75	Schéma de l'extraction et de la constitution du jeu de données biologiques	194
Figure 76	MTw30j sur le réseau RHT à l'échelle de la France	196
Figure 77	<i>Spiderplot</i> de l'ordination basée sur les données poissons des quatre bassins versants	198
Figure 78	<i>Spiderplot</i> de l'ordination basée sur les données macroinvertébrés des quatre bassins versants	199
Figure 79	Lien entre les données biologiques ordonnées par NMDS et la température de l'eau	200
Figure 80	Ordinations par NMDS des données poissons en fonction des groupes d'espèces utilisés	203

Figure 81 Ordinations par NMDS des données macroinvertébrés en fonction des groupes de genres utilisés203

Liste des tableaux

Tableau 1	Liste des espèces de poissons présentes dans le jeu de données du bassin versant de la Loire	54
Tableau 2	Liste des genres de macroinvertébrés	57
Tableau 3	Gamme de valeurs des variables environnementales dans chacun des jeux de données	79
Tableau 4	Gamme environnementales couvertes par chacun des jeux de données .	84
Tableau 5	Importance des variables environnementales pour expliquer ou prédire l'ordination biologique des ensembles de données	96
Tableau 6	Analyse de sensibilité des résultats de la RDA aux incertitudes des variables environnementales dérivées des modèles physiques	99
Tableau 7	Liste des espèces de poissons modélisées ou non	112
Tableau 8	Liste des genres de macroinvertébrés modélisés ou non	113
Tableau 9	Variables environnementales utilisées comme prédicteurs dans les SDM d'abondance	122
Tableau 10	Matrice de confusion d'un modèle multi-classe dans un cas à 4 classes.	123
Tableau 11	Régime trophique des espèces de poissons retenues pour la construction de SDM intégrant leur ressource trophique en prédicteur	145
Tableau 12	Informations concernant les jeux de données biologiques sur les bassins versants Seine, Loire, Rhône et Garonne	194
Tableau 13	Liste des variables environnementales mobilisées pour l'analyse de la similarité entre les quatre grands bassins versants Seine, Loire, Rhône et Garonne	197
Tableau 14	Part de variance expliquée par chaque variable dans la RDA entre les axes des NMDS poissons et macroinvertébrés et les variables environnementales . .	201

Liste des acronymes

ACP	Analyse en Composante Principale
BAM	<i>Biotic Abiotic Mobility</i>
CLC	<i>Corine Land Cover</i>
CNPE	Centres Nucléaires de Production d'Électricité
CPUE	Captures Par Unités d'Effort
EDF	Électricité De France
ENM	<i>Ecological Niche Models</i>
EROS	Ensemble de Rivières Organisé en Sous bassins
INRAE	Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement
INSEE	Institut National de la Statistique et des Études Économiques
JSDM	<i>Joint-Species Distribution Models</i>
MESS	<i>Multivariate Environmental Similarity Surface</i>
ML	<i>Machine Learning</i>
NMDS	<i>Non-metric Multidimensional Scaling</i>
OF	<i>Ordinal Forest</i>
OFB	Office Français pour la Biodiversité
PDP	<i>Partial Dependence Plots</i>
PERMANOVA	<i>PERmutational Multivariate ANalysis Of VAriance</i>
RCC	<i>River Continuum Concept</i>
RCP	<i>Representative Concentration Pathways</i>
RDA	<i>Redundancy Analysis</i>
RF	<i>Random Forest</i>
RHT	Réseau Hydrographique Théorique
ROE	Référentiel des Obstacles à l'Écoulement

SAFRAN	Système d'Analyse Fournissant des Renseignements Adaptés à la Nivologie
SCUT	<i>SMOTE and Cluster-based Undersampling Technique</i>
SDM	<i>Species Distribution Models</i>
SMOTE	<i>Synthetic Minority Oversampling Technique</i>
SIG	Systèmes d'Informations Géographiques
SRC	<i>Species Range Change</i>
TBI	<i>Temporal Beta-diversity Index</i>
T-NET	<i>Temperature NETwork</i>

« It is not necessary in any empirical science to keep an elaborate logicomathematical system always apparent, any more than it is necessary to keep vacuum cleaner conspicuously in the middle of a room at all times. When a lot of irrelevant litter has accumulated the machine must be brought out, used, and then put away. »

G.Evelyn Hutchinson, Concluding remarks, 1957

Chapitre I

Introduction générale

Ce chapitre introductif commence par présenter, dans la section A, une brève histoire de la théorie de la niche écologique. Le lien entre la niche écologique théorique et la structuration spatiale des abondances des communautés dans l'espace géographique est discuté dans la section B. Les enjeux de la modélisation de la distribution des abondances des espèces sont exposés dans la section C. Enfin, la section D présente le plan de ce manuscrit ainsi que les grands objectifs de ce travail de thèse.

A. La théorie de la niche écologique

A.1. Des définitions multiples

Comment expliquer l'aire de répartition d'une espèce ? Pourquoi ne se trouve-t-elle pas dans certains lieux qui lui semblent pourtant favorables ? Quels effets les espèces peuvent-elles avoir les unes sur les autres ? [...] sont autant d'interrogations auxquelles l'écologie tente de répondre. L'écologie cherche à comprendre les relations entre les êtres vivants et leur environnement. La théorie de la niche écologique a été formulée pour expliquer les relations entre les espèces et l'ensemble des facteurs susceptibles de les influencer. Le terme de *niche* est utilisé pour la première fois par Johnson (1910) et fait ensuite l'objet de nombreuses définitions et conceptualisations, comportant chacune leurs particularités (Figure 1). Grinnell (1917) et Elton (1927) ont considéré deux aspects différents pour appréhender ce qui constitue la niche. La définition de Grinnell considère uniquement les variables environnementales nécessaires pour une espèce, au sens de besoins (aussi appelée variables scénopoétiques). Elle s'intéresse donc aux besoins nécessaires au maintien d'une espèce et définit les contours de la niche en explorant les relations entre les taux de croissance ou de mortalité d'une espèce et les différents facteurs environnementaux. La niche d'Elton quant à elle considère l'impact que l'espèce exerce sur son environnement, notamment sur ses ressources, et l'influence de l'environnement sur l'espèce au cours du temps (Peterson *et al.*, 2011 ; Sales *et al.*, 2021). De fait, la niche d'Elton s'applique à une échelle beaucoup plus locale. Hutchinson (1957) a ensuite proposé de définir la niche écologique comme l'hypervolume à n dimensions dont les limites comprennent l'ensemble des facteurs favorables à une espèce, c'est-à-dire l'espace dans lequel une espèce pourrait théoriquement exister et se maintenir indéfiniment (Figure 2). En biologie des populations, cela se traduit par un taux de croissance intrinsèque positif pendant une période de temps définie (Peterson *et al.*, 2011).

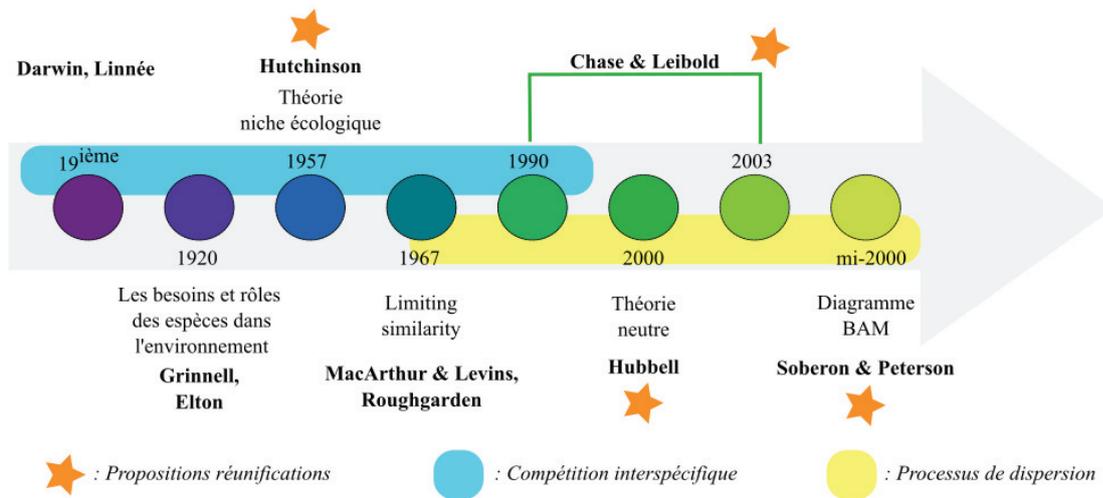


FIGURE 1

Chronologie des conceptualisations majeures de la niche écologique. Adapté de Sales *et al.* (2021).

Dans la réalité, les espèces ne sont pas observées systématiquement dans les habitats regroupant les variables abiotiques favorables (définition grinnellienne). Hutchinson a nommé l'espace effectivement occupé par une espèce la niche réalisée. Il considère que les interactions biotiques qui se produisent dans l'espace géographique restreignent la niche potentielle d'une espèce à un espace qui est communément admis comme plus petit. La première démonstration expérimentale de l'importance du rôle des interactions biotiques à structurer la niche réalisée par une espèce a été attribuée à Gause (1934). Il démontra que deux espèces de paramécies (*Paramecium caudatum* et *Paramecium aurelia*; *i.e.* protozoaires) placées au sein du même environnement ne pouvaient pas cohabiter, et que l'une des deux espèces finissait par évincer l'autre en quelques jours seulement. En effet, les ressources n'étant pas suffisantes pour le maintien des deux espèces dans le même habitat et sans possibilité de se déplacer vers un habitat plus favorable, l'espèce la plus compétitrice, c'est-à-dire celle parvenant le mieux à s'approprier et à exploiter la ressource commune, survivait au détriment de la seconde. Il définit ainsi le *principe d'exclusion compétitive*, qui prévoit que lorsque des espèces partagent la même niche écologique, c'est l'espèce qui exploite la ressource commune limitante le plus efficacement qui finira par supplanter l'autre. Un autre mécanisme peut permettre la co-existence de deux espèces partageant la même niche écologique : le partitionnement de niche (Figure 2; Sales *et al.*, 2021). Afin de coexister, chaque espèce n'occupe qu'une partie de sa niche dans le but de limiter la superposition avec celle de l'autre espèce, là où la compétition a lieu. La restriction de

la niche potentielle donne lieu à une niche réalisée encore plus restreinte, mais qui permet le maintien des deux espèces. Les implications de la compétition interspécifique sur le territoire que peut occuper une espèce, au regard de sa niche écologique et de celle de l'espèce concurrente, ont été le cœur des préoccupations des écologues jusque dans les années 1990 (Figure 1 ; Peterson *et al.*, 2011 ; Sales *et al.*, 2021).

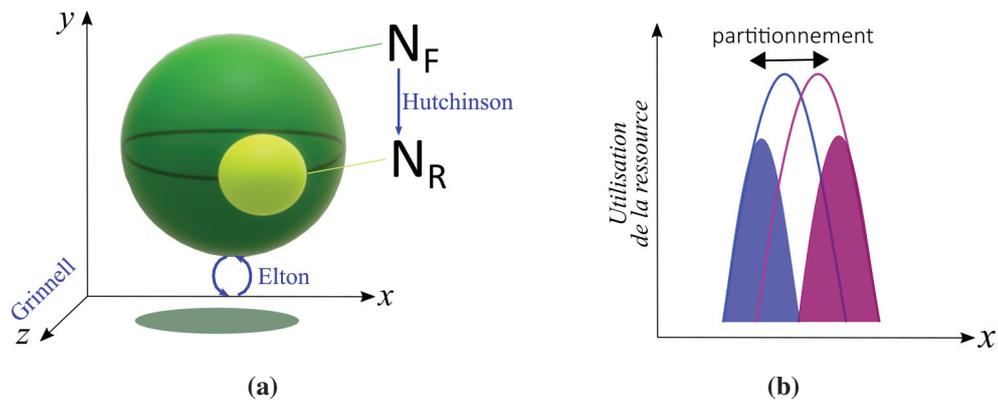


FIGURE 2

Illustrations du concept de niche écologique (a) et du partitionnement de niche écologique en présence d'un compétiteur (b). D'après la définition de Hutchinson la niche écologique fondamentale, notée N_F , est un hypervolume dans l'espace formé par les variables environnementales où une espèce peut se maintenir indéfiniment. Sous l'effet de compétiteurs partageant les mêmes exigences environnementales, N_F est restreinte à la niche réalisée, notée N_R . Grinnell définit la niche écologique à partir des conditions environnementales favorables à une espèce (besoins) tandis qu'Elton considère les rétro-actions que l'espèce et son environnement vont avoir l'une sur l'autre. Lorsque deux espèces utilisent la même ressource, elles sont en compétition (b). L'une des espèces peut évincer l'autre, ou alors chacune restreint la gamme de la ressource utilisée, c'est le partitionnement de la ressource.

L'intérêt des écologues pour le rôle des processus de dispersion dans la structuration spatiale des espèces a mené à la formulation de la théorie neutre (Hubbell, 2001). Cette théorie généralise celle de la biogéographie insulaire de MacArthur & Wilson (1967) à tous les écosystèmes, en considérant l'organisation en métacommunautés des organismes dans un écosystème donné. Les taux de natalité, de mortalité et de spéciation sont similaires, quelle que soit l'espèce et c'est la stochasticité des dynamiques de population et des échanges entre communautés qui définissent les structurations de la biodiversité. La capacité de cette théorie à expliquer différents patrons de la biodiversité en ont fait son succès et elle a souvent été opposée à la théorie de la niche écologique. Il n'est toutefois pas nécessaire d'opposer ces deux théories car elles peuvent être considérées comme deux processus extrêmes d'un même continuum, qui agissent certainement à des échelles spatiales distinctes (Sales *et al.*, 2021).

À ma connaissance, Chase & Leibold (2003) sont les derniers à avoir proposé une réunification des différents concepts de la niche écologique avec les avancées contemporaines de la discipline (Sales *et al.*, 2021). Selon eux, la niche définit la biogéographie d'une espèce en associant les processus de dispersion et sa niche abiotique (concept grinnellien), à toutes les échelles auxquelles elle est présente, ainsi que l'ensemble des interactions qu'elle a avec son environnement (concept eltonnien). Ces différentes composantes illustrent bien la complexité du concept de niche écologique et les difficultés à trouver une définition unificatrice (Scheiner & Willig, 2011 ; Sales *et al.*, 2021). En effet, l'étendue des champs d'études en écologie et la complexité des processus amènent à se concentrer sur une composante particulière de la niche écologique (*i.e.* relation organisme-environnement, interactions biotiques ou processus de dispersion), et donc à opter pour une définition partielle de ce qu'est la niche écologique.

A.2. Les hypothèses liées à la niche écologique

Pour Hutchinson, il ne s'agit pas de voir la niche écologique comme un déterminant binaire de la présence ou de l'absence d'une espèce. En effet, la niche fondamentale exprime la relation entre la valeur sélective (ou *fitness*) et un gradient environnemental. La variation de la *fitness* en fonction du positionnement dans l'hypervolume que forme la niche écologique fait l'objet de plusieurs hypothèses (Marino *et al.*, 2020). L'*abundant-center hypothesis*, ou hypothèse de l'abondance-centrée, stipule que l'abondance d'une espèce est maximale au centre de sa niche écologique et décroît à mesure qu'on approche les « bords » (Hutchinson, 1957). Cela traduit le lien étroit entre la *fitness* d'une espèce et l'environnement. La *niche breadth hypothesis*, ou hypothèse d'étendue de la niche, permet quant à elle d'expliquer l'ubiquité de certaines espèces par une largeur de niche plus importante (espèces généralistes), et donc par la capacité à occuper une large gamme de conditions environnementales (Marino *et al.*, 2020). La diversité des conditions favorables permet à ces espèces de se maintenir sur un territoire plus vaste et de présenter fréquemment des valeurs d'abondance élevées. Dans un contexte de changements globaux, ces espèces sont plus résilientes et ont donc plus de facilité à maintenir leur population car elles ont plus de chances de s'adapter aux nouvelles conditions environnementales (Marino *et al.*, 2020). Enfin, la *niche position hypothesis* propose que les espèces dont la niche se situe proche des conditions environnementales moyennes disponibles sur un territoire

donné seront avantagées car elles ont la possibilité d'occuper une grande partie des habitats disponibles (Marino *et al.*, 2020). Comme dans l'hypothèse précédente, ces espèces sont généralement caractérisées par de fortes occurrences et abondances. Ces deux dernières hypothèses postulent ainsi qu'il existe une relation positive entre la favorabilité d'un habitat et l'abondance des espèces.

La confusion fréquente entre les différents espaces de la niche écologique, l'espace environnemental théorique et l'espace géographique réel, et les différents champs d'application des hypothèses ayant trait à la niche rendent difficile la définition d'un cadre conceptuel clair. En effet, les hypothèses précédentes concernent la niche écologique dans son espace environnemental théorique, et elles ont été largement étendues à l'espace géographique (Peterson *et al.*, 2011). Les deux espaces sont évidemment fortement liés puisque les distributions observées sur un territoire donné sont l'expression de la niche écologique. La modélisation de la niche écologique permet de caractériser et de quantifier ces processus. De nombreux auteurs tentent d'apporter des clarifications sur le cadre conceptuel des modélisations de la niche écologique ou de la distribution des espèces et appellent à préciser dans quel cadre se place chaque étude. En ce sens, de nombreuses publications discutent de la sémantique et des concepts employés en modélisation de la distribution des espèces (voir par exemple Soberon & Peterson 2005 ; Kearney 2006 ; Peterson *et al.* 2011). L'étude de la distribution des espèces vise à caractériser la répartition des espèces dans l'espace géographique et est souvent basée sur l'étude des relations entre l'espèce et son environnement (Peterson *et al.*, 2011 ; Guisan *et al.*, 2017). La plupart du temps, l'objectif est de définir la niche grinnellienne (*i.e.* les variables abiotiques favorables) et éventuellement d'évaluer quels facteurs ou processus contraignent l'espèce à sa niche réalisée. Ces autres facteurs regroupent toutes les influences qui interviennent dans l'espace géographique, comme les interactions biotiques, l'histoire biogéographique ou encore les possibilités de dispersion offertes à l'espèce.

A.3. Comment faire le lien entre la niche écologique théorique et l'espace géographique observable ?

Le diagramme *Biotic Abiotic Mobility* (BAM) a été conceptualisé par Soberon & Peterson (2005) pour illustrer le lien entre la niche fondamentale (*i.e.* niche théorique de l'espace environnemental) et son expression dans l'espace géographique (Figure 3). Il reprend

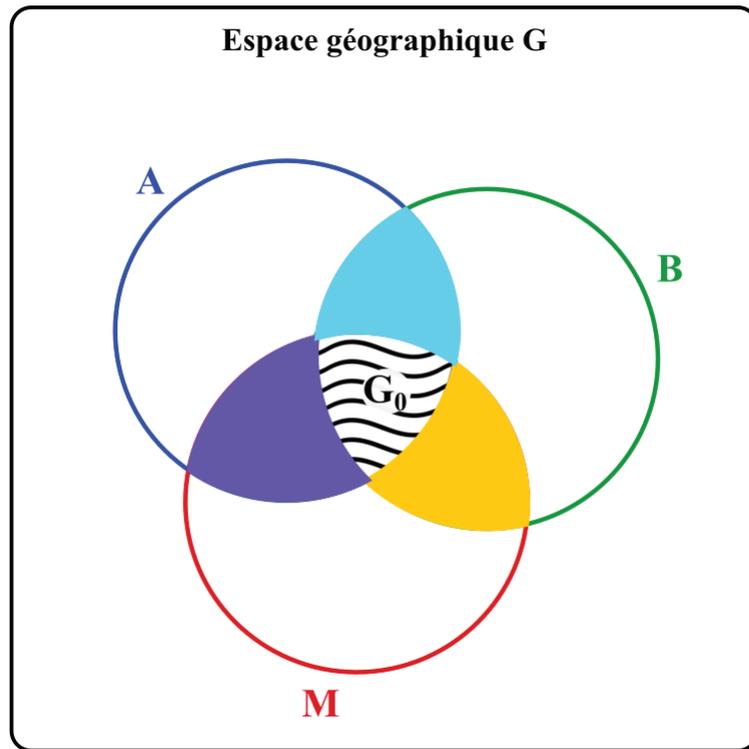


FIGURE 3

Diagramme BAM (Biotic Abiotic Mobility). Ce schéma permet de conceptualiser les différents facteurs qui définissent la présence d'une espèce dans un espace géographique donné, nommé G.

Les conditions abiotiques A sont l'ensemble des conditions environnementales favorables à l'espèce telles que définies dans l'espace théorique de la niche écologique fondamentale. Les conditions biotiques B sont l'ensemble des interactions provenant d'autres espèces qui contraignent ou étendent la niche fondamentale pour former la niche réalisée. Enfin, M représente les capacités de déplacement d'une espèce et, par extension, l'ensemble des zones géographiques atteignables. L'intersection de toutes ces conditions forme la niche observée dans l'espace géographique G₀. Adapté de Peterson & Soberón (2012).

les composantes abiotiques (A), biotiques (B) et les capacités de dispersion (M) évoquées précédemment. L'intersection des espaces A et B forme la niche potentielle d'une espèce et présente l'ensemble des conditions abiotiques et biotiques qui lui sont favorables. L'intersection de cet espace avec l'espace M représente la distribution géographique actuelle d'une espèce, c'est-à-dire les lieux où les conditions biotiques et abiotiques sont favorables à l'espèce dans un habitat qui lui est accessible. L'espace formé par l'intersection de A et B et exclu de M permet d'identifier les zones favorables à une espèce si elle y avait accès. Dans le cadre des changements globaux, l'espace M peut être vu comme l'ensemble des territoires à rendre accessibles pour servir de zones refuges aux espèces pour lesquelles une partie voire l'intégralité du territoire actuellement occupé ne serait plus favorable.

Le diagramme BAM offre une opérationnalisation du lien entre la niche écologique

fondamentale et la distribution spatiale des espèces effectivement observées (Sales *et al.*, 2021). Premièrement, il illustre bien le fait que la niche théorique d'une espèce est composée de combinaisons uniques entre les conditions abiotiques et biotiques favorables, dont chacune n'est pas nécessairement rencontrée dans l'espace géographique. En revanche, un ensemble de conditions favorables de l'espace théorique peut être retrouvé en plusieurs endroits de l'espace géographique ; ce phénomène est appelé la dualité de Hutchinson et peut impliquer un biais de sur-représentation de ces conditions environnementales dans les données d'échantillonnages terrain (Peterson *et al.*, 2011). Ces constats impliquent que la caractérisation complète de la niche fondamentale d'une espèce ne peut être faite à partir des observations de terrain seules. Colwell & Rangel (2009) ont identifié trois raisons qui expliquent l'inaccessibilité à l'intégralité de la niche fondamentale à partir des observations de la distribution géographique : (i) certaines conditions favorables de la niche fondamentale n'existent pas dans l'espace géographique, (ii) certaines parties de la niche fondamentale sont bien présentes dans l'espace géographique mais sont inaccessibles à l'espèce (limitation par la dispersion) et (iii) la distribution géographique potentielle de l'espèce est tronquée par les interactions biotiques (limitation biotique). L'expérimentation en conditions contrôlées permet de s'affranchir des limites à la dispersion ou de l'effet potentiel des interactions biotiques pour établir l'ensemble des gammes environnementales favorables constituant la niche fondamentale d'une espèce. D'autre part, la contribution relative des composantes abiotiques, biotiques et de dispersion dans la structuration spatiale des communautés reste à déterminer (Galiana *et al.*, 2023).

Chaque composante du diagramme BAM intervient donc dans la structuration de la distribution des espèces observée dans l'espace géographique. L'influence de chacune de ces composantes abiotiques, biotiques et de la capacité de dispersion des espèces sur la distribution spatiale des abondances est décrite dans la section ci-après. La manière dont se répartissent les espèces dans l'espace a des conséquences à l'échelle de la population, c'est-à-dire l'ensemble des individus d'une même espèce, et à l'échelle de la communauté. Une communauté est définie comme l'assemblage des populations de différentes espèces en un lieu donné. Ce travail de thèse s'est concentré en particulier sur la distribution spatiale des espèces dulçaquicoles (*i.e.* eau douce) et sur la structuration spatiale des communautés qu'elles forment.

B. Processus à l'origine de la distribution spatiale des abondances des espèces aquatiques

Malgré la profusion d'articles qui étudient la distribution des espèces et les processus qui la sous-tende, il subsiste un manque de connaissance certain sur les processus de structuration spatiale. Cette limite est nommée *Wallacean shortfall* (Whittaker *et al.*, 2005). Un manque encore plus grand existe sur le rôle du réseau d'interactions entre les organismes (échelle de la communauté) et l'effet de l'environnement (limites nommées respectivement *Eltonian and Grinnellian shortfall*; Peterson *et al.*, 2011). La première lacune tend toutefois à se résorber au vu du nombre grandissant de données d'observation accessibles librement (*e.g.* Chase *et al.*, 2019 ; Comte *et al.*, 2021a).

B.1. Influence des processus biogéographiques sur les patrons de structuration spatiaux

Les patrons de structuration spatiaux sont définis comme l'ensemble des caractéristiques spatiales (*e.g.* aire de répartition, distribution des abondances) de l'occupation de l'espace géographique par une espèce. La biogéographie comporte deux sous-disciplines : la biogéographie historique, qui cherche à expliquer la part des patrons de distribution observés qui est due à des processus de dispersion ayant eu lieu par le passé, et la biogéographie interprétative, qui relie les variables géographiques, comme les variations du climat, à la distribution des espèces (Persat & Keith, 2020).

B.1.1. Biogéographie historique des communautés dulçaquicoles françaises

L'histoire biogéographique des communautés dulçaquicoles de la France métropolitaine a laissé une empreinte importante sur les distributions actuelles des espèces (Dias *et al.* 2014; Keith *et al.* 2020). Les dernières glaciations survenues au Quaternaire ont exposé les organismes à des climats de froids extrêmes, et seules certaines espèces ont pu trouver refuge dans les zones méridionales. Ces épisodes glaciaires ont provoqué l'abaissement du niveau marin ce qui a permis de connecter certains bassins et de favoriser des échanges d'espèces habituellement rendus impossibles par le milieu marin salin. Les recolonisations qui en ont suivi proviennent essentiellement du Danube et probablement quelques-unes par les voies maritimes pour les espèces tolérant la salinité (Keith *et al.*, 2020). Elles ont pu avoir lieu entre chaque épisode glaciaire et ont été limitées par la struc-

ture particulière du réseau hydrographique (*i.e.* structure dendritique; Figure 4), au sein de laquelle chaque bassin versant est relativement bien isolé, contrairement au milieu terrestre où, hormis les grandes barrières naturelles (*e.g.* chaînes montagneuses, déserts), les organismes circulent plus facilement. Ainsi, les communautés au sein d'un bassin versant peuvent avoir leur propre histoire évolutive. Les organismes aquatiques peuvent également se retrouver contraints par des conditions non favorables (*e.g.* limite supérieure de température de l'eau comme barrières physico-chimique), et ne pas pouvoir traverser certains milieux. Ces possibilités de dispersion limitées permettent d'expliquer la faible richesse faunistique des eaux européennes, par rapport à d'autres régions du Monde (Oberdorff *et al.*, 2011).

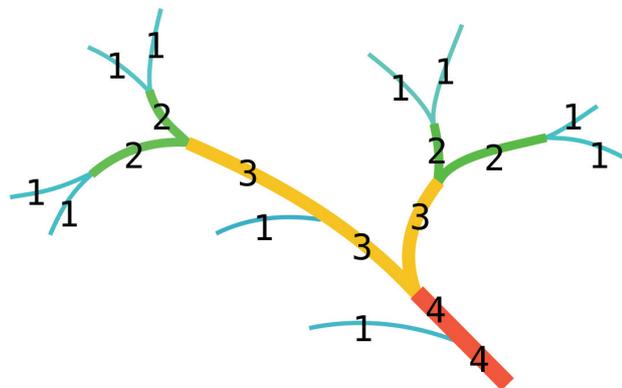


FIGURE 4

Schématisme de la structuration dendritique d'un réseau hydrographique. Le niveau hiérarchique de chaque tronçon de cours d'eau est défini par l'ordre de Strahler, qui prend 1 pour les tronçons de sources et +1 après chaque confluence avec un tronçon de même ordre de Strahler. Cet indicateur permet de rendre compte de la taille du cours d'eau et la complexité du réseau hydrographique en fonction de l'ordre de Strahler maximal (*e.g.* 8 pour la Loire, 12 pour l'Amazonie). Source de l'image : Wikipedia.

Les interventions humaines observées depuis le moyen-âge jouent également un rôle important dans la modification des communautés à travers l'apport d'espèces exogènes et la modification des connexions entre bassins. La croissance économique et le développement du commerce fluvial ont entraîné la construction de nombreux canaux et favorisé le transport des espèces, ce qui a facilité la dispersion de nombreuses espèces aquatiques entre les bassins. L'ouvrage de Keith *et al.* (2020) recense ainsi pas moins de 50 introductions de poissons en France, bien souvent involontaires. Près de deux tiers de ces espèces ont réussi à se maintenir dans le temps au sein des zones colonisées. Certains milieux sont donc aujourd'hui plus riches du fait de l'arrivée d'espèces exogènes. A contrario, ces nou-

veaux apports ont entraîné la disparition d'espèces indigènes auparavant non confrontées à la compétition inter-spécifique ou à certains pathogènes (*e.g.* arrivée d'*Aphanomyces astaci*, la peste de l'écrevisse responsable de la régression des écrevisses autochtones, via l'écrevisse américaine *Orconectes limosus*).

B.1.2. Théorie des filtres environnementaux

Chaque individu de chaque espèce possède des caractéristiques physiologiques, morphologiques et comportementales ainsi que des préférences écologiques qui lui sont propres ; ce sont ses traits autécologiques (Poff, 1997 ; Statzner & Bêche, 2010). L'ensemble de ces caractéristiques est compris dans une certaine gamme de variation à l'échelle de l'espèce qui définit ses capacités de réponse face à son environnement. En écologie fonctionnelle, la théorie des filtres environnementaux apporte des éléments sur les règles d'assemblage des communautés en se basant sur ces traits autécologiques. D'après cette théorie, les espèces doivent "passer" une succession de filtres environnementaux en fonction de leurs traits autécologiques afin de s'établir en une localisation donnée. Ces filtres peuvent être hiérarchisés selon l'échelle spatiale à laquelle ils s'appliquent sur les espèces. Les espèces dont les traits autécologiques ne leur permettent pas de "passer" l'un des filtres voient leur abondance extrêmement limitée à tous les niveaux hiérarchiques inférieurs voire sont absentes au-delà. Dans le cadre de l'hydrobiologie, les échelles successives d'effet de ces filtres vont du bassin versant ($\sim\text{km}^2$) au micro-habitat ($\sim\text{m}^2$), en passant par les cours d'eau (10 - 1 000 km) et le tronçon de rivière (~ 100 m ; Poff, 1997). Il est communément admis qu'à chaque échelle est associée un type de filtre différent : les processus biogéographiques ont lieu à l'échelle du bassin versant, le gradient environnemental intervient à l'échelle des cours d'eau et de leurs tronçons, et enfin les contraintes biologiques (*e.g.* compétition) et de qualité de l'habitat local (*e.g.* vitesse de courant, substrat) s'appliquent à l'échelle locale de l'habitat (Figure 5).

L'*Habitat Templet Concept*, formulé par Townsend & Hildrew (1976) et repris par Southwood (1977), se base sur la théorie des filtres environnementaux et relie le niveau de perturbation d'un habitat donné, défini par sa variabilité spatio-temporelle, et la sélection des traits autécologiques des espèces vivant dans cet habitat (Statzner & Bêche, 2010). Cela permet l'unification de plusieurs concepts tels que le *Flood Pulse Concept* qui lie la distribution des espèces aux régimes de perturbations engendrés par les crues (Junk *et al.*,

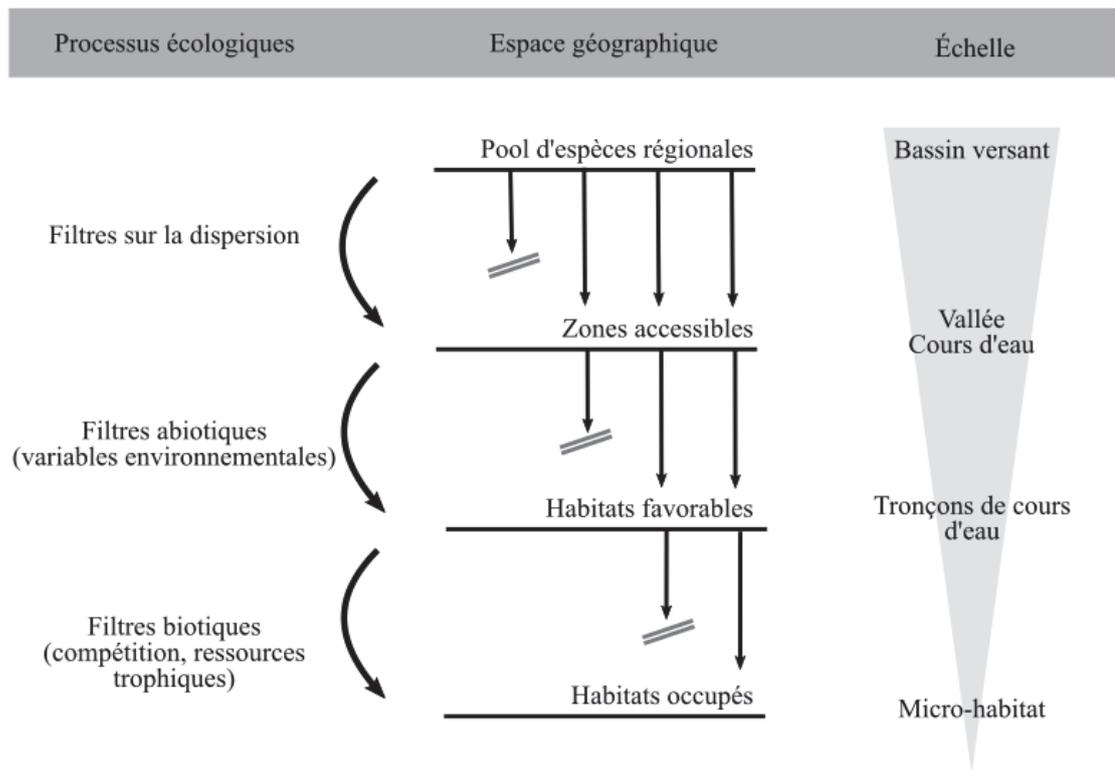


FIGURE 5

Schématisation de la théorie des filtres environnementaux. Différents processus filtrent successivement les espèces de l'échelle régionale à l'échelle locale en fonction des traits autécologiques des espèces. Le bassin versant est le lieu de déroulement des processus biogéographiques, qui limitent les espèces en fonction de leur capacité de dispersion. Les conditions abiotiques des rivières filtrent les espèces en fonction de leurs préférences thermiques et hydrologiques. Enfin, les processus biologiques, comme la compétition interspécifique, définissent les espèces à l'échelle de l'habitat. Adapté de Poff (1997) et Guisan *et al.* (2017).

1989).

B.2. Quelles sont les principales variables environnementales qui structurent les communautés aquatiques dulçaquicoles ?

La structuration spatiale des communautés est la façon dont elle s'organise dans l'espace en terme de composition en espèces et des abondances de chacune, en l'occurrence dans les milieux aquatiques. Les déterminants environnementaux principaux de la structuration spatiale des communautés aquatiques sont bien connus. La structure dendritique particulière des réseaux hydrographiques au sein des bassins versants forme un gradient de conditions environnementales de l'amont des cours d'eau vers l'aval (Ward, 1989). L'étude de ce gradient a donné lieu à différentes classifications des sections de cours d'eau, de l'amont vers l'aval, en fonction des communautés et des conditions environnementales

rencontrées (Wasson, 1989).

B.2.1. Les classifications écologiques de cours d'eau

L'importance du gradient environnemental amont-aval sur la structuration des communautés dulçaquicoles a conduit les hydrobiologistes à classer les zones des cours d'eau longitudinalement. À l'échelle française, la première classification est attribuée à Huet (1949) qui a formulé la règle des pentes (Figure 6). Celle-ci prédit qu'en conditions hydromorphologiques proches (*i.e.* en termes de pente et de largeur du cours d'eau), les communautés piscicoles observées sont similaires. En effet, ces conditions hydromorphologiques définissent la température de l'eau, la saturation en oxygène et la vitesse de courant qui influencent à leur tour les espèces de poissons dans toutes les étapes de leur cycle de vie (*i.e.* reproduction, alimentation, etc. ; Huet 1949). Ainsi, quatre zones sont définies d'amont en aval, chacune identifiable par une espèce indicatrice, d'amont en aval : la zone à truites, la zone à ombres, la zone à barbeaux puis la zone à brèmes. Une cinquième zone a ensuite été rajoutée dans la zone de mélange des eaux douces et salées (estuaire) : la zone à flets.

Illies & Botosaneanu (1963) ont proposé une typologie des eaux courantes basée sur les communautés d'invertébrés benthiques, caractérisée par la succession de trois zones d'amont en aval en fonction de la température de l'eau : le crénon, pour les zones fraîches de sources, le rhitron, pour les cours supérieurs encore soumis à des courants rapides, et le potamon, pour les cours inférieurs avec des températures estivales supérieures à 20°C. Cette typologie universelle est probablement la plus utilisée pour classer les cours d'eau dans les milieux tempérés (Laffaille *et al.*, 2020).

Verneaux (1976) a ensuite élaboré une biotypologie des cours d'eau en se basant sur les communautés de poissons et de macroinvertébrés de 240 stations françaises dont 140 sur le bassin du Doubs. Dix niveaux biotypologiques ont été définis (Figure 7) et déterminent les guildes d'espèces (*i.e.* biocénotypes) attendues en fonction du positionnement sur le gradient environnemental et en l'absence de perturbations d'origine anthropique du milieu (*i.e.* conditions de référence). Les variables considérées sont la température de l'eau, la dureté de l'eau, la section mouillée à l'étiage, la pente et la largeur du cours d'eau (Verneaux, 1977b). Ainsi, l'observation des conditions environnementales d'un tronçon de cours d'eau permet de l'affecter à un niveau typologique et de définir quelles y sont les

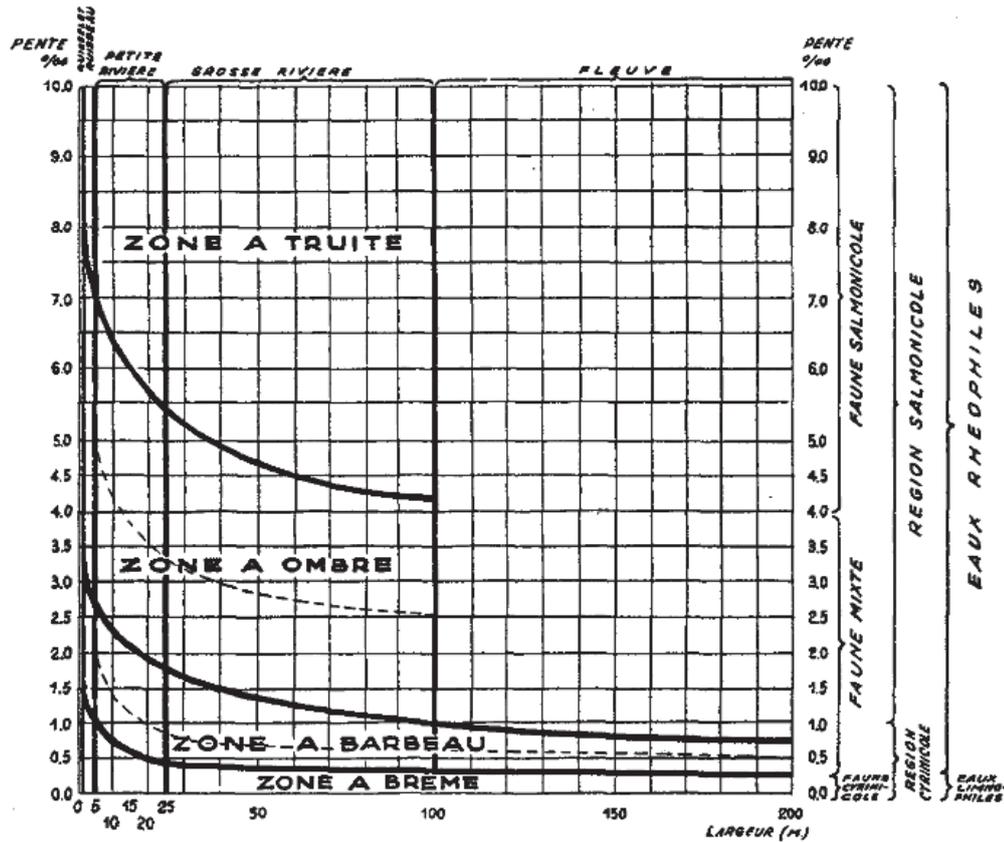


FIGURE 6

La règle des pentes définie par Huet (1949), qui est énoncée comme suit : « Dans une région biogéographique déterminée, des eaux courantes de même importance quant à la largeur et la profondeur, et possédant des pentes comparables, ont des caractères biologiques et spécialement des populations piscicoles analogues ». D'après Huet (1949).

espèces supposées présentes (Verneaux, 1977a). Le cortège d'espèces de poissons réellement observé peut servir au calcul d'une note, qui permet de mesurer l'écart à la composition spécifique attendue en conditions de référence.

L'apport de Vannote *et al.* (1980) avec le *River Continuum Concept* (RCC) se concentre plutôt sur le fonctionnement des cours d'eau, puisqu'il relie la productivité primaire d'un cours d'eau à son niveau énergétique qui évolue de l'amont vers l'aval et influence la structuration des communautés de macroinvertébrés (Vannote *et al.*, 1980). Les différents apports allochtones qui se produisent tout le long d'un cours d'eau dépendent de la géologie et de l'occupation du sol de son bassin versant (notamment la ripisylve), et résulte en un enrichissement et une diversification des ressources. Les communautés de macroinvertébrés s'organisent ainsi en fonction du type de ressource disponible, avec une fréquence relative maximale de taxons broyeur sur la zone amont, assimilant la matière organique primaire, suivis des taxons brouteurs/racleurs dans les zones intermédiaires puis des col-

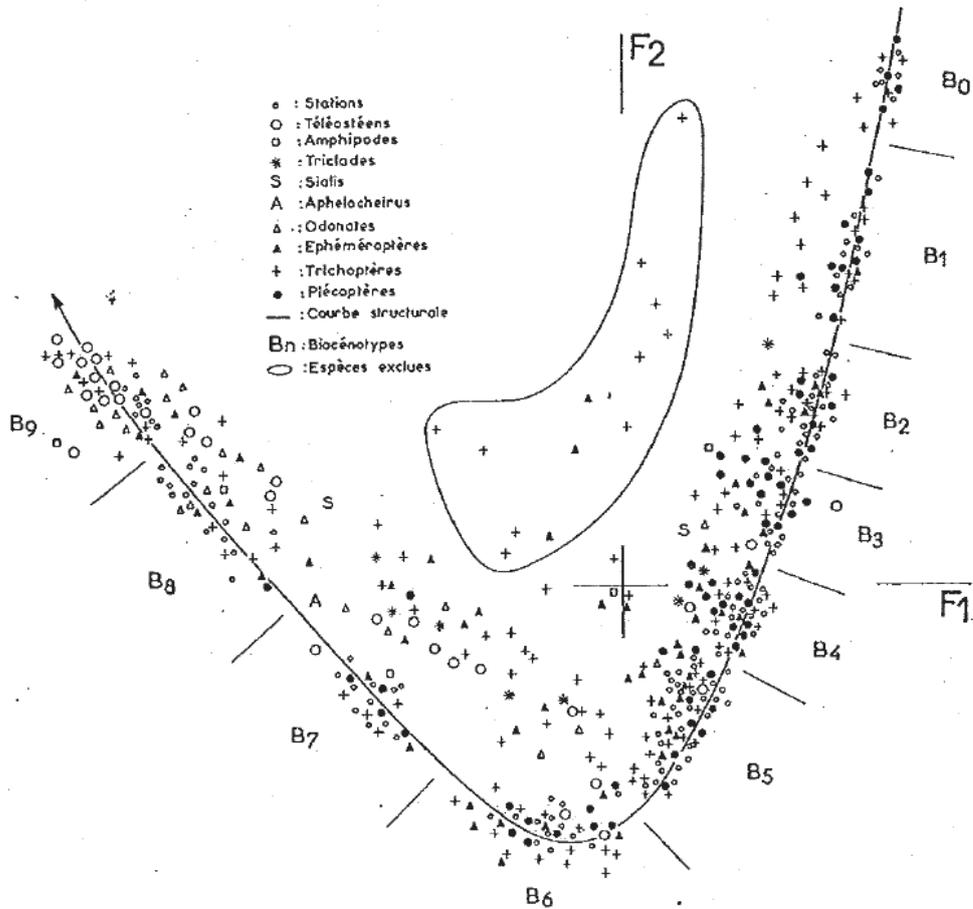


FIGURE 7

Les dix biocénotypes définis dans la biotypologie de Verneaux. Les groupements socio-écologiques, notés B0 à B9, formés par les espèces de poissons et les genres de macroinvertébrés se succèdent le long du gradient amont-aval. On notera l'effet Guttman qui donne sa forme en fer à cheval à l'ordination multifactorielle et suggère de multiple lien entre les réponses que prend la structuration des communautés. D'après Verneaux (1973).

lecteurs, qui s'alimentent sur les biofilms s'étant développés grâce aux ressources riches, dans les zones les plus aval. Les prédateurs de macroinvertébrés se distribuent ensuite le long du gradient longitudinal en fonction de la quantité de proies disponible. Le RCC explique ainsi les patrons de diversité, à savoir l'augmentation de la diversité relative depuis les têtes de bassins vers les zones intermédiaires (ordre de Strahler 4/5) puis une réduction de cette diversité tout à l'aval, par la diversification des habitats, associées à une proportion croissante de macroinvertébrés détritivores et filtreurs vers l'aval.

De nombreuses autres typologies européennes ont été formulées et ne seront pas abordées (*e.g.* Léger, 1909). Les classifications des sections de cours d'eau sont majoritairement issues de travaux européens, bien que quelques exemples existent ailleurs dans le monde. Par exemple, en Alaska et en Nouvelle-Zélande sur la base des macroinvertébrés

par Statzner & Higler (1986), ou aux États-Unis où Ricker (1934) élaborait une typologie piscicole de certains cours d'eau de l'Ontario et Wehrly *et al.* (2003) classèrent les rivières du Michigan en fonction des communautés de poissons et de la température moyenne de l'eau ainsi que sa variabilité. Ces classifications présentent toutes des limites qui ont probablement conduit à la diminution de l'intérêt pour ce champ d'études et poussé à leur moindre utilisation (Lasne *et al.*, 2007 ; Melles *et al.*, 2014). En premier lieu, la définition d'espèces indicatrices de chaque zone pose problème lorsque l'espèce en question n'est pas présente, du fait de son absence naturelle au sein du bassin versant ou d'une disparition liée aux perturbations du milieu (*e.g.* quasi-absence de l'ombre dans le bassin versant de la Loire ; Huet, 1949). En second, le passage net entre deux « boîtes » d'une classification est certainement plus progressif dans la réalité, défini par une zone de transition, ce qui peut rendre difficile l'attribution de certains tronçons de cours d'eau à une zone particulière (Matthews, 1998 ; Wehrly *et al.*, 2003). Enfin, ces travaux présentent des limites géographiques et temporelles fortes (Wasson, 1989). Par exemple, la biotypologie de Verneaux a été élaborée en majeure partie sur les cours d'eau de l'est de la France (karst jurassien), avec une faible capacité de généralisation à d'autres cours d'eau français (Wasson, 1989). De plus, les milieux ont évolué depuis les années 70 et les peuplements actuels sont certainement différents de ceux qui ont servi à construire la biotypologie. Enfin, ces classifications ne considèrent pas la dimension latérale des cours d'eau, avec par exemple l'ensemble des bras morts qui constituent un refuge ou un milieu de reproduction primordial pour certaines espèces (Junk *et al.*, 1989 ; Petts & Amoros, 1996). Néanmoins, du fait de leur facilité d'appréhension et de leur popularité, les classifications sont toujours citées afin de situer un site d'étude sur le gradient amont-aval.

B.2.2. Les déterminants abiotiques majeurs de la distribution des espèces

Les classifications des cours d'eau ont permis d'identifier les variables majeures qui structurent les communautés aquatiques le long du gradient longitudinal.

Premièrement, le positionnement amont-aval définit les conditions topographiques (altitude, pente) à l'origine des conditions abiotiques (Buisson *et al.*, 2008b ; Keith *et al.*, 2020). Le régime thermique sélectionne les espèces en fonction des préférences thermiques nécessaires à l'accomplissement de tous leurs stades de vie (Matthews, 1998 ; Laffaille *et al.*, 2020). En tête de bassin se trouvent les espèces psychrophiles, auxquelles succèdent

des espèces euryèces en zone médiane puis des espèces de plus en plus thermophiles, en accord avec le gradient thermique de plus en plus chaud, généralement en direction de l'estuaire. Cette structuration longitudinale des tolérances thermiques explique les phénomènes de glissement des communautés vers l'amont en réponse au réchauffement de la température de l'eau sous l'influence du changement climatique, avec une régression des espèces d'eau froide et l'extension de la répartition des espèces d'eaux chaudes (Buisson *et al.*, 2008b ; Flourey *et al.*, 2018 ; Maire *et al.*, 2019).

Enfin, le régime hydrologique définit les conditions hydrauliques locales ainsi que la dynamique des crues qui façonne les cours d'eau et leurs communautés aquatiques (Figure 8). Les espèces se distribuent sur le gradient longitudinal en fonction de leur capacité de nage et de leurs exigences en oxygène, avec un gradient d'espèce à forte capacité de nage et haut besoin en oxygène dissout en amont (*e.g.* truite commune) vers des espèces moins mobiles et plus tolérantes (*e.g.* carpe commune) en aval. Le niveau de connectivité latérale définit par l'intensité des crues annuelles et sa relation avec l'assemblage des communautés aquatiques peuvent être décrits par le *Flood Pulse Concept* (Junk *et al.*, 1989 ; Petts & Amoros, 1996 ; Keith *et al.*, 2020). Ce concept met en avant l'importance de la dimension latérale des cours d'eau, juste après le gradient longitudinal (Ward, 1989 ; Petts & Amoros, 1996).

Évidemment, l'ensemble de ces déterminants environnementaux intervient simultanément dans la structuration des communautés, c'est pourquoi il peut être difficile de déterminer leur influence relative.

B.3. Le rôle des interactions biotiques

Historiquement, les écologues se sont d'abord intéressés au rôle des interactions biotiques, en particulier au processus d'exclusion compétitive, dans la structuration des communautés à l'échelle locale (voir section A de ce chapitre). Puis, ils l'ont quelque peu délaissé au profit des études et conceptualisations des effets de la dispersion à large échelle sur les méta-communautés, avec notamment la théorie de la biogéographie insulaire (MacArthur & Wilson, 1967). Plus récemment, le concept de niche écologique a de nouveau suscité l'intérêt des écologues, et les interactions biotiques ont été ré-intégrées au cœur des études sur la structuration des communautés, et particulièrement dans l'étude de la distribution des espèces (Peterson *et al.*, 2011 ; Boulangeat *et al.*, 2012 ; Anderson, 2017 ;

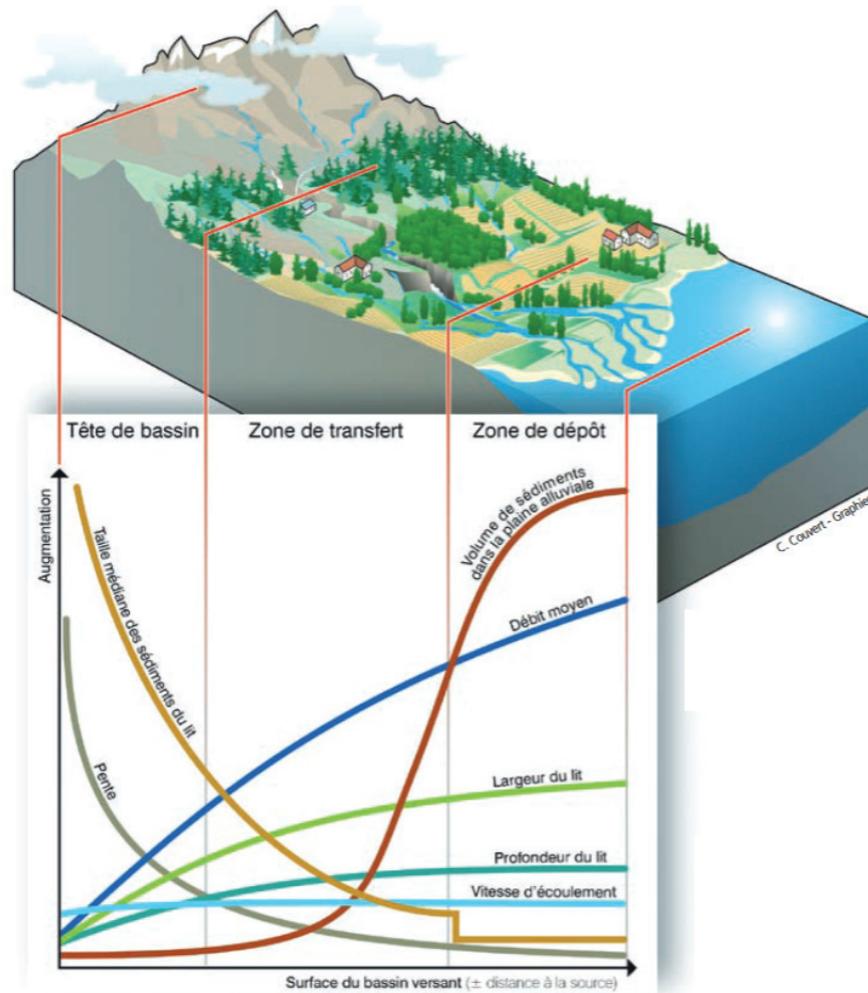


FIGURE 8

Structuration amont-aval des différentes caractéristiques d'un cours d'eau. Les principales caractéristiques hydromorphologiques sont présentées sur ce graphique et montrent une évolution dépendante du positionnement sur le gradient amont-aval. Comme les débits, la température de l'eau augmente vers l'aval. D'après Agence française pour la biodiversité - AFB, Centre de ressources Cours d'Eau (2021).

Dormann *et al.*, 2018). En effet, avec l'avancement de l'étude du rôle des composantes abiotiques et de la capacité de dispersion des espèces, la part de chacune dans la distribution des espèces et, *in fine*, dans la structuration des communautés reste à déterminer (Jackson *et al.*, 2001 ; Galiana *et al.*, 2023 ; Ohlmann *et al.*, 2023).

Il est communément admis que les interactions biotiques n'interviennent qu'à l'échelle très locale (souvent assimilée à l'habitat de l'espèce étudiée ; Chesson, 2000 ; Soberon & Nakamura, 2009). Pourtant, leur intégration dans les études de la distribution des espèces à large échelle (> régionale) semble essentielle (Soberon & Nakamura, 2009 ; Pollock *et al.*, 2014 ; Lucas *et al.*, 2023). On peut supposer que les interactions biotiques intègrent alors une variable environnementale qui est elle-même le facteur causal qui influence la distri-

bution (Dormann *et al.*, 2018). Une explication alternative a été proposée, l'hypothèse du bruit eltonien, qui stipule que la large échelle associée à la faible résolution de la grille d'étude ne permet pas de mettre en évidence l'effet des interactions biotiques au-delà de l'échelle locale (Soberon & Nakamura, 2009). En effet, les effets locaux sont moyennés à l'échelle de la zone géographique d'étude et n'apparaissent plus pertinents face à l'influence des variables environnementales qui interviennent à large échelle. D'autre part, la plupart des études ne considèrent que la présence et l'absence des espèces, ce qui semble assez réducteur pour étudier l'influence des interactions biotiques. En effet, ces dernières influent sur les niveaux d'abondance et, dans quelques cas (*e.g.* exclusion compétitive, pression de prédation), sur la présence des espèces en un lieu donné. En particulier, les interactions biotiques peuvent être un facteur explicatif majeur du faible niveau d'abondance d'une espèce, alors que dans le même temps les prédicteurs abiotiques prévoiraient une association entre la favorabilité et un niveau d'abondance élevés (Gerhardt Braz *et al.*, 2020).

B.4. Les modifications induites par les changements globaux

La niche écologique d'une espèce présente des variabilités temporelles (Scheiner & Willig, 2011). À l'échelle de millions d'années, la niche de chaque espèce se transforme, en fonction des aléas génétiques et des effets de l'environnement, et conduit à l'apparition de nouvelles espèces selon le processus de l'évolution (Scheiner & Willig, 2011). La plupart des écologues s'accordent sur la lenteur des processus évolutifs, à l'origine des adaptations qui permettent aux espèces de subsister dans un environnement en perpétuel changement (Olden *et al.*, 2004). Au cours de ces dernières décennies, la vitesse et l'importance des modifications de l'environnement se sont accrues, avec leur lot de conséquences sur les écosystèmes, comme l'homogénéisation biotique (Olden, 2006 ; Villéger *et al.*, 2011), la perte d'espèces rares (Thuiller *et al.*, 2005 ; Thuiller, 2007 ; Dudgeon, 2019 ; Auber *et al.*, 2022).

Les changements globaux comprennent toutes les modifications durables engendrées par les activités anthropiques (*e.g.* changement d'occupation des sols), dont la composante majeure, le changement climatique, est la modification induite sur le climat (GIEC, 2023). L'augmentation des gaz à effet de serre tel que le CO₂, liée en grande partie aux activités anthropiques, conduit à un réchauffement global de l'atmosphère et des océans, ainsi

qu'une réduction des surfaces enneigées ou gelées (*e.g.* banquises, glaciers). Ces changements, observables dès les années 1950, sont en nette accélération avec une rapidité sans précédent dans l'histoire de la planète Terre. Des projections des conditions climatiques futures en fonction de différents scénarios d'évolutions des émissions de gaz à effet de serre, les *Representative Concentration Pathways* (RCP), permettent d'entrevoir les potentiels changements futurs. En France, les températures moyennes sont amenées à augmenter de 1.5 à 5 °C d'ici à 2100 en fonction du scénario RCP (Figure 9), en association avec une diminution de la fréquence des précipitations, une augmentation de leur intensité et une importante variabilité spatiale du degré des modifications sur le territoire métropolitain (Soubeyroux *et al.*, 2020).

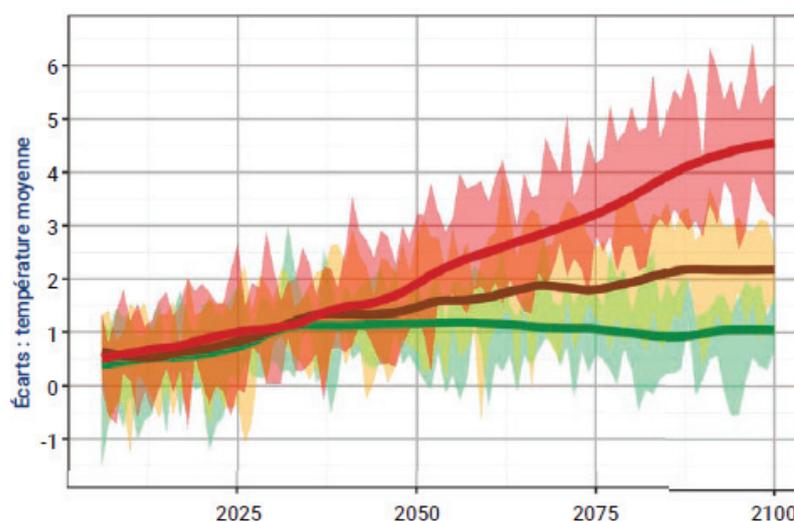


FIGURE 9

Évolution de l'écart de température moyenne annuelle en France (par rapport à la période référence 1976-2005) pour trois scénarios RCPs (*Representative Concentration Pathways*), RCP2.6 en vert, RCP4.5 en orange et RCP8.5 en rouge. Le trait continu correspond à la médiane lissée (moyenne glissante sur 20 ans) et les faisceaux autour de cette médiane correspondent à l'intervalle de confiance à 95 %. D'après Soubeyroux *et al.* (2020).

Le changement climatique influence fortement les conditions abiotiques des milieux aquatiques et modifie les régimes thermiques et hydrologiques des rivières. Globalement, les températures tendent à augmenter sur l'ensemble des cours d'eau, ainsi que la fréquence et l'intensité des événements hydrologiques extrêmes (*i.e.* crues et étiages) en réponse à la modification des températures moyennes et du régime des précipitations (Seyedhashemi *et al.*, 2021b). La température de l'eau des rivières tend à augmenter plus vite que la température de l'air, du fait d'une baisse accrue des débits et d'une réduction des apports souterrains (Leach & Moore, 2019 ; Seyedhashemi *et al.*, 2021b).

Les espèces ont déjà fait face à ce type d'événements climatiques par le passé et se sont adaptées ou éteintes. Or, les modifications actuellement en cours se produisent beaucoup plus rapidement que ce qui a pu avoir lieu par le passé, ce qui réduit les capacités des espèces à se maintenir et altère les processus évolutifs (GIEC, 2023). Le changement climatique, et particulièrement l'augmentation des températures, impacte les organismes à différents niveaux ; il a des effets sur leur physiologie et leur phénologie, sur la distribution des espèces et sur les règles d'assemblage et de structuration des communautés (Daufresne & Boët, 2007 ; Buisson *et al.*, 2008b ; Villéger *et al.*, 2011 ; Comte *et al.*, 2013). Les modifications de distribution peuvent prendre plusieurs formes en fonction de la niche réalisée d'une espèce dans l'espace géographique (Comte & Grenouillet, 2013a). Si les conditions environnementales de la niche réalisée d'une espèce sont toujours rencontrées, et que la quantité d'habitats favorables disponibles ne changent pas, un glissement de sa distribution pour suivre son habitat est attendu. Lorsque de nouvelles conditions favorables apparaissent, une extension de l'aire de répartition de l'espèce peut se produire. En revanche, dans le cas où la quantité d'habitat favorable diminue, une contraction de l'aire de répartition est attendue. Ces modifications de distribution sont généralement en lien avec le déplacement des espèces pour suivre leur préférence thermique (Parmesan *et al.*, 2005). La capacité de ces dernières à suivre leur niche thermique est souvent évaluée en comparant la vitesse climatique (*i.e.* la vitesse de déplacement des habitats thermiques des espèces dans l'espace) à la vitesse biotique (*i.e.* la vitesse de déplacement de l'aire de répartition des espèces) et montre une corrélation spatiale entre les directions de déplacement, mais avec un temps de latence pour les espèces qui pourrait leur être délétère sur le long terme (Buisson *et al.*, 2008b ; Comte & Grenouillet, 2015).

À l'échelle de la communauté, les modifications induites par le changement climatique peuvent se traduire par la perte des espèces sensibles aux températures élevées, et mener à une homogénéisation biotique (Olden, 2006 ; Villéger *et al.*, 2011). L'homogénéisation biotique est la diminution de la diversité β dans le temps, suite à l'augmentation de la similarité entre les communautés d'une zone géographique (Olden, 2006). Les pressions anthropiques sur les écosystèmes sont la première cause de ce processus, suivie par le changement climatique. Bien qu'il soit attendu que ces phénomènes entraînent une perte de biodiversité, un enrichissement des communautés de poissons et d'invertébrés françaises a été constaté récemment (Daufresne & Boët, 2007 ; Villéger *et al.*, 2011). Il résulte

de l'élargissement des aires de répartition des espèces d'eau fraîche et chaude, au détriment des espèces d'eau froide, et de l'arrivée d'espèces exogènes (Buisson *et al.*, 2008b ; Flourey *et al.*, 2018 ; Maire *et al.*, 2019). Ce phénomène est la résultante de la réorganisation rapide des communautés et d'un déséquilibre en faveur de l'arrivée de nouvelles espèces et la persistance de celles vouées à s'éteindre face aux nouvelles conditions favorables pour les premières et délétères à terme pour les secondes (Kuczynski *et al.*, 2023). Cet accroissement de richesse ne serait que transitoire et précéderait le processus d'homogénéisation biotique (Villéger *et al.*, 2011).

La majorité des études prospectives sur les effets potentiels du changement climatique dans le futur portent sur les taxons terrestres et les oiseaux (Guisan *et al.*, 2017 ; Robinson *et al.*, 2018 ; Steen *et al.*, 2021) alors que les organismes aquatiques font partie des plus menacés par les changements globaux (Dudgeon, 2019). Contrairement aux milieux terrestres ou marins, la structure dendritique du réseau hydrographique est une contrainte importante à la dispersion des espèces aquatiques et restreint leur accès à des conditions favorables voire aux zones refuges (Ward, 1989 ; Buisson *et al.*, 2008b). La majorité des études se sont concentrées sur les modifications des aires de répartition des espèces (Comte *et al.*, 2013), et ne considère pas les abondances qui apportent pourtant des informations essentielles sur les potentialités d'adaptation et de risque d'extinction des espèces (Potts & Elith, 2006 ; Waldock *et al.*, 2021). La section suivante présente les différentes approches qui existent pour étudier les relations espèce-environnement et la distribution résultante, ainsi que l'intérêt et les défis quant à la modélisation de la distribution des abondances.

C. Contexte et enjeux de la modélisation de la distribution spatiale des abondances

C.1. Quelle méthodologie pour étudier la distribution spatiale des abondances ?

La relation entre les espèces et leur environnement est complexe. De nombreux facteurs de nature variée et pouvant interagir entre eux interviennent dans la structuration de la biodiversité (voir section B de ce chapitre). À large échelle, ces facteurs présentent des variations spatio-temporelles importantes qui influencent la distribution des espèces (Chase, 2011). Afin d'étudier ces relations organisme-environnement et de prédire les mo-

difications de distribution induits par le changement climatique nous avons eu recours à la modélisation. Un modèle est une représentation simplifiée de la réalité, qui considère le lien entre une variable réponse, la variable à expliquer, et un jeu de prédicteurs restreints. Il tend à se rapprocher le plus possible de la réponse observée d'une espèce ou d'une communauté à certains facteurs et peut apporter des éléments intéressants sur les processus écologiques à large échelle. Néanmoins, les modèles restent des simplifications qui peuvent ne pas tenir compte de l'ensemble des facteurs influents, de leurs interactions, et de la complexité que cela engendre dans la réponse (Peterson *et al.*, 2011 ; Guisan *et al.*, 2017). C'est un élément à avoir en tête au moment de l'interprétation des sorties d'un modèle. Les modèles de distribution d'espèces (notés SDM) permettent de représenter la relation entre la distribution géographique d'une espèce (ou d'une communauté) et les conditions environnementales dans lesquelles on la trouve (Guisan & Zimmermann, 2000).

La majorité des études portant sur la modélisation de la distribution des espèces utilisent des données de présence et d'absence (notée présence-absence) ou de présence seule pour dresser les aires de répartition potentielles des espèces (Peterson *et al.*, 2011 ; Guisan *et al.*, 2017). La modélisation des abondances présente des verrous méthodologiques importants, alors que celles-ci sont bien plus informative (Waldock *et al.*, 2021). Tout d'abord, la compréhension des processus écologiques structurants les abondances permettrait de gagner en finesse vis-à-vis des relations d'ores-et-déjà identifiées entre les espèces et leurs déterminants environnementaux. La relation entre les abondances d'une espèce et ses déterminants environnementaux pourrait être caractérisée dans l'espace géographique. La forme de ces relations pourrait être comparée entre différentes espèces ce qui pourrait permettre, par exemple, de comparer l'occupation des différentes niches à disposition sur la zone d'étude. Il devient ainsi possible de tester empiriquement les hypothèses sur les mécanismes de la structuration des abondances (voir section B de ce chapitre). Par exemple, la modélisation de la distribution des abondances d'une espèce sur son aire de répartition permettrait de tester la validité de l'extension de l'hypothèse d'abondance-centrée à l'espace géographique (voir par exemple Waldock *et al.*, 2019 ; Yañez-Arenas *et al.*, 2020). Pour les gestionnaires, les informations sur les abondances d'une espèce sont également bien plus informatives que la seule présence ou absence. Comme évoqué précédemment, l'abondance informe sur la *fitness* d'une population, son état de santé et ses potentiels évo-

lutifs. Ces informations sont précieuses dans le cadre de suivis de populations, d'élaboration de mesures de conservation ou encore pour évaluer l'influence de différents scénarios de gestion (Waldock *et al.*, 2021). D'un point de vue des métacommunautés, elles peuvent permettre, par exemple, d'identifier les populations sources dans un territoire donné (Waldock *et al.*, 2021). C'est pourquoi l'élaboration de modèles de distribution des abondances performants est un défi scientifique majeur, tant pour l'écologie fondamentale que pour l'écologie de la conservation.

C.2. Quelles sont les méthodes de modélisation à disposition ?

Plusieurs catégories de modèles sont couramment utilisées pour représenter la distribution des espèces en fonction de leur niche écologique (Peterson *et al.*, 2015 ; Tourinho & Vale, 2023) : les modèles mécanistes, basés sur les processus physiologiques et biophysiques, et les modèles statistiques, basés sur des relations statistiques espèce-environnement. Une troisième catégorie, qui gagne en popularité mais qui ne sera pas détaillée ici, regroupe les simulations orientées-processus (*process-based simulations*) où des a priori sur la niche et la dispersion de l'espèce sont posés, afin de simuler virtuellement la distribution des espèces.

C.2.1. Les modèles mécanistes

Les modèles mécanistes visent à reconstruire la relation organisme-environnement en intégrant l'effet des conditions environnementales sur la physiologie et la biophysique des organismes. Cette relation est formalisée mathématiquement au moyen d'équations entre une variable réponse de physiologie/biophysique et les facteurs environnementaux étudiés. L'équation permet d'émettre des prédictions des réponses attendues chez l'espèce dans différentes conditions, qui sont ensuite vérifiées à l'aide de données empiriques (Peterson *et al.*, 2015). La construction de ce type de modèle demande une connaissance avancée des interactions entre l'environnement et la physiologie des organismes, qui dans notre cas semble trop éparsée pour permettre une telle modélisation à l'échelle multi-spécifique des communautés aquatiques (Tourinho & Vale, 2023 ; Wagner *et al.*, 2023). Dans les cas où l'élaboration d'un modèle mécaniste est possible, la tendance observée dans la littérature consiste à combiner différents types de modèles, mécanistes et statistiques par exemple, pour capter l'entière des processus écologiques de la structuration de l'espèce dans l'espace (Caradima *et al.*, 2021 ; Wagner *et al.*, 2023).

C.2.2. Les modèles statistiques de distribution des espèces

Les modèles de distribution d'espèces regroupent un ensemble de méthodes de modélisations statistiques de la distribution spatiale des espèces. Ils permettent de comprendre ces règles de structuration, en modélisant la relation espèce-environnement à partir de données d'observation. Ils sont utilisés dans plusieurs cadre d'application : (i) identifier les déterminants environnementaux de la distribution spatiale des espèces et décrire leur relation (Guisan & Zimmermann, 2000 ; Peterson *et al.*, 2011), (ii) évaluer des mesures de gestion en identifiant des zones favorables à la présence d'une espèce (Floury *et al.*, 2021), afin de prioriser les lieux de prospection d'une espèce d'intérêt (Prié *et al.*, 2014) ou d'identifier des zones propices à des opérations de translocation (Segal *et al.*, 2021) et (iii) de la même manière, ils peuvent permettre d'identifier les zones potentiellement favorables à l'établissement d'espèces invasives (Peterson *et al.*, 2011 ; Guisan *et al.*, 2017), (iv) d'identifier les effets du changement climatique (Buisson *et al.*, 2008b ; Elith *et al.*, 2010). La mise en place de base de données biologiques et environnementales conséquentes et l'incitation à l'*open access* ont favorisé le succès des SDM dont l'utilisation ne cesse de croître depuis les années 2000 (Guisan *et al.*, 2017). Leur développement avancé permet aujourd'hui de modéliser la distribution d'un grand nombre d'organismes à large échelle (*i.e.* régionale, continentale voire mondiale).

La modélisation de la distribution d'une espèce est constituée de plusieurs étapes représentées dans la Figure 10 et détaillées ci-après :

- (1) La première étape consiste à collecter les données de terrain sur l'espèce cible et les variables environnementales susceptibles d'expliquer dans sa distribution spatiale. La disponibilité de base de données libres d'accès facilite ce travail de collecte et permet de mutualiser différentes sources de données, en combinant des bases de données climatiques et biologiques par exemple ;
- (2) La base de données, éventuellement géoréférencée, ainsi constituée sert de données d'entrée aux modèles statistiques de distribution. Ils modélisent la réponse aux variables environnementales utilisées en prédicteurs et par extension, la niche réalisée sur la zone d'étude, ou du moins la niche réalisée dans les sites échantillonnées ;
- (3) Enfin, la projection du modèle sur toutes les unités géographiques permet de tracer la distribution potentielle de l'espèce sur la zone d'étude. Cette projection peut

être réalisée dans les conditions environnementales actuelles ou dans des contextes spatiaux et temporels différents, par exemple (4) sous changement climatique ;

- (4) Les résultats de la modélisation sont ensuite utilisés pour répondre à un objectif précis. Par exemple, dans le cadre du changement climatique, les distributions potentielles actuelles et futures sont généralement comparées afin d'évaluer les effets du changement climatique sur les espèces et la communauté.

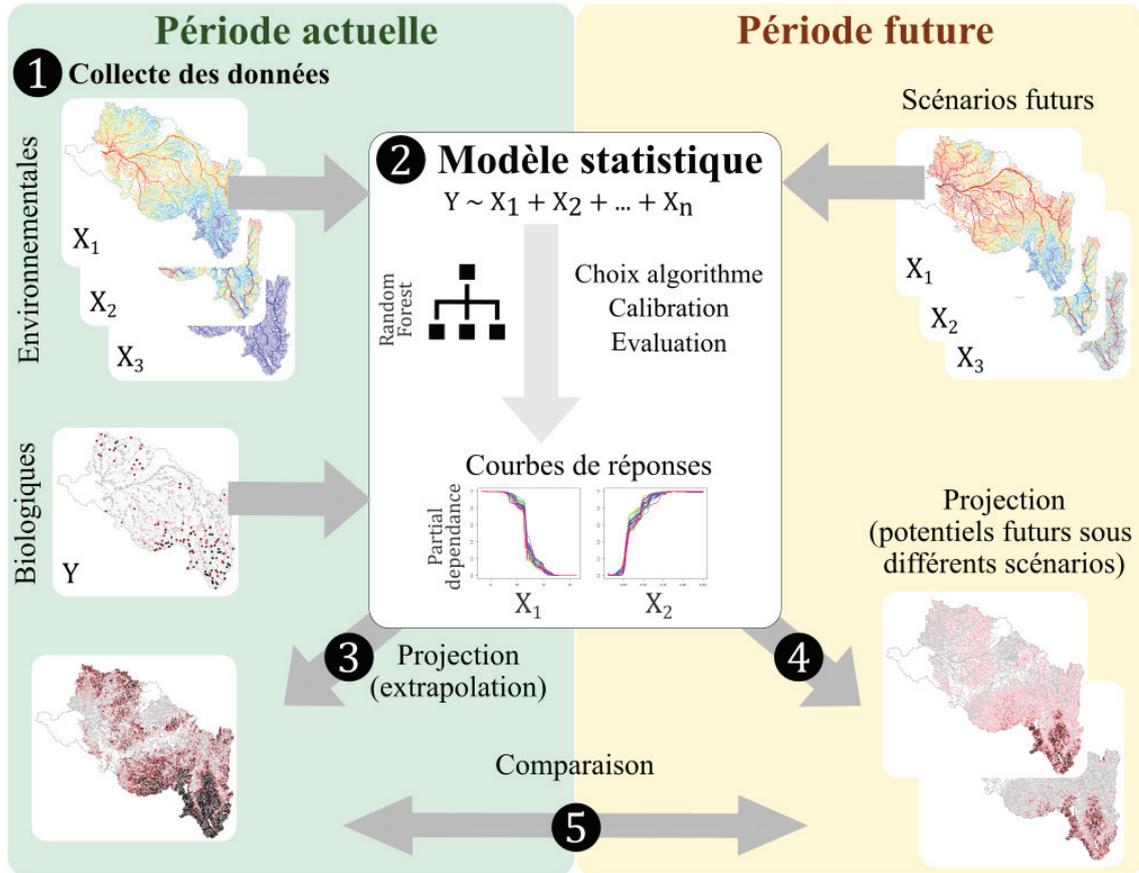


FIGURE 10

Principe des modèles statistiques de distribution d'espèces, basé sur l'exemple de modélisation par *random forest*. Les différentes étapes de la modélisation sont (1) collecte des données, (2) construction du modèle statistique, (3) projections spatiales et (4) temporelles, par exemple sous scénarios de changement climatique et (5) analyse des modifications de distribution. Modifié d'après Guisan *et al.* (2017).

Les modèles de distribution d'espèces permettent de déterminer la niche réalisée d'une espèce sur un territoire donné. Ces modèles sont aussi très utilisés pour projeter les distributions potentielles futures des espèces, afin de comprendre les impacts du changement climatique et d'établir des mesures de conservation adéquates (Guisan *et al.*, 2017). La transférabilité du modèle, c'est-à-dire sa capacité de généralisation à d'autres contextes

spatiaux et/ou temporels, dépend notamment de la représentativité des conditions environnementales du jeu de projection par rapport au jeu de calibration du modèle (Peterson *et al.*, 2011). C'est là qu'intervient la dualité de Hutchinson évoquée précédemment, qui lie l'espace géographique (G) à l'espace environnemental, soit l'espace formé par les conditions environnementales rencontrées dans G (Figure 3). Si les deux espaces environnementaux associés à la calibration et la projection sont similaires, le modèle devrait présenter une transférabilité élevée. Dans le cas contraire, le modèle est projeté dans des conditions qui lui sont inconnues, c'est-à-dire en condition d'extrapolation pure (Elith *et al.*, 2010), ce qui, bien que très pratiqué, est aussi vivement critiqué (Thuiller *et al.*, 2004 ; Buisson *et al.*, 2010 ; Peterson *et al.*, 2011). En effet, la projection est souvent utilisée dans des contextes (spatiaux et/ou temporels) où il n'existe pas de données d'observation permettant l'évaluation des performances prédictives du modèle (Elith & Leathwick, 2009). De plus, la différence entre les conditions environnementales couvertes par le modèle et les conditions de projection sont susceptibles d'entraîner un écart important par rapport à la réalité (Thuiller *et al.*, 2004). Par exemple, Gies *et al.* (2015) ont trouvé que leurs modèles de distribution de 11 espèces de macroinvertébrés avaient une faible capacité de transférabilité entre bassins versants, vraisemblablement induite par des différences d'occupation des sols. Il existe tout de même des cas où la transférabilité des modèles est élevée (*e.g.* Valavi *et al.* 2023). Finalement, quel que soit le cadre de la projection, il s'agira toujours d'une distribution *potentielle* issue du modèle.

Encart 1. SDM, ENM, ... débats sémantiques ou différences conceptuelles ?

Il existe un grand nombre de termes pour désigner les modèles de distribution basés sur le concept de niche écologique : SDM, *Habitat Suitability Models* (HSM), *Ecological Niche Models* (ENM), *Climate Envelope Models* (CEM), etc. (Guisan *et al.*, 2017 ; Peterson *et al.*, 2011). Dans certaines publications, ces termes sont considérés interchangeables. Dans d'autres, le choix du terme dépend de l'objectif du modèle : (i) étudier la distribution géographique et son évolution dans différents contextes (*e.g.* évolutions sous changement climatique ; SDM), (ii) dresser la relation de l'espèce avec son environnement dans une zone géographique donnée (via les courbes de réponses ; HSM ; Guisan *et al.*, 2017), ou (iii) approcher la niche écologique (ENM ; Peterson *et al.*, 2011) . Les auteurs T.Peterson et J.Soberón mettent un point d'honneur à faire la distinction entre SDM et ENM au travers de nombreuses publications (voir notamment Soberon & Peterson 2005 ; Soberon & Nakamura 2009 ; Peterson *et al.* 2011). Selon eux, bien que les principes de la modélisation en tant que tels soient les mêmes, ces deux types de modèles sont conceptuellement bien différents. Un ENM a pour vocation de modéliser la niche écologique fondamentale d'une espèce (Soberón, 2007). Leur définition d'ENM semble inclure tous les modèles de distribution basés sur l'environnement et des données *presence-only* qui seraient les seules qui permettraient d'approcher la niche fondamentale. Néanmoins, ce dernier point est problématique car les auteurs n'expliquent pas clairement en quoi des données de *presence-only* (par rapport aux données de présence/absence) permettraient de surmonter l'impossibilité d'approcher la niche fondamentale (espace environnemental) à partir de données d'échantillonnage provenant du terrain (espace géographique ; Elith & Leathwick, 2009). D'autre part, l'utilisation du terme « niche » dans le contexte de modèles corrélatifs est critiquée car une approche plus mécaniste est requise pour comprendre le lien entre la *fitness* et l'environnement d'une espèce (Kearney, 2006 ; Wagner *et al.*, 2023).

Je considère donc que l'espace théorique de la niche fondamentale n'est pas atteignable à partir de données d'observations seules. J'ai donc choisi de retenir le terme SDM, qui est aussi le terme le plus générique selon Elith & Leathwick (2009), pour nommer la modélisation de la distribution des espèces employée au cours de cette thèse.

Dans le cadre de cette thèse, je me suis attachée à décortiquer les relations entre les organismes aquatiques et leur environnement. La modélisation de la distribution des abondances et l'identification de ces déterminants environnementaux au sens large (*i.e.* biotiques et abiotiques) à partir de données d'observation dans un espace géographique, le tout dans un contexte de changement climatique, a pour objectif de décrire la niche occupée actuelle et la distribution potentielle future des espèces (au sens de niche réalisée ; voir

Encart 1). Ainsi, les modélisations présentées dans ce manuscrit sont basées sur le concept de niche réalisée d'Hutchinson. Afin d'obtenir la vision la plus complète de la niche réalisée, nous avons adopté une approche plus contemporaine en intégrant des prédicteurs représentatifs des différentes composantes qui la façonne, soit des facteurs abiotiques, biotiques et représentatifs de la dispersion des espèces.

C.3. Hypothèses en modélisation de la distribution des espèces

Les SDM reposent sur des hypothèses théoriques (hyp. 1 à 7) et méthodologiques (hyp. 4 à 7) fortes (Guisan *et al.*, 2017) :

- **Hyp. 1 : équilibre de la relation espèce-environnement** - les SDM sont basés sur le postulat que l'espèce d'intérêt est à l'équilibre démographique dans son environnement à l'échelle de la zone d'étude. Ceci implique qu'elle a colonisé l'ensemble des conditions qui lui sont favorables et que sa distribution et ses abondances sont stables dans l'espace et dans le temps ;
- **Hyp. 2 : les déterminants environnementaux majeurs sont intégrés comme prédicteurs dans le modèle** - l'absence d'un prédicteur important induirait un manque de performance, voir un biais dans le modèle ;
- **Hyp. 3 : les données biologiques observées sont appropriées à l'objectif de la modélisation** - elles doivent être représentatives de la niche écologique de l'espèce sur la zone d'étude ;
- **Hyp. 4 : pertinence de la méthode de modélisation** - en fonction des objectifs appliqués de la modélisation (*e.g.* performances prédictives), de la qualité des données d'entrées ;
- **Hyp. 5 : erreur de mesures des prédicteurs limitée**, dans le cas contraire, l'erreur se répercuterait dans les prédictions du modèle ;
- **Hyp. 6 : données biologiques peu biaisées** - l'influence des erreurs induites par les biais d'échantillonnage a fait l'objet de plusieurs études dans le cadre de la modélisation de la distribution des espèces aquatiques et montre un impact sur les prédictions des modèles qui dépend notamment de la détectabilité des espèces (Comte & Grenouillet, 2013b ; Vaudor *et al.*, 2015) ;

- **Hyp. 7 : indépendance des données observées pour les espèces** - en théorie des probabilités, l'indépendance est l'absence de lien entre la survenue de deux événements. L'indépendance totale des données est difficile à vérifier. La violation de l'hypothèse d'indépendance peut être induite par plusieurs facteurs (*e.g.* auto-corrélation spatiale) et peut limiter les performances des modèles. Néanmoins, des données totalement indépendantes sont rares, c'est pourquoi il existe des méthodes de correction.

Malgré ce cadre théorique clair, la réunion de l'ensemble de ces hypothèses méthodologiques se vérifie rarement dans les jeux de données à disposition (Guisan *et al.*, 2017).

C.3.1. Modèles de distribution à l'échelle de la communauté d'espèces

Il existe plusieurs méthodes pour prédire la communauté à partir de données à l'échelle de l'espèce (Ferrier & Guisan, 2006) : (i) prédire en premier et assembler ensuite, (ii) assembler les données puis prédire la communauté et (iii) assembler et prédire simultanément les données à l'échelle de la communauté. Lorsqu'il s'agit de prédire les effets du changement climatique sur les organismes, l'approche la plus courante consiste à prédire les projections par espèce puis à les assembler pour former la communauté (Guisan *et al.*, 2017). En effet, la plupart des études portent sur les réponses à l'échelle de l'espèce, qui sont plus difficiles à retranscrire avec un modèle à l'échelle de la communauté (Ferrier & Guisan, 2006). De plus, les réponses aux changements climatiques sembleraient plutôt se produire à l'échelle spécifique (Guisan *et al.*, 2017). Mais les modèles à l'espèce ne tiennent pas compte de la structuration de la communauté et des liens entre les espèces, qui sont amenés à être modifiés sous l'effet du changement climatique (Jackson *et al.*, 2001 ; Lucas *et al.*, 2023). En général, les prédicteurs mobilisés dans les SDM se limitent aux facteurs abiotiques, en négligeant les interactions biotiques qui peuvent pourtant présenter des effets confondants avec l'environnement (Elith & Leathwick, 2009). Les relations inter-spécifiques sont inhérentes au fonctionnement écologique de la communauté et bien que leur intégration suscite un intérêt grandissant actuellement, les données nécessaires sont difficiles à acquérir (Lucas *et al.*, 2023).

Les *Joint-Species Distribution Models* (JSDM) font partie des approches permettant de prédire directement la communauté dans son ensemble en tenant compte des interactions potentielles (Ovaskainen, 2020). Cette catégorie de modèles, plus récente que les SDM, permet de modéliser l'ensemble des espèces d'une communauté simultanément, en

considérant les liens entre toutes les espèces constituant la communauté (Pollock *et al.*, 2014). Contrairement aux SDM qui assument des réponses à l'environnement propres à l'espèce, les JSDM assument des patrons de réponses communs, en fonction de traits autécologiques similaires et/ou de relations phylogénétiques. À partir des données d'un assemblage d'espèces, ces modèles bayésiens construisent une matrice de co-occurrence des espèces qui est reliée aux prédicteurs environnementaux (Ovaskainen, 2020). Ils permettent la décomposition des co-occurrences entre espèces en patrons de réponses communs face à l'environnement et en co-occurrence résiduelle, dont certaines peuvent retranscrire les interactions biologiques. Ces modèles deviennent de plus en plus courant pour l'étude des communautés marines, et commencent aussi à être appliqués avec succès aux espèces aquatiques dulçaquicoles (voir Wagner *et al.* 2020 pour un exemple sur les poissons d'eau douce). Néanmoins, le développement de ces modèles et notamment l'interprétation des corrélations résiduelles sont en cours, et il apparaît toujours difficile de démêler les simples co-occurrence des réelles interactions inter-spécifiques (Pollock *et al.*, 2014).

L'un des objectifs de ce travail étant de comprendre les modifications de distribution induites à l'échelle spécifique, les modèles à l'espèce semblent tout indiqué. De plus, les modifications induites par les effets du changement climatique opèrent différemment sur chaque espèce et ne sont pas homogènes à l'échelle de la communauté (Buisson *et al.*, 2008b ; Comte & Grenouillet, 2015 ; Floury *et al.*, 2018). C'est pourquoi nous avons préféré une modélisation espèce par espèce.

C.4. Modélisation de la distribution des abondances des espèces

C.4.1. Problématiques spécifiques à la modélisation de la distribution des abondances

La modélisation d'une donnée binaire comme la présence/absence d'une espèce est devenue relativement classique, et les méthodes existantes offrent des résultats très performants (Guisan *et al.*, 2017). La littérature sur la modélisation de la distribution des espèces en présence/absence, c'est-à-dire l'identification des aires de répartition, est foisonnante, tant en publications qu'en outils informatiques. Par exemple, un récent recensement des packages R dédiés à la modélisation en dénombre un total de 60, dont certains sont dédiés à la modélisation de la distribution des abondances (Sillero *et al.*, 2023). En revanche, le développement des modèles de distribution des abondances qui soient performants et applicables à un large panel d'espèces fait face à plusieurs verrous scientifiques (Potts &

Elith, 2006 ; Waldock *et al.*, 2021).

La première limite réside dans la disponibilité des données d'abondance de qualité à large échelle (Thuiller, 2007). Les bases de données disposant d'une couverture spatiale et d'un effort d'échantillonnage suffisants à large échelle (> 1 000 km²) pour plusieurs espèces ne sont pas nombreuses. Néanmoins, grâce à certains suivis imposés par la réglementation et à d'autres projets de suivis de la biodiversité, ce type de base de données tend à se développer et à être mis à disposition publiquement (Chase *et al.*, 2019 ; Comte *et al.*, 2021a). La qualité des données, liée à leur acquisition, peut également jouer sur les performances des modèles d'abondance. Par exemple, des problèmes de détection peuvent conduire le modèle à considérer qu'un habitat n'est pas favorable à une espèce alors qu'elle n'a juste pas été contactée lors de l'échantillonnage (Vaudor *et al.*, 2015 ; Guisan *et al.*, 2017). Des méthodes existent pour évaluer ces limites de détection et leurs impacts dans la modélisation (Buisson & Grenouillet, 2009 ; Comte & Grenouillet, 2013b).

La seconde limite la plus importante est liée à la donnée d'abondance en elle-même. Dans de très nombreux cas, la distribution des abondances présente une sur-inflation en zéros, c'est-à-dire un grand nombre d'absence par rapport aux abondances >0 (Figure 11 ; Potts & Elith, 2006). Cette asymétrie dans la distribution limite les performances d'un grand nombre de catégories de modèles. Outre la sur-inflation en zéros, certaines espèces peuvent présenter une sur-dispersion de leurs abondances, avec une queue de distribution très étendue (*i.e.* quelques occurrences avec des abondances très élevées ; Figure 11), que les modèles classiques ont du mal à reproduire. La sur-inflation en zéros et la sur-dispersion vont souvent de pair, et la première est très importante chez les espèces rares (Jeliazkov *et al.*, 2022). En général, les modèles sont adaptés au cas par cas, en fonction de la distribution observée, et sont difficilement généralisables à un grand nombre d'espèces (Waldock *et al.*, 2021).

C.4.2. Les différentes stratégies de modélisation de la distribution des abondances

Les concepts écologiques expliquant la distribution des abondances des espèces considèrent que les processus à l'origine de la présence d'une espèce et ceux expliquant son abondance sont différents (Chesson, 2000). En effet, les abondances seraient expliquées par l'intervention de mécanismes très locaux, comme les interactions biotiques ou les processus stochastiques de la dynamique de la population. Sous l'hypothèse de processus dis-

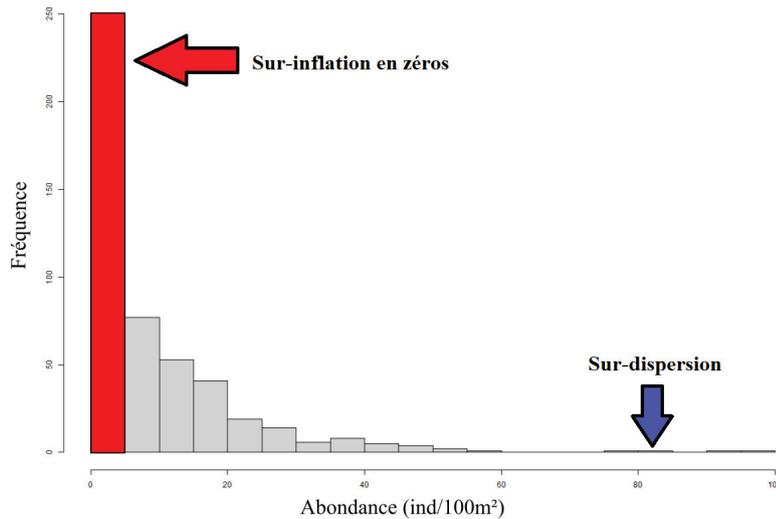


FIGURE 11

Exemple de distribution avec une sur-inflation en zéros. Cet exemple montre aussi une sur-dispersion de la distribution des abondances, avec quelques occurrences de fortes abondances au-delà de 60 ind/100m².

jointes à l'origine des présences et des niveaux d'abondance, une approche possible est une modélisation en deux étapes (Potts & Elith, 2006). Un premier modèle est dédié à prédire les présences/absences de l'espèce, puis un second modélise les abondances sur les sites où l'espèce est prédite présente par le premier modèle. Deux sous-catégories de ce type de modèles sont fréquemment utilisés : les modèles *delta* et les modèles *hurdle*. Le modèle *delta* considère que les absences sont générées par un processus unique, liée à la favorabilité de l'habitat, tandis que le modèle *hurdle* considère que les absences proviennent d'une composante liée à la favorabilité de l'habitat et d'une composante stochastique (Potts & Elith, 2006 ; Waldock *et al.*, 2021).

Une autre stratégie considère les abondances, absences incluses, directement comme variable réponse du modèle. Cela suppose que les mêmes processus déterminent l'aire de répartition et les niveaux d'abondance de l'espèce modélisée. L'idée de réaliser un modèle en un seul temps est plutôt séduisante, notamment de par sa simplicité et parce qu'elle limite la propagation de l'erreur du premier modèle de présence/absence au second modèle d'abondance.

C.4.3. Les différentes approches algorithmiques

Il existe de nombreux algorithmes pour la modélisation de la distribution des espèces (Peterson *et al.*, 2011 ; Guisan *et al.*, 2017). Ils sont généralement regroupés en cinq caté-

gories : les enveloppes géographiques ou climatiques, les méthodes de régression, les méthodes de classification, les méthodes d'apprentissage (*machine learning*) et les méthodes bayésiennes ; une sixième catégorie regroupe les approches mixtes (*i.e.* combinant plusieurs des premières approches ; Guisan & Zimmermann 2000). Plusieurs études ont évalué les performances de différentes approches et algorithmes pour modéliser la distribution des abondances (*e.g.* Potts & Elith 2006 ; Waldock *et al.* 2021 ; Jeliakov *et al.* 2022). Globalement, les performances des modèles de distribution d'abondance n'atteignent pas ceux de la modélisation des présences/absences et des développements sont encore nécessaires. Au vu de l'état de l'art actuel, les méthodes de *Machine Learning* (ML) semblent offrir les modèles de distribution des abondances avec les meilleures performances prédictives potentielles (Waldock *et al.*, 2021 ; Valavi *et al.*, 2023). La distribution des abondances est la résultante de nombreux facteurs, dont des processus stochastiques, qui se traduisent par une relation complexe avec les conditions environnementales. Les méthodes de ML permettent justement de modéliser des relations simples à complexes sans *a priori* sur leur forme et sont plus flexibles (*i.e.* capables de s'adapter à différents jeux de données) que les approches de régression (Valavi *et al.*, 2023). Par exemple, elles sont capables de gérer dans une certaine mesure la sur-inflation en zéros et la sur-dispersion de l'abondance sans demander une paramétrisation au cas par cas. Elles ont par ailleurs montré de meilleures capacités prédictives en-dehors du seul cadre écologique (Snelder *et al.*, 2013). De plus, le côté « boîte noire » de ces approches tend à se résorber avec le développement de méthodes d'interprétation dédiées qui permettent de comprendre la relation entre la variable réponse et les prédicteurs construite par le modèle (Olden & Jackson, 2002 ; Cha *et al.*, 2021 ; Molnar, 2022).

D. Cadrage de l'intérêt et de l'échelle d'étude de la thèse

D.1. Objectifs de compréhension des processus structurants les communautés aquatiques dans un contexte de changements globaux

Les conséquences des changements globaux, notamment du changement climatique, sur la biodiversité ont suscité un vif intérêt dans la compréhension des processus écologiques en jeu dans la structuration des communautés (Baptist *et al.*, 2014 ; GIEC, 2023). La rapidité des modifications observées rend d'autant plus urgent ce besoin de connaissances. La menace du changement climatique vient s'ajouter aux nombreuses pressions

déjà existantes sur les milieux aquatiques dulçaquicoles : changement d'usage des sols, contrôle de l'hydrologie, pollution, espèces invasives et sur-exploitation (Dudgeon, 2019). Ces milieux regroupent près d'un tiers de la faune vertébrée mondiale et font partie des écosystèmes les plus menacés (Dudgeon, 2019).

D.2. Quelle échelle d'étude idéale ? Le bassin versant

La construction de SDM à large échelle sont nécessaires pour produire des modèles généralisables et comprendre les patrons de structuration globaux (Thuiller, 2007). En hydrobiologie, plusieurs unités géographiques peuvent être considérées : le micro-habitat, la portion de cours d'eau, le cours d'eau de sa source à son embouchure et le bassin versant. Le bassin versant apparaît la meilleure échelle d'étude pour comprendre les processus qui structurent les abondances à large échelle pour plusieurs raisons. (i) Il constitue une unité écologique à part entière, un biome, qui peut être considéré relativement indépendant des échanges exogènes car, bien que des flux entre bassins versants existent, ils restent minimes par rapport aux flux intra-bassin. (ii) Du fait de son isolement, le bassin versant possède sa propre histoire biogéographique et donc ses propres communautés. (iii) L'échelle du bassin versant est plus adaptée qu'une étude à une échelle supra-régionale (Alabia *et al.*, 2018). En effet, les SDM construits sur des échelles très larges souffrent souvent d'un biais spatial important, et ils sont encore plus limités dans l'exploration des patrons de distributions aux échelles inférieures. De plus, le changement climatique présente une variabilité spatiale importante, de même que la réponse des espèces à ce dernier, ce qui justifie d'autant de travailler à une échelle régionale.

L'absence de données disponibles d'une résolution spatiale suffisante et ayant un lien causal dans la distribution spatiale des espèces est souvent évoquée comme cause au manque de performance des SDM d'abondance (Guisan *et al.*, 2017). Par exemple, dans le cas des organismes aquatiques, l'utilisation de la température de l'eau au lieu de la température de l'air représenterait mieux les conditions de vie des organismes aquatiques (Caissie, 2006). Ce type de données fines et à haute résolution spatiale sont disponibles sur tout le réseau hydrographique du bassin versant de la Loire ($\sim 10^5$ km²), ainsi que les données d'abondances de nombreux taxons. Par ailleurs ce bassin subit moins de pressions anthropiques, notamment sur son hydromorphologie, par rapport à d'autres bassins français (voir chapitre II; Moatar *et al.*, 2022). Le bassin versant de la Loire a donc été

choisi comme modèle d'étude.

D.3. Objectifs de la thèse et organisation du manuscrit

La suite de ce manuscrit s'articule autour de six chapitres, chacun étant associé à un objectif et à un axe de travail particulier (Figure 12). Le **chapitre II** présente le bassin versant de la Loire, qui est la zone d'étude principale de ce travail, ainsi que la constitution des jeux de données mobilisés. Le **chapitre III** a pour objectif de déterminer les variables environnementales principales qui structurent les communautés d'organismes aquatiques du bassin versant de la Loire, en évaluant d'une part l'intérêt d'inclure des variables directes par rapport à des proxys (*e.g.* température de l'eau *vs.* température de l'air), et d'autre part l'apport d'une approche multi-compartiments biologiques (poissons et macroinvertébrés). La stratégie de modélisation des abondances des espèces de poissons et de macroinvertébrés ligériennes ainsi que l'apport des relations trophiques dans les modèles sont présentés en **chapitre IV**. L'objectif principal est de proposer des modèles de la distribution des abondances flexibles, et de construire des modèles performants pour un maximum d'espèces de poissons et de genres de macroinvertébrés. L'objectif secondaire est d'évaluer l'intérêt d'intégrer des descripteurs des interactions biotiques dans la modélisation. Le **chapitre V** évalue l'impact du changement climatique sur la distribution potentielle des abondances à partir des modèles construits dans le chapitre précédent. Ce chapitre présente ainsi les projection des distributions des abondances des espèces de poissons et de genres de macroinvertébrés sous divers scénarios de changement climatique et en évalue les effets sur la structuration spatiale des communautés aquatiques projetées. Le chapitre VI clôture la présentation des analyses réalisées pendant la thèse par une évaluation de la transférabilité des patrons spatiaux obtenus à l'échelle du bassin versant de la Loire vers d'autres grands bassins versants français (*i.e.* la Seine, la Rhône et la Garonne). Enfin, le **chapitre VII** présente la discussion générale et les perspectives de ce travail.

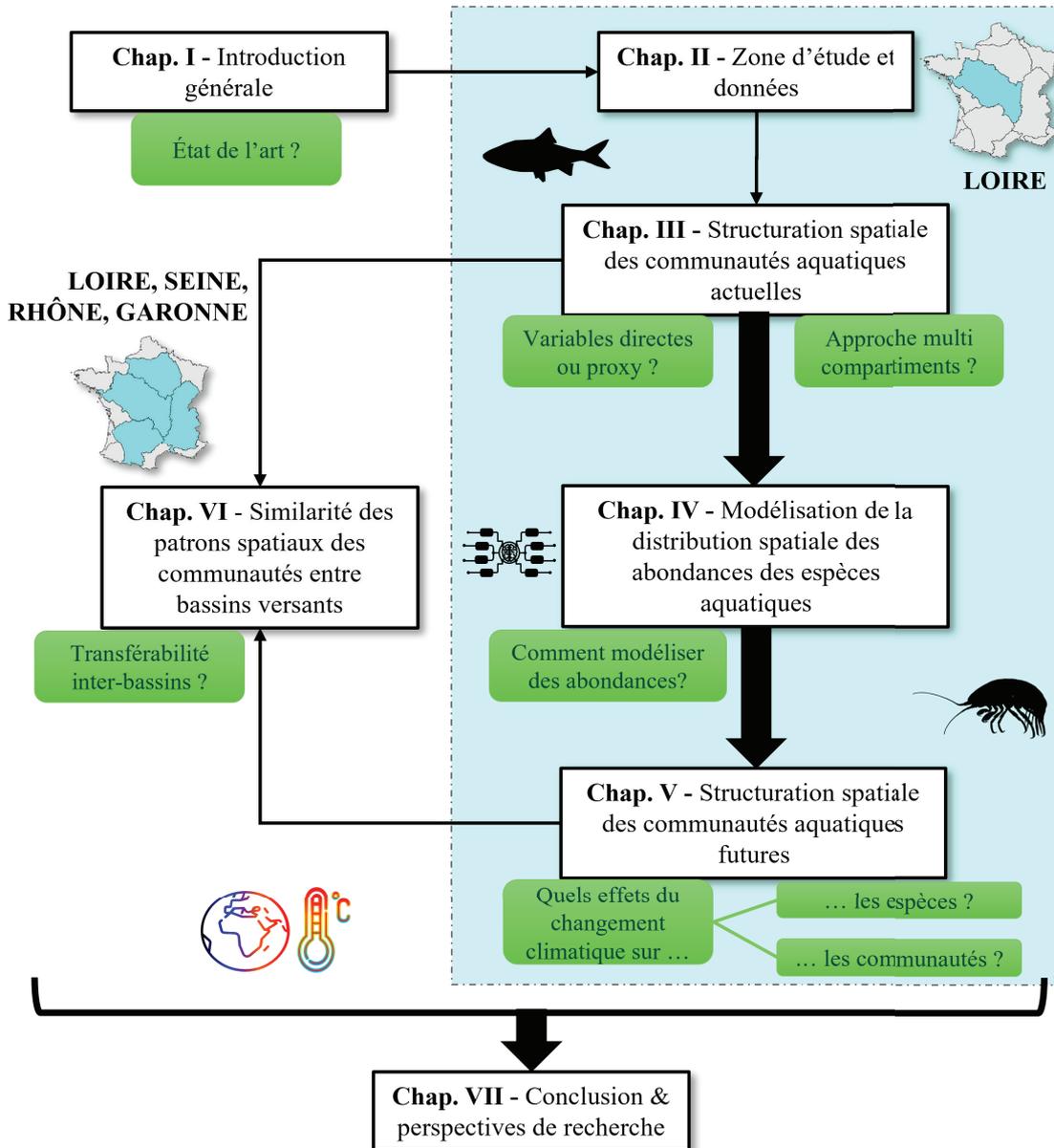


FIGURE 12

Résumé graphique du plan du manuscrit. Les principales questions attenantes à chaque chapitre sont illustrées par les bulles vertes. Les chapitres II à V portent exclusivement sur le bassin versant de la Loire et le chapitre VI est étendu à la Seine, la Garonne et le Rhône.

Les analyses et résultats du chapitre III ont fait l'objet d'une publication dans la revue Plos One en 2022. Le texte complet de l'article est disponible en annexe A.

* Picard, C., Flourey, M., Seyedhashemi, H., Morel, M., Pella, H., Lamouroux, N., Buisson, L., Moatar, F., & Maire, A. (2022). Direct habitat descriptors improve the understanding of the organization of fish and macroinvertebrate communities across a large catchment. *PLOS ONE*, **17** (9), e0274167. doi : 10.1371/journal.pone.0274167

D.4. Contexte et enjeux industriels

Ce travail de thèse s'inscrit dans un contrat CIFRE (Convention Industrielle de Formation par la Recherche) entre l'unité de recherche RiverLy de l'Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE) (Centre de Lyon-Grenoble-Auvergne - Rhône-Alpes à Villeurbanne), et le Laboratoire National d'Hydraulique et Environnement (LNHE) de la R&D d'Électricité De France (EDF) (Chatou) qui travaillent en partenariat dans le cadre de l'équipe de recherche commune HYNES. En tant qu'organisme de recherche public, INRAE vise à apporter des éléments scientifiques de compréhension des écosystèmes relatifs à leur qualité, leur fonctionnement et leur dynamique. En tant qu'utilisateur de la ressource en eau, EDF s'intéresse à l'incidence environnementale des activités de production d'électricité et à l'évolution des communautés aquatiques aux droits des ouvrages de production d'électricité sur les milieux aquatiques et terrestres. D'autre part, l'entreprise est soumise à la réglementation concernant l'exploitation de la ressource en eau (e.g. seuil de température pour les rejets thermiques des Centres Nucléaires de Production d'Électricité (CNPE)). Ces règles ont été établies à partir de travaux scientifiques (Verneaux, 1977a) qui présentent un certain nombre de limites au vu de leur ancienneté, de la zone géographique d'étude et des moyens analytiques mis en place (Wasson, 1989). Dans le contexte du changement climatique, les hydrosystèmes et les communautés aquatiques qui s'y trouvent ont commencé à changer, et ses modifications vont se poursuivre. La multiplication des récents événements extrêmes qui se sont produits en France (e.g. canicules et étiages historiques de la rivière Loire lors de l'été 2022) et ailleurs peuvent déjà en témoigner. Ces événements provoquent un échauffement des eaux sans précédents aux abords des CNPE et les limites réglementaires des rejets thermiques sont de plus en plus souvent approchées, voire atteintes, et ce parfois en amont des sites eux-mêmes. La survenue de ces changements appelle de plus en plus régulièrement à une adaptation de la réglementation contexte actuel du changement climatique. L'objectif industriel de ce travail de thèse est d'apporter des éléments de compréhension quant au lien entre l'environnement changeant et les modifications attendues au niveau des communautés aquatiques, afin de participer à l'anticipation de ces changements et à l'adaptation des activités en lien avec ces hydrosystèmes.

Chapitre II

Présentation de la zone d'étude et des données utilisées

Afin de répondre aux objectifs de ce travail de thèse, deux échelles d'étude ont été choisies : (i) le bassin de la Loire pour la modélisation de la distribution spatiale des abondances des espèces aquatiques et leur projection sous divers scénarios de changement climatique (chapitres IV et V), et (ii) les quatre grands bassins de France métropolitaine (aussi appelée échelle nationale ci-après) de la Loire, la Seine, le Rhône et la Garonne pour l'évaluation de la similarité des patrons de structuration des abondances des communautés aquatiques entre bassins versants (chapitre VI). Les données biologiques et variables environnementales utilisées à l'échelle nationale pour l'évaluation de la transférabilité sont présentées dans le chapitre dédié (chapitre VI). Le bassin versant de la Loire a été choisi pour la modélisation de la distribution des abondances car différentes données environnementales à haute résolution spatiale et temporelle y sont disponibles. En effet, un grain spatial suffisamment fin et l'utilisation de variables ayant un rôle causal (*i.e.* variables directes, par opposition aux variables proximales, voir chapitre III) comme prédicteurs dans les modèles sont susceptibles d'améliorer la performance des SDM d'abondance (Irving *et al.*, 2022 ; Haesen *et al.*, 2023). Ces variables à haute résolution ne sont pas disponibles sur les autres bassins.

Dans ce chapitre, une première section présente les contextes biologiques et hydroclimatique de la zone d'étude principale, le bassin versant de la Loire. Une seconde section présente la constitution du jeu de données biologique, de l'extraction de la base de données Naiades à l'application de différents filtres en vue des chapitres IV et V. Les variables environnementales qui y ont été associées sont également présentées ici.

A. Présentation de la zone d'étude principale : le bassin versant de la Loire

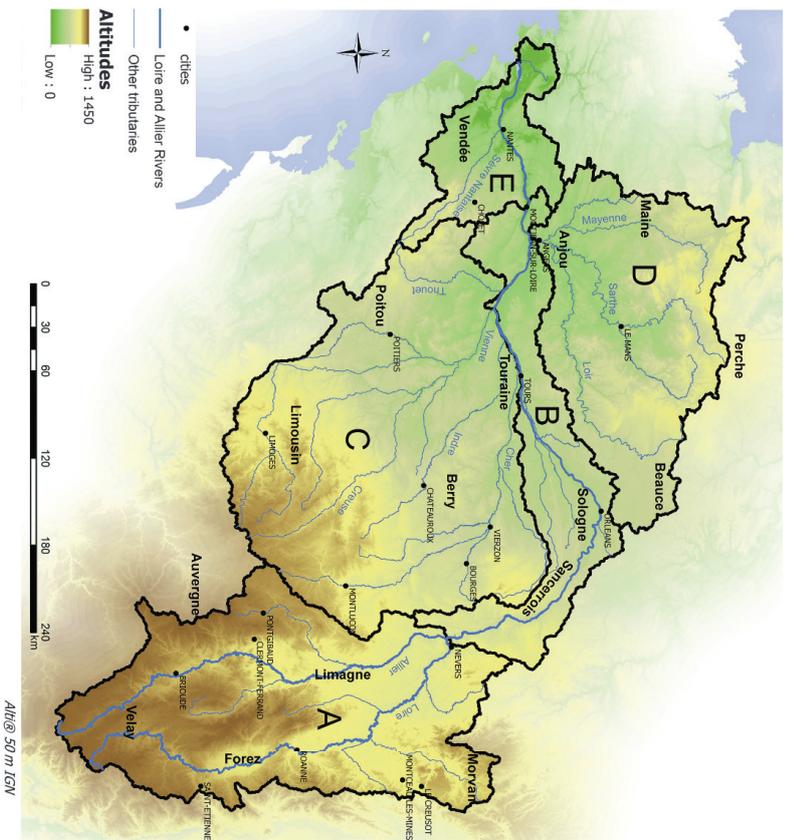
Le bassin versant de la Loire (ou bassin ligérien) est le plus grand de France, avec près de 117 000 km² de surface recouvrant environ 20% du territoire métropolitain et près de 88 000 km de réseau hydrographique (Descy, 2009 ; Beaufort *et al.*, 2016). Le fleuve Loire parcourt 1 012 km entre Gerbier-de-Jonc (Ardèche), sa source, au sein du Massif Central, et Saint-Nazaire (Loire-Atlantique) son estuaire où il se jette dans l'océan Atlantique. Le cours du fleuve suit un axe sud-nord jusqu'à Orléans où il s'infléchit sur un axe est-ouest jusqu'à son estuaire (Figure 13). La Loire est le seul fleuve français à présenter ces deux orientations cardinales perpendiculaires, ce qui a eu un impact important sur son histoire

biogéographique durant les périodes glaciaires (Oberdorff *et al.*, 2011). Ses principaux affluents sont, par ordre d'importance des modules, la Vienne, l'Allier, la Maine (formée par la confluence de la Sarthe et de la Mayenne et seul affluent important de la rive droite de la Loire) et le Cher (Moatar et Dupont, 2016, Tableau 1.3). Elle traverse plusieurs grandes agglomérations françaises telles que Roanne, Nevers, Orléans, Tours ou encore Nantes (Figure 13). Au cours de son périple elle traverse une grande variété de géologies, de paysages et de climats qui lui donnent son caractère unique en Europe. Le bassin de la Loire bénéficie d'un climat continental et méditerranéen sur sa partie sud et d'un climat océanique à mesure qu'on approche de son embouchure. Les précipitations annuelles cumulées sont comprises entre 500 et 1 800 mm (1971- 2000) et les températures de l'air moyennes entre 6 et 12,5 °C. Le bassin comprend des altitudes jusqu'à 1 850 m, et près d'un quart de son linéaire hydrographique est situé à une altitude supérieure à 600 m (Beaufort *et al.*, 2016). Au niveau du Massif Central (Figure 13 - région A) le bassin est majoritairement dédié à la sylviculture et à l'élevage, la topographie étant peu favorable à un autre type d'exploitation (European Environment Agency, 2012). L'agriculture s'intensifie à mesure qu'on s'éloigne de la source de la Loire (Figure 13 - régions B, C et D ; Moatar & Dupont (2016)), avec la présence de grandes plaines céréalières dans le bassin sédimentaire parisien (27 % de la surface agricole utile de blé tendre, 11 % de colza et 5 % de maïs) et un important vignoble le long des coteaux ligérien. L'élevage ainsi que des polycultures sont également présents dans ces zones.

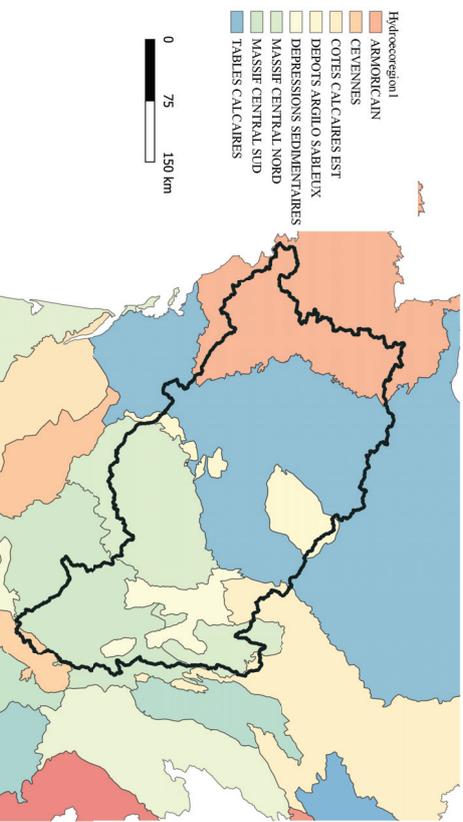
Les données de température de l'eau qui ont été mobilisées dans ce travail n'existent que pour la partie du bassin qui n'est pas soumise à l'influence des marées, c'est-à-dire jusqu'à Montjean-sur-Loire. À ce niveau, exutoire du bassin, le module de la Loire atteint 830 m³-s à Montjean-sur-Loire (1863 - 2023). La partie aval du bassin n'est donc pas considérée (exclusion de la région E sur la figure 13.a).

A.1. Régimes hydrologique et thermique

La variabilité spatiale du climat, de la géologie et des paysages, décrite à travers les 8 hydroécotémoins de niveau 1 (Figure 13) conditionne les régimes hydrologique et thermique des cours d'eau du bassin de la Loire.



(a)



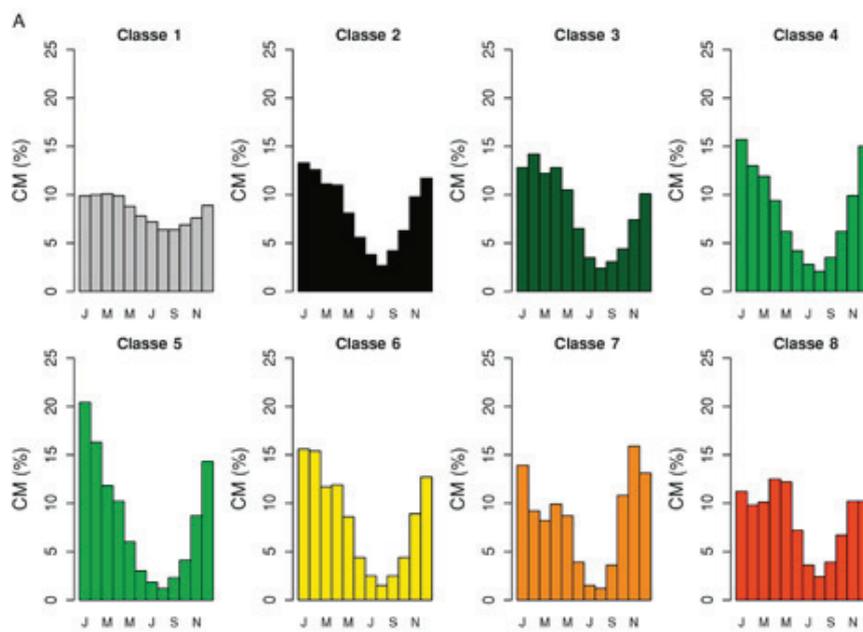
(b)

FIGURE 13
 (a) Carte altitudinale du bassin versant de la Loire (modèle numérique de terrain à 50 m, BD Alt@50 m IGN), ses principales régions hydrologiques et ses agglomérations. (A) Loire amont et Allier ; (B) vallée de la Loire ; (C) affluents rive gauche (Cher, Indre, Vienne) ; (D) sous-bassin de la Maine ; (E) sous-bassin de l'estuaire (d'après Moatar & Dupont (2016)). (b) Hydro-écorégions françaises et zoom sur le bassin versant de la Loire (d'après (Wasson *et al.*, 2002)).

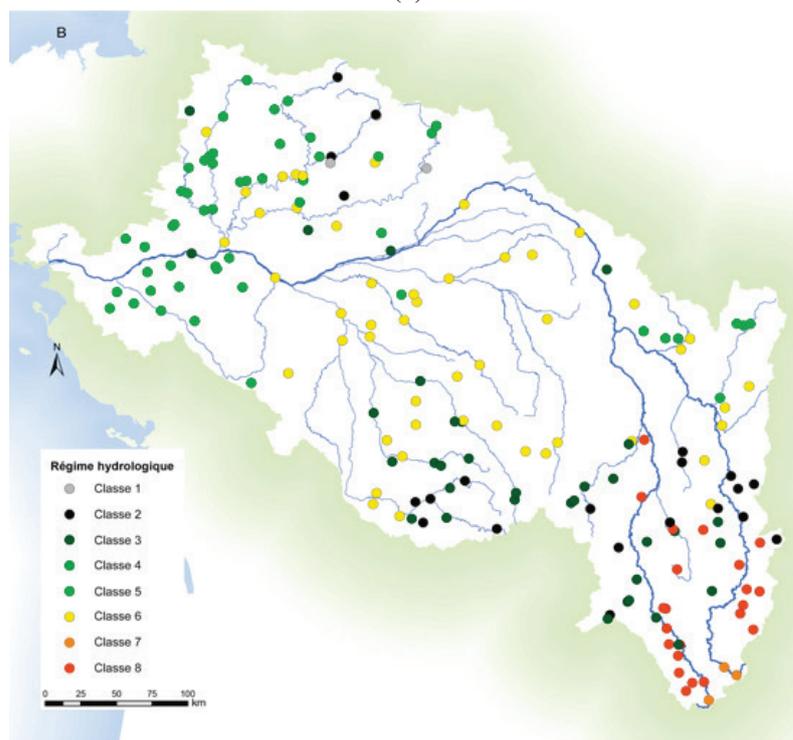
A.1.1. Hydrologie

Comme tous les bassins versants, le bassin ligérien est composé en majeure partie de cours d'eau de tête de bassins avec un ordre de Strahler inférieur à 2, parmi lesquels 70% des tronçons sont d'ordre 1. À l'exception des régimes d'influence nivale et glaciaire, le bassin présente huit des douze régimes hydrologiques identifiés en France (Figure 14), ce qui témoigne d'une grande diversité de conditions hydrologiques (Sauquet *et al.*, 2000 ; Moatar & Dupont, 2016). La haute Loire est caractérisée par un régime pluvio-nival, du fait de la présence du Massif Central (altitude moyenne de la zone A de 1 009 m, débit spécifique moyen de 11,5 l/s/km²), tandis que la basse Loire présente un régime pluvial comme en témoigne les épisodes de crues et d'étiages intenses observés sur le fleuve (Moatar & Dupont, 2016). L'hydrologie du fleuve montre une grande variabilité à court terme mais aussi à long terme, due aux conditions météorologiques changeantes (Figure 14). Les apports d'eaux souterraines, qui ont une influence importante sur l'hydrologie et la thermie des cours d'eau, sont importants notamment dans le bassin sédimentaire parisien (Moatar & Dupont, 2016).

En comparaison avec d'autres bassins, les contraintes d'origine anthropique rencontrées dans le bassin sont en moyenne moins élevées. Le nombre d'ouvrages contraignant la dynamique fluviale n'est pas très élevé, et son espace de liberté au sein de sa plaine alluviale est relativement bien conservé (Moatar *et al.*, 2022). Parmi les ouvrages majeurs existants qui régulent les écoulements de la Loire moyenne et aval, il convient de citer les barrages de Villerest (volume de la retenue 68 à 232 millions de m³, hauteur 59 m) et de Naussac (volume de la retenue 190 millions de m³ - hauteur 50 m ; voir www.eptb-loire.fr) avec un rôle d'écrêtement des crues et de soutien d'étiage. L'hydrologie puissante du fleuve ainsi que sa dynamique sédimentaire intense lui valent le qualificatif de « dernier fleuve sauvage d'Europe » (Moatar & Dupont, 2016).



(a)



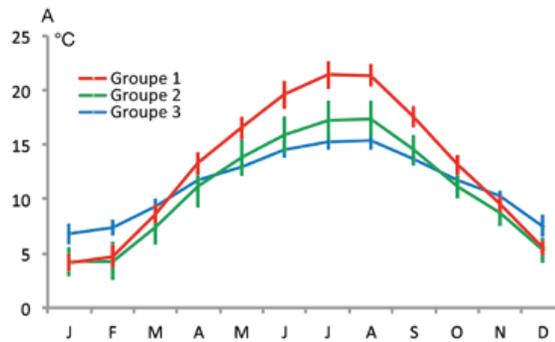
(b)

FIGURE 14

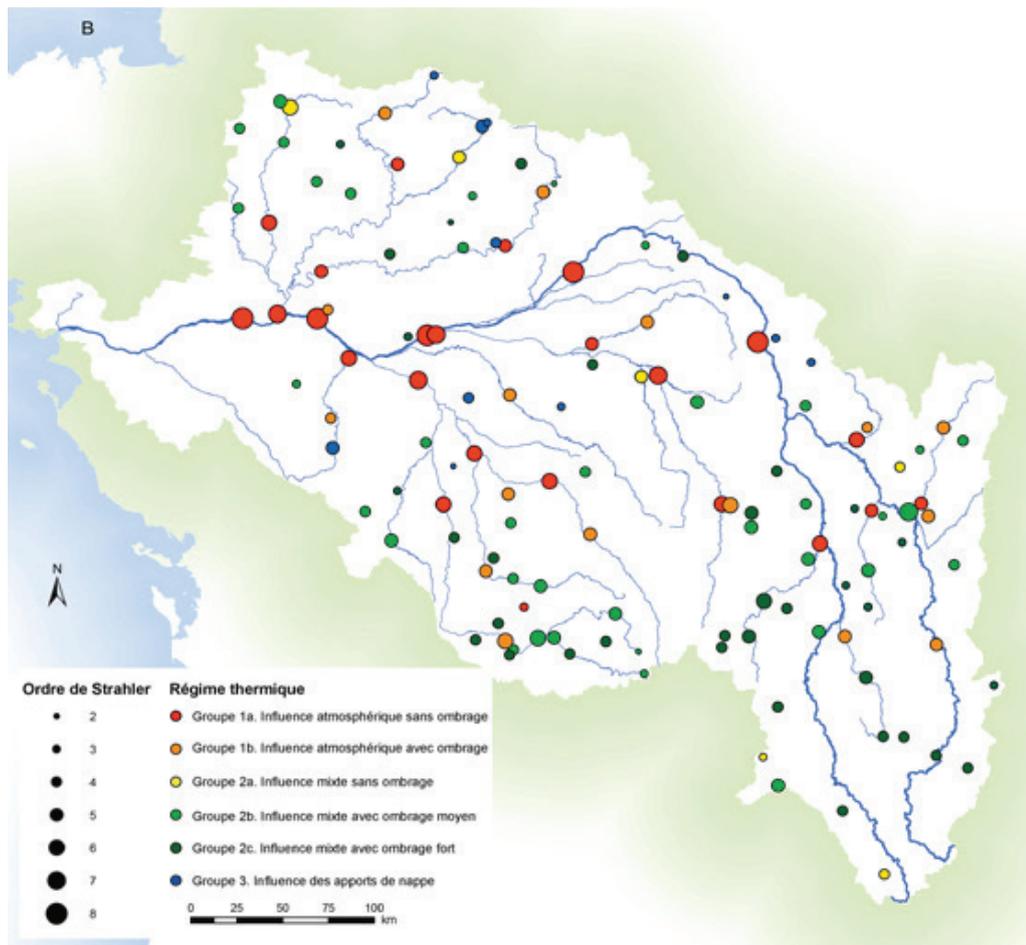
Régimes hydrologiques observés dans le bassin versant de la Loire. (a) Huit classes principales des régimes, les débits mensuels sont exprimés sous forme de rapports des débits mensuels moyens interannuels (CM, en %) (b) localisation des stations avec leur régime hydrologique indiqué. D'après Moatar & Dupont (2016).

A.1.2. Thermie

Les températures de l'eau observées sur le bassin versant de la Loire sont soumises à la diversité des conditions climatiques rencontrées : climat tempéré à l'est et océanique à l'ouest, avec influence du Massif Central sur la haute Loire. Les plus faibles températures sont observées sur les plus petits ordres de Strahler, en têtes de bassins, où la température annuelle ne dépasse pas 13 °C (Beaufort *et al.*, 2016). Les plus hautes températures sont observées pour les grands cours d'eau, d'ordre de Strahler 8, comme le cours principal de la Loire aval. Les régimes thermiques peuvent être classés en trois catégories en fonction des facteurs climatique, lithologique et de la distance à la source (Figure 15 ; (Beaufort *et al.*, 2020b)). Le groupe 1 montre une variation annuelle importante (jusqu'à 18 °C d'amplitude à l'échelle de l'année) et des températures estivales qui excèdent 21 °C en moyenne, contrairement au groupe 3 pour lequel les apports des nappes tamponnent cette variation (amplitude annuelle de 9 °C et températures moyennes estivales autour de 15 °C); le groupe 2 présente un comportement intermédiaire.



(a)



(b)

FIGURE 15

Régimes thermiques observés dans le bassin versant de la Loire. (a) Température de l'eau moyenne mensuelle normalisée par la moyenne sur les trois groupes de stations identifiés. (b) carte des régimes thermiques en fonction de leur facteur de contrôle identifié. La taille des cercles est proportionnelle à l'ordre de Strahler. D'après Moatar & Dupont (2016) et Beaufort *et al.* (2020b).

A.2. La faune aquatique ligérienne et ses particularités

Le bassin ligérien fait partie des écorégions mondiales les moins riches en espèces piscicoles (richesse spécifique entre 12 et 35 en moyenne sur la zone), les zones les plus riches étant situées au niveau de l'équateur (avec une richesse pouvant dépasser 1 000 espèces ; Oberdorff *et al.* (2011)). L'histoire biogéographique du bassin ligérien (voir chapitre I.B.1) et la faible diversité d'habitats peuvent expliquer ce niveau de richesse inférieur par rapport à d'autres régions du monde. Ces observations sont en accord avec le gradient latitudinal observé par ailleurs chez les espèces terrestres, bien qu'elles soient moins contraintes dans leurs déplacements (Griffiths, 2015). Ainsi, le bassin comporte plus de 60 espèces de poissons et plus de 200 taxons de macroinvertébrés dont certains sont non natifs (Moatar & Dupont, 2016 ; Keith *et al.*, 2020). Les conditions hydro-climatiques particulières du bassin lui valent quand même d'abriter certaines espèces remarquables. Les bassins de la Vienne et de la Creuse abritent des populations de la grande mulette (*Margaritifera auricularia*) et de la moule perlière (*M. margaritifera*) deux espèces patrimoniales très rares à l'échelle mondiale et bénéficiant d'un statut de protection (annexe IV directive Habitats et annexe II & III de la Convention de Berne ; Prié *et al.*, 2014. La majorité des espèces migratrices amphihalines présentes en France sont observées sur le cours de la Loire et de nombreuses mesures de protection et de gestion viennent en soutien aux populations d'espèces emblématiques comme le saumon Atlantique (*Salmo salar* ; Directive Habitats-Faune-Flore & Convention de Berne).

Des couloirs de circulation entre les bassins versants et des apports exogènes, volontaires ou non, ont permis à des espèces non natives du bassin de s'y installer. Ces apports d'espèces ont été observés dès l'Antiquité, avec l'introduction de la tanche et de la carpe (*Tinca tinca* et *Cyprinus carpio*). Plus récemment, la circulation des espèces a été facilitée par la mise en place de canaux reliant des bassins versants autrefois non connectés. Par exemple, le blageon, espèce endémique du Rhône, colonise actuellement le bassin ligérien (Keith *et al.*, 2020). L'arrivée de certaines espèces exogènes peut avoir un impact sur les communautés natives de la Loire mais aussi sur les différents usages du bassin. Par exemple, la moule zébrée (*Dreissena polymorpha*) a envahi les milieux jusqu'à coloniser complètement les habitats disponibles, au détriment des espèces présentes. La corbicule (*Corbicula fluminea/fluminalis*) est envahissante au même titre que la moule zébrée, et, grâce à sa grande capacité de filtration, elle participerait à l'oligotrophisation de

la Loire (Diamond *et al.*, 2022). Certaines espèces envahissantes peuvent interagir entre elles, comme la crevette du Danube (*i.e. Dikerogammarus villosus*) qui voit son impact sur les chaînes trophiques ligériennes s'accroître du fait de la présence de la moule zébrée (*invasional meltdown*; Gergs & Rothhaupt (2015)). Il semblerait que la colonisation des milieux par ces espèces soit encore limitée par les conditions hydro-climatiques et l'équilibre des écosystèmes (Moatar & Dupont, 2016) mais ces conditions sont susceptibles d'évoluer, notamment sous l'effet du changement climatique.

A.3. Tendances d'évolution sous l'effet du changement climatique

A.3.1. Les effets observés du changement climatique

L'influence du changement climatique se fait déjà sentir et risque de profondément modifier les conditions hydro-climatiques du bassin versant de la Loire. Depuis 1963, la température moyenne de l'eau estivale a augmenté de 0,44 °C par décennies, soit une augmentation globale de + 2,5 °C entre 1963 et 2019 (Seyedhashemi *et al.*, 2023). Le réchauffement observé montre une nette accélération à partir de 1980 (Seyedhashemi *et al.*, 2023).

La complexité des processus structurants les communautés aquatiques et la multiplicité des pressions (*e.g.* anthropiques) qui engendrent leur modification rendent difficile l'établissement d'un lien causal entre les évolutions hydroclimatiques et les changements observés dans les communautés aquatiques. Néanmoins, l'analyse des tendances passées montre que le changement climatique opère déjà sur les écosystèmes et entraînent des modifications sur tous les niveaux d'organisation de la biodiversité depuis l'échelle cellulaire avec des modifications physiologiques jusqu'à l'échelle des assemblages. L'impact du changement climatique surpasse même parfois celui des pressions anthropiques, et va jusqu'à modifier la structure des tailles des communautés (Daufresne & Boët, 2007). Les espèces aquatiques montrent des changements dans leur distribution géographique, entraînant une réorganisation importante des communautés. Les communautés piscicoles se déplacent pour suivre leur préférence thermique, ce qui conduit à une augmentation de la richesse spécifique locale en faveur des espèces thermophiles (Poulet *et al.*, 2011 ; Souchon & Tissot, 2012 ; Comte *et al.*, 2013 ; Baptist *et al.*, 2014). Les mêmes tendances d'accroissement des abondances de taxons thermophiles et polluo-tolérants sont observées chez les macroinvertébrés, au détriment des taxons typiques des milieux froids à forts cou-

rants (Floury *et al.*, 2013).

A.3.2. Projections des évolutions climatiques futures

De récentes études permettent de dresser la gamme de ces possibles évolutions (Seyed-hashemi *et al.*, 2022). Une diminution de la ressource généralisée en eau est attendue, avec une élévation de la fréquence et de l'intensité des épisodes d'étiages sévères. Une augmentation globale des températures de l'eau est également attendue, avec une augmentation moyenne de 2 à 3 °C (Moatar & Dupont, 2016).

Ces changements environnementaux à venir auront certainement un effet important sur les communautés aquatiques actuelles. Les projections futures laissent supposer une poursuite de ces tendances, avec un déplacement des communautés vers l'amont pour tenter de suivre leur optimum thermique (Comte *et al.*, 2013 ; Buisson *et al.*, 2008b).

Faute de modèles d'abondance performants, les projections d'évolution des communautés aquatiques réalisées jusqu'ici se sont concentrées sur la présence ou l'absence des espèces, qui prévoient une augmentation de la richesse spécifique locale du fait du glissement des communautés plus riches de l'aval vers l'amont. Les informations supplémentaires qui pourraient être apportées par des données d'abondance pourraient permettre de mieux présager de la capacité des populations à perdurer dans le temps. De plus, l'utilisation de la richesse spécifique seule ne permet pas de retranscrire les changements de structuration des communautés et peut masquer certains processus (Kuczynski *et al.*, 2023). Enfin, les modifications observées sur la richesse spécifique peuvent être décorrélés de celles des abondances (Waldock *et al.*, 2021).

B. Constitution des jeux de données biologiques et environnementaux

Cette section présente : (i) l'extraction et la mise en qualité des jeux de données biologiques (bulle bleue sur la figure 16), (ii) le réseau qui constitue l'unité géographique pour la modélisation sur le bassin de la Loire, et (iii) enfin, les variables environnementales mobilisées (bulle verte et orange sur la figure 16) et leur raccord sur le réseau H.

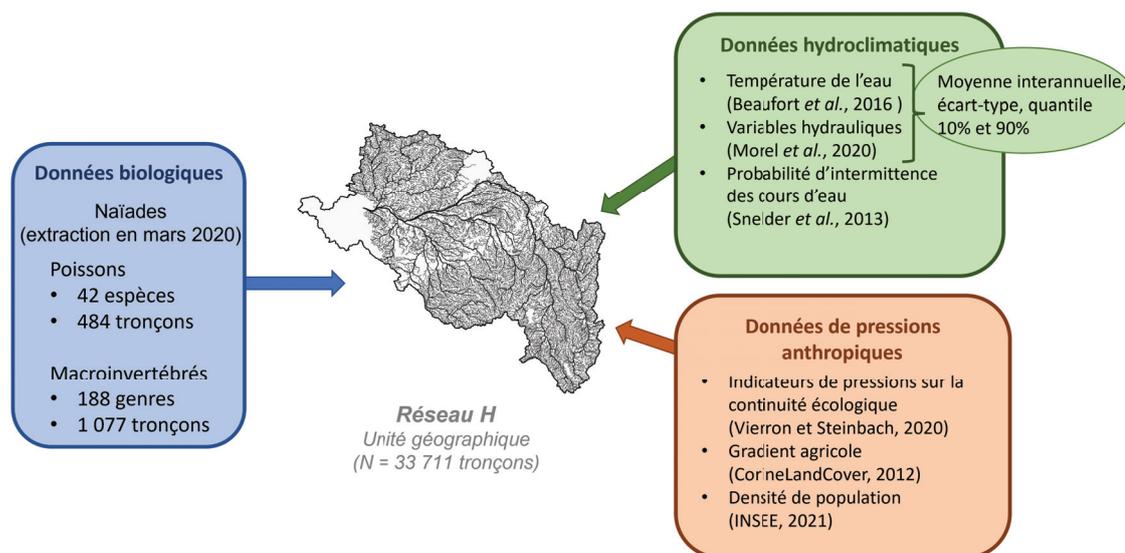


FIGURE 16

Constitution du jeu de données pour la modélisation de la distribution des abondances des poissons et des macroinvertébrés sur le bassin versant de la Loire. Les informations sur leur nombre de taxons et de tronçons échantillonnés à partir de la base Naïades sont précisées. Les variables sont regroupées par catégorie (*i.e.* biologique, hydroclimatique et pression anthropique) et leur source est précisée.

B.1. Les données biologiques

La modélisation de la distribution des abondances à large échelle requiert des données d'abondance à haute résolution et représentatives des communautés rencontrées sur le bassin (*i.e.* dans le cas présent le bassin versant d'un grand fleuve). La réglementation mise en place suite à la Directive Cadre sur l'Eau (DCE ; Commission Européenne, 2000) a imposé le suivi de l'état des masses d'eau sur les plans biologiques, physico-chimiques et morphologiques à l'échelle européenne. En France, c'est à l'Office Français pour la Biodiversité (OFB), aux Directions Régionales de l'Environnement, de l'Aménagement et du Logement (DREAL) et aux Agences de l'eau que cette mission de suivi a été confiée, de même que la gestion et la diffusion des données collectées. Les échantillonnages sont

réalisés suivant des protocoles standardisés afin d'assurer la reproductibilité et la comparabilité spatiale et temporelle des suivis (Poulet *et al.* 2011).

Les données biologiques mobilisées au cours de cette thèse portent sur deux compartiments : les poissons et les macroinvertébrés. Les macroinvertébrés désignent l'ensemble des invertébrés aquatiques mesurant plus de 0,5 mm (taille du vide de maille du filet Surber, le plus couramment utilisé pour échantillonner les macroinvertébrés) durant leur dernier stade de développement et au mode de vie épibenthiques, c'est-à-dire vivant en surface ou dans les premiers centimètres des sédiments (Tachet *et al.*, 2010). Outre la disponibilité des données, le choix s'est porté sur ces deux compartiments biologiques pour plusieurs raisons. Les poissons et les macroinvertébrés sont des organismes clés pour la bioindication de l'état des cours d'eau au travers de différents systèmes de notations du bon état du système étudié. En effet, ils sont de bons intégrateurs des différentes perturbations que subissent les écosystèmes aquatiques (Alric *et al.*, 2020). D'une part, les espèces des deux compartiments présentent différents préférences concernant les conditions abiotiques de leur milieu de vie (*e.g.* différentes préférences vis-à-vis du régime thermique, différentes sensibilités aux polluants chimiques et organiques ; Tachet *et al.* 2010 ; Keith *et al.* 2020). D'autre part, les poissons présentent des durées de vie assez longues, de l'ordre de quelques années, et une mobilité à longue distance rapide. Cela leur permet d'éviter certains dangers tout en intégrant certaines perturbations dont les effets sont plus lents (modifications hydromorphologiques, pollutions chimiques, etc.). Les macroinvertébrés quant à eux présentent des cycles de vie plus rapides, avec plusieurs générations annuelles pour certains taxons, ce qui les rend complémentaires des poissons car ils intègrent les perturbations survenues sur un temps plus court. Par ailleurs, l'un des objectifs étant d'évaluer l'apport des interactions biotiques pour la modélisation de la distribution des abondances, la combinaison de données sur les poissons et les macroinvertébrés permet de tester différents types d'interactions (*e.g.* relations trophiques). En effet, les macroinvertébrés constituent une ressource alimentaire essentielle pour un grand nombre d'espèces piscicoles (Matthews, 1998 ; Gerking, 1994). Enfin, ces compartiments ont déjà fait l'objet d'un certain nombre d'études dans un cadre similaire (*e.g.* Buisson *et al.* 2008a ; Comte & Grenouillet 2013b ; Maire *et al.* 2017).

Les communautés de poissons et de macroinvertébrés sont échantillonnées dans le cadre de réseaux de suivis nationaux selon des protocoles normalisés (Poulet *et al.*, 2011)

dans le but de calculer différents indices de qualité biologique et de surveiller l'évolution de ces communautés et de leurs milieux dans le temps (Commission Européenne, 2000). Ces suivis permettent d'accumuler un grand nombre de données dont l'abondance des taxons recensés. Les données sur les poissons ont régulièrement été mobilisées pour différents travaux sur la distribution des espèces et ses potentielles modifications sous l'effet du changement climatique (Buisson & Grenouillet, 2009 ; Comte *et al.*, 2013).

B.1.1. Les données piscicoles

Les données piscicoles sont issues de la base nationale publique de suivis réalisés par l'OFB et mise à disposition sur le site Naiades (naiades.eaufrance.fr; Poulet *et al.* 2011). Les données de la base ont été extraites à l'échelle de la France métropolitaine le 10 mars 2020 (Office Français pour la Biodiversité, mars 2020). Elles ont été enrichies des données issues des suivis hydroécologiques réglementaires des CNPE d'EDF. Différents protocoles d'échantillonnage sont utilisés pour suivre les communautés des poissons. Tous les protocoles sont basés sur la pêche à l'électricité et le nombre de points d'échantillonnage dépend du gabarit de cours d'eau échantillonnés : pêche complète, pêche par ambiances ou faciès, protocole grand milieu (accompagné de pêche partielle en berge) et échantillonnage ponctuel d'abondance (EPA). La quasi-totalité des opérations (> 95 %) utilisées dans le cadre de cette thèse a été réalisée en pêche complète ou en protocole grand milieu. L'exhaustivité de l'échantillonnage varie en fonction du protocole, comportant plus ou moins de points d'échantillonnage. Seule la pêche complète peut être considérée comme quasiment exhaustive puisque la station est délimitée par des filets en l'absence d'obstacles infranchissables pour les poissons, et toute sa surface est parcourue avec l'anode deux à trois fois. Ce protocole permet de contacter l'ensemble des espèces de poissons présentes sur la station et d'obtenir une image représentative des abondances de chaque population, bien que la détectabilité varie en fonction des espèces (Vaudor *et al.*, 2015 ; Comte & Grenouillet, 2013b). Chaque individu capturé est identifié à l'espèce, mesuré et pesé avant d'être relâché à la fin de l'épisode de pêche (Tableau 1). Les opérations de pêche pour lesquelles la surface échantillonnée n'était pas renseignée (86 opérations) ou semblait erronée (surface d'échantillonnage > 10 000 m² laissant penser à une erreur de saisie) ont été corrigées quand cela était possible (recalcul de la surface échantillonnée avec la longueur et la largeur prospectées durant l'échantillonnage ; approximation par la recherche d'autres

opérations sur la même station), ou ont été retirées du jeu de données.

TABEAU 1

Liste des espèces de poissons présentes dans le jeu de données du bassin versant de la Loire.
La colonne « Code » présente les codes qui sont utilisés ensuite dans le manuscrit pour désigner chacune des espèces de poissons.

Code	Famille	Espèce	Nom latin
ANG	Anguillidae	Anguille	<i>Anguilla anguilla</i>
BBG	Centrarchidae	Black-bass à grande bouche	<i>Micropterus salmoides</i>
CDR	Centrarchidae	Crapet de Roche	<i>Ambloplites rupestris</i>
PES	Centrarchidae	Perche soleil	<i>Lepomis gibbosus</i>
LOR	Cobitidae	Loche épineuse	<i>Barbatula barbatula</i>
CHA	Cottidae	Chabot commun	<i>Cottus gobio</i>
ABH	Cyprinidae	Able de Heckel	<i>Leucaspius delineatus</i>
ABL	Cyprinidae	Ablette	<i>Alburnus alburnus</i>
ASP	Cyprinidae	Aspe	<i>Aspius aspius</i>
BAF	Cyprinidae	Barbeau fluviatile	<i>Barbus barbus</i>
BLN	Cyprinidae	Blageon	<i>Telestes souffia</i>
BOU	Cyprinidae	Bouvière	<i>Rhodeus amarus</i>
BRE	Cyprinidae	Brème commune	<i>Abramis brama</i>
CAA	Cyprinidae	Carassin	<i>Carassius sp.</i>
CCO	Cyprinidae	Carpe commune	<i>Cyprinus carpio</i>
CHE	Cyprinidae	Chevaine	<i>Squalius cephalus</i>
GAR	Cyprinidae	Gardon	<i>Rutilus rutilus</i>
GOU	Cyprinidae	Goujon	<i>Gobio gobio</i>
HOT	Cyprinidae	Hotu	<i>Chondrostoma nasus</i>
IDE	Cyprinidae	Ide	<i>Leuciscus idus</i>
PSR	Cyprinidae	Pseudorasbora	<i>Pseudorasbora parva</i>
ROT	Cyprinidae	Rotengle	<i>Scardinius erythrophthalmus</i>
SPI	Cyprinidae	Sprilin	<i>Alburnoides bipunctatus</i>
TAN	Cyprinidae	Tanche	<i>Tinca tinca</i>
TOX	Cyprinidae	Toxostome	<i>Parachondrostoma toxostoma</i>
VAI	Cyprinidae	Vairon commun	<i>Phoxinus phoxinus</i>
VAN	Cyprinidae	Vandoise commune	<i>Leuciscus leuciscus</i>
BRO	Esocidae	Brochet	<i>Esox lucius</i>
LOT	Gadidae	Lote	<i>Lota lota</i>
EPI	Gasterosteidae	Épinoche	<i>Gasterosteus gymnurus</i>
EPT	Gasterosteidae	Épinochette	<i>Pungitius laevis</i>
PCH	Ictaluridae	Poisson chat	<i>Ameiurus melas</i>
MUP	Mugilidae	Mulet porc	<i>Liza ramada</i>
LOF	Nemacheilidae	Loche franche	<i>Barbatula barbatula</i>
GRE	Percidae	Grémille	<i>Gymnocephalus cernuus</i>
PER	Percidae	Perche	<i>Perca fluviatilis</i>
SAN	Percidae	Sandre	<i>Sander lucioperca</i>
LPP	Petromyzontidae	Lamproie de Planer	<i>Lampetra planeri</i>
FLE	Pleuronectidae	Flet	<i>Platichthys flesus</i>
OBR	Salmonidae	Ombre commun	<i>Thymallus thymallus</i>
TRF	Salmonidae	Truite commune	<i>Salmo trutta</i>
SIL	Siluridae	Silure glane	<i>Silurus glanis</i>

B.1.2. Les données macroinvertébrés

Les données macroinvertébrés à l'échelle nationale ont été obtenues auprès des Directions Régionales de l'Environnement, de l'Aménagement et du Logement (DREAL), puis bancarisées et fournies par INRAE. Ces données sont désormais disponibles dans la base Naiades au même titre que les données poissons. Les macroinvertébrés sont suivis selon plusieurs protocoles d'échantillonnage dont la différence majeure réside dans le nombre d'échantillons collectés lors d'une opération. Le jeu de données macroinvertébrés a été constitué sur la base de deux techniques d'échantillonnage : l'Indice Biologique Global Normalisé (IBGN) et l'Indice Biologique Global adapté à la Directive Cadre sur l'Eau (IBG-DCE ; norme AFNOR 2009 XP T90-333). Les données d'abondance ont été standardisées par la surface échantillonnée afin d'être comparables. L'identification des macroinvertébrés est réalisée jusqu'au niveau du genre, voire de la famille pour certains taxons dont l'identification au genre se révèle trop incertaine (*e.g.* Diptera, Dytiscidae).

En fonction des analyses réalisées au cours de la thèse, différents ordres de macroinvertébrés sont considérés. L'ensemble des ordres est considéré pour les analyses de la structuration observée des communautés ainsi que pour l'analyse de la transférabilité entre bassins versants (chapitres III et VI). La modélisation de la distribution étant basée sur le principe de la niche écologique propre à une espèce (voir le chapitre I section C) ; une sous-sélection de taxons est opérée en vue de la modélisation et est présentée dans le chapitre dédié (chapitre IV). Le terme « taxon » est utilisé dans la suite du manuscrit pour désigner les macroinvertébrés ou les données poissons et macroinvertébrés considérées ensemble.

Le jeu de données macroinvertébrés a fait l'objet d'une vérification sur la base des données de distribution connues et d'avis d'experts avec un focus particulier sur les ordres qui ont été considérés pour la modélisation de la distribution de leur abondance, à savoir les Éphéméroptères, Plécoptères, Trichoptères et Odonates (EPTO). En effet, ce sont pour ces ordre bio-indicateurs que les connaissances ont permis la validation d'un grand nombre de genres pour la modélisation, contrairement à d'autres ordres ; il a donc été jugé préférable de se concentrer sur les ordres comportant le plus grand nombre de genres validés pour la modélisation. De plus, les EPTO sont souvent considérés car ils sont présents tout le long du gradient amont-aval des cours d'eau et présentent une grande diversité de genres et d'écologies. Pour chaque genre, la distribution observée dans le jeu de données a été comparée aux distributions des différentes espèces du genre répertoriées dans l'inventaire

de l'Office Pour les Insectes et leur Environnement (OPIE) (voir www.opie-benthos.fr). Lorsque la distribution observée dans le jeu de données était trop différente de la distribution connue dans les données de l'OPIE, le genre était écarté du jeu de données. La comparaison des distributions entre le jeu de données et les données de l'OPIE a aussi permis d'identifier les genres mono-spécifiques en vue de la modélisation de la distribution réalisée en chapitre IV. Cette première sélection a été validée par des experts (communication personnelle avec Maxence Forcellini, Bertrand Launay et Julien Barnasson; observatoire-rhodanien.fr). Le tableau 2 présente la liste des genres EPTO présents dans le jeu de données et leur statut de validation. Au total, 188 genres tous ordres confondus ont été validés dont 99 EPTO : 22 Éphéméroptères, 15 Plécoptères, 44 Trichoptères et 18 Odonates.

TABLEAU 2

Liste des genres de macroinvertébrés pour les EPTO et du statut de validation de la donnée pour chaque genre. ok : les données pour ce genre sont considérées validées, expert : donnée invalidée par expertise, manque d'infos : le genre est retiré faute d'informations suffisante pour valider la donnée.

Ordre	Famille	Genre	Validation
Ephemeroptera	Ameletidae	<i>Ameletus</i>	expert
Ephemeroptera	Ametropodidae	<i>Ametropus</i>	expert
Ephemeroptera	Baetidae	<i>Acentrella</i>	expert
Ephemeroptera	Baetidae	<i>Baetis</i>	ok
Ephemeroptera	Baetidae	<i>Centroptilum</i>	ok
Ephemeroptera	Baetidae	<i>Cloeon</i>	ok
Ephemeroptera	Baetidae	<i>Procloeon</i>	ok
Ephemeroptera	Caenidae	<i>Caenis</i>	ok
Ephemeroptera	Ephemerellidae	<i>Ephemerella</i>	ok
Ephemeroptera	Ephemerellidae	<i>Torleya</i>	ok
Ephemeroptera	Ephemeridae	<i>Ephemera</i>	ok
Ephemeroptera	Heptageniidae	<i>Electrogena</i>	ok
Ephemeroptera	Heptageniidae	<i>Epeorus</i>	ok
Ephemeroptera	Heptageniidae	<i>Heptagenia</i>	ok
Ephemeroptera	Heptageniidae	<i>Rhithrogena</i>	ok
Ephemeroptera	Isonychiidae	<i>Isonychia</i>	ok
Ephemeroptera	Leptophlebiidae	<i>Choroterpes</i>	ok
Ephemeroptera	Leptophlebiidae	<i>Habrophlebia</i>	ok
Ephemeroptera	Leptophlebiidae	<i>Leptophlebia</i>	ok
Ephemeroptera	Leptophlebiidae	<i>Paraleptophlebia</i>	ok
Ephemeroptera	Leptophlebiidae	<i>Thraululus</i>	ok
Ephemeroptera	Oligoneuriidae	<i>Oligoneuriella</i>	ok
Ephemeroptera	Polymitarcyidae	<i>Ephoron</i>	ok
Ephemeroptera	Potamanthidae	<i>Potamanthus</i>	ok
Ephemeroptera	Siphonuridae	<i>Siphonurus</i>	ok
Odonata	Aeshnidae	<i>Aeshna</i>	ok
Odonata	Aeshnidae	<i>Anax</i>	ok
Odonata	Aeshnidae	<i>Boyeria</i>	ok
Odonata	Aeshnidae	<i>Brachytron</i>	ok
Odonata	Aeshnidae	<i>Hemianax</i>	ok
Odonata	Calopterygidae	<i>Calopteryx</i>	manque d'infos
Odonata	Cordulegasteridae	<i>Cordulegaster</i>	manque d'infos
Odonata	Corduliidae	<i>Cordulia</i>	ok
Odonata	Corduliidae	<i>Epithea</i>	ok
Odonata	Corduliidae	<i>Somatochlora</i>	manque d'infos
Odonata	Gomphidae	<i>Gomphus</i>	manque d'infos
Odonata	Gomphidae	<i>Onychogomphus</i>	manque d'infos
Odonata	Gomphidae	<i>Ophiogomphus</i>	ok
Odonata	Gomphidae	<i>Paragomphus</i>	ok
Odonata	Lestidae	<i>Chalcolestes</i>	ok
Odonata	Lestidae	<i>Lestes</i>	ok
Odonata	Lestidae	<i>Sympecma</i>	ok
Odonata	Libellulidae	<i>Brachythemis</i>	ok
Odonata	Libellulidae	<i>Crocothemis</i>	ok
Odonata	Libellulidae	<i>Diplacodes</i>	ok
Odonata	Libellulidae	<i>Leucorrhinia</i>	expert
Odonata	Libellulidae	<i>Libellula</i>	ok
Odonata	Libellulidae	<i>Orthetrum</i>	ok
Odonata	Libellulidae	<i>Sympetrum</i>	ok

Ordre	Famille	Genre	Validation
Odonata	Platycnemididae	<i>Platycnemis</i>	manque d'infos
Plecoptera	Capniidae	<i>Capnia</i>	expert
Plecoptera	Capniidae	<i>Capnioneura</i>	ok
Plecoptera	Chloroperlidae	<i>Chloroperla</i>	expert
Plecoptera	Chloroperlidae	<i>Isoptena</i>	expert
Plecoptera	Chloroperlidae	<i>Xanthoperla</i>	ok
Plecoptera	Leuctridae	<i>Euleuctra</i>	ok
Plecoptera	Leuctridae	<i>Leuctra</i>	ok
Plecoptera	Nemouridae	<i>Amphinemura</i>	ok
Plecoptera	Nemouridae	<i>Nemoura</i>	ok
Plecoptera	Nemouridae	<i>Nemurella</i>	expert
Plecoptera	Nemouridae	<i>Protonemura</i>	ok
Plecoptera	Perlidae	<i>Dinocras</i>	ok
Plecoptera	Perlidae	<i>Perla</i>	ok
Plecoptera	Perlodidae	<i>Arcynopteryx</i>	expert
Plecoptera	Perlodidae	<i>Isogenus</i>	ok
Plecoptera	Perlodidae	<i>Isoperla</i>	ok
Plecoptera	Perlodidae	<i>Perlodes</i>	ok
Plecoptera	Taeniopterygidae	<i>Brachyptera</i>	ok
Plecoptera	Taeniopterygidae	<i>Rhabdiopteryx</i>	ok
Plecoptera	Taeniopterygidae	<i>Taeniopteryx</i>	ok
Trichoptera	Beraeidae	<i>Beraea</i>	ok
Trichoptera	Beraeidae	<i>Beraeodes</i>	ok
Trichoptera	Brachycentridae	<i>Micrasema</i>	ok
Trichoptera	Brachycentridae	<i>Oligoplectrum</i>	ok
Trichoptera	Calamoceratidae	<i>Calamoceras</i>	ok
Trichoptera	Dipseudopsidae	<i>Pseudoneureclipsis</i>	ok
Trichoptera	Glossosomatidae	<i>Agapetus</i>	ok
Trichoptera	Glossosomatidae	<i>Glossosoma</i>	ok
Trichoptera	Glossosomatidae	<i>Synagapetus</i>	ok
Trichoptera	Goeridae	<i>Goera</i>	ok
Trichoptera	Goeridae	<i>Lithax</i>	ok
Trichoptera	Goceridae	<i>Silo</i>	ok
Trichoptera	Hydropsychidae	<i>Cheumatopsyche</i>	ok
Trichoptera	Hydropsychidae	<i>Diplectrona</i>	ok
Trichoptera	Hydropsychidae	<i>Hydropsyche</i>	ok
Trichoptera	Hydroptilidae	<i>Agraylea</i>	ok
Trichoptera	Hydroptilidae	<i>Allotrichia</i>	expert
Trichoptera	Hydroptilidae	<i>Hydroptila</i>	ok
Trichoptera	Hydroptilidae	<i>Ithytrichia</i>	ok
Trichoptera	Hydroptilidae	<i>Orthotrichia</i>	ok
Trichoptera	Hydroptilidae	<i>Ptilocolepus</i>	ok
Trichoptera	Hydroptilidae	<i>Stactobiella</i>	expert
Trichoptera	Lepidostomatidae	<i>Crunoecia</i>	ok
Trichoptera	Lepidostomatidae	<i>Lepidostoma</i>	expert
Trichoptera	Leptoceridae	<i>Adicella</i>	ok
Trichoptera	Leptoceridae	<i>Ceraclea</i>	ok
Trichoptera	Leptoceridae	<i>Erotosis</i>	expert
Trichoptera	Leptoceridae	<i>Leptocerus</i>	expert

Ordre	Famille	Genre	Validation
Trichoptera	Leptoceridae	<i>Mystacides</i>	ok
Trichoptera	Leptoceridae	<i>Setodes</i>	ok
Trichoptera	Leptoceridae	<i>Triaenodes</i>	ok
Trichoptera	Leptoceridae	<i>Ylodes</i>	ok
Trichoptera	Limnephilidae	<i>Apataniinae</i>	ok
Trichoptera	Molannidae	<i>Molanna</i>	ok
Trichoptera	Molannidae	<i>Molannodes</i>	expert
Trichoptera	Philopotamidae	<i>Philopotamus</i>	ok
Trichoptera	Philopotamidae	<i>Wormaldia</i>	ok
Trichoptera	Phryganeidae	<i>Agrypnia</i>	ok
Trichoptera	Phryganeidae	<i>Phryganea</i>	expert
Trichoptera	Phryganeidae	<i>Trichostegia</i>	ok
Trichoptera	Polycentropodidae	<i>Cyrnus</i>	ok
Trichoptera	Polycentropodidae	<i>Holocentropus</i>	expert
Trichoptera	Polycentropodidae	<i>Neureclipsis</i>	ok
Trichoptera	Polycentropodidae	<i>Plectrocnemia</i>	ok
Trichoptera	Polycentropodidae	<i>Polycentropus</i>	ok
Trichoptera	Psychomyiidae	<i>Lype</i>	ok
Trichoptera	Psychomyiidae	<i>Metalype</i>	expert
Trichoptera	Psychomyiidae	<i>Paduniella</i>	expert
Trichoptera	Psychomyiidae	<i>Psychomyia</i>	ok
Trichoptera	Psychomyiidae	<i>Tinodes</i>	ok
Trichoptera	Rhyacophilidae	<i>Rhyacophila</i>	ok
Trichoptera	Sericostomatidae	<i>Notidobia</i>	ok
Trichoptera	Sericostomatidae	<i>Oecismus</i>	expert
Trichoptera	Sericostomatidae	<i>Schizopelex</i>	expert
Trichoptera	Sericostomatidae	<i>Sericostoma</i>	ok
Trichoptera	Uenoidae	<i>Thremma</i>	ok

B.1.3. Sélection de la période temporelle sur les jeux de données biologiques

Les deux jeux de données biologiques sont constitués d'opérations d'échantillonnage réalisées sur plusieurs années. Afin de limiter l'influence des dynamiques temporelles long-terme, une restriction de la période temporelle a été appliquée sur les deux jeux de données biologiques. Au final, une période de 11 ans, de 2007 à 2017 a été retenue (Figure 17). L'échelle décennale est également assez courte pour s'affranchir des processus évolutifs qui pourraient avoir lieu dans les communautés étudiées (Tison-Rosebery *et al.*, 2022). Cette sélection permet de réduire le nombre de stations qui n'ont été échantillonnées qu'une seule fois ($n = 263$ contre 546 stations entre 1994 et 2017) au cours de la période et pour lesquelles le risque est plus élevé d'avoir une image moins représentative de la communautés en place sur la station. Dans la même logique, les échantillonnages poissons et macroinvertébrés se répartissent sur différents mois de l'année et la majorité ont lieu durant les périodes estivales et automnales. Le choix a été fait de restreindre les données aux échantillonnages entre les mois de juin et octobre inclus, période comportant la majorité des opérations d'échantillonnage.

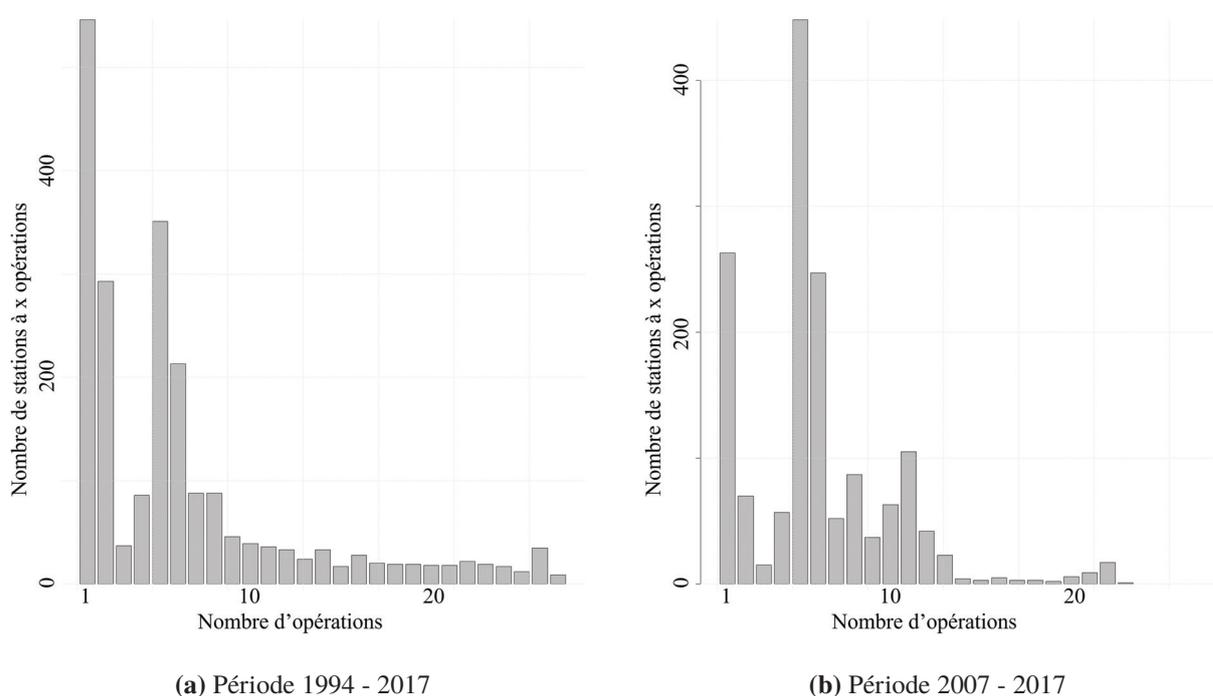


FIGURE 17
Répartition du nombre de stations en fonction de leur nombre d'opérations d'échantillonnage de la faune piscicole pour la période initiale 1994-2017 (a) et la période retenue 2007-2017 (b) pour les analyses des communautés de poissons et de macroinvertébrés.

B.2. Le réseau H comme unité géographique

Les données biologiques ont été raccordées au réseau H qui constitue l'entité géographique d'étude. Le réseau H est basé sur une simplification du réseau *Temperature Network* (T-NET) (Beaufort *et al.*, 2016) qui a dû être opérée dans le but de raccorder les données biologiques à un plus grand nombre de variables environnementales (*e.g.* probabilité d'intermittence des cours d'eau disponible sur le réseau RHT, qui ne tient pas compte des cours d'eau d'ordre 1 ; pour plus de détails sur la jointure des réseaux T-NET et RHT voir l'Annexe C). Il représente le réseau hydrographique du bassin versant de la Loire, de la source du fleuve jusqu'à Montjean-sur-Loire qui est la limite amont de la zone d'influence des marées. Il comprend 33 711 tronçons de cours d'eau d'une longueur moyenne de $1,85 \pm 1,63$ km (Figure 18).

La première étape a consisté à projeter les stations de suivis biologiques sur le réseau H. L'ensemble des traitements Systèmes d'Informations Géographiques (SIG) présentés ci-après ont été réalisés dans le logiciel de programmation R avec l'aide des packages *sf* et *maptools* (Pebesma, 2018 ; Bivand & Lewin-Koh, 2022 ; R Core Team, 2022). Les stations de suivis hydrobiologiques poissons et macroinvertébrés ont été projetées sur le tronçon de H le plus proche dans un rayon limite de 600 m pour les poissons et 200 m pour les macroinvertébrés. 20 stations poissons et 65 stations macroinvertébrés n'ont pas pu être projetées car elles sont toutes situées en-dehors de l'emprise du réseau H, dans la zone d'embouchure de la Loire en aval de Montjean-sur-Loire. Les stations projetées à une distance supérieure à 50 m ont été vérifiées une par une (poissons $n = 113$, macroinvertébrés $n = 10$), et si nécessaire leur position a été corrigée (*i.e.* ré-attribution au bon tronçon H) voire supprimée lorsqu'aucun tronçon H ne semblait correspondre. Cette vérification a permis d'identifier 35 mauvaises projections de stations poissons et 1 station macroinvertébrés dans des canaux, ou des bras morts. Au total, 492 stations poissons et 1 093 stations macroinvertébrés ont été raccordées au réseau H (Figure 20).

Afin d'étudier les deux compartiments biologiques simultanément, un troisième jeu de données a été constitué en identifiant les tronçons portant à la fois au moins une station d'échantillonnage piscicole et au moins une station d'échantillonnage macroinvertébrés. Il peut s'agir de la même station de suivi ou de plusieurs stations proches portées par le même tronçon. Ce troisième jeu de données, appelé « commun », concerne 408 tronçons, ce qui représente 84 % du jeu poisson et 38 % du jeu macroinvertébrés.

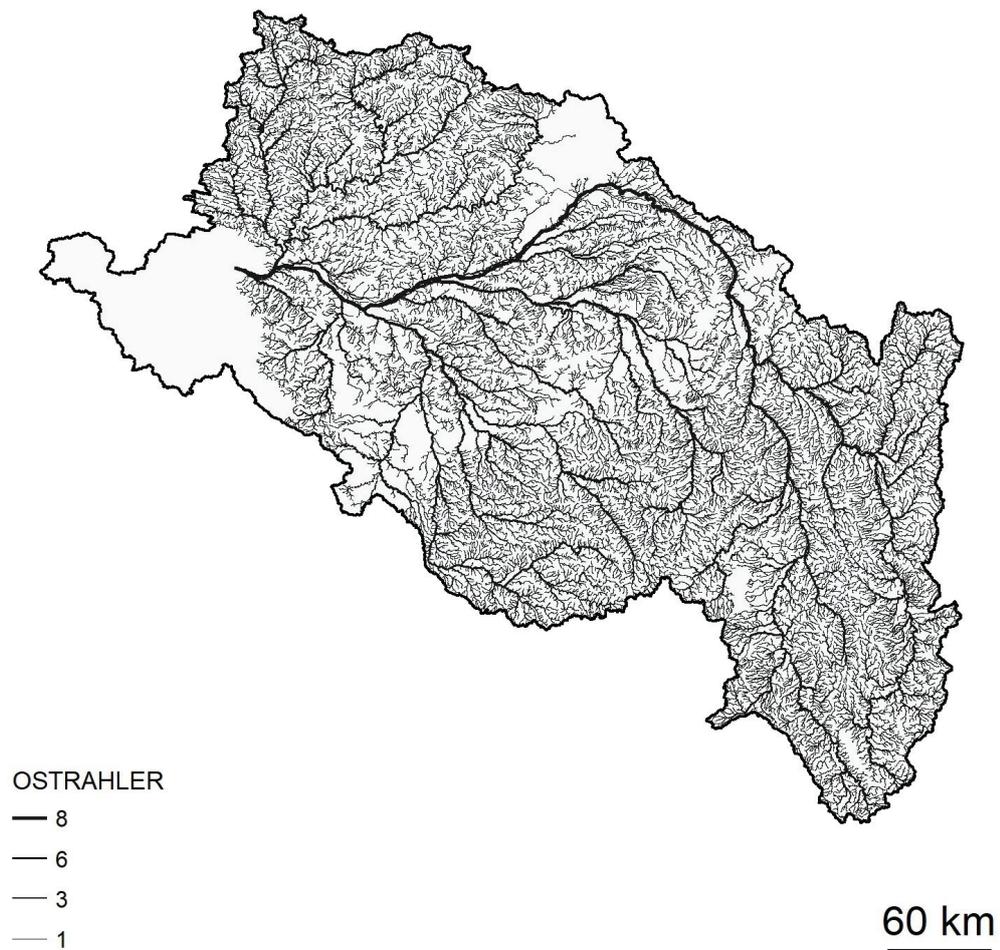


FIGURE 18

Représentation du réseau H sur le bassin versant de la Loire (N = 33 711 tronçons). Les tronçons de ce réseau constituent notre unité géographique pour la modélisation de la distribution spatiale des abondances. *Remarque* : bien que le réseau H ne couvre pas la partie du bassin de la Loire influencé par la marée, en aval de Montjean-sur-Loire, l'ensemble du bassin est représenté dans les cartographies présentées dans ce manuscrit.

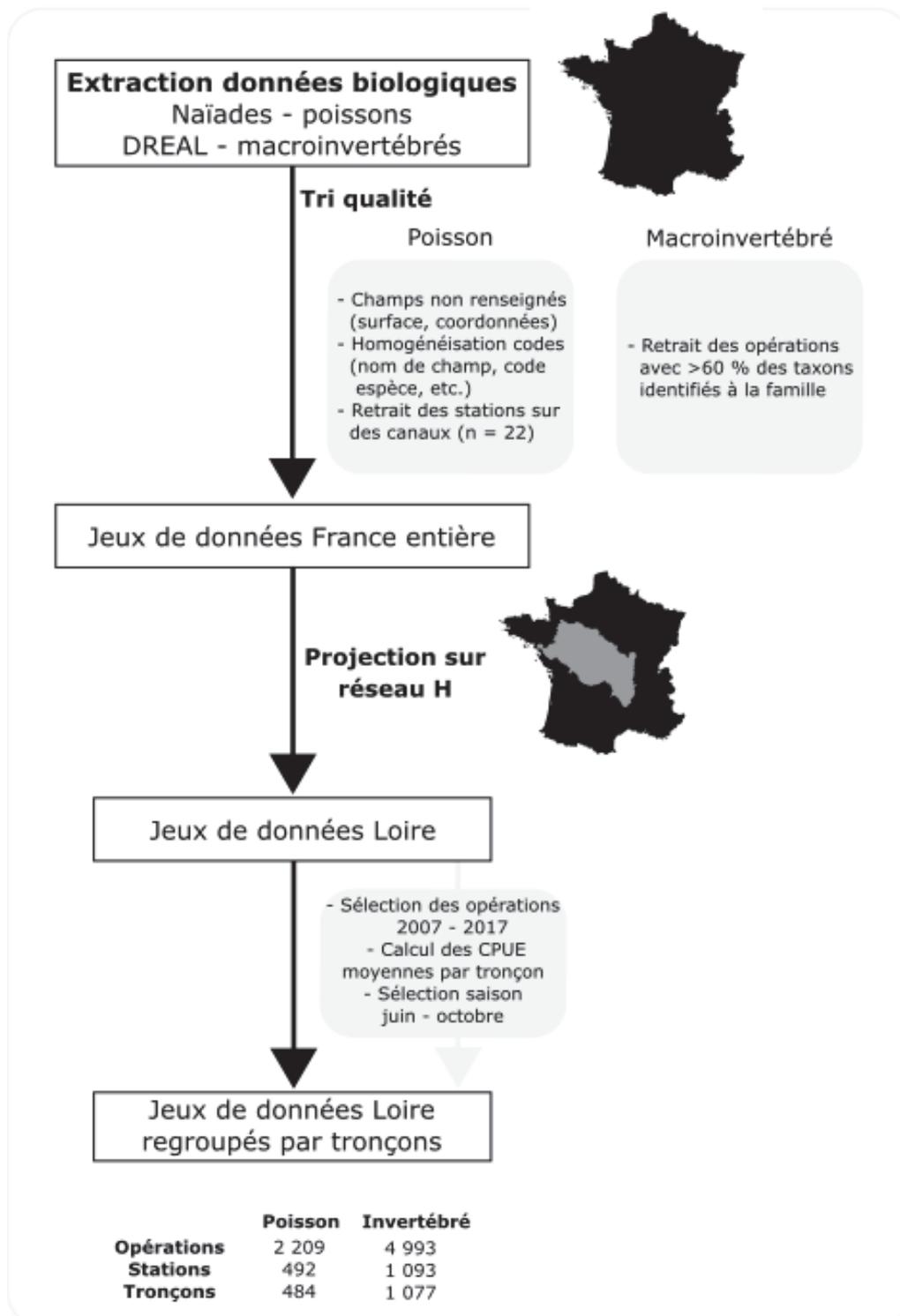


FIGURE 19
Schéma de l'extraction et du traitement des données biologiques.

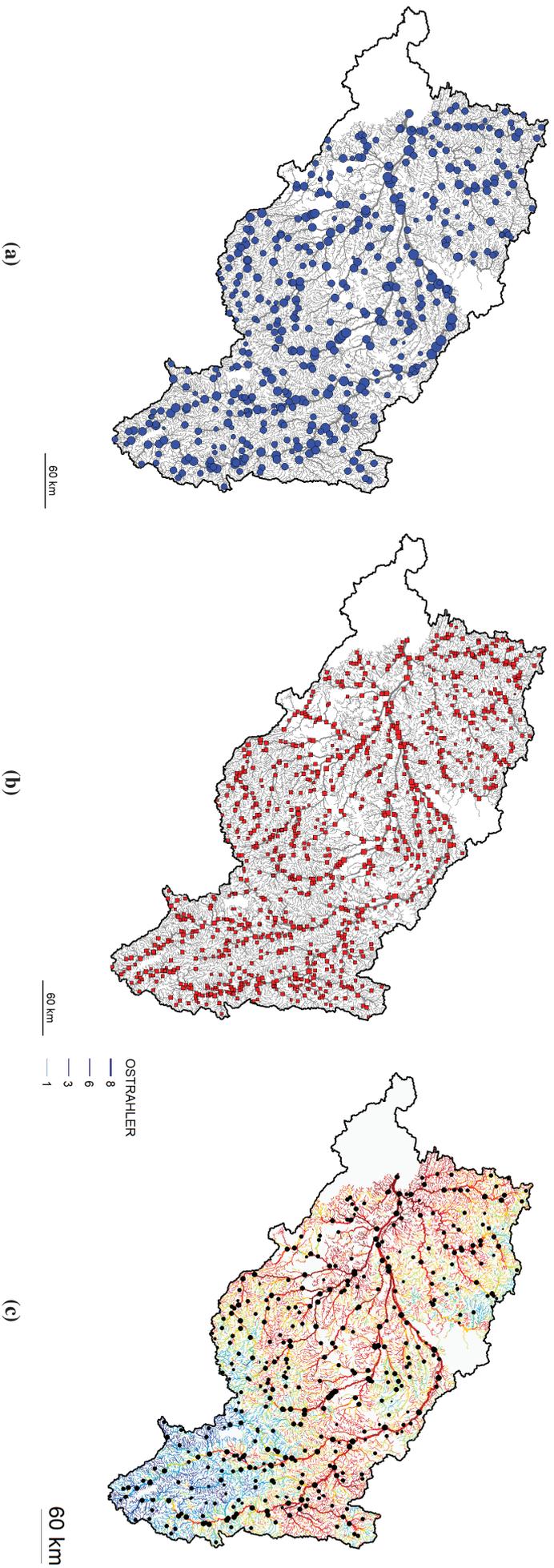


FIGURE 20 Localisation des tronçons portant une station de suivi biologique. (a) Tronçons portant au moins une station de suivi poissons (N = 484), (b) tronçons portant au moins une station de suivi macroinvertébrés (N = 1 077) et (c) présentation du réseau H avec illustration de la température moyenne interannuelle de l'eau sur la période 2007 - 2017 et localisation des tronçons portants au moins une station poissons et au moins une station macroinvertébrés (N = 408 tronçons).

B.2.1. Calcul du peuplement moyen par tronçon

Les données sur les peuplements de poissons et de macroinvertébrés sont issues de suivis pluriannuels dont la fréquence d'échantillonnage varie en fonction du type de suivi réalisé sur la station, généralement entre deux à cinq ans. Par conséquent, l'effort d'échantillonnage varie d'une station à l'autre sur la période 2007-2017 (entre 1 et 23 opérations d'échantillonnage par station). Afin d'obtenir une vision du peuplement moyen sur cette période et sur tous les tronçons du réseau échantillonnés, les abondances pour chaque espèce ont été rapportées en Captures Par Unités d'Effort (CPUE) moyennes interannuelles (voir calcul ci-dessous). L'unité spatiale des analyses étant le tronçon, les données des stations se trouvant sur un même tronçon ont été rassemblées par calcul de la CPUE moyenne pour chaque taxon sur la période considérée. Le calcul d'un peuplement moyen sur chaque tronçon permet de lisser les effets d'événements exceptionnels (*i.e.* hyperabondance ponctuelle) ou les incertitudes liées à l'échantillonnage (*e.g.* multiplie les chances de rencontrer toutes les espèces) et d'obtenir l'image la plus représentative de la structure des abondances des peuplements en place sur les tronçons sur la période d'étude.

$$CPUE_{taxa} = \frac{\sum_1^N CPUE_{operation}}{N}$$

où N : nombre d'opération d'échantillonnage sur le tronçon.

B.3. Les données environnementales du bassin versant de la Loire

Les données environnementales telles que la température de l'eau, le pH, le débit sont mesurées à l'occasion des pêches et aussi sur d'autres réseaux dédiés (*e.g.* réseau national thermique de l'OFB), mais les couvertures spatiale et temporelle de ces données reste limitée à l'instant et au lieu de la mesure. Pour pallier ce manque, des modèles simulant certaines de ces variables environnementales ont été développés sur des réseaux hydrographiques à l'échelle de bassin versant. Ils permettent d'obtenir une vision spatialement complète des conditions environnementales rencontrées par les communautés aquatiques tout au long de l'année, parfois sur plusieurs décennies. Les données mobilisées ici pour retranscrire la thermie, l'hydrologie et des informations sur la topographie (*i.e.* pente, altitude, distance à la source) des cours d'eau proviennent de ce type de modèles. Enfin, les données de pressions d'origine anthropique comme l'occupation des sols, les obstacles à

la continuité écologique ou l'urbanisation ont été collectées auprès de différentes bases de données ou organismes.

B.3.1. Descripteurs du débit et de l'habitat hydraulique local

L'hydrologie et l'habitat hydraulique participe à sélectionner les espèces présentes en un endroit donné de la rivière, en fonction de leur capacité de nage ou de leur besoin en oxygène (Huet, 1949 ; Morel *et al.*, 2020). Des descripteurs des débits ont été calculés à partir des simulations hydrologiques réalisées avec le modèle semi-distribué EROS du Bureau de Recherches Géologiques et Minières (BRGM) (Sauquet *et al.*, 2000 ; Seyedhashemi *et al.*, 2022). Les travaux de Morel *et al.* (2020) ont ensuite permis de caractériser les géométries hydrauliques de la rivière hauteur (h) et largeur (l), à partir des modèles statistiques de géométrie hydraulique et du débit journalier simulé par EROS en assumant une section de rivière rectangulaire et constante pendant 24 heures. Les équations de calcul de h et l sont :

$$l = L50_{me} \left(\frac{Q}{Q50} \right)^{b_{me}}$$
$$h = H50_{me} \left(\frac{Q}{Q50} \right)^{f_{me}}$$

où Q est le débit journalier moyen fourni par le modèle hydrologique EROS, Q50 le débit médian et L50_{me} et H50_{me} sont respectivement la largeur et la hauteur en eau associées, et b_{me} et f_{me} sont respectivement les exposants spécifiques à chaque tronçon, définis d'après le climat, l'hydrologie, la topographie et des descripteurs de l'occupation des sols définis par Morel *et al.* (2020). La vitesse de courant est ensuite calculée comme suit :

$$v = \frac{Q}{l + h}$$

B.3.2. Descripteurs de la température de l'eau

La température de l'eau joue un rôle majeur dans la structuration des communautés aquatiques, notamment ses valeurs extrêmes (Verneaux, 1977a ; Laffaille *et al.*, 2020 ; Matthews, 1998 ; voir chapitre I). Néanmoins, comme pour d'autres variables, la mesure de la température de l'eau est impossible à obtenir sur l'ensemble du réseau hydrogra-

phique d'un bassin versant car cela demanderait des moyens humains et financiers colossaux. Elle est donc bien souvent approchée par la température de l'air qui ne permet pas de retranscrire toute la variabilité spatio-temporelle des régimes thermiques des cours d'eau (Caissie, 2006 ; Culler *et al.*, 2018). Par exemple, des changements très locaux de la température de l'eau peuvent être induits par des apports d'eau plus fraîche provenant des nappes souterraines (Seyedhashemi *et al.*, 2021a ; Beaufort *et al.*, 2020a). Récemment, des modèles à base physique ont été développés dans le but de simuler la température de l'eau et sa variabilité (Beaufort *et al.*, 2016) mais ils n'ont été que rarement mobilisés dans un cadre écologique (Ficklin *et al.*, 2012 ; Leach & Moore, 2019).

Les données de température de l'eau proviennent des simulations journalières issues du modèle T-NET sur la période 2007-2017 (Beaufort *et al.*, 2016 ; Seyedhashemi *et al.*, 2023 ; Figure 21). Le modèle T-NET est basé sur le concept d'équilibre thermique et la propagation du signal thermique de l'amont vers l'aval (Figure 21). Il utilise des variables de forçage climatique et hydrologique (base Système d'Analyse Fournissant des Renseignements Adaptés à la Nivologie (SAFRAN) et modèle Ensemble de Rivières Organisé en Sous bassins (EROS)), des paramètres topographiques et représentatifs de la végétation rivulaire (Loicq *et al.*, 2018) pour calculer la température de l'eau sur le réseau du même nom (N = 52 278 tronçons).

B.3.3. Méthode de calcul des métriques de la température de l'eau et des variables hydrauliques sur la période 2007 - 2017

Différentes métriques ont été calculées pour capter les régimes moyens et les variabilités de la température de l'eau et des variables hydrauliques sur la période d'étude, soit 2007 - 2017. Les métriques utilisées sont la moyenne interannuelle des températures de l'eau, l'écart-type ainsi que les percentiles 10 et 90 (*i.e.* respectivement les valeurs inférieures aux 10% et 90% des valeurs rencontrées). Elles ont été calculées à partir des simulations journalières de la température de l'eau et des débits issus respectivement des modèles T-NET et EROS. Les moyennes de l'ensemble des valeurs journalières sur une année ont été calculées, puis ont servi au calcul de la moyenne interannuelle (*i.e.* moyenne interannuelle = moyenne des moyennes annuelles).

Les cartographies des moyennes interannuelles de l'eau et de la hauteur en eau sont présentées en Figure 22. Elles illustrent l'évolution des valeurs de chacune des variables le long du gradient amont-aval, avec un accroissement des valeurs à mesure que la taille

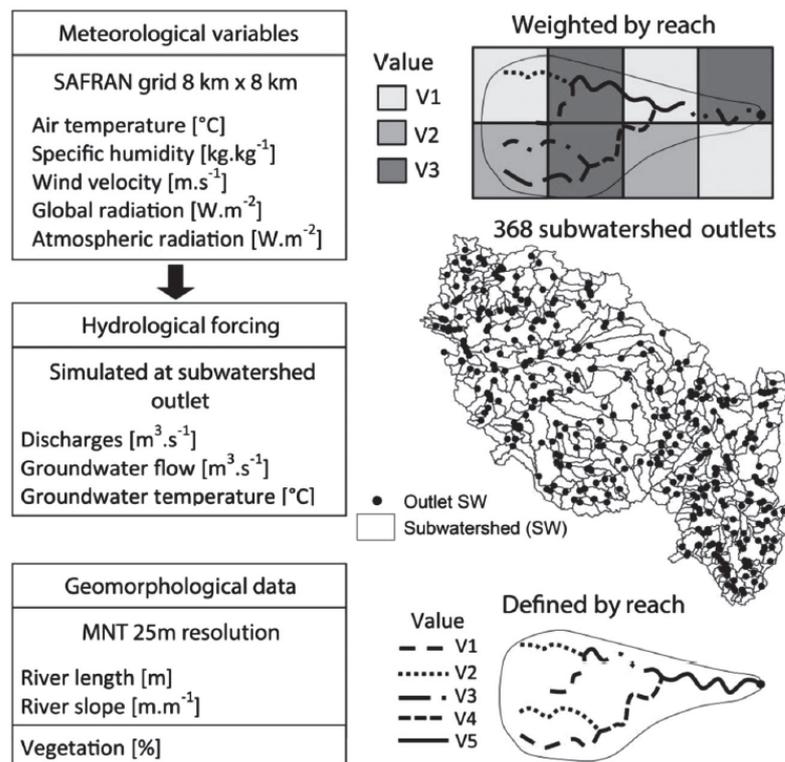
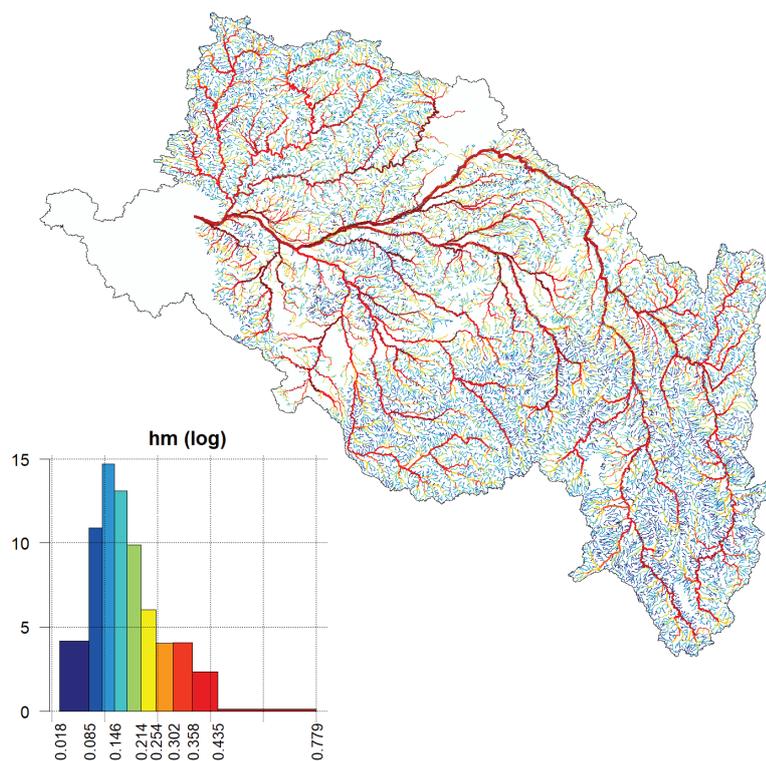


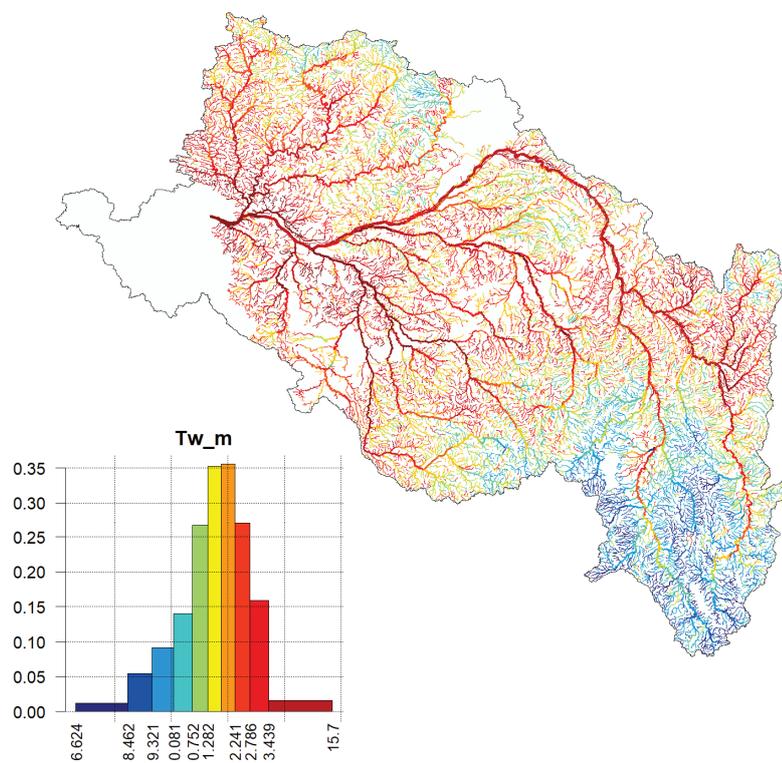
FIGURE 21

Données d'entrée du modèle T-NET. Les variables météorologiques proviennent de la base SAFRAN, les variables de forçages hydrauliques du modèle EROS, les données géomorphologiques du MNT 25 m et les données sur la végétation rivulaire de SYRAH. EROS simulent les débits en sorties des 368 sous-bassins ; ils servent à la reconstruction des débits sur chacun des tronçons. D'après Beaufort *et al.* (2016).

du cours d'eau augmente.



(a)



(b)

FIGURE 22

Cartographie des moyennes interannuelles de la hauteur en eau (en m ; a) et de la température de l'eau (en °C ; b) représentées sur le réseau H (N = 33 711 tronçons).

B.3.4. La probabilité d'intermittence des cours d'eau

L'intermittence des cours d'eau est variable d'une rivière à l'autre et participe aux régimes des perturbations qui façonnent les communautés aquatiques (Statzner & Higler, 1986 ; Allen *et al.*, 2020). Snelder *et al.* (2013) ont établi une probabilité d'intermittence annuelle des cours d'eau du réseau hydrographique français à partir d'un modèle de classification (*Random Forest* (RF)). Le modèle repose sur les données de jaugeage (N = 628 stations) issues de rivières où les régimes hydrologiques sont considérés faiblement perturbés. Les données proviennent des stations suivies sur une période de 23 à 35 ans, entre 1975 à 2009, et sans interruption de suivi supérieur à 20 jours consécutifs dans l'année.

La proportion de jours sans écoulement est la plus élevée entre les mois de juillet et octobre. L'intermittence des cours d'eau dépend de multiples facteurs physiques et est la plus élevée dans les zones avec de faibles précipitations annuelles et de fortes températures de l'air, ainsi que pour des bassins de pentes abruptes, petits et de forme allongée, ce qui correspond majoritairement aux cours d'eau de petits ordres (Snelder *et al.*, 2013). Les probabilités d'intermittence des cours d'eau ont été extraites du réseau RHT et associées à notre réseau H (Figure 23).

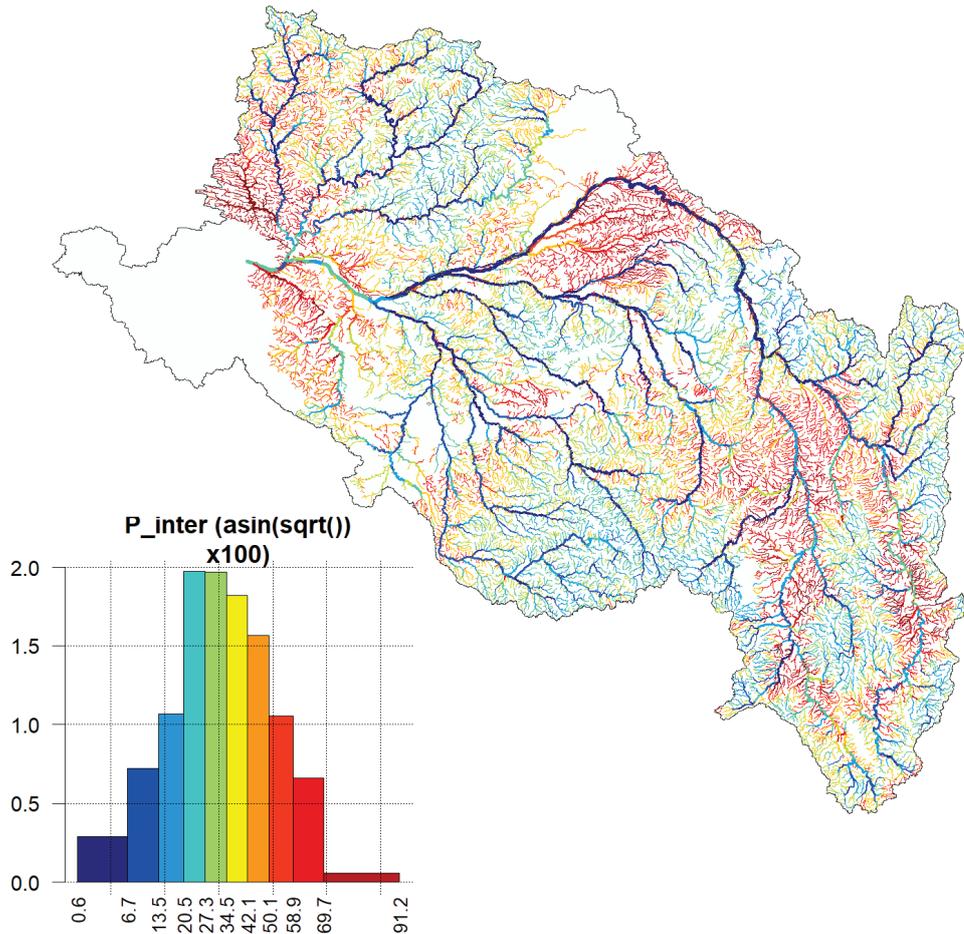


FIGURE 23
Cartographie de la probabilité d'intermittence des cours d'eau.

B.3.5. Intégration des pressions anthropiques du bassin ligérien

Les modèles EROS et T-NET nous fournissent respectivement les simulations des régimes hydrologiques et thermiques des cours d'eau sans tenir compte des influences anthropiques comme les barrages, les grandes retenues d'eau ou encore l'occupation des sols. Des travaux sont nécessaires pour intégrer ce type d'influence dans la modélisation de la température (Seyedhashemi *et al.*, 2021a). Dans le but de tenir compte des influences anthropiques directes sur nos données biologiques, diverses variables environnementales ont été intégrées, chacune représentant différents types de pressions : (i) les pressions liées à l'occupation du sol, comme l'agriculture ou les milieux urbains, via *Corine Land Cover* (CLC), (ii) les données de recensement des populations humaines issues de l'Institut National de la Statistique et des Études Économiques (INSEE) et (iii) les pressions sur la continuité écologique via des indices calculés sur la base du Référentiel des Obstacles à

l'Écoulement (ROE).

(i) L'occupation des sols

Les données d'occupation des sols, comme l'agriculture ou les surfaces forestières, ont été extraites depuis la base CLC, (version de 2012 ; European Environment Agency 2012). Les taux d'urbanisation et d'agriculture sont les paramètres les plus discriminants dans la composition des communautés aquatiques car ils sont à l'origine d'apports de matières potentiellement néfastes pour les organismes aquatiques (matières organiques, pollutions d'origine chimique ; Gao *et al.* 2023 ; Park *et al.* 2006 ; Malmqvist & Rundle 2002). Un gradient d'occupation des sols a donc été construit à partir des données de CLC et de la méthode suivante. Une zone tampon d'un rayon de 100 m a été construite autour de chaque tronçon T-NET afin d'intersecter l'occupation du sol (code de niveau 3) aux abords directs du cours d'eau. Les codes de CLC ont été regroupés par catégorie d'anthropisation comme suit :

- Agriculture intensive : 211, 212, 213, 221, 222, 223, 241, 242
- Espaces naturels ou peu anthropisés : 311, 312, 313, 321, 322, 323, 324, 331, 332, 333, 334, 335, 411, 412, 421, 422, 423, 511, 512, 521, 522, 523

Une Analyse en Composante Principale (ACP) a ensuite été réalisée sur la base des groupes « agriculture intensive » et « espaces naturels ou peu anthropisés » afin de représenter un gradient du niveau d'anthropisation des sols aux abords des cours d'eau. Sur la base de cette ACP, une unique variable, nommée *agrsyn*, synthétisant le niveau d'anthropisation aux abords directs des cours d'eau a été retenue (Figure 24 ; 97% de variabilité portés par l'axe 1). Cette variable synthétique caractérise un gradient de milieux forestiers et naturels sur les faibles valeurs de l'axe d'ACP, et des milieux très anthropisés comme les milieux d'agriculture intensive vers les fortes valeurs d'axe de l'ACP.

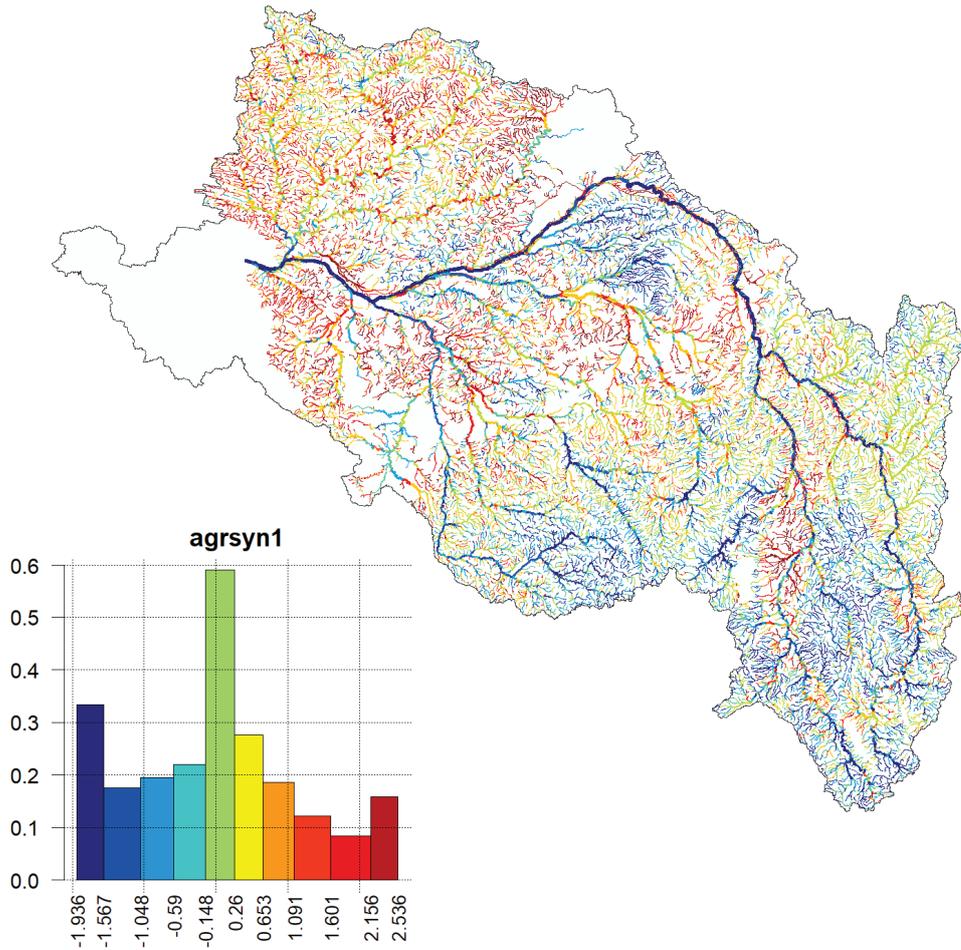


FIGURE 24
Cartographie de la variable synthétique, *agsyn*, issue d'une ACP caractérisant l'anthropisation des sols aux abords directs des cours d'eau.

(ii) Densité de population aux abords des cours d'eau

Afin d'approcher plus spécifiquement l'occupation urbaine aux abords des cours d'eau, les données issues de l'INSEE ont été mobilisées (INSEE, 2021). L'intersection des couches SIG comportant le nombre d'habitants avec un buffer de 100 m autour de chaque tronçon T-NET a été effectuée. Pour chaque tronçon qui intersecte plusieurs polygones de différentes densités de populations, la densité moyenne pondérée par la surface intersectée est calculée. La variable finale est exprimée en nombre moyen d'habitant/km² dans le buffer pour chaque tronçon (Figure 25).

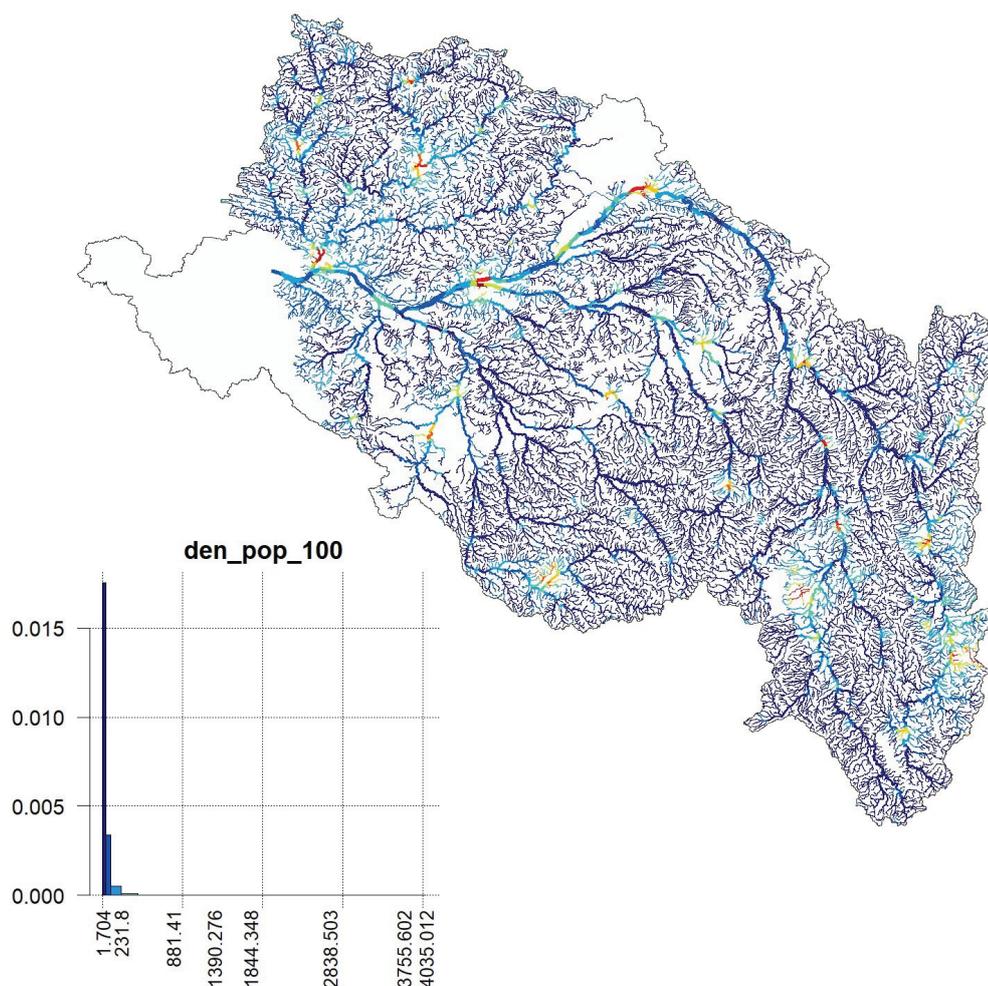


FIGURE 25
Cartographie de la densité de population aux abords directs des cours d'eau (dans un buffer de 100 m).

(iii) Les obstacles à la continuité écologique

Le référentiel des obstacles à l'écoulement recense l'ensemble des ouvrages en long ou en travers du lit des cours d'eau sur l'ensemble du réseau hydrographique français. Sur le bassin de la Loire, 28 766 ouvrages y sont recensés (extraction réalisée le 17 novembre 2021 ; Figure 26) dont des seuils en rivière, des barrages et des obstacles induits par des ponts, avec une hauteur de chute moyenne de 1,6 m à l'échelle de l'ensemble du bassin (www.eptb-loire.fr). Leur susceptibilité à entraver la continuité écologique peut être estimée par différents indices, comme l'Information sur la Continuité Écologique (ICE) développée par l'Office National de l'Eau et des Milieux Aquatiques (ONEMA ; Baudoin *et al.* 2014), qui est basée sur la hauteur de chute et les capacités de franchissement des espèces. Cependant, les informations nécessaires à l'attribution d'une note de «franchissa-

bilité » ne sont pas disponibles pour l'ensemble des ouvrages recensés dans le référentiel, notamment sur les ouvrages de moindre importance localisés en tête de bassin. De plus, le recensement des ouvrages n'est pas exhaustif bien que le bassin de la Loire soit l'un des mieux renseignés au niveau national.

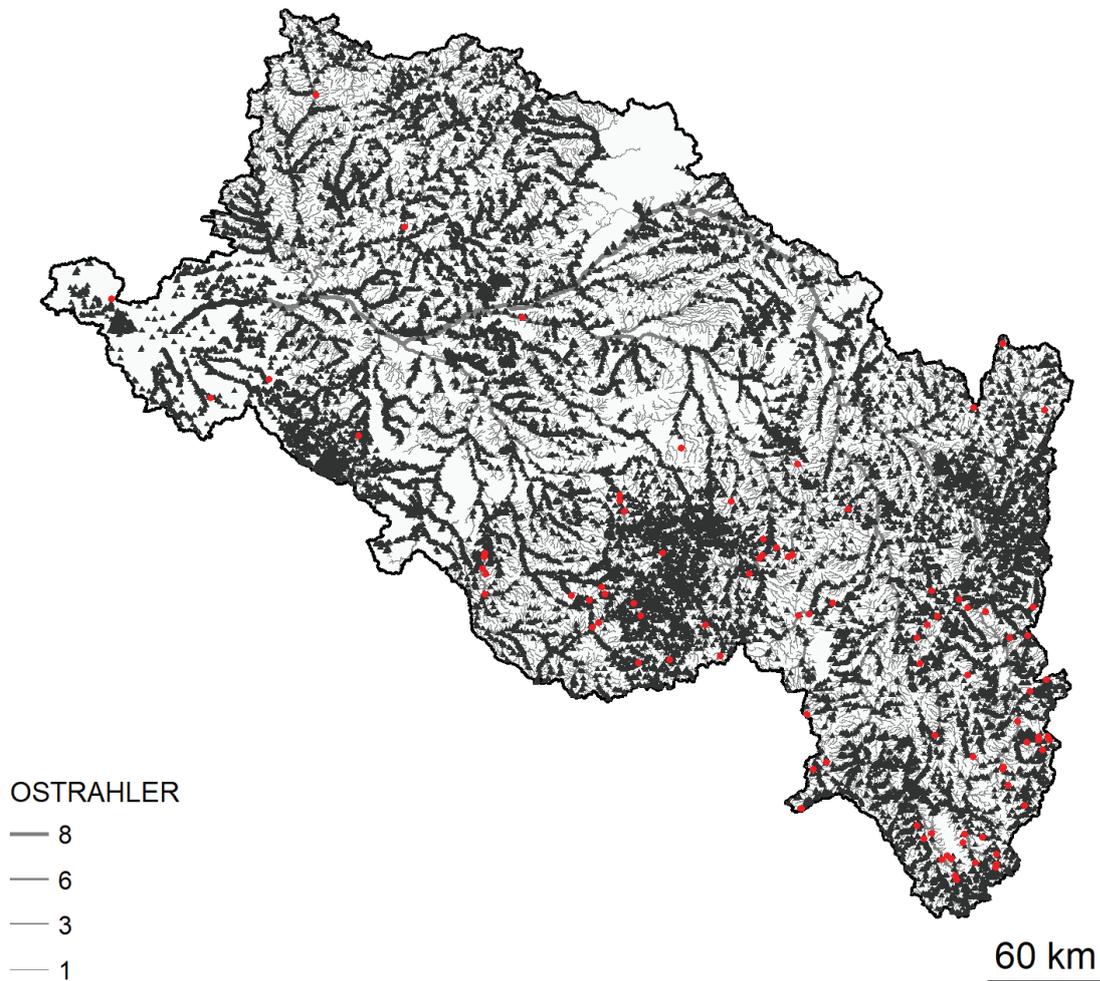


FIGURE 26

Ouvrages référencés dans le référentiel des obstacles à l'écoulement (n = 28 766). Les ouvrages d'une hauteur supérieure à 10 m sont représentés en rouge.

Des indices facilement mobilisables ont donc été développés dans le but de mesurer l'influence des obstacles sur la continuité écologique : (i) la densité d'ouvrage, (ii) le taux de fractionnement et (iii) le taux d'étagement (Vierron & Steinbach, 2020). La densité d'ouvrage mesure le taux d'ouvrage par linéaire hydrographique. Le taux de fractionnement rapporte la somme des chutes artificielles sur le linéaire hydrographique, afin de caractériser l'effet barrière à la migration induit par les ouvrages. Il représente en quelque sorte la densité d'ouvrages pondérée par leur hauteur de chute. Le taux d'étagement rapporte la somme des chutes artificielles sur la dénivellation naturelle du cours d'eau afin

de tenir compte de l'effet retenue des ouvrages. Considérant que les ouvrages majeurs les plus impactants pour les communautés aquatiques ont été recensés dans le ROE, les ouvrages potentiellement manquants sur les tronçons ne portant pas de note (majoritairement des tronçons de têtes de bassins) ont probablement un impact plus limité. La note de 0 a donc été affectée à ces tronçons, au contraire de ce qu'on choisit les auteurs de la méthode initiale. Des détails supplémentaires sur ces indices et leur modalité de calcul peuvent être trouvés dans le rapport de Vierron & Steinbach (2020).

L'utilisation de ces indicateurs de pressions à la continuité écologique dans les SDM permet d'intégrer un aspect de la composante dispersion parmi les facteurs structurants la distribution spatiale des espèces (*i.e.* composantes abiotiques, biotiques et dispersion du diagramme BAM; Figure 3).

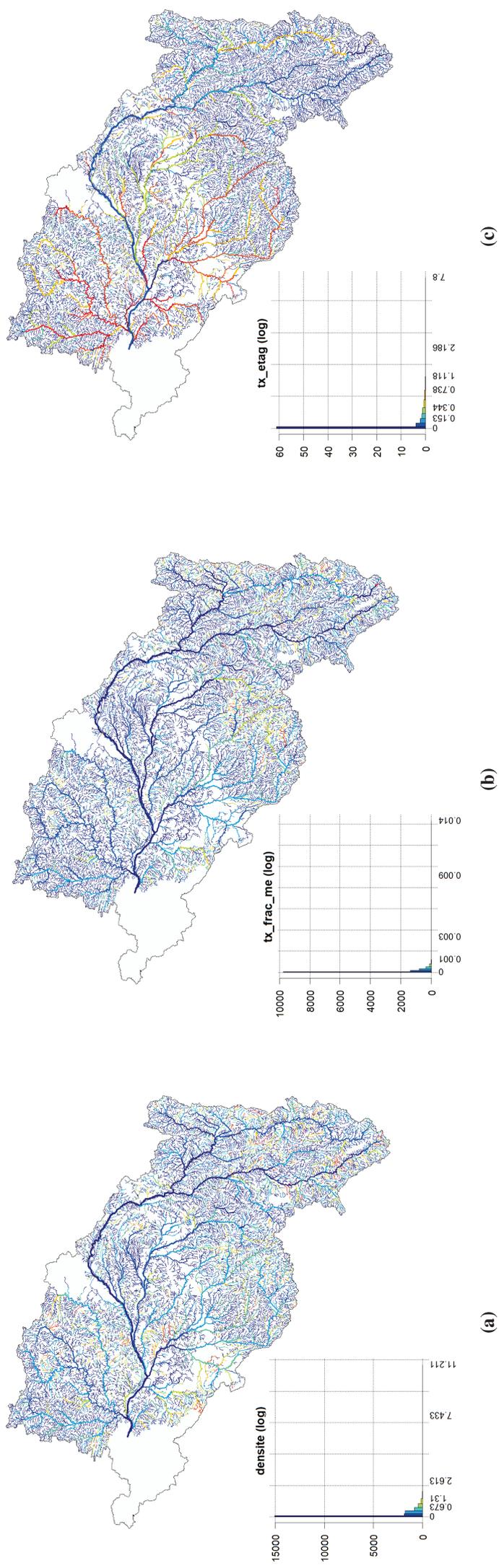


FIGURE 27 Indicateurs de pression sur la continuité écologique : (a) densité d'obstacles, (b) taux de fractionnement et (c) taux d'étagement. Les données sont issues du traitement de la base du référentiel des obstacles à l'écoulement réalisé par Vierron & Steinbach (2020).

B.4. Représentativité des conditions environnementales du bassin dans les jeux de données biologiques

Les jeux de données biologiques sont issus d'un échantillonnage des cours d'eau répartis sur l'ensemble du réseau hydrographique ligérien (Figure 20). Les localisations géographiques des stations ont été choisies pour être représentatives de l'ensemble des conditions biologiques et environnementales rencontrées sur le bassin versant de la Loire. Afin de s'assurer de la représentativité des données issues des stations de suivi, les gammes de valeurs de chaque paramètres rencontrés sur les stations sont comparés à celles du réseau. Ici, une comparaison supplémentaire de la représentativité des ordres de Strahler est réalisée.

Les jeux de données constitués à partir des tronçons portant une station de suivi hydrobiologique sont jugés représentatifs des conditions moyennes rencontrées sur le bassin versant de la Loire (Tableau 3 et Figure 28). En effet, les gammes de valeurs couvertes par chacun de ces deux jeux de données (*i.e.* poissons et macroinvertébrés) sont proches de celles observées sur l'ensemble du réseau H, hormis pour les valeurs extrêmes qui correspondent aux extrêmes amont et aval du gradient environnemental, et qui n'ont été que rarement échantillonnés (Tableau 3). Le faible nombre de tronçons échantillonnés avec des valeurs de pente $> 0,2$ et des taux d'étagement $> 0,4$, notamment pour les poissons, est lié à la répartition des tronçons échantillonnés, dont certains sont localisés sur des portions très amonts et abrupts du bassin ligérien. Ces tronçons représentent une minorité du réseau hydrographique considéré (quelques dizaines de tronçons sur les 33 711).

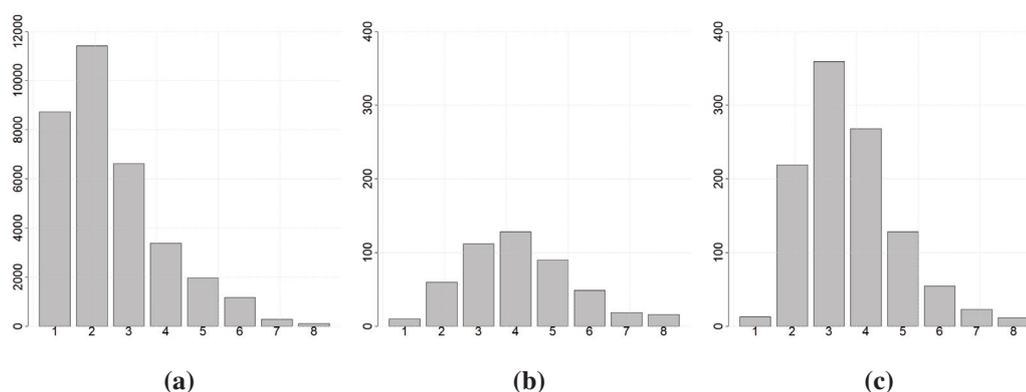


FIGURE 28

Représentativité des ordres de Strahler dans chacun des jeux de données : (a) réseau H complet (N = 33 711), (b) poisson (N = 484) et (c) macroinvertébrés (N = 1077).

TABLEAU 3
Gamme de valeurs des variables environnementales dans chacun des jeux de données.

Variable	Unité	Réseau H			Poisson			Macroinvertébrés		
		Min	Moyenne	Max	Min	Moyenne	Max	Min	Moyenne	Max
Ordre de Strahler	-	1	2,5184	8	1	4,0971	8	1	3,5534	8
Pente	%	0,000	0,013	0,284	0,000	0,005	0,062	0,000	0,006	0,081
Tw_m	°C	6,624	11,499	15,700	7,261	11,652	14,535	7,261	11,569	14,535
Tw_q10	°C	1,027	6,084	11,155	1,780	5,227	8,948	1,082	5,379	9,499
Tw_q90	°C	9,004	17,528	26,605	10,647	18,647	23,001	10,647	18,296	23,541
Tw_sd	°C	0,723	4,407	7,677	2,436	5,142	7,069	2,349	4,955	7,069
lm	m	0,083	5,251	122,217	1,212	14,158	109,628	1,056	9,036	109,628
l10	m	0,074	4,881	120,433	1,071	13,449	108,059	0,918	8,468	108,059
l90	m	0,087	5,478	123,664	1,284	14,627	110,885	1,102	9,396	110,885
hm	m	0,018	0,185	0,779	0,075	0,290	0,641	0,074	0,261	0,779
h10	m	0,012	0,133	0,689	0,048	0,225	0,513	0,044	0,195	0,689
h90	m	0,021	0,222	0,881	0,093	0,338	0,726	0,087	0,309	0,881
vm	m/s	0,000	0,019	0,071	0,004	0,033	0,065	0,002	0,029	0,065
v10	m/s	0,000	0,008	0,054	0,001	0,017	0,052	0,001	0,013	0,052
v90	m/s	0,001	0,033	0,097	0,007	0,053	0,088	0,003	0,049	0,089
P_inter	-	0,006	0,336	0,912	0,006	0,236	0,852	0,008	0,287	0,912
tx_etag	-	0,000	0,087	7,800	0,000	0,232	4,283	0,000	0,186	4,283
Tx_fractio	-	0,000	0,000	0,034	0,000	0,001	0,006	0,000	0,001	0,006
D_ouvrage	nb/km	0,000	0,000	0,011	0,000	0,000	0,004	0,000	0,000	0,004
den_pop_100	hab/km ²	1,704	57,207	4035,012	2,572	91,507	2219,433	2,572	106,631	3955,704
agrsyn1	-	-1,936	0,000	2,536	-1,936	-0,182	2,527	-1,936	-0,060	2,536

En résumé ...

- ★ Le bassin versant de la Loire regroupe une diversité de conditions hydro-climatiques qui façonne le gradient amont-aval et la succession des communautés aquatiques ;
- ★ Le réseau H constitue l'unité géographique d'étude (33 711 tronçons d'une longueur moyenne de $1,85 \pm 1,63$ km) et regroupe des données sur les abondances de 42 espèces de poissons (N = 484 tronçons échantillonnés) et 188 genres de macroinvertébrés (dont 55 d'EPTO ; N = 1 077 tronçons échantillonnés)
- ★ Les variables environnementales suivantes ont été intégrées : la topographie (pente), la température de l'eau et les variables hydrauliques (hauteur et largeur en eau, vitesse du courant ; régimes moyens et amplitude), la probabilité d'intermittence des cours d'eau, le gradient agricole, la densité de population et des indicateurs de la pression sur la continuité écologique ;
- ★ Les conditions environnementales rencontrées sur les stations biologiques sont représentatives des conditions moyennes rencontrées sur le bassin versant ligérien.

Chapitre III

**Apport des variables directes et
de l'approche
multi-compartiments pour
l'analyse de la structuration des
communautés aquatiques à
l'échelle d'un grand bassin
versant**

A. Introduction

Un postulat très répandu stipule qu'à large échelle les différents compartiments biologiques des milieux aquatiques (*e.g.* poissons, macroinvertébrés, macrophytes) s'assemblent et se structurent en fonction de filtres environnementaux similaires, et forment ainsi des patrons spatiaux semblables (Paavola *et al.*, 2006 ; Dolph *et al.*, 2011). La plupart des études se concentrent donc sur un unique compartiment aquatique, en assumant sa représentativité des autres groupes (Heino, 2010 ; Stewart *et al.*, 2018 ; Vehanen *et al.*, 2020). Cependant, de récentes études ont questionné ce postulat et évalué la concordance entre compartiments biologiques à différentes échelles spatiales. Par exemple, Backus-Freer et Pyron (2015) ont montré que la prédiction de la distribution spatiale des macroinvertébrés est meilleure avec les variations environnementales locales (*e.g.* turbidité) contrairement aux poissons, d'où la faible concordance entre ces deux compartiments. D'autres études ont suggéré que la concordance des patrons spatiaux entre organismes aquatiques pourrait augmenter avec l'échelle spatiale, probablement du fait d'une moindre contribution des processus locaux à large échelle (*i.e.* régionale). De plus, la concordance spatiale entre communautés pourrait résulter d'autres processus que les filtres environnementaux, même à large échelle. Par exemple, les interactions biotiques, les dynamiques de dispersion (voir le diagramme BAM présenté au chapitre I section A), l'effet d'échelle ou encore l'homogénéisation des assemblages entraînée par la perte de taxons (Jackson *et al.*, 2001 ; Olden, 2006 ; Dolph *et al.*, 2011). La considération de multiple compartiments biologiques dans l'investigation des processus écologiques à large échelle apparaît alors pertinente, voire nécessaire, pour assurer des résultats intégrateurs et généralisables. Cela permet d'identifier des patrons à l'échelle de l'écosystème, et pas seulement de la population ou de la communauté.

L'utilisation de variables dites « proxies » peut également limiter la compréhension des écosystèmes aquatiques et leur fonctionnement. Les *variables proxies* sont les variables classiquement utilisées pour approcher les conditions réellement endurées par les organismes, lorsque ces dernières ne sont pas accessibles (*e.g.* température de l'air pour approximer celle de l'eau) ou qui mesure une caractéristique générale du cours d'eau qui n'est pas en relation directe avec l'habitat des organismes (*e.g.* les débits pour la vitesse de courant) (Caissie, 2006). De nombreuses études doivent se reposer sur ces variables proxies en l'absence de variables plus fines, précises et disponibles à large échelle (*e.g.*

couvrant l'ensemble d'un réseau hydrographique) ou sur une période temporelle suffisante (Maire *et al.*, 2017 ; Flourey *et al.*, 2021). Par opposition, les *variables directes* sont celles qui constituent directement les conditions environnementales endurées par les organismes, c'est-à-dire leur habitat (Irving *et al.*, 2022).

Dans ce chapitre, la concordance entre les patrons de structuration des communautés de poissons et de macroinvertébrés est évaluée, en supposant, au regard de ce qui a été évoqué précédemment, une faible concordance à l'échelle du bassin versant de la Loire. L'intérêt des variables directes par rapport aux variables proxies dans leur capacité à expliquer les patrons de structuration des communautés aquatiques à l'échelle d'un grand bassin versant (*i.e.* Loire) est également exploré. Ainsi, nous avons testé si les variables directes, qui correspondent aux conditions d'habitat des organismes et à leurs préférences écologiques, retranscrivent mieux ces patrons de structuration spatiaux que les variables proxies (Caissie, 2006). Étant donné la plus grande sensibilité des macroinvertébrés aux conditions locales, du fait de leur aire de vie plus restreinte et de leur capacité de dispersion limitée, nous avons aussi supposé que ce compartiment serait mieux expliqué par des variables directes que les poissons. Chaque compartiment biologique a ses particularités de structuration spatiale et, sous l'hypothèse d'une faible concordance entre eux, nous avons testé si la combinaison des données des deux compartiments biologiques permettait une meilleure retranscription du gradient biologique et de ses liens avec les prédicteurs environnementaux (comparaison d'une analyse séparée des compartiments).

B. Données et méthodologie

Dans ce chapitre, les données utilisées sont celle présentées au chapitre II à l'échelle du bassin versant de la Loire (Figure 19), pour la période temporelle de 1994 à 2017 et le réseau T-NET comme unité géographique (N = 52 278 tronçons, de longueur moyenne $1,7 \pm 1,4$ km). La période temporelle sélectionnée dans ce chapitre est antérieure à la décision de restreindre les données à la période 2007-2017, d'où les deux périodes utilisées entre ce chapitre et les suivants. Les jeux de données n'étant pas exactement les mêmes, une présentation détaillée de ceux utilisés dans ce chapitre est faite ci-après.

Les analyses présentées ci-après ont été réalisées avec le logiciel R (R Core Team, 2022). Les packages *vegan* (Oksanen *et al.*, 2019) et (*Bivand & Lewin-Koh*, 2022) ont été utilisés pour les analyses statistiques et géomatiques, respectivement.

TABLEAU 4

Gamme environnementales couvertes par chacun des jeux de données. Valeurs moyennes [minimum : maximum] pour les 23 variables environnementales calculées sur : (1) l'ensemble du réseau T-NET (N = 52 278 tronçons), (2) tronçons portant de la donnée poissons (N = 1 191), (3) les tronçons portant de la donnée macroinvertébrés (N = 1 195) et (4) les tronçons portant de la donnée sur les deux compartiments, poissons et macroinvertébrés (jeu de données "commun"; N = 532). *Ta* et *Tw* font référence à la température de l'air et de l'eau, respectivement. Les données de *Ta* proviennent de la base SAFRAN, un système d'analyse atmosphérique à méso-échelle pour les variables de surface au pas de temps horaire, utilisant des observations de données au sol à la résolution spatiale d'une grille de 8 x 8 km (Loicq *et al.*, 2018). À l'échelle annuelle, la RMSE entre les observations et les sorties correspondantes de SAFRAN est de 1 °C (Vidal *et al.*, 2010). Les suffixes *m*, *sd*, *10* et *90* font référence à la moyenne, l'écart-type, les percentiles 10 et 90, respectivement.

	Variables environnementales	Réseau T-NET	Poissons	Macroinvertébrés	Commun
Variables proxy	Débit m (m ³ /s)	2.9 [<0.01 : 820.3]	10.2 [<0.01 : 675.4]	8.5 [0.01 : 675.4]	14.7 [0.01 : 675.4]
	Débit 10 (m ³ /s)	0.7 [<0.01 : 212.1]	2.5 [<0.01 : 171.1]	2.1 [<0.01 : 171.1]	3.6 [<0.01 : 171.1]
	Débit 90 (m ³ /s)	6.2 [<0.01 : 1704.9]	21.6 [<0.01 : 1408.7]	18 [0.02 : 1408.7]	31.2 [0.02 : 1408.7]
	Débit sd (m ³ /s)	2.7 [<0.01 : 667.4]	9.2 [<0.01 : 547.4]	7.9 [0.01 : 547.4]	13.4 [0.01 : 547.4]
	Ta m (°C)	10.7 [6.1 : 12.5]	10.3 [6.2 : 12.5]	10.9 [6.2 : 12.5]	10.8 [6.2 : 12.5]
	Ta 10 (°C)	1.9 [-2.5 : 4.2]	1.4 [-2.5 : 4.2]	2.0 [-2.5 : 4.2]	2.0 [-2.1 : 4.2]
	Ta 90 (°C)	19.4 [15.3 : 21.1]	19.2 [15.3 : 21.1]	19.6 [15.3 : 21.1]	19.6 [15.3 : 21.1]
	Ta sd (°C)	6.6 [5.5 : 7.3]	6.7 [5.7 : 7.3]	6.6 [5.7 : 7.3]	6.7 [5.7 : 7.3]
Variables directes	Hauteur en eau m (m)	0.24 [0.02 : 2.86]	0.42 [0.05 : 2.61]	0.46 [0.07 : 2.61]	0.55 [0.07 : 2.61]
	Hauteur en eau 10 (m)	0.15 [0.01 : 1.79]	0.26 [0.04 : 1.65]	0.28 [0.04 : 1.65]	0.34 [0.05 : 1.65]
	Hauteur en eau 90 (m)	0.30 [0.03 : 3.70]	0.53 [0.06 : 3.34]	0.58 [0.08 : 3.34]	0.70 [0.08 : 3.34]
	Vitesse de courant m (m/s)	0.08 [<0.01 : 4.14]	0.20 [<0.01 : 3.78]	0.20 [0.01 : 3.78]	0.28 [0.01 : 3.78]
	Vitesse de courant 10 (m/s)	0.02 [<0.01 : 1.23]	0.05 [<0.01 : 1.08]	0.05 [<0.01 : 1.08]	0.08 [<0.01 : 1.08]
	Vitesse de courant 90 (m/s)	0.15 [<0.01 : 7.97]	0.40 [0.01 : 7.37]	0.40 [0.01 : 7.37]	0.56 [0.01 : 7.37]
	Largeur en eau m (m)	4.6 [0.1 : 204.9]	11.2 [0.5 : 176.3]	11.3 [0.7 : 176.3]	15.6 [0.7 : 176.3]
	Largeur en eau 10 (m)	4.0 [0.08 : 181.2]	9.7 [0.4 : 156.7]	9.8 [0.6 : 156.7]	13.5 [0.6 : 156.7]
	Largeur en eau 90 (m)	5.0 [0.1 : 219.0]	12.1 [0.5 : 187.8]	12.2 [0.7 : 187.8]	16.8 [0.7 : 187.8]
	Tw m (°C)	11.2 [5.4 : 16.2]	11.0 [5.9 : 14.1]	11.5 [6.8 : 14.1]	11.5 [6.8 : 14.1]
	Tw 10 (°C)	6.3 [0.1 : 11.5]	4.3 [0.1 : 9.5]	4.8 [0.2 : 8.7]	4.6 [0.5 : 8.4]
	Tw 90 (°C)	17.2 [8.4 : 27.6]	18.5 [10.2 : 24.5]	19.1 [11.4 : 24.5]	19.3 [11.4 : 24.5]
	Tw sd (°C)	4.1 [0.3 : 8.0]	5.3 [2.0 : 7.7]	5.4 [2.7 : 7.7]	5.5 [2.7 : 7.7]
Autres	Distance à l'estuaire (km)	514.5 [110.1 : 1004.5]	601.0 [114.7 : 999.5]	512.0 [120.2 : 999.5]	511.4 [138.2 : 999.5]
	Pente (%)	0.02 [<0.01 : 0.40]	0.02 [<0.01 : 0.17]	0.01 [<0.01 : 0.12]	0.01 [<0.01 : 0.12]

B.1. Variables environnementales directes et proxies

Un ensemble de vingt-trois variables environnementales connues pour influencer les communautés de poissons et de macroinvertébrés ont été considérées, incluant des paramètres de la température de l'air et de l'eau, des débits, des conditions hydrauliques locales et de la topographie (voir le tableau 4 pour la liste complète). Les métriques moyenne, percentiles 10 et 90 et écart-type sont utilisés pour chaque paramètre environnemental pour rendre compte du régime moyen, de son amplitude et de la variabilité au travers des extrêmes.

B.2. Données poissons et macroinvertébrés

Les données biologiques proviennent de la base nationale Naiades (voir chapitre II section B.1) et des suivis hydrobiologiques de cinq CNPE réalisés par EDF. La période d'échantillonnage retenue pour ce chapitre va de 1994 à 2017. Ainsi, 50 espèces de poissons collectées au cours de 5 225 opérations de pêche réparties sur 1 393 stations et 230 genres de macroinvertébrés récoltés au cours de 4 645 échantillonnages répartis sur 1 215 stations constituent les jeux de données. Cela représente, 2 950 799 individus de poissons parmi dix-huit familles, avec une majorité de cyprinidés (20 espèces) et 8 334 023 individus de genres macroinvertébré parmi 186 genres dont les plus représentés sont les trichoptères (N = 55 genres), les éphéméroptères (N = 27), les coléoptères (N = 27), les gastropodes (N = 17) les odonates (N = 17) et les plécoptères (N = 14). Elles ont été standardisées en CPUE (ind/100m²) pour chaque tronçon du réseau T-NET où au moins une station d'échantillonnage, soit 1 191 tronçons pour les poissons et 1 195 tronçons pour les macroinvertébrés ; respectivement nommés jeux « poisson » et « macroinvertébré ». Cela représente en moyenne pour les poissons, 97 [11 ; 253] ind/100m² (moyenne [5^{ième} et 95^{ième} percentiles]) et 10 [1 ; 26] espèces par tronçons T-NET, et pour les macroinvertébrés 3 851 [541 ; 10 123] ind/100m² et 43 [16 ; 71] genres par tronçons T-NET. Un troisième jeu de données ne contenant que les tronçons portant à la fois au moins une station de suivi hydrobiologique poisson et une station macroinvertébré simultanément a été constitué, soit 532 tronçons T-NET ; il est ensuite nommé jeu « commun ». Afin de limiter l'influence des taxons trop rares, tous les taxons dont l'occurrence totale est inférieure à 0,5 % ont été retirés des jeux de données. Au final, 43 espèces de poissons et 186 genres de macroinvertébrés ont été considérés pour les analyses.

B.3. Analyse de la structuration des communautés par ordination multivariée

L'ordination par *Non-metric Multidimensional Scaling* (NMDS) a été utilisée pour examiner la structuration des communautés aquatiques à l'échelle du bassin versant de la Loire. Cette méthode d'ordination itérative se base sur une matrice de dissimilarité pour ordonner les individus statistiques (ici le tronçon T-NET) dans un espace factoriel (Legendre & Legendre, 2012). La NMDS est adaptée aux données biologiques comportant des taxons rares et/ou peu abondants, en particulier lorsque plusieurs compartiments biologiques sont considérés (Minchin, 1987). De plus, la NMDS est performante pour montrer les patrons de structuration clés à l'échelle de la communauté, en évitant l'emphase sur les spécificités de certains taxons.

Pour chacun des trois jeux de données biologiques, une matrice de dissimilarité basée sur les CPUE transformée par *Wisconsin* et la distance de Bray-Curtis entre chaque tronçon a été construite. La meilleure ordination par NMDS a été sélectionnée après 200 itérations, en partant d'une ordination aléatoire. La qualité de la NMDS a été évaluée par la mesure de stress, qui correspond à la qualité de l'ajustement entre le rang des dissimilarités observées et la distance d'ordination. Un stress inférieur à 0,2 correspond à une ordination correcte, précise en-dessous de 0,1 (Clarke, 1993). Sur la base de cette mesure, trois axes ont été conservés pour l'espace factoriel.

B.4. Évaluation de la concordance des patrons spatiaux entre les communautés aquatiques

La concordance entre les ordinations par NMDS des différents jeux de données biologiques (*i.e.* poissons, macroinvertébrés et communs), et donc entre les patrons de structuration des communautés poissons et macroinvertébrés, est évaluée par une analyse procustéenne. L'analyse procustéenne a été préférée au test de Mantel pour son meilleur pouvoir statistique (Peres-Neto & Jackson, 2001). L'analyse résulte en une valeur de concordance m^2 , comprise entre 0 et 1 (de forte à aucune concordance ; Peres-Neto & Jackson, 2001). De par sa construction, l'analyse procustéenne requiert l'utilisation des mêmes sites entre les ordinations, des NMDS ont donc été construites sur les données biologiques uniquement pour les 532 tronçons communs. Enfin, une quatrième NMDS a été construite à partir des vingt-trois variables environnementales afin de comparer la concordance de la

structuration des données biologiques, formant le gradient biologique, à celle des variables environnementales formant le gradient environnemental.

B.5. Comparaison de l'apport des variables environnementales directes par rapport aux variables proxies

Afin d'évaluer l'influence des variables environnementales dans la structuration des communautés aquatiques mise en évidence par NMDS, une *Redundancy Analysis* (RDA) est réalisée. La RDA a été utilisée comme une extension à la régression linéaire autorisant l'utilisation d'une variable réponse multi-variées et capable de surpasser la colinéarité inhérente aux différentes métriques calculées pour chaque variables environnementales (*i.e.* moyenne, quantile 10 et 90 %, et écart-type) (Peres-Neto *et al.*, 2006). Premièrement, pour chaque ordination NMDS sur la base des trois jeux de données biologiques (*i.e.* poisson, macroinvertébré et commun) et chaque type de variables environnementales (*i.e.* température de l'eau et de l'air, débits et variables hydrauliques), une RDA est réalisée avec les trois axes de NMDS comme variables réponses et les groupes de 4 métriques de variables environnementales (*i.e.* moyenne, quantiles 10 et 90 % et écart-type) en variables explicatives. Deuxièmement, afin de comparer l'apport des variables directes *vs.* proxies, un partitionnement de variance est réalisé à partir de la RDA pour chaque groupes biologiques et métriques de variables environnementales. Le pouvoir explicatif respectif de chaque variable (coefficient de détermination R^2) est calculé pour chaque compartiment biologique. Pour les variables hdyrauliques, qui comportent neuf métriques (pour chacun des trois variables, vitesse de courant, hauteur et largeur en eau), toutes les combinaisons de quatre parmi neufs ont été testées, pour conserver le même nombre de variables explicatives et degrés de liberté entre les RDA. Pour minimiser l'influence de l'autocorrélation spatiale inhérente aux données dans les mesures statistiques, la RDA a été combinée à une procédure de randomisations de Monte-Carlo. Dans un premier temps, les tronçons contenus dans 70% des sous-bassins versants (*i.e.* parmi les 368 sous-bassins considérés dans le modèle EROS) ont servi à la calibration de la RDA. La procédure est répétée 999 fois. L'évaluation des pouvoirs explicatifs et de leur significativité entre groupes de variables environnementales est réalisée par comparaison du chevauchement des intervalles de confiance, calculés comme égaux à deux écart-types.

C. Patrons de structuration obtenus à partir des ordinations multivariées

La NMDS basée sur les données poissons a un stress de 0,11, indiquant une bonne représentation de la dissimilarité inter-tronçons (Figure 29). Les communautés piscicoles présentent aux valeurs positives le long de l'axe 1 ont une faible richesse spécifique et sont dominées par la truite commune (*Salmo trutta*), tandis que celles du côté négatif de l'axe ont une forte richesse spécifique et sont dominées par les cyprinidés (Figure 32).

La NMDS basée sur les données macroinvertébrés a un stress de 0,16, indiquant une bonne représentation de la dissimilarité inter-tronçons (Figure 30). Les communautés macroinvertébrés situées sur les valeurs positives du second axe montre des assemblages à forte richesse générique, tandis que celles sur les valeurs négatives ont une faible richesse générique (figure 32).

La NMDS basée sur les données communes a un stress de 0,13, indiquant une bonne représentation de la dissimilarité inter-tronçons (Figure 31). L'ordination des centroïdes des espèces de poissons sur le premier plan factoriel est très similaire à celui observé dans l'ordination basée uniquement sur les données poissons. Le patron de structuration est bien moins clair pour les macroinvertébrés, où les communautés riches en macroinvertébrés incluent des taxons abondants et communs vers le centre de l'espace factoriel et une déclinaison d'assemblages plus particuliers avec des taxons plus rares localisés dans toute la périphérie du nuage de points (Figure 32).

Les NMDS basées sur les trois jeux de données biologiques soulignent le rôle du gradient longitudinal dans la structuration des communautés aquatiques à l'échelle d'un grand bassin versant. La NMDS basée sur les densités de poissons retranscrit clairement ce gradient longitudinal en indiquant une opposition entre les variables thermiques et hydrauliques-hydrologiques (valeurs croissantes de l'amont vers l'aval) avec les variables topographiques (décroissantes). La communauté poisson s'organise le long de cet axe amont-aval, depuis des communautés dominées par les salmonidés dans les tronçons amont vers des communautés dominées par les cyprinidés vers l'aval, associée à l'augmentation de la richesse spécifique et de la densité totale. L'organisation des poissons qui en résulte est cohérente avec celle généralement décrite dans la littérature (Matthews, 1998 ; Laffaille *et al.*, 2020). La relation entre le gradient amont-aval et les densités de macroinvertébrés est moins claire mais reste visible sur le second axe de la NMDS. Dans l'ensemble, pour

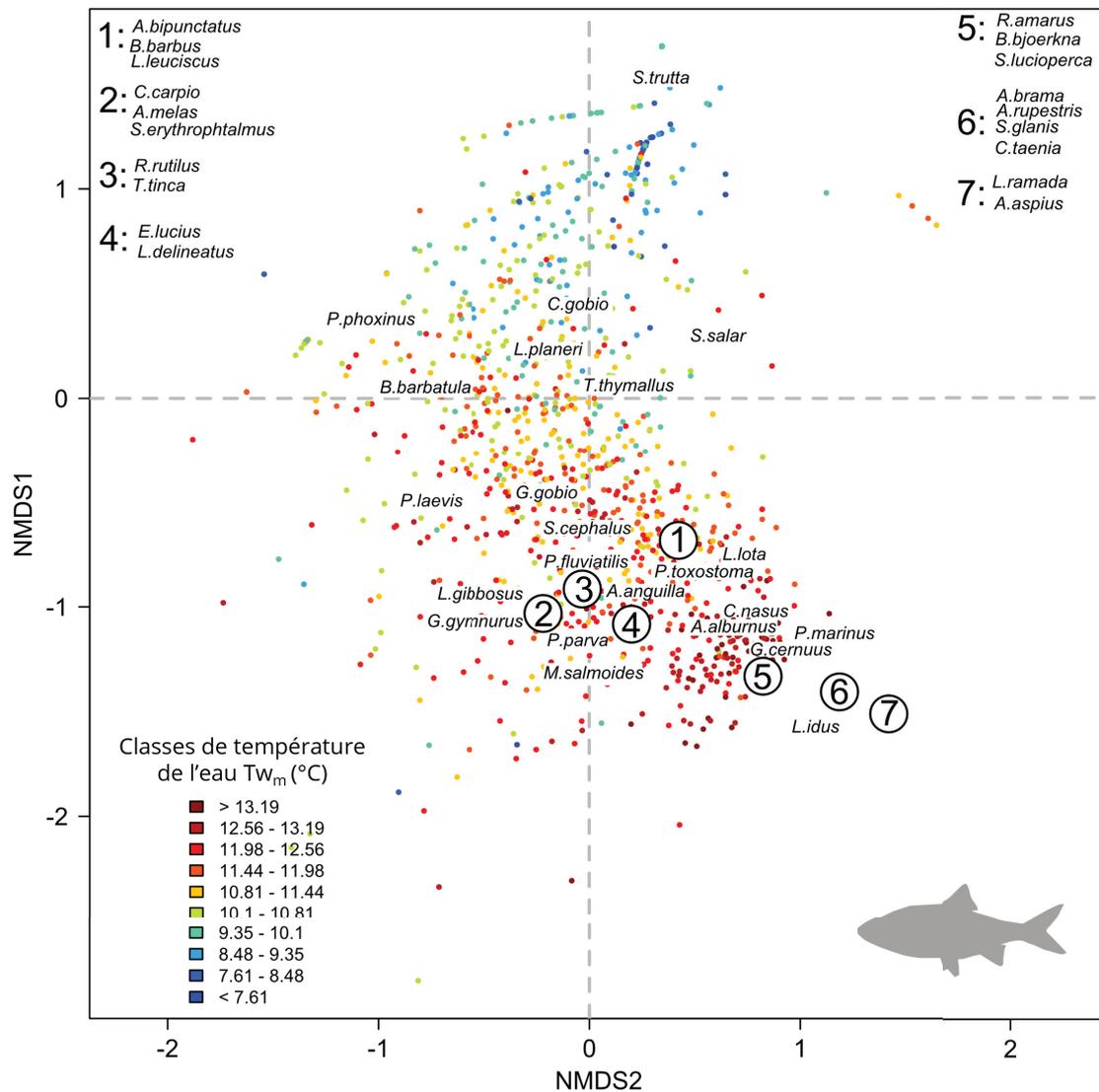


FIGURE 29

Ordination biologique basée sur les données poissons. Positionnement des tronçons T-NET dans le premier plan factoriel formé par les axes 1 et 2 (N = 1 191 tronçons). Chaque tronçons (= point) est coloré en fonction de sa température moyenne de l'eau ($T_w m$; en °C). Les bulles chiffrées indiquent les groupes d'espèces proches dans l'espace factoriel, dont les noms ont été listés dans les coins supérieurs à des fins de lisibilité (Voir les annexes de l'article disponible en annexe A pour les noms complets des espèces et les coordonnées du centroïde des espèces sur les trois axes de la NMDS).

les trois NMDS basées sur les données biologiques, aucun regroupement distinct de tronçons n'est observé, bien que les échantillonnages couvrent des entités géographiques bien différentes (e.g. Massif Central, vallée de la Loire). Cela soutient l'idée que la succession entre les communautés est graduelle et continue, avec des zones de transition où des espèces avec différentes exigences écologiques coexistent (comme formulé dans le RCC) (Vannote *et al.*, 1980).

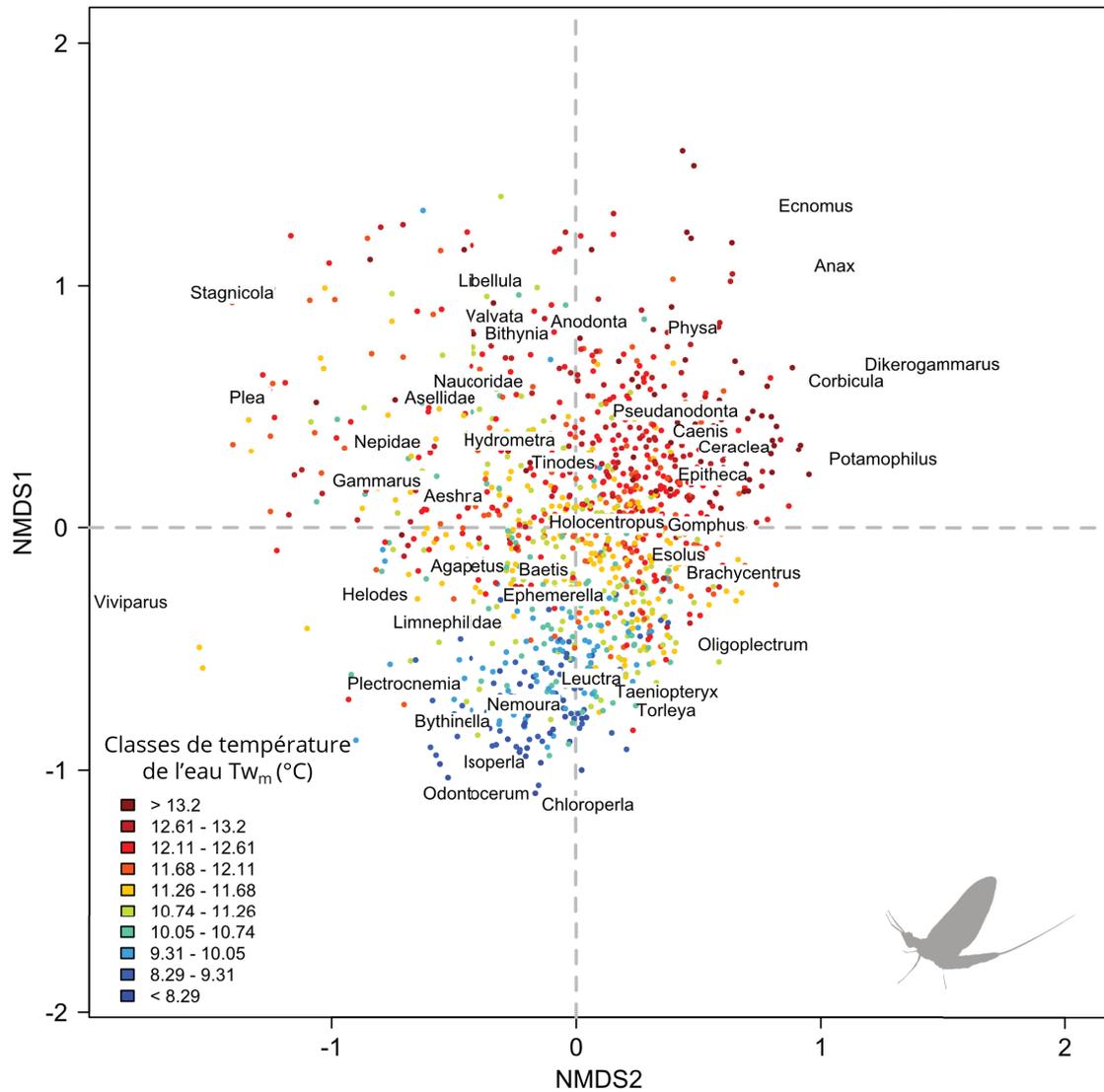


FIGURE 30

Ordinations biologiques basées sur les données macroinvertébrés. Positionnement des tronçons T-NET dans le premier plan factoriel formé par les axes 1 et 2 ($N = 1\,195$ tronçons). Chaque tronçons (= point) est coloré en fonction de sa température moyenne de l'eau (Tw_m ; en °C). Les bulles chiffrées indiquent les groupes d'espèces proches dans l'espace factoriel, dont les noms ont été listés dans les coins supérieurs à des fins de lisibilité (Voir les annexes de l'article disponible en annexe A pour les noms complets des espèces et les coordonnées du centroïde des genres sur les trois axes de la NMDS).

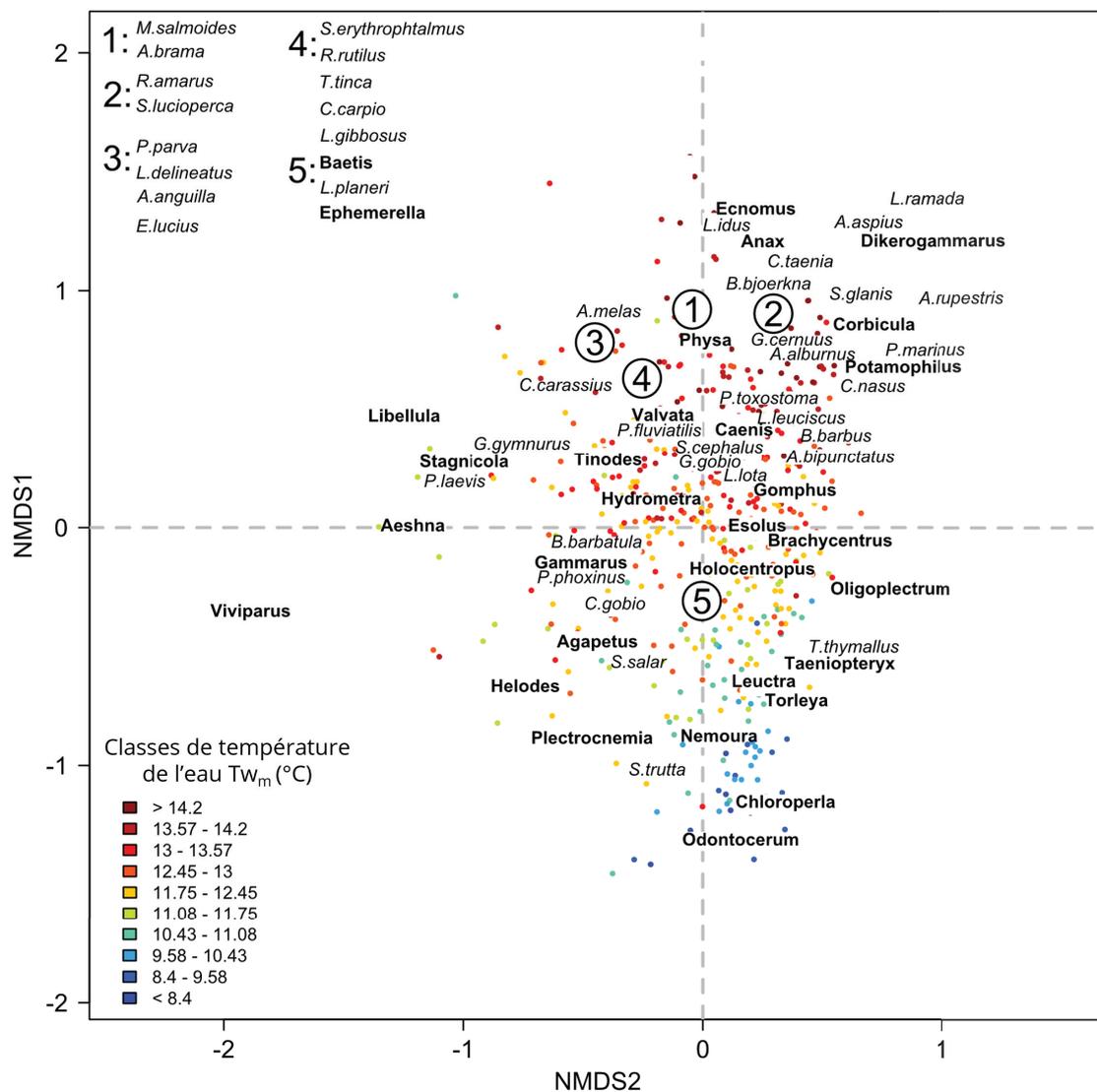


FIGURE 31

Ordination biologique basée sur les données poissons et macroinvertébrés. Positionnement des tronçons T-NET dans le premier plan factoriel formé par les axes 1 et 2 ($N = 532$ tronçons). Chaque tronçons (= point) est coloré en fonction de sa température moyenne de l'eau (T_{w_m} ; en °C). Les bulles chiffrées indiquent les groupes d'espèces de poissons (en *italique*) et genres de macroinvertébrés (en **gras**) proches dans l'espace factoriel, dont les noms ont été listés dans les coins supérieurs à des fins de lisibilité (Voir les annexes de l'article disponible en annexe A pour les noms complets des espèces et les coordonnées du centroïde des taxons sur les trois axes de la NMDS).

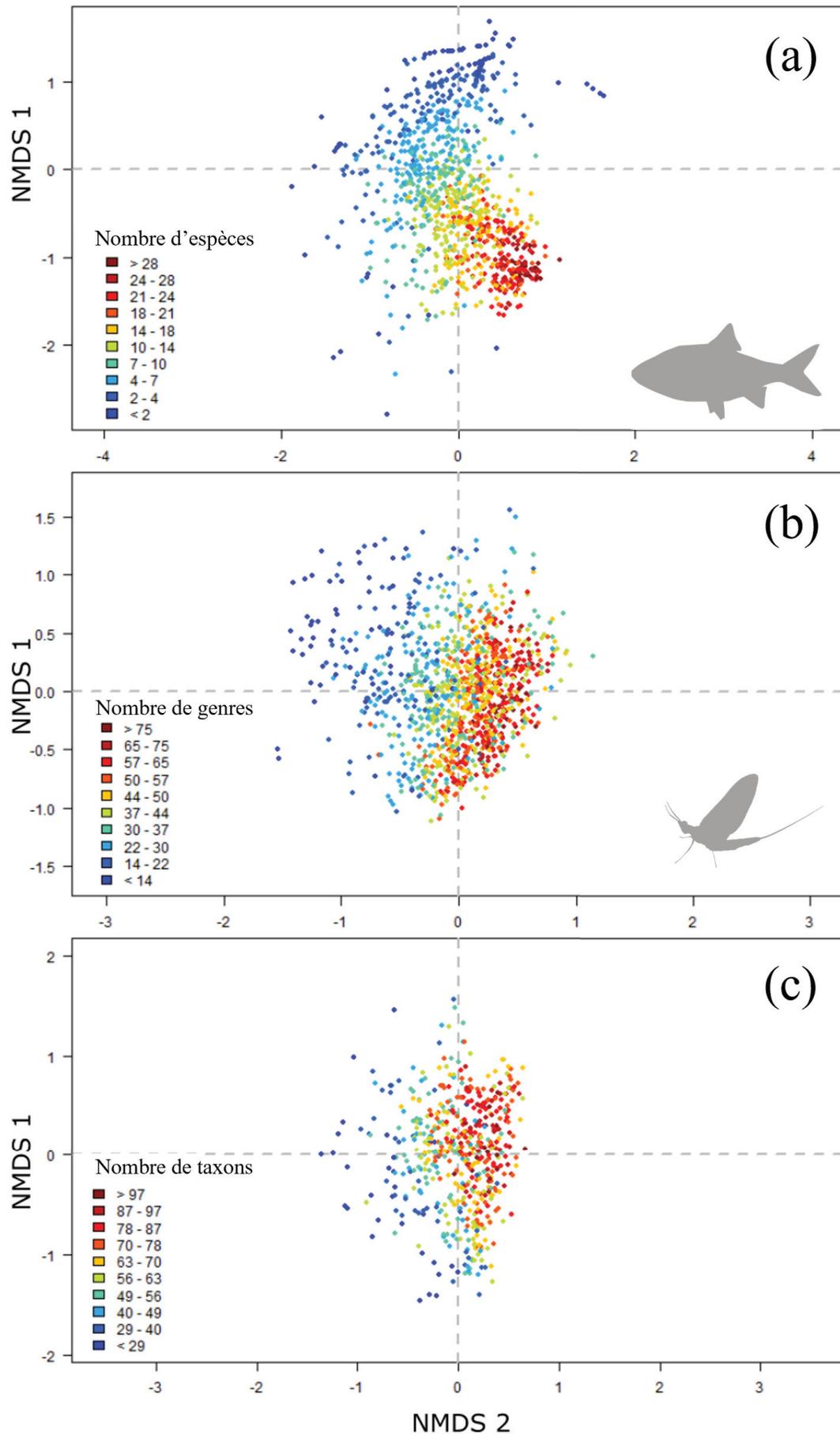


FIGURE 32

Richesse taxonomique illustrée dans les ordinations biologiques poissons (a), macroinvertébrés (b) et communes (c).

D. Intérêt des variables environnementales directes par rapport aux proxies dans la structuration des communautés aquatiques

L'ordination poisson se structure principalement le long du gradient amont-aval comme l'illustre la température de l'eau en Figure 29. Les résultats du partitionnement de variance de la RDA sont présentés en Figure 33. Les variables structurantes de l'ordination poisson sont, par ordre d'importance, la température de l'eau (R^2 explicatifs moyenné sur 999 itérations = 47,1 %), les variables hydrauliques ($R^2 = 43,9$), la pente et la distance à l'estuaire ($R^2 = 43,3$), la température de l'air ($R^2 = 41,0$) et le débit ($R^2 = 28,1$). Considérées ensemble, les variables directes expliquent l'ordination poisson avec un $R^2 = 54,7$ et les variables proxies avec un $R^2 = 51,8$.

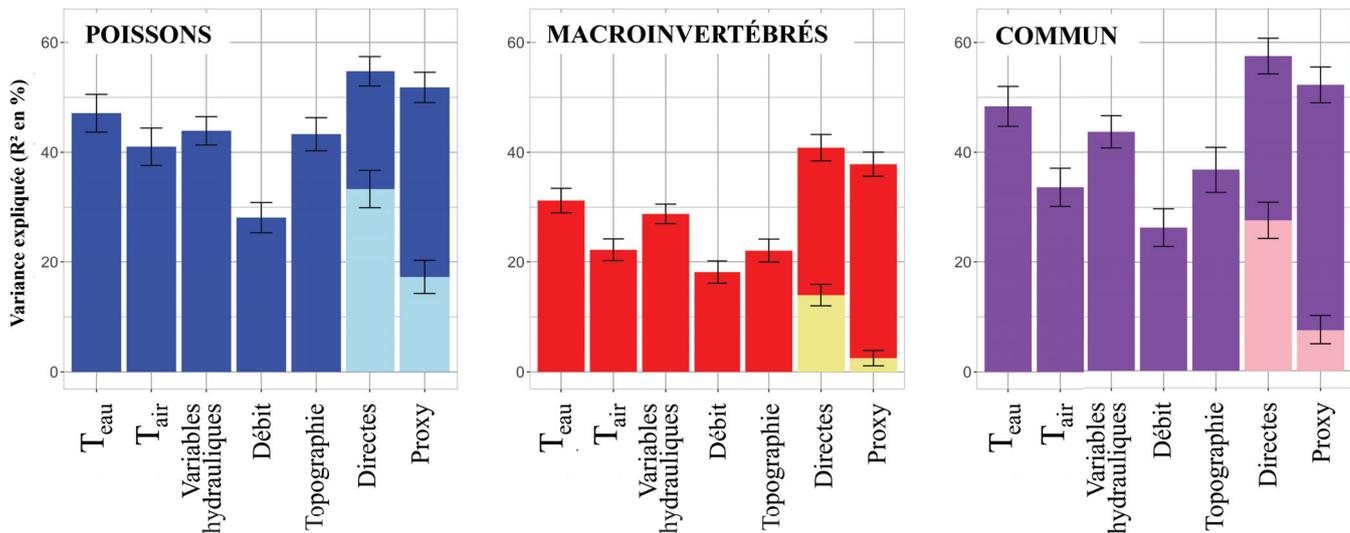


FIGURE 33

Importance de chaque variable environnementale dans la structuration des ordinations

biologiques (A) poissons, (B) macroinvertébrés et (C) jeu commun : Les coefficients de détermination moyens (R^2 en %) et l'intervalle de confiance à 95% associé proviennent d'une

RDA (combinée à un partitionnement de variance pour les modèles intégrant les variables directes et proxies) et obtenu pour 999 sous-échantillonnage aléatoire des tronçons contenus dans 70% des sous-bassins. Les fractions de variance expliquée partagées sont en couleur plus claires dans les barres correspondants aux modèles intégrant les variables directes et proxies.

La structuration de l'ordination macroinvertébrés par le gradient environnemental principal illustré par la température de l'eau est aussi visible mais moins apparente que pour les poissons, en Figure 30. Les résultats du partitionnement de variance de la RDA sont présentés en Figure 33. Les variables structurantes de l'ordination macroinvertébrés sont,

D. Intérêt des variables environnementales directes par rapport aux proxies dans la structuration des communautés aquatiques

par ordre d'importance, la température de l'eau ($R^2 = 31,2$), les variables hydrauliques ($R^2 = 28,8$), la température de l'air ($R^2 = 22,2$), la pente et la distance à l'estuaire ($R^2 = 22,1$) et le débit ($R^2 = 18,1$). Considérée ensemble, les variables directes expliquent l'ordination macroinvertébré avec un $R^2 = 40,8$ et les variables proxies avec un $R^2 = 37,8$.

L'ordination produite sur les données communes est présentée en Figure 31. Elle présente des similarités avec les ordinations poisson et macroinvertébré au niveau des placements des tronçons les uns par rapport aux autres et de coordonnées des centroïdes des taxons. Les résultats du partitionnement de variance de la RDA sont présentés en Figure 33. Les variables structurantes de l'ordination commune sont, par ordre d'importance, la température de l'eau ($R^2 = 48,4$), les variables hydrauliques ($R^2 = 43,7$), la pente et la distance à l'estuaire ($R^2 = 36,8$), la température de l'air ($R^2 = 33,6$) et le débit ($R^2 = 26,3$). Considérées ensemble, les variables directes expliquent l'ordination commune avec un $R^2 = 57,5$ et les variables proxies avec un $R^2 = 52,3$.

Concernant le pouvoir explicatif de la RDA, toutes les différences de R^2 entre chaque paire de variables directes ou proxies (*i.e.* température de l'eau vs. de l'air, variables hydrauliques vs. débit, et directes vs. proxies combinées) est significative, comme le montre l'absence de chevauchement entre les intervalles de confiance, hormis avec les moyennes des combinaisons directes et proxies. Concernant le pouvoir prédictif de la RDA, la plupart des R^2 sont supérieurs avec l'utilisation des variables directes comparées aux variables proxies, bien que ces différences soient pour la plupart non significatives. Par ailleurs, l'analyse de sensibilité aux incertitudes liées à la provenance des variables environnementales de modèles à base physique montre une robustesse certaine des résultats présentés ici.

D.1. Procédure de validation croisée de Monte-Carlo et du pouvoir prédictif de la RDA

En complément des résultats concernant le pouvoir explicatif de la RDA, nous avons également testé leur pouvoir prédictif sur les 30 % de sous-bassins restants, qui n'ont pas été pris en compte dans la calibration de la RDA (Tableau 5). Comme pour les résultats concernant le pouvoir explicatif de la RDA, toutes les variables directes (c'est-à-dire les variables hydrauliques, la température de l'eau et les deux combinées) ont montré un R^2 prédictif plus élevé en moyenne que leurs équivalents proxies (c'est-à-dire le débit, la tem-

pérature de l'air et les deux combinés) pour les trois groupes biologiques, à l'exception du débit par rapport aux variables hydrauliques pour l'ensemble de données commun pour lequel la différence était légèrement opposée. Pour les macroinvertébrés, les différences de pouvoir prédictif entre la température de l'eau et la température de l'air et entre les variables directes combinées et les variables de substitution combinées n'étaient pas significatives, comme pour l'ensemble de données commun, à l'exception de la différence entre la température de l'eau et la température de l'air (c'est-à-dire que la moyenne du R^2 de substitution n'est pas incluse dans l'intervalle de confiance du R^2 direct, mais les intervalles de confiance se chevauchent), les variables directes prédisant mieux les ordinations biologiques que les variables de substitution. La principale différence avec les résultats concernant le pouvoir explicatif est l'élargissement des intervalles de confiance autour du R^2 , ce qui rend la différence entre les paires de variables environnementales non significative en raison du chevauchement des intervalles de confiance. En outre, les R^2 prédictifs étaient plus élevés que leurs équivalents explicatifs. Cela a mis en évidence le rôle prédominant des gradients environnementaux dans la structuration de l'organisation spatiale des communautés aquatiques à l'échelle du bassin versant.

TABLEAU 5

Importance des variables environnementales pour expliquer ou prédire l'ordination biologique des ensembles de données sur les poissons (A), les macroinvertébrés (B) et les espèces communes (C) : coefficients de détermination moyens (R^2) et intervalles de confiance à 95 % associés, dérivés de la RDA. Le R^2 explicatif a été dérivé de la RDA réalisée sur les tronçons contenus dans 70 % des sous-bassins ; le R^2 prédictif a été dérivé de la RDA prédictive réalisée sur les tronçons contenus dans les 30 % restants des sous-bassins.

Tableau A <i>Poissons</i>	Variables environnementales	R^2 explicatif	R^2 prédictif
Régime hydrologiques	Débites	28.1 [25.4; 30.9]	52.4 [43.2; 61.6]
	Variables hydrauliques	43.9 [41.3; 46.5]	58.7 [53.0; 64.4]
Régime thermique	Température de l'air	41.0 [37.6; 44.4]	54.1 [44.9; 63.3]
	Température de l'eau	47.1 [43.7; 50.5]	58.2 [49.4; 67.0]
Variables proxy	Température de l'air + débits	51.8 [49.1; 54.6]	62.3 [55.5; 69.1]
Variables directes	Température de l'eau + variables hydrauliques	54.7 [52.1; 57.4]	64.5 [58.3; 70.7]

Tableau B <i>Macroinv.</i>	Variables environnementales	R^2 explicatif	R^2 prédictif
Régime hydrologiques	Débites	18.1 [16.1; 20.2]	47.9 [38.8; 57.0]
	Variables hydrauliques	28.8 [27.0; 30.6]	47.5 [42.6; 52.4]
Régime thermique	Température de l'air	22.2 [20.0; 24.2]	46.2 [37.9; 54.5]
	Température de l'eau	31.2 [29.0; 33.4]	53.9 [46.9; 60.9]
Variables proxy	Température de l'air + débits	37.8 [35.6; 40.0]	50.7 [45.2; 56.2]
Variables directes	Température de l'eau + variables hydrauliques	40.8 [38.4; 43.2]	54.7 [48.6; 60.8]

Tableau C <i>Commun</i>	Variables environnementales	R^2 explicatif	R^2 prédictif
Régime hydrologiques	Débites	26.3 [22.9; 29.7]	61.7 [47.4; 74.4]
	Variables hydrauliques	43.7 [40.7; 46.6]	61.3 [53.7; 68.9]
Régime thermique	Température de l'air	33.6 [30.2; 37.1]	49.4 [39.8; 59.0]
	Température de l'eau	48.4 [44.7; 52.0]	63.7 [54.8; 72.6]
Variables proxy	Température de l'air + débits	52.3 [49.0; 55.5]	61.6 [54.7; 68.5]
Variables directes	Température de l'eau + variables hydrauliques	57.5 [54.3; 60.8]	67.8 [61.2; 74.4]

D.2. Analyse de sensibilité des résultats de la RDA aux incertitudes des variables environnementales dérivées de modèles physiques

Des variables environnementales dérivées de modèles physiques avec des incertitudes connues ont été utilisées. Afin d'évaluer l'influence de ces incertitudes sur les résultats de la RDA, nous avons effectué une analyse de sensibilité comme suit. Nous avons utilisé les valeurs de la littérature pour les incertitudes sur les variables de substitution et les variables hydrauliques. Des biais de 0,3 m, 0,3 m/s et 1°C ont été utilisés pour la profondeur et la largeur de l'eau, la vitesse du courant et la température de l'air, respectivement (Quintana-Segui *et al.*, 2008 ; Morel *et al.*, 2020). Pour la température de l'eau, un biais de 0,8°C a été considéré, correspondant à l'écart-type des erreurs entre les moyennes annuelles observées et prédites (Figure 34. a ; Seyedhashemi *et al.* 2022). Pour le débit, un biais relatif de 10,6 % a été considéré, correspondant à l'écart-type des erreurs entre les moyennes annuelles observées et prédites (Figure 34. b ; Seyedhashemi *et al.* 2022). Comme nous supposons que l'erreur est centrée sur zéro (comme le montre la Figure 1 pour la température de l'eau et le débit), les biais précédemment spécifiés associés à chaque variable modélisée ont été utilisés comme écart type d'une distribution normale centrée sur 0 pour échantillonner les erreurs de manière aléatoire.

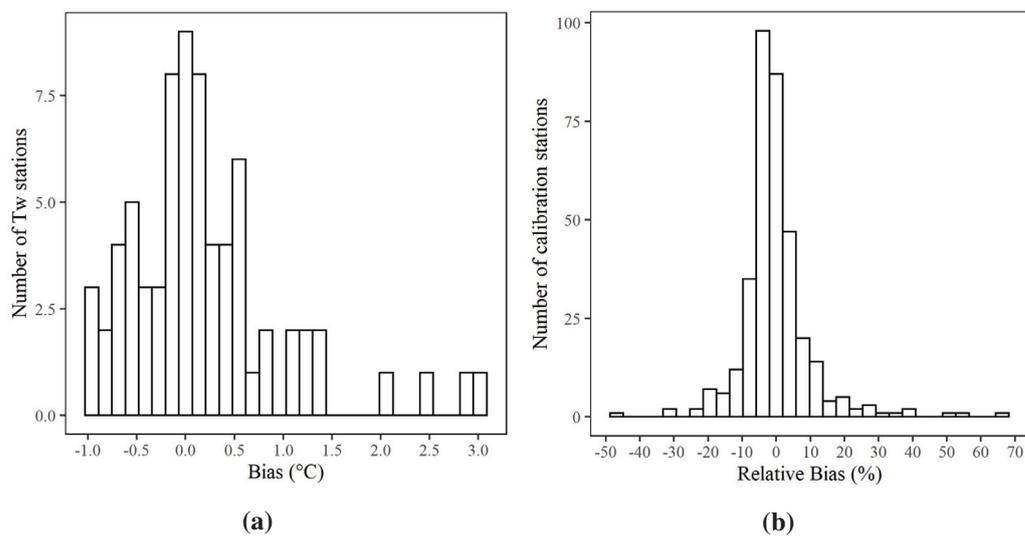


FIGURE 34

Distribution des biais annuels entre les valeurs observées et simulées pour la température de l'eau (A) et le débit (B) fournies par les modèles TNET et EROS, respectivement. Adapté de Seyedhashemi *et al.* (2022).

Pour réaliser cette analyse de sensibilité, nous avons reproduit la procédure de validation croisée décrite précédemment (voir chapitre III, section B.5), sauf que chaque valeur

D. Intérêt des variables environnementales directes par rapport aux proxies dans la structuration des communautés aquatiques

de la température de l'eau, de la température de l'air, du débit et des variables hydrauliques a été modifiée par l'ajout d'une erreur obtenue de manière aléatoire comme décrit ci-dessus (c'est-à-dire une distribution normale centrée sur 0 avec un écart type égal au biais du modèle du paramètre considéré). La procédure de validation croisée a consisté à répéter l'analyse 999 fois avec un sous-échantillonnage aléatoire de 70 % des sous-bassins.

Il convient de noter que nous avons utilisé des valeurs moyennes interannuelles multi-décennales (calculées sur la période 1990-2010) des variables environnementales dérivées de modèles physiques dans les analyses présentées dans le manuscrit. Mécaniquement, les biais pris en compte dans l'analyse de sensibilité dérivée des moyennes annuelles sont plus élevés que les biais associés aux moyennes interannuelles multidécennales. Par conséquent, le degré de sensibilité de nos résultats aux incertitudes des modèles environnementaux est probablement surestimé comparé à la sensibilité réellement associée à l'usage des moyennes interannuelles.

Les résultats de l'analyse de sensibilité pour chacun des trois groupes biologiques sont présentés dans le tableau 2. Pour les trois groupes biologiques, la différence de R^2 entre chaque paire de variables directes et proxies (c'est-à-dire la température de l'eau par rapport à la température de l'air, les variables hydrauliques par rapport au débit, et les variables directes par rapport aux variables proxies combinées) est restée statistiquement significative lorsque les incertitudes des variables environnementales dérivées des modèles physiques ont été prises en compte, comme le montrent les intervalles de confiance qui ne se chevauchent pas. Pour quelques comparaisons (variables hydrauliques par rapport au débit pour les ensembles de données sur les macroinvertébrés et les données communes, variables directes par rapport aux variables proxies combinées pour l'ensemble de données sur les macroinvertébrés), les différences étaient statistiquement significatives, mais seulement de façon marginale (c'est-à-dire que les intervalles de confiance se chevauchaient). Dans tous les cas, les variables directes expliquent mieux les trois ordinations biologiques que les variables proxies.

TABLEAU 6

Analyse de sensibilité des résultats de la RDA aux incertitudes des variables environnementales dérivées des modèles physiques. L'importance des différentes variables environnementales dans l'explication de l'ordination biologique des ensembles de données sur les poissons, les macroinvertébrés et les espèces communes : R^2 explicatif et intervalles de confiance à 95 % associés a été dérivée de la RDA réalisée sur des tronçons contenus dans 70 % des sous-bassins. Les données environnementales ont été affectées par un échantillonnage aléatoire d'erreurs dans une distribution normale centrée sur zéro et avec les incertitudes des modèles environnementaux comme écart-type.

	Variables environnementales	Jeux de données biologiques		
		Poissons	Macroinvertébrés	Commun
Régime hydrologiques	Débits	28.4 [25.7 ; 31.1]	18.5 [16.5 ; 20.5]	27.1 [23.7 ; 30.5]
	Variables hydrauliques	35.7 [33.1 ; 38.3]	21.6 [19.7 ; 23.5]	33.2 [29.8 ; 36.6]
Régime thermique	Température de l'air	34.1 [30.8 ; 37.4]	20.0 [17.8 ; 22.2]	29.0 [25.3 ; 32.7]
	Température de l'eau	44.2 [41.0 ; 47.4]	29.5 [27.4 ; 31.6]	46.8 [43.2 ; 50.4]
Variables proxy	Température de l'air + débits	43.4 [40.4 ; 46.4]	22.3 [20.2 ; 24.4]	37.1 [33.0 ; 41.2]
Variables directes	Température de l'eau + variables hydrauliques	47.9 [45.3 ; 50.5]	36.5 [34.2 ; 38.8]	49.9 [46.4 ; 53.4]
Régime hydrologiques	Débits	53.8 [50.8 ; 56.8]	39.5 [37.2 ; 41.8]	57.6 [54.4 ; 60.8]

E. Discussion

E.1. Quel type de variables pour expliquer la structuration spatiale des communautés aquatiques ?

Les variables environnementales directes expliquent significativement mieux la structuration des communautés aquatiques à l'échelle du bassin versant de la Loire que les variables proxies, ce qui soutient la première hypothèse. Une différence significative dans la capacité à expliquer les patrons de structuration des poissons et des macroinvertébrés a été trouvée entre la température de l'air et celle de l'eau. En général, les études s'accordent sur la forte corrélation entre température de l'eau et celle de l'air, en particulier lorsque la résolution spatiale de l'étude est large (Caissie, 2006). Comme les valeurs de température de l'eau sont rarement disponibles à large échelle, la température de l'air est souvent prise comme proxies dans les études écologiques (*e.g.* Buisson et al. (2008a) ; Maire et al. (2017)). Cependant, la température de l'air ne capture pas toute la variabi-

lité de la température de l'eau (Caissie, 2006). Les gammes de valeurs prises entre les deux descripteurs à l'échelle du réseau T-NET (voir tableau 4) appuient ce constat, avec par exemple la température moyenne de l'air qui varie de 6,1 à 12,5 °C et la température moyenne de l'eau qui varie de 5,4 à 16,2 °C. Les différences entre les valeurs des quantiles sont encore plus marquées. Par conséquent, la température de l'air ne se substitue pas bien à la température de l'eau dans certaines conditions, par exemple à proximité d'apports de nappes souterraines (Caissie, 2006 ; Beaufort *et al.*, 2016).

Pour les trois NMDS biologiques, la relation entre l'ordination et le débit est plus faible qu'avec les variables hydrauliques correspondante. Cela supporte l'idée que la résolution à laquelle les métriques de débit ont été calculées n'est pas la plus appropriée pour retranscrire correctement la relation organisme-débit, contrairement aux variables hydrauliques qui décrivent mieux l'habitat hydraulique. En effet, les variables hydrauliques offrent des dimensions supplémentaires par rapport aux débits car elles reflètent mieux les variations locales de l'hydromorphologie qui influencent l'habitat (Statzner & Higler, 1986 ; Morel *et al.*, 2020).

Finalement, lorsque les variables directes sont considérées ensemble, elles expliquent encore mieux la structuration des communautés de poissons et de macroinvertébrés que l'ensemble des variables proxies. De plus, ces différences sont encore plus marquées pour le jeu de données communs. Ces deux résultats renforcent l'hypothèse que la combinaison de plusieurs compartiments biologiques associés à l'utilisation de variables environnementales directes est plus appropriée pour établir avec précision la relation organisme-environnement dans les milieux lotiques. Néanmoins, l'utilisation de variables proxies en l'absence de variables directes restent intéressante, en particulier à large échelle où les différences trouvées peuvent être faibles, comme dans les résultats présentés ici. En particulier, les variables directes sont susceptibles d'être plus performantes à une étendue spatiale plus fine et peuvent être plus intéressantes dans le cadre de gestion environnementale comme leviers d'action directs (Irving *et al.*, 2022). Ces différences sont encore plus réduites lorsqu'il s'agit du pouvoir prédictif des deux variables. Par conséquent, le choix d'utiliser des variables directes ou proxies varie en fonction de l'objectif qui vise soit à expliquer et comprendre les déterminants environnementaux majeurs de la structuration spatiale des communautés aquatiques soit à projeter les patrons spatiaux sous différents scénarios (*e.g.* changements climatiques).

L'organisation des communautés aquatiques étudiées est essentiellement expliquée par le même ensemble de variables environnementales. Les déterminants environnementaux les plus influents sont la température de l'eau (notamment la moyenne et les quantiles 90%), les variables hydrauliques (*i.e.* la profondeur en eau, la largeur en eau et la vitesse de courant) et la pente associées à la distance à l'estuaire. Ces résultats sont cohérents avec la littérature (Matthews, 1998 ; Irving *et al.*, 2022). Premièrement, la température de l'eau, particulièrement les fortes températures, limite la distribution spatiale des espèces aquatiques en fonction de leur tolérance thermique et de leurs exigences en oxygène (Caissie, 2006 ; Souchon & Tissot, 2012). Cependant, bien que cela soit fréquemment reconnu dans la littérature, peu d'études ont montré le rôle primordial de la température de l'eau dans la structuration des communautés aquatiques par rapport à la température de l'air. Deuxièmement, les conditions hydrauliques représentent un filtre environnemental important qui contraint les espèces en fonction de leur préférence d'habitat hydraulique, qui joue aussi sur les autres caractéristiques de l'habitat (*e.g.* granulométrie du substrat, type de végétation) (Lamouroux *et al.*, 2004).

E.2. Limites à l'explication de la structuration spatiale des communautés

Les variables environnementales utilisées ici expliquent seulement partiellement la structuration des communautés de poissons et de macroinvertébrés. L'unité géographique des analyses est le tronçon T-NET (de longueur moyenne 1,7 km ; N = 52 278) et il est supposé qu'une telle précision, densité et couverture à l'échelle d'un réseau hydrographique large (longueur cumulée de cours d'eau = 88 058 km) le rendent très approprié pour étudier la structuration des communautés aquatiques ligériennes (Pletterbauer *et al.*, 2016). Le choix pour une étendue spatiale large permet de capturer les réponses aux gradients environnementaux majeurs qui découlent du climat, de la géologie ou de la topographie (Infante *et al.*, 2009). Cependant, certains processus qui se produisent à d'autres échelles spatiales peuvent être brouillés à une telle étendue spatiale (Wiens, 2002). Par exemple, une large étendue spatiale peut inclure des unités biogéographiques distinctes, qui sont connues pour être la résultante d'autres facteurs comme les événements historiques (*e.g.* périodes de glaciations majeures) ou les phénomènes de dispersion des espèces (Dias *et al.*, 2014 ; Oberdorff *et al.*, 2011).

Tandis que l'échelle du bassin versant est reconnue comme appropriée pour étudier la structuration des communautés de poisson et l'application de mesures de gestion et de conservation, l'étendue appropriée pour les macroinvertébrés fait encore débat (Dovciak & Perry, 2002). Par exemple, Hawkins et Norris (2000) et Johnson *et al.* (2004) montrent que la structuration des communautés de macroinvertébrés est mieux expliquée par des facteurs d'habitats locaux (*e.g.* substrat, végétation) que régionaux (*e.g.* occupation du sol dans le bassin versant amont, positionnement sur le réseau hydrographique) tandis que Townsend *et al.* (2003) ont trouvé le contraire. Plus généralement, il semblerait que la structuration des communautés aquatiques soit régie par l'interaction de différentes échelles spatiales, où les processus à large échelle sont influencés par les processus plus locaux (Poff, 1997 ; Johnson *et al.*, 2004 ; Lamouroux *et al.*, 2004 ; Villeneuve *et al.*, 2018). Par ailleurs, tous les R^2 entre l'ordination macroinvertébrés et les variables environnementales sont légèrement moins élevés que ceux trouvés pour les poissons, indiquant que la plus forte variabilité des communautés de macroinvertébrés est un peu moins bien expliquée par l'ensemble des variables environnementales utilisées ici. Cela suggère que les macroinvertébrés ne sont pas structurés aussi fortement que les poissons par le gradient longitudinal à cette échelle spatiale. Associés à la NMDS, ces résultats suggèrent l'influence de processus opérant à deux échelles spatiales : le bassin versant, principalement gouverné par les régimes thermiques et hydrologiques, et le tronçon où les variables hydrauliques montrent d'importantes variations.

La contribution des variables environnementales à expliquer la structuration des communautés peut aussi être limitée par le fait qu'elles proviennent de modèles. L'analyse de sensibilité montrent que les résultats sont robustes à une certaine incertitude des modèles à base physique. Néanmoins, les modèles EROS et T-NET, d'où proviennent la température de l'eau et le débit, n'intègrent pas les altérations d'origine anthropique, notamment les barrières physiques (*e.g.* barrages) qui sont connues pour influencer les régimes thermiques et hydrologiques des rivières (Seyedhashemi *et al.*, 2022).

E.3. Concordance entre les compartiments poissons et macroinvertébrés

L'analyse procustéenne indique de très faibles concordances entre chaque paire d'ordinations biologiques (poisson-macroinvertébrés : $m^2 = 0,99$; poisson-commun : $m^2 = 0,99$;

macroinvertébrés-commun : $m^2 = 0,99$). La concordance avec l'ordination environnementale est plus importante pour l'ordination sur les données communes ($m^2 = 0,61$) que pour les ordinations poissons ($m^2 = 0,97$) ou macroinvertébrés ($m^2 = 0,95$). Les communautés de poissons et de macroinvertébrés n'ont donc pas l'air de partager des patrons spatiaux similaires à l'échelle du bassin versant de la Loire. Bien que les communautés soient façonnées par les mêmes déterminants environnementaux, leur réponse diffère face au gradient environnemental. Ce constat renforce l'hypothèse que les communautés de macroinvertébrés et de poissons réagissent au gradient environnemental à différentes échelles spatiales.

La faible concordance entre les deux compartiments biologiques renforce la pertinence d'étudier les poissons et les macroinvertébrés ensemble pour identifier les patrons de distribution généraux plutôt que de n'en considérer qu'un. La plupart des études qui utilisent une approche similaire (*i.e.* ordination multivariée et évaluation de la concordance) ont aussi trouvé une faible concordance entre les compartiments biologiques à une échelle similaire au bassin versant de la Loire (*e.g.* poisson et macroinvertébré dans Kilgour & Barton (1999); plusieurs compartiments dans la revue de Heino (2010); diatomées, macroinvertébrés, macrophytes et poissons dans Johnson & Hering (2010)). D'autres études ont évalué le lien entre la concordance de plusieurs compartiments et l'échelle spatiale. Dolph *et al.* (2011) et Paavola *et al.* (2006) ont trouvé une augmentation de la concordance à large échelle (*i.e.* étendue similaire au bassin de la Loire) tandis que Bae *et al.* (2014) montre l'inverse. Les deux premières montrent que les communautés de poisson et de macroinvertébrés ont des patrons concordants à large échelle mais elles présentent des différences notables avec l'analyse menée ici. L'étude de Paavola 2006 concerne des communautés aquatiques boréales, habitant dans des conditions bien plus froides et avec une richesse spécifique et des densités moins élevées. Dolph 2011 ont basé leur résultat sur des communautés nord-américaines à faible diversité spécifique avec un total de 5 espèces de poissons et 15 taxons de macroinvertébré. De plus, l'étendue spatiale est composée de plusieurs petites aires géographiques de sous-bassins distincts, et donc ne considèrent pas l'ensemble des conditions environnementales rencontrées à l'échelle d'un même grand bassin versant.

Par ailleurs, la faible concordance trouvée entre l'ordination basée sur les données communes et les autres ordinations indique que ni les poissons ni les macroinvertébrés ne sont plus représentatifs de la structuration spatiale des communautés combinées. L'in-

fluence des variables environnementales diffère : les macroinvertébrés montrent une plus faible corrélation avec l'ensemble des variables environnementales que les poissons, malgré la seconde hypothèse. La concordance entre les ordinations environnementales et celle sur les données biologiques communes est plus élevée que qu'avec celle des poissons ou des macroinvertébrés, ce qui est probablement imputable à une meilleure intégration du gradient environnemental dans l'ordination combinant les données poissons et macroinvertébrés. Ceci est corroboré par une légère augmentation de R^2 entre la plupart des variables environnementales et l'ordination commune. Ceci supporte la troisième hypothèse qui prédit un meilleur ajustement de la combinaison des compartiments biologiques avec les variables environnementales, bien que la variabilité biologique soit accrue.

Certaines hypothèses peuvent expliquer les différences trouvées entre les compartiments biologiques. Premièrement, les jeux de données macroinvertébrés et poissons comportent 186 genres et 43 espèces, respectivement. Les communautés macroinvertébrés sont donc plus diverses que les poissons avec une moyenne de 43 genres par tronçon contre 10 espèces de poissons. Les CPUE observées sont aussi plus variables (CPUE moyenne = $3\ 851 \pm 3\ 121$ individus pour $100\ m^2$ par tronçon T-NET) que pour les poissons (CPUE moyenne = 97 ± 140 individus pour $100\ m^2$ par tronçon T-NET). Ce qui peut conduire à une plus grande variabilité à expliquer pour les macroinvertébrés que pour les poissons. De plus, l'occupation spatiale de chacun des genres de macroinvertébrés est au global plus variable que pour les poissons, avec des taxons ubiquistes répartis tout le long du gradient amont-aval (*e.g.* le genre *Baetis* est présent sur plus de 95% des tronçons ; les occurrences des taxons peuvent être consultés à l'annexe S2 de l'article complet à l'Annexe A de ce manuscrit) et d'autres montrant des préférences écologiques plutôt restreintes et donc une distribution très localisée (*e.g.* le genre *Leptophlebia* est présent sur moins de 2% des tronçons, majoritairement en tête de bassin ; les occurrences des taxons peuvent être consultés à l'annexe S2 de l'article complet à l'Annexe A de ce manuscrit). Au contraire, les ordinations poissons mettent en avant des espèces indicatrices typiques comme *Salmo trutta* ou *Thymallus thymallus* dont les préférences écologiques sont bien connues et sont trouvées en des endroits spécifiques du bassin de la Loire. Considérant le grand nombre de genres macroinvertébrés avec une large étendue spatiale, il est très probable que différentes unités biogéographiques soient incluses. Ainsi, l'organisation de ces communautés macroinvertébrés est moins influencée par le gradient amont-aval.

Deuxièmement, l'identification des macroinvertébrés est limitée au genre, qui peuvent regrouper différentes préférences écologiques (Hawkins & Norris, 2000 ; Tachet *et al.*, 2010). De précédentes études ont montré que l'identification à la famille pouvait suffire pour évaluer les patrons spatiaux généraux, bien que des patrons plus précis peuvent être décrits avec des niveaux d'identification plus fins (Heino & Soininen, 2007). De plus, la sensibilité des méthodes d'ordinations multivariées au niveau d'identification semble réduite au-delà de la famille (Kilgour & Barton, 1999 ; Grenouillet *et al.*, 2008). D'après ces études, le niveau du genre est un bon compromis entre un niveau approprié pour les analyses écologiques et des considérations pratiques de temps d'échantillonnage et de temps alloué à l'identification.

Enfin, les poissons occupent des aires spatiales plus larges pour accomplir leur cycle de vie que les macroinvertébrés : de l'ordre du kilomètre pour la plupart des espèces contre quelques mètres pour les genres macroinvertébrés (Kilgour & Barton, 1999 ; Backus-Freer & Pyron, 2015). Les deux compartiments sont sensibles aux mêmes variables environnementales mais à deux grains spatiaux différents, les macroinvertébrés étant plus sensibles à des variables d'habitat. Ils montrent aussi une réponse plus marquée aux perturbations locales que les poissons du fait de leur capacité de dispersion limitée et de protection limitée lors du stade larvaire (Lamouroux *et al.*, 2004). Cela implique que l'identification des patrons de structuration à large échelle peut être brouillée par les processus locaux. Les poissons sont probablement moins affectés par des perturbations locales car ils ont une plus grande capacité de déplacement et d'évitement lorsque l'altération de l'habitat est temporaire. Les processus stochastiques de colonisations comme la dérive ou les déplacements lors des phases adultes aériennes peuvent également jouer un rôle plus important pour les macroinvertébrés que pour les poissons.

En résumé ...

- ★ Les ordinations par NMDS (*Non Metric Multidimensional Scaling*) des communautés de poissons et de macroinvertébrés ont montré que leur structuration spatiale s'effectue principalement le long du gradient amont-aval formé par la température de l'eau, l'hydrologie et la topographie ;
- ★ L'influence du gradient amont-aval sur les macroinvertébrés semblent moins élevée que sur les poissons, probablement car l'influence d'autres facteurs plus locaux (*e.g.* ; composition et qualité du substrat) est plus importante pour les macroinvertébrés ;
- ★ La concordance entre les patrons de structuration des communautés de poissons et de macroinvertébrés mise en avant par NMDS et mesurée par une analyse procustéenne est faible. Il est donc préférable d'éviter l'utilisation d'un compartiment biologique comme indicateur d'un autre. Une approche multi-compartiments semble ainsi plus pertinente pour étudier les processus de structuration spatiale des communautés aquatiques ;
- ★ Les variables environnementales directes (*i.e.* température de l'eau et variables hydrauliques) apportent un niveau d'explication supplémentaire par rapport aux variables proxies (*i.e.* température de l'air et débit), mais qui n'est pas aussi élevé qu'attendu. En l'absence de variables directes l'utilisation de variables proxies peut suffire en première intention d'analyse de la structuration des communautés aquatiques ;
- ★ La part d'explication de la structuration spatiale des communautés qui n'est pas expliquée par les variables environnementales testées ici laisse supposer que d'autres facteurs non pris en compte interviennent certainement, comme les processus biogéographiques historiques ou les dynamiques de dispersion des organismes, ainsi que des phénomènes plus locaux comme les interactions biologiques.

Chapitre IV

Modélisation de la distribution des abondances de multiples taxons

A. Introduction

La modélisation de la distribution des espèces (*i.e.* *Species Distribution Models* (SDM)) permet d'apporter des éléments de compréhension des processus écologiques structurants et de répondre à différents objectifs : (i) identifier les déterminants environnementaux de la distribution spatiale des espèces et caractériser la relation organisme-environnement (Peterson *et al.*, 2011 ; Guisan *et al.*, 2017), (ii) en vue de l'élaboration d'un plan de prospection pour une espèce cible, déterminer des zones potentiellement favorables à échantillonner (Prié *et al.*, 2014), (iii) déterminer les potentiels effets d'une modification de l'environnement sur la distribution, comme l'application d'une mesure de conservation (Floury *et al.*, 2021) ou les effets du changement climatique (Elith *et al.*, 2010 ; Comte & Grenouillet, 2015). Les méthodes actuelles permettent d'estimer les aires de répartition des espèces, mais le besoin de considérer les niveaux d'abondance a mené à différentes tentatives de modélisation des abondances (Potts & Elith, 2006 ; Waldock *et al.*, 2021). Contrairement à la modélisation de la distribution des présence-absence, la recherche des modèles avec les meilleures performances prédictives et généralisables à un grand nombre d'espèce n'en est qu'à ses débuts (Potts & Elith, 2006 ; Fernandes *et al.*, 2020).

La modélisation des abondances présente plusieurs verrous méthodologiques. Premièrement, le grain spatial des données biologiques et environnementales doit être suffisamment fin pour capter les relations abondance-environnement et notamment leur variabilité spatiale (Johnson *et al.*, 2004 ; Filz *et al.*, 2013 ; Haesen *et al.*, 2023). Par ailleurs, du fait de l'indisponibilité de données sur les variables jugées structurantes pour les abondances, les variables environnementales mobilisées en tant que prédicteurs ne présentent pas toujours une relation causale (Guisan *et al.*, 2017). Par exemple, l'utilisation de variables proxies pour expliquer la structuration spatiale des communautés aquatiques (*e.g.* température de l'air et débits) ne sont pas représentatives des conditions de vie réellement endurées par les organismes (chapitre III). Ensuite, la modélisation de données d'abondance présente des difficultés techniques importantes : (i) sur-inflation des données d'absence, (ii) sur-dispersion, (iii) distribution qui ne correspond pas à une distribution théorique connue et (iv) gamme d'abondance très variée en fonction des taxons. Enfin, dans la majorité des cas, seuls les facteurs abiotiques sont considérés pour estimer les abondances, alors que les composantes biotiques et liées à la capacité de dispersion des taxons jouent un rôle certainement non négligeable (Soberon & Nakamura, 2009), même à large échelle (Bou-

langeat *et al.*, 2012 ; Lucas *et al.*, 2023). La non prise en compte des autres composantes pourrait expliquer les faibles performances des modèles d'abondance existants (Potts & Elith, 2006 ; Waldock *et al.*, 2021).

L'objectif de ce chapitre est de trouver une méthodologie de modélisation de la distribution spatiale des abondances qui soit performante et applicable au plus grand nombre de taxons, issues de différents compartiments biologiques (*i.e.* poissons et macroinvertébrés). L'objectif secondaire est de tester l'apport de l'intégration des interactions biotiques dans les modèles. Le but final de cette modélisation est de produire des scénarios potentiels d'évolution des communautés aquatiques à l'échelle du bassin de la Loire sous changement climatique (chapitre V). Cela nécessite de modéliser le plus grand nombre de taxons, afin de se rapprocher le plus possible de la communauté réelle. Les données mobilisées concernent plus de 200 taxons et il apparaît difficile de calibrer chaque modèle individuellement, d'où l'objectif de trouver une méthode applicable au plus grand nombre de taxons. De plus, une méthode généralisable aura plus de chances d'être efficace sur d'autres compartiments biologiques (*e.g.* macrophytes, mammifères). Une méthode innovante de modélisation des abondances des espèces aquatiques a été développée et est proposée dans ce chapitre. Elle consiste en la transformation des CPUE en classes d'abondance puis modélisation par une méthode de *Machine Learning* (ML), l'*Ordinal Forest* (OF). Différents choix et stratégies de modélisation ont été explorés et comparés au cours de ce développement (section B), comme le choix du nombre de classe ou des méthodes de transformation des données pour limiter l'influence de la sur-inflation sur les performances prédictives. L'analyse des performances des modèles OF ainsi que leur interprétation écologique sont ensuite présentées (section C). Enfin, un indicateur de ressource alimentaire basé sur les macroinvertébrés a été intégré dans les modèles poissons et son apport pour la modélisation des classes d'abondance a été évalué (section D).

B. Données et méthodologie de modélisation des abondances des espèces aquatiques

B.1. Sélection des données biologiques

Les modèles de distribution des abondances ont été construits à partir du jeu de données du bassin versant de la Loire (chapitre II), car nous disposons à cette échelle des données à haute résolution spatiale pour les principaux déterminants environnementaux de la structuration des abondances des espèces aquatiques (*i.e.* température de l'eau et variables hydrauliques ; voir chapitre III). Nous disposons également des données d'abondance pour 48 espèces de poissons et 214 genres de macroinvertébrés, répartis respectivement sur 484 et 1 077 tronçons du réseau H, dont 408 tronçons portant simultanément des données poissons et macroinvertébrés (voir chapitre II pour la constitution du jeu de données). Les 262 taxons du jeu de données ne sont pas tous de bons candidats à la modélisation pour plusieurs raisons. Certains taxons ne respectent pas l'hypothèse d'équilibre de la relation organisme-environnement nécessaire à la construction des SDM. Les espèces de poissons suivantes ont été exclues de la modélisation car leur distribution est en forte dynamique (*i.e.* espèce exogène supposée en dynamique de colonisation sur le bassin, espèce amphihaline, migratrice ou non, non résidente des milieux d'eau douce) sur la zone d'étude : blageon, pseudorasbora, saumon atlantique, flet, mulot porc.

Dans le jeu de données, le niveau d'identification des macroinvertébrés s'arrête au genre, voire à la famille pour certains taxons. Seuls les taxons identifiés au niveau du genre et appartenant aux ordre Éphéméroptères, Plécoptères, Trichoptères et Odonates (EPTO) ont été conservés pour la modélisation. Conformément aux explications données le chapitre II, les EPTO sont les ordres pour lequel les données du plus grand nombre de genres ont été validés par expertise, et comportant les genres dont les espèces sont proches écologiquement. De plus, les EPTO sont souvent considérés pour leur bonne représentation du compartiment macroinvertébrés, du fait de leur grande diversité écologique et de leur présence tout le long du gradient amont-aval des cours d'eau. Les opérations d'échantillonnage pour lesquelles la proportion de taxons identifiés au niveau de la famille représente plus de 60 % des taxons échantillonnés ont été supprimées (N = 149 opérations). Les autres opérations ont été conservées puis les taxons identifiés à la famille ont été supprimés du jeu de données (N = 10 taxons). Afin de respecter le postulat des SDM qui nécessite de

construire un modèle sur une entité qui utilise les mêmes conditions d'habitats (réponse à l'environnement homogène au sein du genre ; voir hypothèses des SDM en chapitre I), les genres contenant des espèces aux préférences écologiques pouvant être trop disparates ont fait l'objet d'un tri. À cette fin, les informations utilisées pour la sélection des genres de macroinvertébrés candidats à la modélisation proviennent des bases de données de l'OPIE (www.opie-benthos.fr) et du site [freshwaterecology](http://freshwaterecology.info) (www.freshwaterecology.info). Pour chaque genre, l'identification des espèces présentes sur le bassin de la Loire a permis d'effectuer une première sélection des genres ne comportant qu'une seule espèce (*i.e.* genres monospécifiques ; N = 38 ; Tableau 8). Pour chaque genre multispécifique (N = 62), la préférence de mésohabitats a été considérée comme proxy de la similarité écologique et a permis de retenir pour la modélisation 18 genres parmi les 62 (Tableau 8).

Les modèles ont besoin d'un minimum de données d'abondance non nulles. Les taxons pour lesquels un nombre insuffisant d'occurrence (< 5 %) dans au moins l'une des classes ne permettaient pas la modélisation ont été retirés. La plupart des taxons concernés ont une fréquence d'occurrence supérieure à 5%, hormis *Cloeon* et *Ephemerella* qui ont pu être conservés grâce à une répartition suffisante d'occurrence dans chacune des classes d'abondance. Au total, 26 espèces de poissons et 31 genres de macroinvertébrés ont été considérés pour la modélisation de la distribution spatiale de leur classe d'abondance (voir Tableaux 7 et 8 pour les détails).

TABLEAU 7

Liste des espèces de poissons modélisées ou non (N = 42 espèces candidates). Les espèces suivantes sont exclues de la modélisation : celles dont (i) l'occurrence est trop faible (< 5 % ; « occurrence »), (ii) la gamme d'abondance ne permet pas la création de quatre classes d'abondance homogènes (« gamme abondance »), et (iii) les espèces dont la distribution n'est pas à l'équilibre dans le milieu car non résident (« amphihaline non résident ») ou en dynamique de colonisation (« exogène »). Les espèces codées « SDM » ont été conservées pour la modélisation. L'occurrence renseignée est en % de tronçons occupés (sur N = 484 tronçons).

Code	Occurrence (%)	Sélection
ANG	48,97	SDM
BBG	5,99	gamme abondance
CDR	1,86	occurrence
PES	58,26	SDM
LOR	4,75	occurrence
CHA	57,23	SDM
ABH	11,57	gamme abondance
ABL	52,48	SDM
ASP	3,51	occurrence
BAF	47,93	SDM
BLN	0,41	exogène
BOU	45,45	SDM
BRE	39,46	SDM
CAA	30,37	SDM
CCO	35,74	SDM
CHE	79,75	SDM
GAR	76,24	SDM
GOU	82,23	SDM
HOT	27,07	SDM
IDE	1,45	occurrence
PSR	35,54	exogène
ROT	47,93	SDM
SPI	48,35	SDM
TAN	39,46	SDM
TOX	0,83	exogène
VAI	75	SDM
VAN	40,7	SDM
BRO	45,25	SDM
LOT	6,2	gamme abondance
EPI	6,82	gamme abondance
EPT	20,45	gamme abondance
PCH	31,82	SDM
MUP	2,89	amphihaline non résident
LOF	82,64	SDM
GRE	30,58	SDM
PER	71,28	SDM
SAN	23,76	gamme abondance
LPP	41,94	SDM
FLE	0,21	amphihaline non résident
OBR	4,96	occurrence
TRF	50	SDM
SIL	23,35	SDM

TABLEAU 8

Liste des genres de macroinvertébrés modélisés ou non (N = 99 genres d'EPTO candidats à la modélisation). Les genres suivants n'ont pas été modélisés : ceux dont l'occurrence était trop faible (< 5% ; « occurrence ») et ceux contenant des espèces à l'écologie trop différentes (« écologie »). Les taxons codés « SDM » ont été conservés pour la modélisation. L'occurrence renseignée est en % de tronçons occupés (sur N = 1 077 tronçons).

Ordre	Famille	Genre	Nombre espèces France	Nombre espèces Loire	Occurrence (%)	Sélection
Ephemeroptera	Baetidae	<i>Baetis</i>	>20	>20	0,56	écologie
Ephemeroptera	Baetidae	<i>Centroptilum</i>	1	1	0,74	occurrence
Ephemeroptera	Baetidae	<i>Cloeon</i>	2	2	3,99	SDM
Ephemeroptera	Baetidae	<i>Procloeon</i>	4	4	20,15	SDM
Ephemeroptera	Caenidae	<i>Caenis</i>	10	9	0,09	occurrence
Ephemeroptera	Ephemerellidae	<i>Ephemerella</i>	2	1	4,18	SDM
Ephemeroptera	Ephemerellidae	<i>Torleya</i>	1	1	25,07	SDM
Ephemeroptera	Ephemeridae	<i>Ephemerella</i>	4	4	36,12	écologie
Ephemeroptera	Heptageniidae	<i>Electrogena</i>	5	3	0,84	écologie
Ephemeroptera	Heptageniidae	<i>Epeorus</i>	4	3	25,26	écologie
Ephemeroptera	Heptageniidae	<i>Heptagenia</i>	3	3	4,74	écologie
Ephemeroptera	Heptageniidae	<i>Rhithrogena</i>	26	NR	29,53	écologie
Ephemeroptera	Isonychiidae	<i>Isonychia</i>	1	1	24,88	SDM
Ephemeroptera	Leptophlebiidae	<i>Choroterpes</i>	1	1	27,39	SDM
Ephemeroptera	Leptophlebiidae	<i>Habrophlebia</i>	3	2	0,19	écologie
Ephemeroptera	Leptophlebiidae	<i>Leptophlebia</i>	2	2	23,31	SDM
Ephemeroptera	Leptophlebiidae	<i>Paraleptophlebia</i>	3	2	24,41	écologie
Ephemeroptera	Leptophlebiidae	<i>Thraululus</i>	1	1	50,79	SDM
Ephemeroptera	Oligoneuriidae	<i>Oligoneuriella</i>	1	1	0,84	occurrence
Ephemeroptera	Polymitarcyidae	<i>Ephoron</i>	1	1	40,85	SDM
Ephemeroptera	Potamanthidae	<i>Potamanthus</i>	1	1	10,49	SDM
Ephemeroptera	Siphonuridae	<i>Siphonurus</i>	2	2	13,18	SDM
Odonata	Aeshnidae	<i>Aeshna</i>	>10	NR	0,84	écologie
Odonata	Aeshnidae	<i>Anax</i>	2	NR	3,62	occurrence
Odonata	Aeshnidae	<i>Boyeria</i>	1	1	18,2	SDM
Odonata	Aeshnidae	<i>Brachytron</i>	1	1	33,43	SDM
Odonata	Aeshnidae	<i>Hemianax</i>	1	1	22,75	SDM
Odonata	Corduliidae	<i>Cordulia</i>	1	1	36,86	SDM
Odonata	Corduliidae	<i>Epithea</i>	1	1	43,92	SDM
Odonata	Gomphidae	<i>Ophiogomphus</i>	1	1	9,84	SDM
Odonata	Gomphidae	<i>Paragomphus</i>	NR	NR	0,28	occurrence
Odonata	Lestidae	<i>Chalcolestes</i>	1	NR	0,28	occurrence
Odonata	Lestidae	<i>Lestes</i>	5	NR	32,21	écologie
Odonata	Lestidae	<i>Sympetma</i>	2	NR	1,21	occurrence
Odonata	Libellulidae	<i>Brachythemis</i>	1	1	24,42	SDM
Odonata	Libellulidae	<i>Crocothemis</i>	1	1	18,66	SDM
Odonata	Libellulidae	<i>Diplacodes</i>	NR	NR	0,19	occurrence
Odonata	Libellulidae	<i>Libellula</i>	3	NR	63,97	écologie
Odonata	Libellulidae	<i>Orthetrum</i>	4	NR	0,56	écologie
Odonata	Libellulidae	<i>Sympetrum</i>	9	NR	10,58	écologie
Plecoptera	Capniidae	<i>Capnioneura</i>	6	2	70,19	écologie
Plecoptera	Chloroperlidae	<i>Xanthoperla</i>	1	1	40,3	SDM
Plecoptera	Leuctridae	<i>Euleuctra</i>	1	1	87	SDM
Plecoptera	Leuctridae	<i>Leuctra</i>	63	NR	88,39	écologie
Plecoptera	Nemouridae	<i>Amphinemura</i>	4	3	4,27	écologie
Plecoptera	Nemouridae	<i>Nemoura</i>	21	NR	0,37	écologie
Plecoptera	Nemouridae	<i>Protonemura</i>	26	NR	0,09	occurrence
Plecoptera	Perlidae	<i>Dinocras</i>	3	1	1,39	occurrence
Plecoptera	Perlidae	<i>Perla</i>	4	2	10,49	écologie

Ordre	Famille	Genre	Nombre espèces France	Nombre espèces Loire	Occurrence (%)	Sélection
Plecoptera	Perlodidae	<i>Isogenus</i>	1	1	21,08	SDM
Plecoptera	Perlodidae	<i>Isoperla</i>	14	NR	0,19	occurrence
Plecoptera	Perlodidae	<i>Perlodes</i>	4	3	45,31	écologie
Plecoptera	Taeniopterygidae	<i>Brachyptera</i>	6	6	39,46	écologie
Plecoptera	Taeniopterygidae	<i>Rhabdiopteryx</i>	4	1	36,12	SDM
Plecoptera	Taeniopterygidae	<i>Taeniopteryx</i>	4	3	36,95	écologie
Trichoptera	Beraeidae	<i>Beraea</i>	5	3	96,29	écologie
Trichoptera	Beraeidae	<i>Beraeodes</i>	1	1	2,51	occurrence
Trichoptera	Brachycentridae	<i>Micrasema</i>	12	NR	58,22	écologie
Trichoptera	Brachycentridae	<i>Oligoplectrum</i>	1	1	21,17	SDM
Trichoptera	Calamoceratidae	<i>Calamoceras</i>	1	1	2,88	occurrence
Trichoptera	Dipseudopsidae	<i>Pseudoneureclipsis</i>	1	1	1,39	occurrence
Trichoptera	Glossosomatidae	<i>Agapetus</i>	8	4	20,06	écologie
Trichoptera	Glossosomatidae	<i>Glossosoma</i>	4	4	80,69	écologie
Trichoptera	Glossosomatidae	<i>Synagapetus</i>	8	6	0,19	occurrence
Trichoptera	Goeridae	<i>Goera</i>	1	1	9,37	SDM
Trichoptera	Goeridae	<i>Lithax</i>	2	1	0,28	occurrence
Trichoptera	Goeridae	<i>Silo</i>	5	NR	15,32	écologie
Trichoptera	Hydropsychidae	<i>Cheumatopsyche</i>	1	1	0,09	occurrence
Trichoptera	Hydropsychidae	<i>Diplectrona</i>	4	1	1,11	occurrence
Trichoptera	Hydropsychidae	<i>Hydropsyche</i>	26	NR	32,5	écologie
Trichoptera	Hydroptilidae	<i>Agraylea</i>	2	2	4,18	occurrence
Trichoptera	Hydroptilidae	<i>Hydroptila</i>	26	NR	29,16	écologie
Trichoptera	Hydroptilidae	<i>Ithytrichia</i>	2	2	13,18	écologie
Trichoptera	Hydroptilidae	<i>Orthotrichia</i>	3	3	30,73	écologie
Trichoptera	Hydroptilidae	<i>Ptilocolepus</i>	1	1	14,11	SDM
Trichoptera	Lepidostomatidae	<i>Crunoecia</i>	1	1	0,29	occurrence
Trichoptera	Leptoceridae	<i>Adicella</i>	3	2	24,61	écologie
Trichoptera	Leptoceridae	<i>Ceraclea</i>	10	7	87,37	écologie
Trichoptera	Leptoceridae	<i>Mystacides</i>	3	3	78,46	SDM
Trichoptera	Leptoceridae	<i>Setodes</i>	4	3	69,45	écologie
Trichoptera	Leptoceridae	<i>Triaenodes</i>	5	NR	65,55	écologie
Trichoptera	Leptoceridae	<i>Ylodes</i>	NR	NR	0,84	écologie
Trichoptera	Limnephilidae	<i>Apataniinae</i>	8	NR	0,28	occurrence
Trichoptera	Molannidae	<i>Molanna</i>	2	1	6,96	SDM
Trichoptera	Philopotamidae	<i>Philopotamus</i>	8	NR	0,56	écologie
Trichoptera	Philopotamidae	<i>Wormaldia</i>	20	NR	8,82	écologie
Trichoptera	Phryganeidae	<i>Agrypnia</i>	3	3	0,09	occurrence
Trichoptera	Phryganeidae	<i>Trichostegia</i>	1	1	23,49	SDM
Trichoptera	Polycentropodidae	<i>Cyrnus</i>	5	5	9,38	écologie
Trichoptera	Polycentropodidae	<i>Neureclipsis</i>	1	1	50,32	SDM
Trichoptera	Polycentropodidae	<i>Plectrocnemia</i>	9	NR	26,28	écologie
Trichoptera	Polycentropodidae	<i>Polycentropus</i>	11	NR	44,57	écologie
Trichoptera	Psychomyiidae	<i>Lype</i>	3	3	40,01	écologie
Trichoptera	Psychomyiidae	<i>Psychomyia</i>	1	1	14,21	SDM
Trichoptera	Psychomyiidae	<i>Tinodes</i>	17	NR	0,37	écologie
Trichoptera	Rhyacophilidae	<i>Rhyacophila</i>	53	53	91,55	écologie
Trichoptera	Sericostomatidae	<i>Notidobia</i>	1	1	0,37	occurrence
Trichoptera	Sericostomatidae	<i>Sericostoma</i>	4	2	6,96	écologie
Trichoptera	Uenoidae	<i>Thremma</i>	3	1	3,34	occurrence

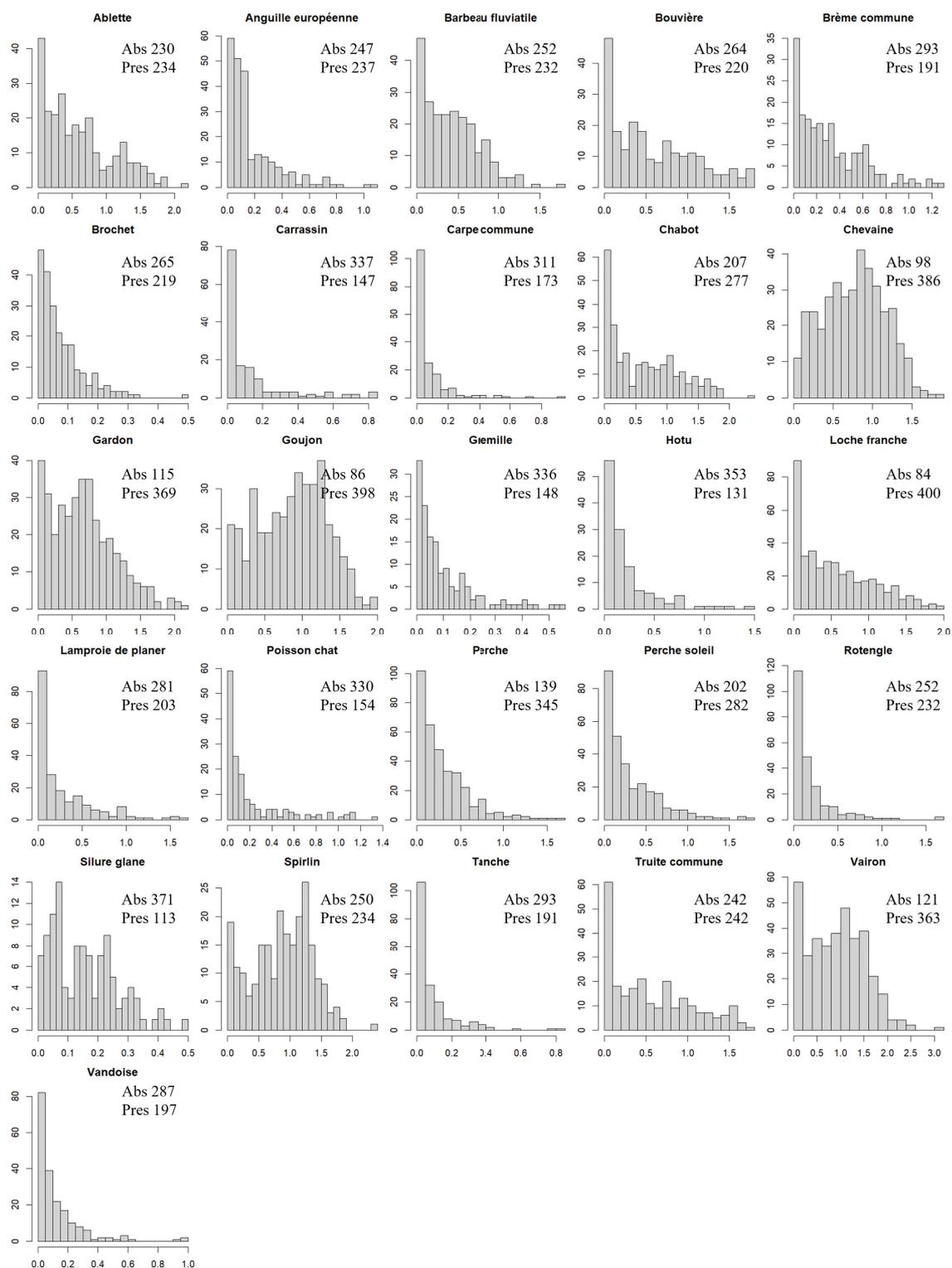


FIGURE 35

Histogrammes des CPUE > 0 observées sur les tronçons échantillonnés (N = 484) pour les **espèces de poissons** modélisées (N = 26). Les CPUE ont été log-transformées (\log_{10}) et les absences ont été retirées à des fins de lisibilité. Les nombres d'absence et de présence sont précisés pour chaque espèce.

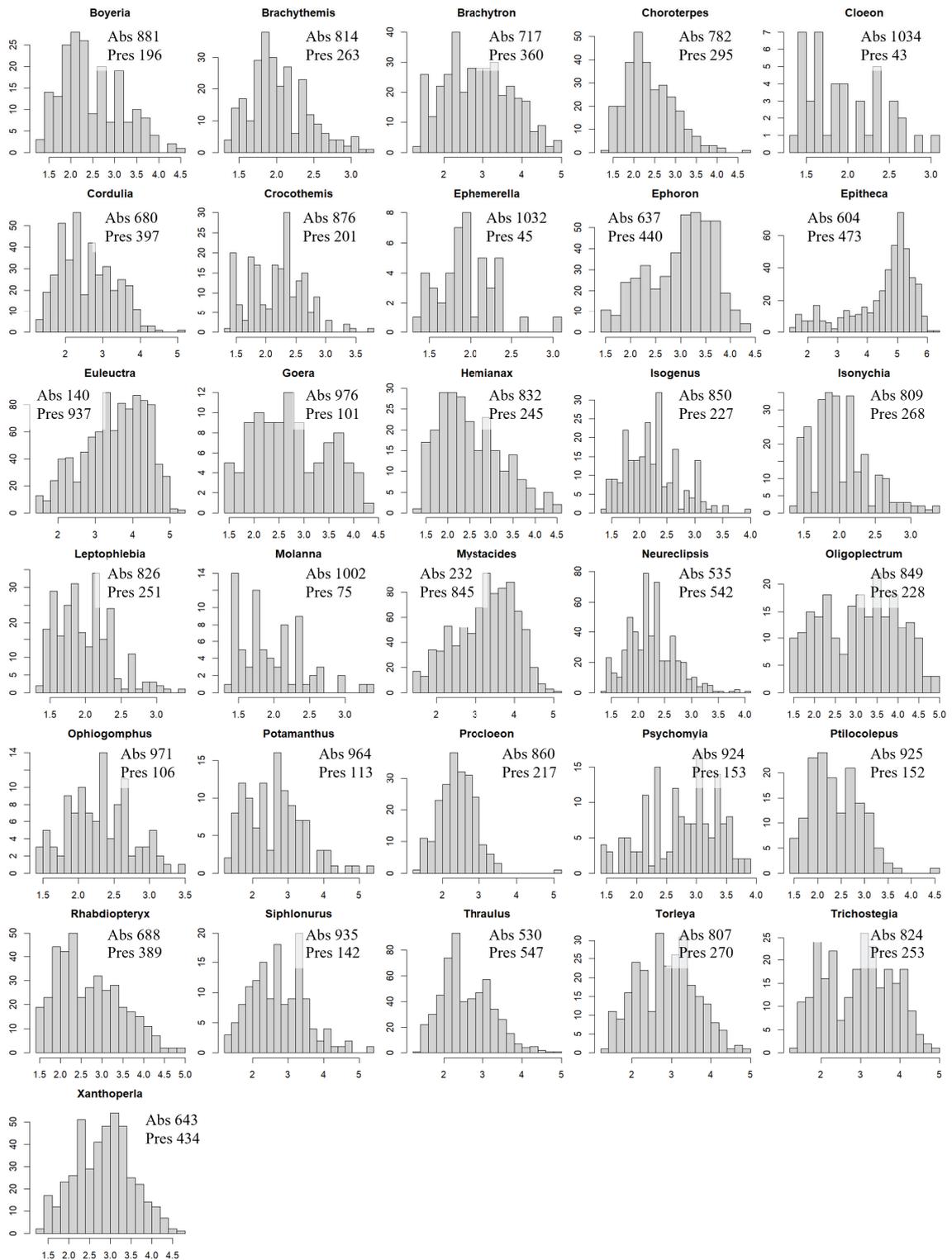


FIGURE 36

Histogrammes des CPUE > 0 observées sur les tronçons échantillonnés (N = 1077) pour les genres de macroinvertébrés modélisés (N = 31). Les CPUE ont été log-transformées (log₁₀) et les absences ont été retirées à des fins de lisibilité. Les nombres d'absence et de présence sont précisées pour chaque taxon.

B.2. Méthodologie déployée pour modéliser la distribution spatiale des abondances

Les méthodes de *machine learning* ont été choisies pour leur flexibilité à modéliser des relations variable réponse-prédicteurs à la fois simples et complexes (Valavi *et al.*, 2023). Parmi les méthodes de *machine learning* à disposition les méthodes de RF ont été choisies pour leur plus grande performance prédictive (Snelder *et al.*, 2013 ; Waldock *et al.*, 2021 ; Valavi *et al.*, 2023). De plus, des méthodes d'analyses interprétatives de RF existent et permettent d'apporter des éléments de compréhension dans les relations organisme-environnement (Snelder *et al.*, 2013 ; Molnar, 2022), ce qui correspond à l'un de nos objectifs.

Les *random forest* sont une méthode d'apprentissage automatique basée sur le regroupement de multiples arbres de décisions (Breiman, 2001). L'algorithme combine les arbres de forêts aléatoires de manière aléatoire par des méthodes dites de « *bagging* » (*bootstrap aggregating*), qui permettent d'améliorer la stabilité des méthodes d'apprentissage ensembliste et d'éviter le sur-apprentissage.

B.2.1. Passage en classes d'abondance

Les abondances de chaque taxon se présente sous la forme de densités CPUE (ind/100m²) dont la modélisation pose un certain nombre de difficultés techniques (voir chapitre I section C). Les termes "abondances" et "densités" sont ensuite utilisés indifféremment pour désigner la modélisation des CPUE. Les données d'abondance continues sont transformées en classes d'abondance. L'utilisation de classes d'abondance permet de s'affranchir d'une limite importante liée à la détectabilité des espèces et aux variabilités induites par les protocoles d'échantillonnage, connue pour influencer les performances des SDM (Comte & Grenouillet, 2013b ; Vaudor *et al.*, 2015 ; Guisan *et al.*, 2017). L'utilisation de classes permet aussi d'atténuer l'impact de la sur-dispersion en incluant les effectifs sur-dispersés dans la classe d'abondance forte. Enfin, il permet aussi de faciliter l'exploitation des sorties de modélisation par les gestionnaires de la biodiversité. L'interprétation des abondances continues nécessite une connaissance fine des gammes de valeurs habituelles pour chaque espèce en fonction de la zone géographique considérée afin de déterminer quelles sont les valeurs faibles et fortes d'abondance. Les classes permettent de communiquer directement cette information, du moins relativement par rapport à la gamme d'abondance que

présente l'espèce sur la zone d'étude.

Le choix du nombre de classes a des implications importantes dans la modélisation et l'évaluation des performances prédictives. Pour les espèces rares, trop peu d'occurrences sont observées pour pouvoir être réparties dans un grand nombre de classes. Et l'application d'un trop petit nombre de classes fait perdre en finesse dans la structuration des abondances des espèces plus abondantes et/ou largement répandues. De plus, mécaniquement, les métriques de performances sont dépendantes du nombre de classes modélisés, rendant incomparables deux modèles basés sur un nombre de classes différent (Hornung, *communication personnelle*). Il n'a donc pas été possible de construire des modèles avec un nombre variable de classes pour pouvoir ensuite conserver le modèle avec le nombre de classes la plus approprié et performant pour chaque taxon. À des fins de simplification et de compromis entre un nombre de classes suffisamment petit pour les espèces rares et suffisamment grand pour les autres, quatre de classes d'abondance ont été définies : absence, faible, moyenne et forte abondance, annotés de 0 à 3 respectivement (Figure 37). La conversion des CPUE \log_{10} -transformées strictement positives en classes d'abondance a été effectuée via l'algorithme de découpage automatique de *Jenks* (package R *ClassInt*, version 0.4-8).

B.2.2. La méthode de modélisation choisie : l'ordinal forest

L'ordination de la classe absence vers les classes d'abondance plus fortes traduit une information écologique qu'il semble essentiel de conserver. Les modèles OF sont basés sur les RF et sont dédiés spécifiquement à la modélisation d'une variable réponse ordinale (Hornung, 2020). La comparaison avec d'autres méthodes de régression ordinale que sont les RF multi-classes, les régression de forêts et la régression ordinale probit, a montré que les OF présentaient de meilleures performances prédictives (Hornung, 2020). De même, les OF surpassent les autres méthodes dans l'ordination des prédicteurs en fonction de leur importance à expliquer la variable réponse (voir Hornung (2020) pour plus de détails). Cette méthode, comme plusieurs méthodes de ML, présente notamment une meilleure performance en présence d'une répartition asymétrique des occurrences entre les classes (Steen *et al.*, 2021), ce qui est le cas de la plupart des classes d'abondance des taxons que j'ai cherché à modéliser.

Lors du développement de la méthode OF, les hyperparamètres (*i.e.* nombre d'arbres,

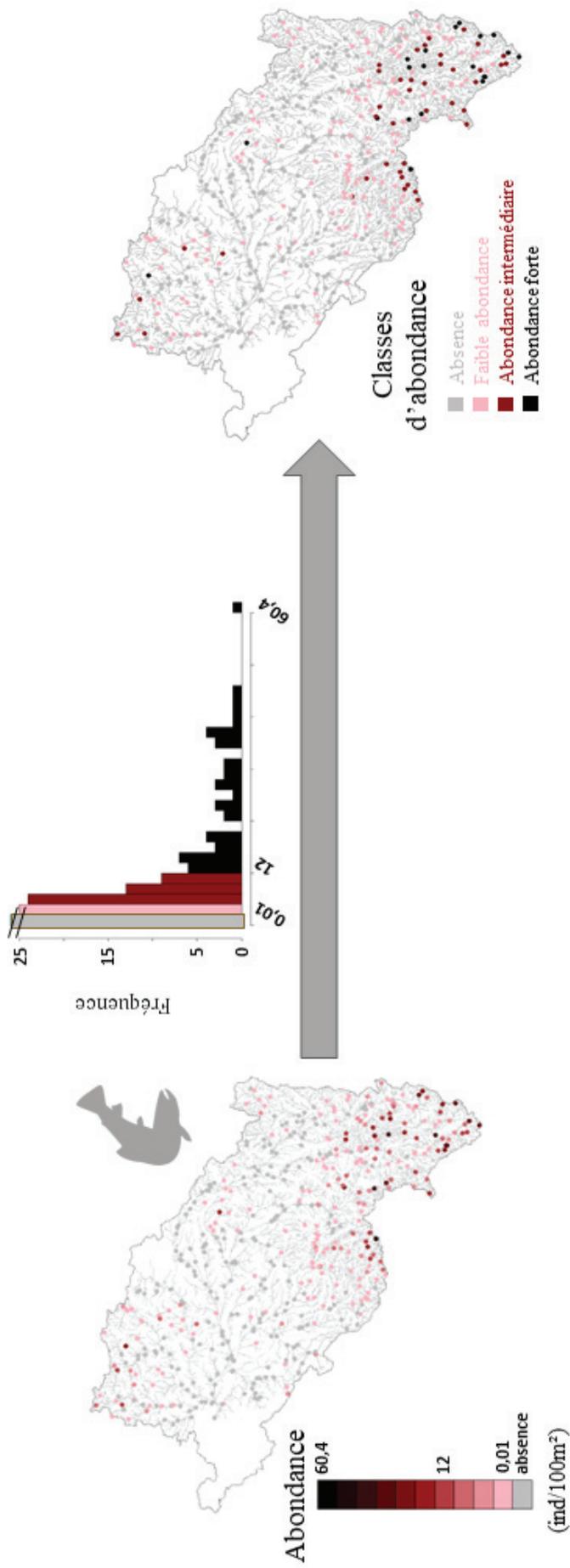


FIGURE 37

Transformation des valeurs de CPUE en classes d'abondance via l'algorithme *Jenks*. Exemple réalisé à partir des données d'abondance de la truite commune (*Salmo trutta*). La barre des absences est tronquée, le nombre total d'absence est de 243, sur un total de 484 tronçons échantillonnés.

nombre de noeuds, etc.) ont fait l'objet d'une optimisation dans différentes configurations de modélisation afin d'obtenir le meilleur compromis entre temps de calcul et performances (Hornung, 2020). Conformément aux recommandations de l'auteur, les valeurs par défaut des hyperparamètres de la méthode OF ont été conservées (Hornung, 2022). La méthode offre la possibilité d'obtenir des prédictions probabilistes, c'est-à-dire d'obtenir pour chaque tronçon la probabilité d'appartenir à chacune des classes d'abondance. La version 2.4-3 du package *OrdinalForest* a été utilisée durant toute la phase de modélisation et d'analyse (Hornung, 2022).

B.2.3. Choix des prédicteurs environnementaux

Les corrélations entre les variables environnementales disponibles dans notre jeu de données ont été calculées afin de sélectionner un ensemble de prédicteurs peu corrélés (coefficient de corrélation de Spearman $< 0,7$; Figure 38). Le chapitre III a mis en évidence l'importance des régimes thermiques et hydrauliques dans la structuration spatiale des abondances des communautés de poissons et de macroinvertébrés (voir chapitre III). Des variables liées aux pressions anthropiques, également connues pour influencer les communautés aquatiques, ont été ajoutées pour la modélisation (Rumschlag *et al.*, 2023). Il s'agit notamment d'indicateurs de la qualité de l'eau au travers des activités anthropiques aux abords des cours d'eau. Afin de pallier le manque de prise en compte du rôle de la composante dispersion dans la distribution spatiale des espèces, deux indicateurs liés à la pression anthropique sur la continuité écologique ont été considérés. À l'instar des données biologiques, l'objectif était de retranscrire les conditions environnementales moyennes et leur variabilité globales à l'échelle du bassin ligérien, pour obtenir une représentation médiane des conditions rencontrées sur une période donnée. Ainsi, les modèles construits mettent en relation les classes d'abondance de chaque taxon avec dix prédicteurs environnementaux :

Classes d'abondance taxon $i \sim$ pente + Tw_m + Tw_{9010} + v_m + v_{9010} + densité de population + gradient d'anthropisation + densité d'ouvrages + taux d'étagement + probabilité d'intermittence

où *pente* est la pente moyenne du tronçon de cours d'eau, Tw_m est la température moyenne interannuelle de l'eau (en °C), Tw_{9010} est le ratio entre les percentiles 90 et 10 des températures de l'eau (adimensionnel), v_m est la vitesse moyenne interannuelle du courant (m/s),

v_{9010} est le ratio entre les percentiles 90 et 10 de vitesse de courant (adimensionnel), *densité de population* représente la densité de population aux abords directs du tronçon (en hab/km²), *gradient d'anthropisation* est la variable synthétique issue de CLC pour le gradient agricole (sans unité), *densité d'ouvrages* est le nombre moyen d'ouvrages présents sur l'ensemble du linéaire de même ordre de Strahler (maille de calcul de cette variable ; en nombre d'ouvrages/m), *taux d'étagement* est le rapport de la hauteur de chute sur la dénivellation naturelle (adimensionnel) et *probabilité d'intermittence* est la probabilité d'assec du cours d'eau (en %). Les métriques v_m , v_{9010} , densité de population, densité d'ouvrages et taux d'étagement ont été normalisées par transformation $\log(x + 1)$, la pente et la probabilité d'intermittence ont été normalisées par transformation $\text{asin}(\sqrt{x})$.

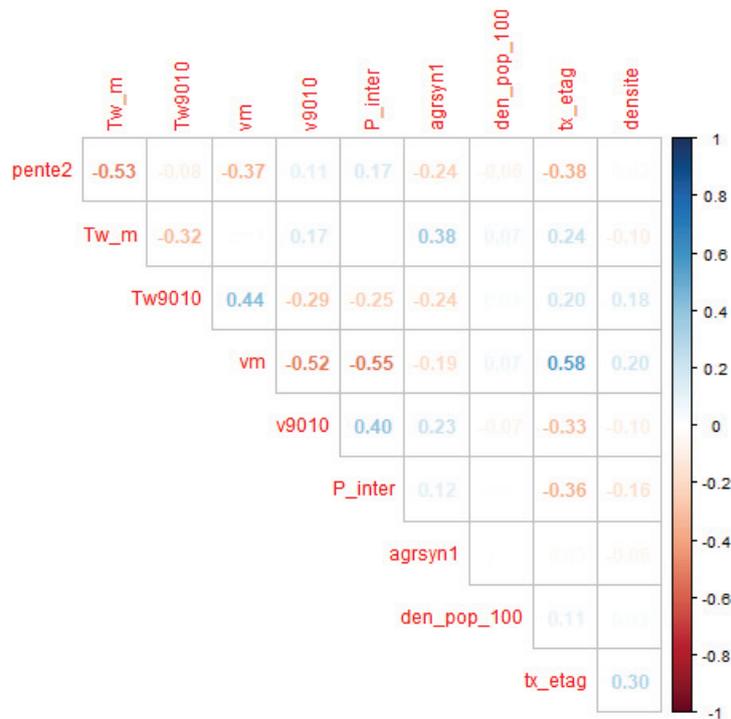


FIGURE 38

Corrélogramme entre les 10 prédicteurs environnementaux utilisés dans la modélisation de la distribution des abondances. Les valeurs présentées sont le coefficient de corrélation de Spearman. La correspondance avec les codes prédicteurs peut être consulté dans le tableau 9.

TABLEAU 9

Variables environnementales utilisées comme prédicteurs dans les SDM d'abondance. Les valeurs sont extraites à partir des tronçons portant une donnée biologique observée (poissons, N = 482 tronçons ; macroinvertébrés, N = 1077 tronçons). Les valeurs moyennes [minimum ; maximum] pour chaque jeu de données et pour l'ensemble du réseau H sont présentées.

Variable	Unité	Code	Réseau	Poisson	Macroinvertébrés
Pente	-	pen	0,01 [0,00 ; 0,28]	0,01 [0,00 ; 0,06]	0,01 [0,00 ; 0,08]
Moyenne interannuelle de la température de l'eau	°C	Tw _m	11,5 [6,6 ; 18,7]	11,7 [7,3 ; 14,5]	11,6 [7,2 ; 14,5]
Amplitude de la température de l'eau	-	Tw ₉₀₁₀	3,1 [0,3 ; 1,7]	3,8 [1,8 ; 11,1]	3,7 [1,9 ; 16,0]
Moyenne interannuelle de la vitesse de courant	m/s	v _m	0,02 [0 ; 0,07]	0,03 [0 ; 0,06]	0,02 [0 ; 0,06]
Amplitude de vitesse de courant	-	v ₉₀₁₀	0,9 [0,3 ; 1,7]	0,7 [0,3 ; 1,5]	0,8 [0,3 ; 1,5]
Densité de population	hab/km ²	den _{pop_100}	57,2 [1,7 ; 4035,0]	91,5 [2,6 ; 2219,4]	106,6 [2,6 ; 3955,7]
Pression agricole	-	agrsyn1	0 [-7,9 ; 2,5]	91,5 [-1,93 ; 2,5]	-0,1 [-1,9 ; 2,5]
Densité d'ouvrages	nb d'ouvrages/m	densite	0,2 [0 ; 11,2]	0,4 [0 ; 3,8]	0,4 [0 ; 3,7]
Taux d'étagement	-	Tx _{etag}	0,01 [0 ; 7,8]	0,2 [0 ; 4,3]	0,2 [0 ; 4,3]
Probabilité d'intermittence	-	P _{inter}	0,3 [0 ; 0,9]	0,2 [0 ; 0,9]	0,3 [0 ; 0,9]

B.2.4. Limites de l'utilisation des classes d'abondance et solution proposée

Bien que le passage des valeurs de CPUE en classes d'abondance permette de limiter certains problèmes comme la sur-dispersion de la distribution des abondances, ils ne sont pas tous résolus. En effet, la sur-inflation en zéros de certains taxons est maintenue puisque tous les zéros forment la classe "absence" (Figure 39). Pour quelques taxons la sur-inflation d'une autre classe d'abondance (que la classe absence) peut être observée (e.g. *Euleuctra*, goujon). Ce problème est bien connu dans d'autres domaines de recherche où l'utili-

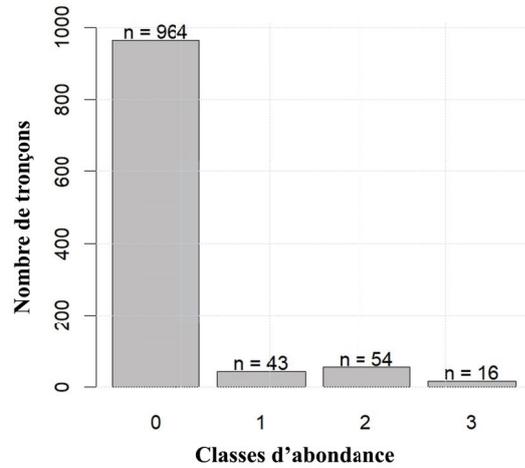


FIGURE 39
Exemple de sur-inflation en zéros illustrée à partir des classes d'abondance observées pour le genre *Potamanthus* (n = 1 077 tronçons).

sation de modèles multi-classes est courante, comme par exemple dans le développement de méthodes de ML et leurs applications en recherches médicales ou en traitement de l'image et du signal. La sur-représentation d'une classe par rapport aux autres induit une asymétrie, c'est-à-dire un ratio non-équilibré entre les effectifs des classes d'abondance (voir par exemple le cas du genre *Potamanthus* en Figure 39). Cette asymétrie est plus ou moins prononcée en fonction de la rareté du taxon et de la distribution de ses abondances.

Elle peut provoquer le sur-apprentissage du modèle de ML sur les classes les plus représentées (Valavi *et al.*, 2021). En effet, lors de la phase d'entraînement, le modèle cherche à maximiser le nombre de prédictions exactes (*accuracy*¹; TP sur la matrice de confusion en tableau 10), sans tenir compte de l'ampleur de l'écart de prédiction pour les prédictions erronées (ec 1, ec 2 ou ec 3 en tableau 10), favorisant ainsi les prédictions sur la classe sur-représentée au

TABLEAU 10
Matrice de confusion d'un modèle multi-classe dans un cas à 4 classes. TP : vrais positifs, ec x : écart de x classes entre observé et prédit.

		Classes observées			
		0	1	2	3
Classes prédites	0	TP	ec 1	ec 2	ec 3
	1	ec 1	TP	ec 1	ec 2
	2	ec 2	ec 1	TP	ec 1
	3	ec 3	ec 2	ec 1	TP

1. Le terme anglais a été conservé afin d'éviter la confusion avec la *precision*.

détriment des classes minoritaires. De plus, un modèle qui a sur-appris sur son jeu d'entraînement risque d'être peu performant face à de nouvelles conditions, ce qui limite son pouvoir de généralisation (Guisan *et al.*, 2017). Lors des tests préliminaires, ce phénomène de sur-apprentissage a été observé, notamment pour les taxons macroinvertébrés rares et/ou peu abondants.

Des méthodes existent pour limiter l'influence des classes sur-représentées dans la construction du modèle. Elles peuvent être appliquées sur une seule classe (*i.e.* majoritaire ou minoritaire) ou sur plusieurs classes à la fois. Il existe quatre catégories de méthodes :

- (i) **Modèle pondéré par les effectifs au sein des classes.** La pondération par le nombre d'effectifs au sein de chaque classe permet au modèle de maximiser le nombre de prédictions exactes (*i.e.* *accuracy*) au sein de chaque classe (Hornung, 2020). Néanmoins, cette méthode semble offrir les moins bons résultats lorsqu'elle est utilisée pour réaliser des SDM, et particulièrement pour les espèces rares (Robinson *et al.*, 2018);
- (ii) **Sous-échantillonnage de la classe majoritaire.** Le sous-échantillonnage permet de réduire l'effectif dans la classe majoritaire et donc de limiter son influence dans la construction du modèle. Il a l'avantage d'être basé sur les données réelles mais présente le désavantage de réduire leur nombre et de limiter la quantité de données mobilisées par le modèle pour s'entraîner, ce qui peut impacter ses performances et lui faire, par exemple, manquer un patron majeur dans la structuration des données. Ceci est d'autant plus vrai avec les méthodes de ML qui ont besoin d'une quantité importante de données pour être performants (Molnar, 2022);
- (iii) **Bootstrap ou Synthetic Minority Oversampling Technique (SMOTE) de la classe minoritaire.** Les méthodes de *bootstrap* ou SMOTE s'appliquent à la classe minoritaire (ou aux classes minoritaires) et reconstituent artificiellement des observations pour augmenter les effectifs au sein de ces classes. Le *bootstrap* ré-échantillonne avec remise dans les observations de la classe minoritaire. Cela permet de réduire l'écart entre les effectifs des classes majoritaires et minoritaires mais dans une certaine mesure seulement. De plus, le *bootstrap* n'apporte pas de nouvelles données et augmente le risque de sur-apprentissage du modèle sur la classe boostannée;
- (iv) **La méthode combinée SMOTE and Cluster-based Undersampling Technique (SCUT).**

Cette méthode est une approche intermédiaire entre augmentation de l'effort d'échantillonnage et le *bootstrap* dont l'essence est basée sur l'intérêt que représentent des données supplémentaires issues d'observations réelles (Chawla *et al.*, 2011). Contrairement au *bootstrap* qui multiplie les mêmes données observées, le SMOTE construit des individus "artificiels" en modifiant très légèrement l'ensemble des valeurs de la variable d'intérêt et des ces prédicteurs associés (Figure 40), comme si de nouvelles données d'échantillonnage étaient obtenues. La méthode SCUT combine une méthode de sous-échantillonnage à la méthode SMOTE, afin de combiner les avantages de ces deux méthodes (Agrawal *et al.*, 2015).

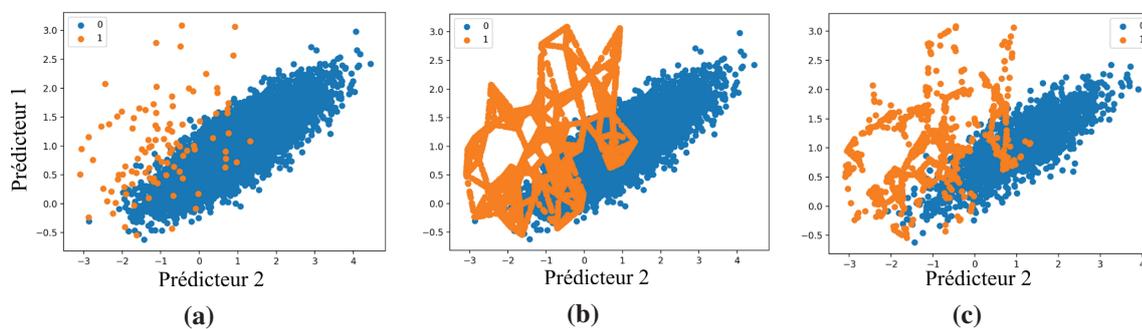


FIGURE 40

Principe de la méthode SMOTE. Exemple basé sur une variable à 2 classes 0 et 1 (la méthode s'applique de manière similaire à une variables de plus de 2 classes), les données de base sont représentées en fonction de deux prédicteurs (40a). La méthode SMOTE reconstruit tout d'abord des individus artificiels (40b) puis en extrait un sous-échantillon en fonction des effectifs finaux souhaités dans chacune des classes (40c).

Illustrations d'après machinelearningmastery.com

Parmi les techniques de sous-échantillonnage qui existent (différents types de *clustering* ou méthode du plus proche voisin), la méthode par *clustering* hiérarchique a été utilisée dans notre cas (Ganz, 2021). Elle permet de réduire l'effectif de la classe majoritaire et d'augmenter les effectifs de la classe minoritaire dans une mesure suffisante pour limiter le biais de sur-apprentissage du modèle tout en conservant le maximum d'information statistique issue de la donnée observée (Agrawal *et al.*, 2015).

Par défaut, la méthode SCUT attribue les mêmes effectifs à toutes les classes de la variable à modéliser (Ganz, 2021). Par exemple, si l'effectif total de toutes les classes confondues est de 40, l'application de la méthode SCUT pour une modélisation à quatre classes équilibrera les effectifs entre les classes en conservant 10 valeurs pour chaque classe. La méthode SMOTE est alors appliquée pour les classes dont l'effectif initial est inférieur à l'effectif recherché (10 dans cet exemple), et un sous-échantillonnage est appliqué pour

les classes dont l'effectif est supérieur à 10. J'ai évalué l'influence du choix de l'effectif à attribuer à chaque classe a été testé avec différentes configurations de nombre d'observations par classe. Par exemple, j'ai tenté d'appliquer un SMOTE et un sous-échantillonnage plus modéré que les méthodes par défaut sur les classes à augmenter et réduire, respectivement. Les effectifs de la classe minoritaire ont été multipliés par 2 au maximum et les effectifs de la classe majoritaire ont été divisés par 2 au maximum (voir par exemple Robinson *et al.* (2018)). D'autres effectifs ont été testés, mais cela n'a pas modifié l'efficacité de la méthode, ce qui est cohérent avec les résultats d'autres études similaires (Steen *et al.*, 2021). Par conséquent, la méthode SCUT par défaut a été conservée.

B.2.5. Principe de modélisation des classes d'abondance

Les étapes de construction des modèles de classes d'abondance sont schématisées en figure 41 et détaillées ci-après.

- Étape 1 :** Les données de CPUE log-tranformées (\log_{10}) sont converties en quatre classes d'abondance via l'algorithme *Jenks* ;
- Étape 2 :** Afin de pouvoir ensuite évaluer les performances du modèle construit, le jeu de données est divisé en un jeu d'entraînement et un jeu de test contenant respectivement 70% et 30% des tronçons échantillonnés (Guisan *et al.*, 2017). Afin de conserver la répartition des effectifs entre les classes d'abondance dans chacun des jeux d'entraînement et de test, un tirage aléatoire stratifié par les classes d'abondance a été réalisé. Pour évaluer l'intérêt de la méthode SCUT, deux modèles ont été construits à partir du même jeu d'entraînement : l'un sur les données brutes (nommé non-SCUT), l'autre sur les données transformées par SCUT (nommé SCUT). Les classes prédites par chacun des modèles sur les données du jeu test sont ensuite comparées aux classes observées afin d'évaluer les performances de chacun des modèles ;
- Étape 3 :** Afin de limiter l'effet de la stochasticité du modèle OF et un potentiel biais induit par le choix des jeux de données d'entraînement et de test, cette procédure est répétée 20 fois ;
- Étape 4 :** Chaque modèle ($N = 20$) ainsi obtenu est ensuite utilisé pour projeter la distribution potentielle des classes d'abondance sur l'ensemble de la zone d'étude (réseau H). Les projections sont ensuite moyennées sur chaque tronçon.

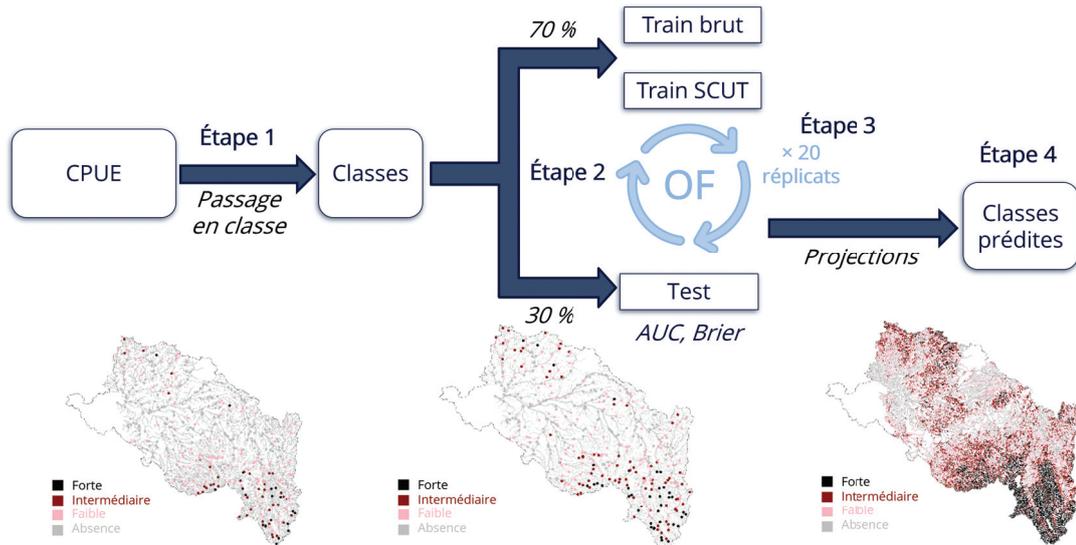


FIGURE 41

Schéma des différentes étapes de la modélisation de la distribution des abondances d'une espèce. Sur chaque tronçon, les CPUE sont transformées en 4 classes via l'algorithme *Jenks*. Un *Ordinal Forest* est entraîné sur 70% des données et testé sur les 30% restants, la procédure est répétée 20 fois (20 réplicats). Les 20 classes prédites sont ensuite moyennées sur chaque tronçon pour obtenir la prédiction moyenne.

C. Intérêt de l'application des OF pour la modélisation de la distribution spatiale des classes d'abondance

C.1. Évaluation des performances de deux méthodologies de modélisation des classes d'abondance par *ordinal forest*

L'objectif de cette section est de présenter les performances de deux méthodologies de modélisation et de les comparer. À notre connaissance, la méthode de transformation des données d'entraînement SCUT n'a été appliquée qu'à quelques cas de SDM de présence/absence (Robinson *et al.*, 2018 ; Steen *et al.*, 2021), bien que cette méthode ait été identifiée comme une des solutions à l'asymétrie depuis un certain temps (Johnson *et al.*, 2012). Cependant, à notre connaissance, il n'existe pas d'exemple d'application pour des SDM d'abondance.

C.1.1. Comparaisons cartographiques

La figure 42 présente les classes d'abondance observées et prédites par les deux modèles SCUT et non-SCUT testés uniquement sur les tronçons portant de la donnée observée. La comparaison visuelle des prédictions issues des modèles non-SCUT et SCUT

C. Intérêt de l'application des ordinal forest pour la modélisation de la distribution spatiale des classes d'abondance

sur la base des cartographies à l'échelle du bassin versant de la Loire semble indiquer la capacité de la méthode SCUT à limiter le phénomène de sur-apprentissage sur la classe majoritaire, en particulier pour les taxons pour lesquels l'asymétrie entre les classes est forte (e.g. *Trichostegia*; Figure 42). En effet, le sur-apprentissage du modèle non-SCUT en faveur de la classe absence induit une sur-prédiction de cette classe, particulièrement visible pour *Trichostegia* (Figure 42), et limite l'information que peut nous apporter le modèle sur les classes non nulles, ainsi que sa capacité de discrimination. Les patrons de distribution des classes d'abondance semblent donc mieux représentés par le modèle SCUT, même si la cartographie des prédictions semble indiquer une légère tendance à la sur-prédiction sur ces classes.

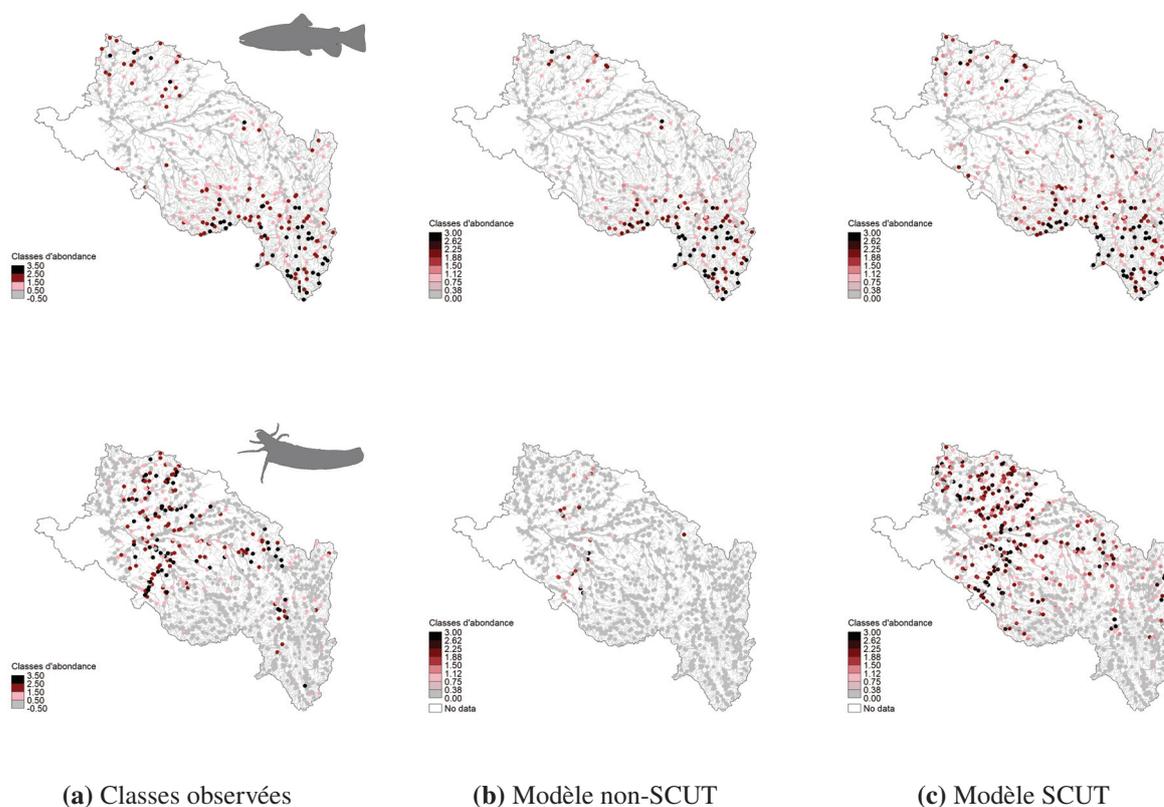


FIGURE 42
Exemples de comparaisons cartographiques des modèles non-SCUT vs SCUT sur les tronçons portant les données observées pour la truite commune en haut (n = 484 tronçons), et *Trichostegia* en bas (n = 1 077 tronçons). Les absences sont représentées en gris.

Bien que cette comparaison visuelle illustre le comportement des deux modèles testés, elle peut être biaisée par la résolution de l'image, le choix de l'échelle des classes prédites ou encore les couleurs utilisées (voir aussi Annexe G).

C.1.2. Métriques de performance

Les modèles de *machine learning* basés sur une variable réponse multi-classes ordinaire nécessitent des méthodes d'analyses dédiées. Différentes métriques ont été considérées et testées pour évaluer les performances des modèles OF de classes d'abondance. Conformément aux recommandations en matière d'évaluation des performances (voir Encart 2), deux métriques non corrélées ont été choisies : le score de Brier et l'AUC multi-classes. Le score de Brier est une mesure de la calibration d'un modèle, c'est-à-dire de la capacité du modèle à estimer correctement les ratios entre les classes dans tous les sites. L'AUC ROC est une mesure de la discrimination d'un modèle, soit sa capacité à classer correctement les sites en fonction de leur niveau d'abondance. L'AUC ROC a par ailleurs montré sa pertinence à évaluer un modèle dans le cadre de données asymétriques (Johnson *et al.*, 2012).

Encart 2. Quelle(s) métrique(s) de performance choisir ?

Il existe plusieurs métriques pour évaluer les performances des modèles multi-classes (> 2 classes). On distingue deux catégories de métriques pour en évaluer les performances prédictives : (i) les extensions de métriques initialement développées pour les modèles binaires (*e.g.* AUC multi-classes, *accuracy*, coefficient de corrélation de Matthews) et (ii) les métriques développées spécifiquement pour les modèles multi-classes (*e.g.* score de Brier, Kappa de Cohen) (Peterson *et al.*, 2011). Le choix de la métrique doit être fait en fonction des objectifs de la modélisation et de sa pertinence avec les données et le type de modélisation (Collart & Guisan, 2023). Pourtant, il existe des cas pour lesquels une métrique peut échouer à montrer de meilleures performances. Par exemple, Delgado & Tibau (2019) ont montré une contradiction entre le Kappa de Cohen et le coefficient de corrélation de Matthew, alors qu'elles sont toutes les deux une mesure de la calibration. Plus généralement, chaque métrique est basée sur un aspect particulier de la performance de modélisation (*i.e.* discrimination, calibration), et lorsque cet aspect est différent elles sont peu corrélées entre elles (Konowalik & Nosol, 2021). En fonction des objectifs d'utilisation du modèle et de ses sorties, ce ne sont pas les mêmes indicateurs qui seront considérés (Guisan *et al.*, 2017). Tout l'objet de la modélisation est de trouver le meilleur compromis entre les différents aspects de la performance du modèle pour que celui-ci réponde aux objectifs attendus. Il est donc recommandé d'utiliser plusieurs métriques de performance, qui soient adaptées aux données et au type de modèle, qui répondent aux objectifs de la modélisation (Peterson *et al.*, 2011) et d'avoir recours à l'expertise quant à la performance du modèle étudié (Konowalik & Nosol, 2021).

C. Intérêt de l'application des ordinal forest pour la modélisation de la distribution spatiale des classes d'abondance

De par sa construction, le score de Brier est très influencé par l'*accuracy* du modèle (Steen *et al.*, 2021) (*i.e.* le nombre de prédictions exactes, notées **TP** dans le tableau 10; voir formule ci-dessous) et par la distribution des probabilités d'appartenir à chacune des classes. L'AUC multi-classes retranscrit lui l'ensemble des écarts entre classes observées et classes prédites, soit l'erreur moyenne du modèle, et est moins sensible à l'*accuracy*.

Le score de Brier BS est calculé comme suit :

$$BS = 1 - \frac{1}{2n} \sum_{i=1}^n \sum_{k=1}^K (c_{ik} - p_{ik})^2$$

où n est le nombre d'observations, K le nombre de classes de la variable réponse, c_{ik} un indicateur qui prend 1 si la classe prédite k correspond à la classe observée pour l'observation i sinon 0 et p_{ik} la probabilité estimée par le modèle de l'observation i d'appartenir à la classe k . Le score de Brier prend des valeurs entre 0 (très mauvaise performance) et 1 (classes observées = classes prédites).

Concernant l'AUC, il existe plusieurs généralisations au cadre multi-classe. Un AUC multi-classes est une moyenne de plusieurs AUC qui peuvent être calculées (i) sur toutes les combinaisons des classes prises deux à deux (*one-vs-one*) ou (ii) en rapportant chaque classe à un cas binaire, la classe d'intérêt prenant la valeur 1 et toutes les autres 0 (*one-vs-rest*). Ici, j'ai choisi d'utiliser la généralisation proposée par Hand & Till (2001) (package R pROC) :

$$AUC \text{ multi-classe} = \frac{2}{K(K-1)} \sum auc_s$$

où K est le nombre de classe de la variable réponse et auc_s les AUC calculés entre les classes prises deux à deux (*one-vs-one*). L'AUC multi-classe ainsi définie prend des valeurs entre 0 (très mauvaise adéquation entre observations et prédictions) et 1 (parfaite adéquation). Le calcul de la métrique étant dépendant du nombre de classes modélisées (voir Encart 2) les valeurs d'AUC multi-classes ne sont pas comparables à un AUC calculé sur un modèle binaire comme un SDM en présence/absence.

L'évaluation des performances par l'AUC multi-classe et le score de Brier est présentée dans la Figure 43. Les gammes de valeurs pour l'AUC vont de 0,5 (équivalent au modèle nul; *e.g.* lamproie de Planer, *Ephemerella*) à 0,85 (truite commune) et, pour le score de Brier, de 0,29 (*Mystacides*) à 0,92 (*Cloeon*). Toutes les valeurs sont représentées entre ces gammes et les performances sont très variables entre taxons, avec des modèles relativement

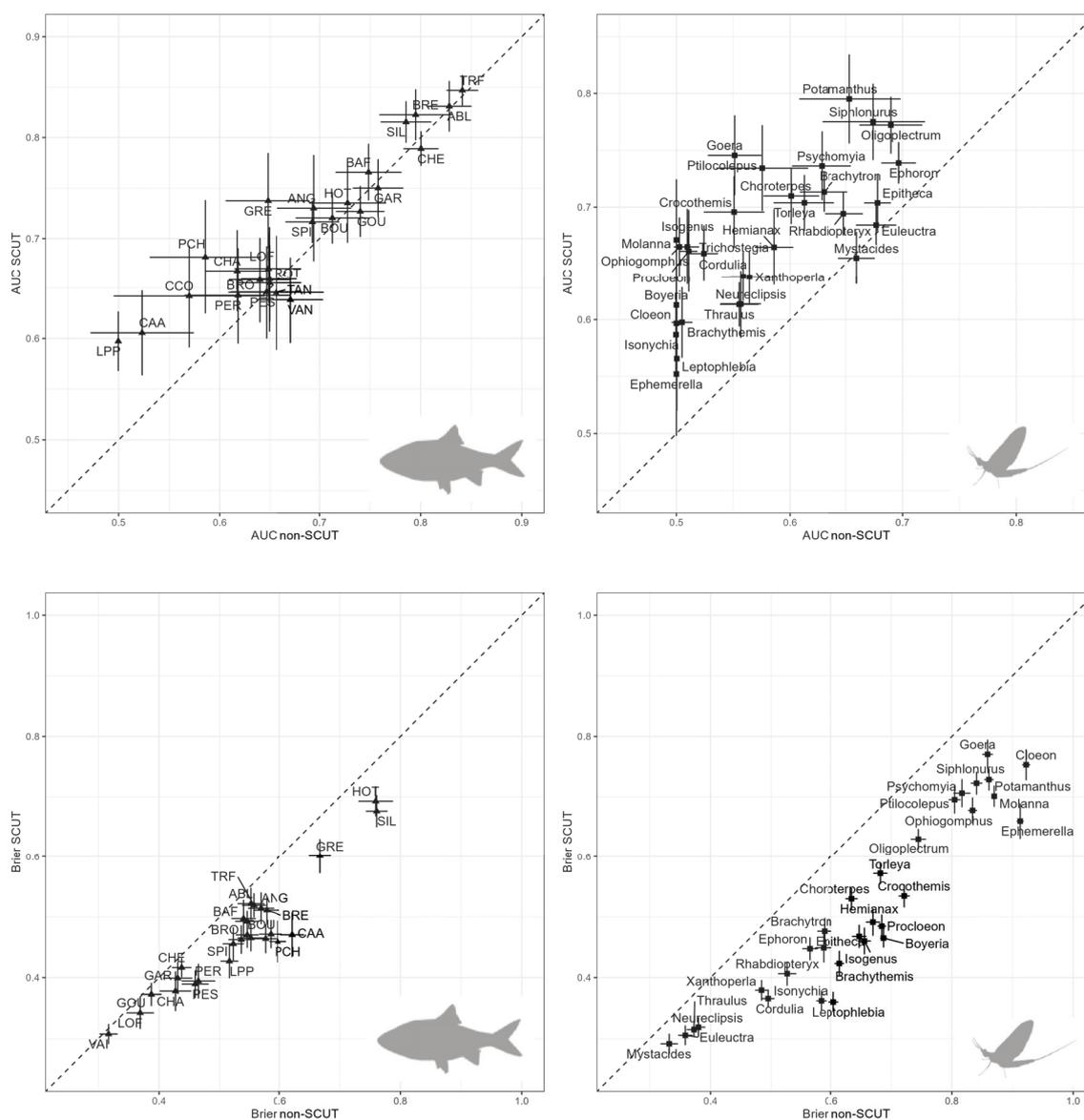


FIGURE 43

Comparaison des performances des modèles sur données non-SCUT vs. SCUT, basée sur l’AUC multi-classe (haut) et le score de Brier (en bas) pour les modèles poissons (à gauche) et pour les macroinvertébrés (à droite). Les valeurs présentées sont la moyenne des métriques sur les 20 réplicats, et les écarts-type associés. La correspondance avec les codes espèces poissons peut être consultées au tableau 1.

bien performants et d’autres non. D’après les écart-types, les valeurs de chaque métrique sont relativement stables entre les réplicats (voir aussi Annexe D).

La Figure 43 permet aussi de comparer les performances respectives des modèles SCUT et non-SCUT pour l’ensemble des 57 taxons modélisés. Le gain d’AUC induit par l’application d’un SCUT sur les données d’entraînement est plus important pour les macroinvertébrés que pour les poissons (Figure 43), de même que la diminution du score

C. Intérêt de l'application des ordinal forest pour la modélisation de la distribution spatiale des classes d'abondance

de Brier. Les modèles non-SCUT pour les genres de macroinvertébrés *Boyeria*, *Brachythemis*, *Cloeon*, *Ephemerella*, *Isogenus*, *Leptophlebia*, *Molanna* et *Procloeon* et pour la lamproie de Planer (LPP) ne prédisent que des absences, d'où la valeur d'AUC de 0,5. La transformation SCUT permet d'améliorer l'AUC jusqu'à 0,16 point pour ces taxons. Une apparente baisse d'AUC ($< 0,05$) du modèle SCUT par rapport au modèle non-SCUT est observée chez certains poissons (*e.g.* tanche et vandoise) mais elle reste marginale (voir les barres d'erreurs). Le score de Brier indique une baisse du nombre de prédictions exactes avec les modèles SCUT pour l'ensemble des taxons par rapport aux modèles non-SCUT. L'augmentation d'AUC conjointe à la diminution du score du Brier semble indiquer un phénomène de sur-apprentissage sur la classe majoritaire de la part des modèles non SCUT. En effet, un modèle qui a sur-appris sur ces données prédira en grande majorité la classe majoritaire, ici la classe absence, et présentera en apparence de bonnes performances de calibration. Cela pose deux problèmes : (i) ce modèle sera peu performant face à de nouveaux jeux de données (faible capacité de généralisation) et (ii) le biais de sur-apprentissage va limiter l'information interprétative qui sera tirée du modèle, ici les relations organisme-environnement. Idéalement, les jeux tests utilisés pour évaluer les performances des modèles doivent être indépendants des jeux d'entraînement, ce qui permet de révéler facilement un modèle sur-entraîné par une faible performance (Guisan & Zimmermann, 2000). En écologie, il est rare d'avoir à disposition des jeux de données parfaitement indépendants, d'où la stratégie de séparation du jeu de données global en un jeu d'entraînement et un jeu de test et la répétition de cette procédure (Guisan *et al.*, 2017). Au final, les différences d'AUC et de score de Brier entre modèles non-SCUT *vs.* modèles SCUT indiquent que l'application de la méthode SCUT augmente la discrimination du modèle au détriment de sa calibration. Ainsi, les niveaux d'abondance prédits sur les tronçons sont plus exacts mais la répartition entre les niveaux d'abondance observés dans le jeu de données test n'est pas maintenue.

Pour plus de détails, les matrices de confusion permettent de juger la performance des modèles et les effets de la transformation SCUT des données d'entraînement sur leurs prédictions et le taux, ainsi que l'amplitude des erreurs (Figure 44). D'après nos observations, elles retranscrivent bien la limitation du phénomène de sur-apprentissage sur la classe majoritaire, ce pour quoi a été conçu le SCUT, avec une redistribution de l'erreur de prédiction en faveur des classes > 0 . Au final, la combinaison d'un modèle OF à la mé-

thode SCUT a permis de surmonter les limites liées à l'asymétrie des classes d'abondance. L'apport de la méthode SMOTE est en accord avec les exemples recensés dans la littérature (Johnson *et al.*, 2012 ; Robinson *et al.*, 2018 ; Steen *et al.*, 2021). Ici, cette méthode a été complétée par l'association d'un sous-échantillonnage de la classe majoritaire. Woods *et al.* (2023) ont utilisé des RF pour modéliser des classes d'abondance mais l'article ne fournit pas de mesure de la performance de leur modèle. Aucun autre exemple de modélisation de la distribution via des classes d'abondance n'a été trouvé dans la littérature, il n'existe donc pas de point de comparaison. Les valeurs d'AUC multi-classes et de score de Brier ont néanmoins été jugées satisfaisantes en tant que telles pour exploiter les modèles sur chaque espèce.

 <i>Salmo trutta</i>	Observations		Prédictions					Prédictions					
				0	1	2	3			0	1	2	3
		0	72	8	3	0	0	30	2	1	0		
		1	10	12	5	1	2	20	7	1			
		2	6	3	12	2	0	7	29	3			
3	0	0	4	8	0	1	4	39					
 <i>Molanna</i>	Observations		0	1	2	3		0	1	2	3		
		0	37	10	14	3	0	57	1	7	10		
		1	43	27	37	12	1	26	53	1	5		
		2	23	13	43	13	2	20	1	58	10		
		3	2	2	13	32	3	8	1	2	63		
		Non-SCUT				SCUT							

FIGURE 44

Exemples de matrice de confusion de deux modèles calibrés sur des données non-SCUT (gauche) vs SCUT (droite). Les matrices du haut correspondent à des modèles pour la truite commune, ceux du bas à des modèles sur *Molanna*. Pour la truite, l'amélioration des prédictions (augmentation des valeurs sur la diagonale et diminution dans les coins supérieur droit et inférieur gauche) avec la méthode SCUT est nette. C'est moins le cas pour *Molanna* pour qui l'asymétrie était plus marquée.

C.1.3. Influence de la fréquence d'occurrence et du niveau d'abondance sur la performance des modèles

Les caractéristiques de la distribution des abondances de chaque espèce peuvent influencer les performances du modèle (Filz *et al.*, 2013 ; Waldock *et al.*, 2021 ; Collart *et al.*, 2023). La différence entre l'AUC multi-classe du modèle non-SCUT et du modèle SCUT a été mise en relation avec l'abondance moyenne et la fréquence des occurrences (Figure 45 haut) puis avec le niveau d'asymétrie (Figure 45 bas) afin d'évaluer leur potentiel impact sur la performance des modèles. L'asymétrie des effectifs entre les classes

C. Intérêt de l'application des ordinal forest pour la modélisation de la distribution spatiale des classes d'abondance

d'abondance a été mesurée par le ratio des effectifs de la classe majoritaire à ceux de la classe minoritaire (Chawla *et al.*, 2011). La fréquence d'occurrence du taxon est inversement corrélée à l'asymétrie des effectifs entre les classes d'abondances (coefficient de corrélation de Spearman $\rho = -0.79$) mais peu à l'abondance moyenne ($\rho = -0.28$). L'asymétrie n'est pas corrélée à l'abondance moyenne ($\rho = 0.19$). L'absence de corrélation entre le niveau moyen de l'abondance et la fréquence d'occurrence montre qu'il n'existe pas une relation simple entre ces deux métriques, il existe des taxons très fréquents et peu abondants ou inversement (*e.g.* *Neureclipsis* et le spirilin, respectivement).

Le gain de performance du modèle SCUT par rapport au modèle non-SCUT est net pour les taxons très peu fréquents (fréquence d'occurrence $< 10\%$ pour les macroinvertés). Il semble limité au-delà d'un certain niveau d'asymétrie dont le seuil semble se situer autour d'une asymétrie de 30. Pour les modèles poissons, le gain est moindre par rapport aux macroinvertés (respectivement en valeur médiane sur l'ensemble des taxons modélisés), mais les espèces de poissons ont toutes des fréquences d'occurrence supérieures à 20% (minimum pour le silure glane avec 23%). Ceci pourrait expliquer la tendance moins nette entre les performances et les caractéristiques des espèces de poissons. Pour les macroinvertés en revanche, la relation entre gain de performance et caractéristiques des taxons est claire : moins un taxon est fréquent ou plus son asymétrie est forte, plus le gain de performance associé à l'utilisation de la méthode SCUT est grand.

Sur la base des résultats présentés ici, les performances des modèles entraînés sur des données SCUT semblent suffisantes pour les interpréter. La combinaison d'un modèle OF à la méthode SCUT a permis de surmonter les limites liées à l'asymétrie des classes d'abondance. Ce résultat est en accord avec les exemples recensés dans la littérature (Johnson *et al.*, 2012 ; Robinson *et al.*, 2018 ; Steen *et al.*, 2021). Les analyses interprétatives présentées ci-après concernent donc uniquement les modèles SCUT.

La figure 45 illustre bien la complexité de la relation entre les performances d'un modèle et la distribution des abondances de l'espèce modélisée. Ces résultats remettent en question la pertinence de ces modèles pour les espèces rares et/ou peu abondantes et leur capacité de généralisation à un grand nombre d'espèce. Ces résultats sont cohérents avec la littérature sur le sujet. Waldock *et al.* (2021) ont montré que les modèles basés sur les espèces avec de faibles abondances étaient les moins performants et qu'il existait une rela-

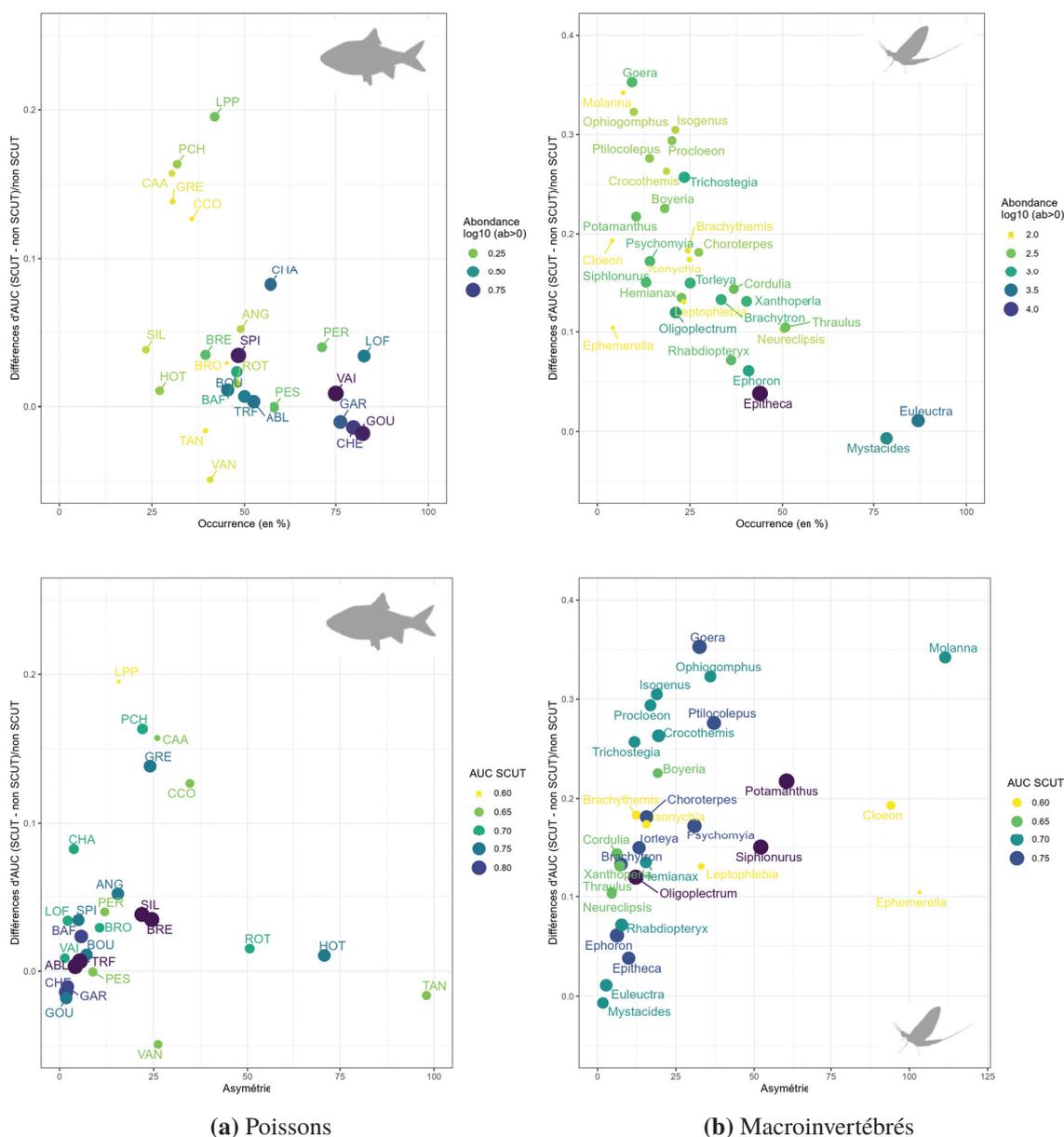


FIGURE 45

Relation entre le gain de performance entre les modèles SCUT et non-SCUT et les caractéristiques des espèces. La fréquence d'occurrence (en haut) et l'asymétrie de la distribution des classes d'abondance (en bas) en fonction de l'AUC multi-classe (moyenne sur les 20 répliquats de modèles) sont présentées pour les modèles poissons (a) et macroinvertébrés (b). L'abondance moyenne a été calculée pour les tronçons avec une occurrence > 0 du taxon, puis log-transformée. La légende des figures du haut est basée sur l'abondance moyenne, la légende des figures du bas est basée sur l'AUC des modèles SCUT. La correspondance avec les codes espèces poissons peut être consultées au tableau 1.

tion entre les performances prédictives des modèles et le niveau d'abondance. Les modèles proposés dans Waldock *et al.* (2021) sont aussi basés sur des RF; il semblerait que cette méthode, même associée à l'application d'un SCUT sur les données d'entraînement, ne suffise pas pour obtenir des modèles performants sur les espèces rares. D'autres caracté-

ristiques des espèces (*i.e.* traits biologiques) sont également susceptibles d'influencer les performances des modèles. Par exemple, Collart *et al.* (2023) ont montré que les traits d'histoire de vie caractérisant 411 espèces de bryophytes peuvent influencer l'*accuracy* de SDM (*accuracy* évaluée au moyen de l'AUC, du Kappa et du TSS max, de l'indice de Boyce et de la *specificity* et *sensitivity*). Il semble donc que d'autres approches de modélisation soient nécessaires pour modéliser la distribution spatiale des abondances des espèces rares (Jeliaskov *et al.*, 2022 ; Collart & Guisan, 2023).

C.2. Interprétation de la relation entre les classes d'abondance et l'environnement

Les méthodes d'apprentissage par ML ont longtemps été critiquées pour leur côté « boîte noire » et pour les difficultés liées à l'interprétation des relations organismes-environnement, qui sont plus aisément mises en évidence par d'autres catégories de modèles (*e.g.* modèle mécanistes ou statistiques purs). Actuellement, cette affirmation n'est plus vraie, car il existe un panel de méthodes pour interpréter les modèles issus d'approches de ML (Olden & Jackson, 2002 ; Molnar, 2022), certains auteurs soutiennent qu'elles permettraient de montrer des relations de causalité comme avec des méthodes de régression plus classiques (Elith *et al.*, 2008 ; Zhao & Hastie, 2021). Ici, l'interprétation des modèles OF est réalisée par l'analyse de l'ordre d'importance des prédicteurs (nommé ensuite rang) et de l'influence des prédicteurs majeurs sur les abondances au moyen de *Partial Dependence Plots* (PDP).

C.2.1. Ordre d'importance des prédicteurs

L'ordre d'importance des prédicteurs, ou rang des prédicteurs, dans les modèles a été évalué de la manière suivante. Pour chaque modèle, l'importance de chaque prédicteur telle que définie dans le package *OrdinalForest* a été utilisée pour déterminer le rang de chacun des dix prédicteurs dans le modèle (Hornung, 2020, 2022). Ce rang retranscrit à quel niveau le prédicteur intervient dans la structuration des branches de l'OF : plus le prédicteur est important plus il intervient tôt et donc plus son rang est faible (Figures 46 et 47).

Les taxons pour lesquels les modèles sont les moins performants montrent de plus grands écarts-types, ce qui témoigne d'une plus forte instabilité du rang des prédicteurs et du modèle en général entre réplicats. Pour les modèles les plus performants en revanche,

les prédictors les plus importants, en général les trois premiers, sont stables entre les réplicats de modèles. Dans tous les cas, le type et l'ordre des prédictors diffèrent en fonction du taxon, témoignant de différences dans la réponse à l'environnement en fonction de l'écologie de chaque taxon. Les déterminants environnementaux les plus importants (*i.e.* ceux le plus fréquemment classés aux trois premiers) dans la distribution spatiale des abondances de poissons et de macroinvertébrés sont la topographie (pente), les indicateurs des régimes thermiques et hydrologiques moyens (Tw_m, v_m) ainsi que le taux d'étagement (tx_etage) et la probabilité d'intermittence (P_inter). Parmi les prédictors considérés, la probabilité d'intermittence et le taux d'étagement ont rarement été considérés dans des études ayant réalisées des SDM, et les résultats présentés ici montrent qu'ils apportent en explicativité des patrons spatiaux de plusieurs espèces. La probabilité d'intermittence du cours d'eau est plus importante dans les modèles macroinvertébrés que poissons (cette variable est retrouvée dans les premiers rangs des prédictors pour 9 genres de macroinvertébrés contre 1 espèce de poisson, le poisson chat). Le taux d'étagement est retrouvé dans les premiers rangs des modèles de 6 espèces de poissons et 10 genres de macroinvertébrés. À notre connaissance, l'intégration d'indicateur de la continuité, plus souvent au moyen du ROE (*e.g.* densité d'ouvrages), montre difficilement un lien avec la distribution des espèces non migratrices (*e.g.* Maire *et al.* 2017). L'importance attribuée au taux d'étagement par les modèles présentés ici montrent bien la pertinence d'utiliser des métriques qui retranscrivent une information portant un sens écologique vis-à-vis des capacités de franchissement des espèces (Vierron & Steinbach, 2020). Quant à elle, la densité d'ouvrage a rarement été sélectionnée parmi les déterminants principaux de la distribution spatiale des abondances des taxons modélisés. Néanmoins, les informations précises sur les ouvrages (*e.g.* hauteur de chute), nécessaires au calcul de ce type d'indicateur, ne sont pas toujours disponibles (Vierron & Steinbach, 2020). D'autre part, l'intégration de paramètres supplémentaires, comme l'existence de passes à poisson, serait susceptible d'accroître le pouvoir explicatif de ces indicateurs, en affinant la représentativité des difficultés de franchissement sur chaque tronçon du réseau hydrographique considéré (Vierron & Steinbach, 2020).



FIGURE 46

Rang des prédicteurs dans les modèles d'abondances. Les valeurs présentées sont la moyenne du rang des prédicteurs dans les modèles d'abondance pour les modèles des espèces de poissons (sur 20 réplicats), ainsi que l'écart-type. Plus ce rang est bas, plus le prédicteur a d'importance dans le modèle. La correspondance avec les noms des prédicteurs peut être consultée au tableau 9.

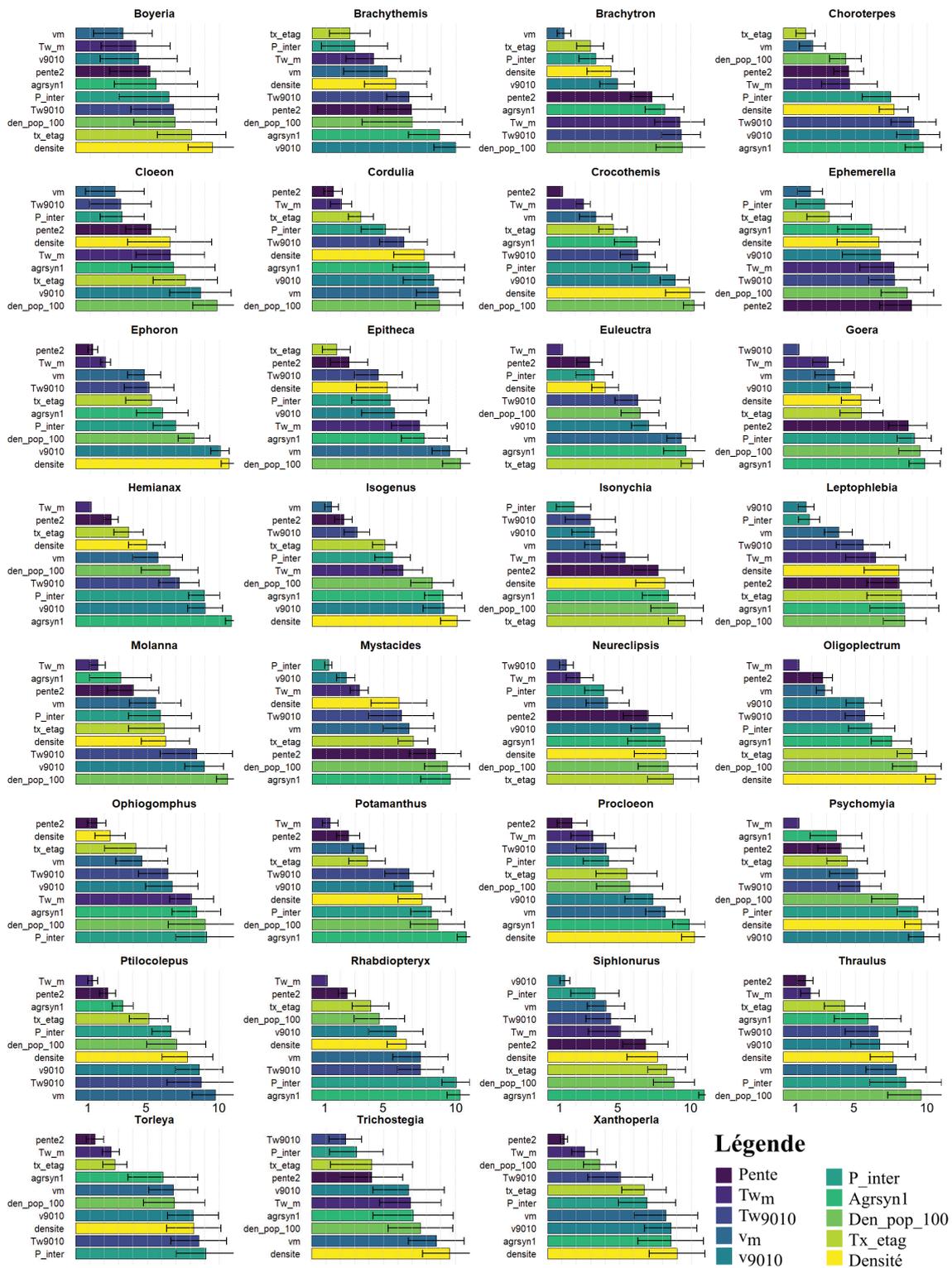


FIGURE 47

Rang des prédicteurs dans les modèles d'abondances. Les valeurs présentées sont la moyenne du rang des prédicteurs dans les modèles d'abondance pour les modèles des genres de macroinvertébrés (sur 20 réplicats), ainsi que l'écart-type. Plus ce rang est bas, plus le prédicteur a d'importance dans le modèle. La correspondance avec les noms des prédicteurs peut être consultée au tableau 9.

C.2.2. Influence des valeurs des prédicteurs sur les classes prédites

L'influence des prédicteurs environnementaux sur les classes d'abondance peut être illustrée au moyen de PDP (Molnar, 2022). Cette méthode permet d'évaluer l'influence d'un prédicteur donné sur la variable réponse, après avoir pris en compte les effets moyens de tous les autres prédicteurs. Les PDP posent donc l'hypothèse d'indépendance des prédicteurs entre eux, ce qui ne leur permet pas de représenter l'effet de variables en interactions fortes. Les PDP permettent au maximum d'illustrer l'effet de deux prédicteurs simultanément (limite induit par la 2D). Les PDP constituent néanmoins une base utile à l'interprétation des modèles de ML.

La fonction dépendance \hat{f}_s est estimée en calculant les prédictions moyennes à partir des données d'entraînement, selon la méthode de Monte Carlo :

$$\hat{f}_s = \frac{1}{n} \sum_{i=1}^n \hat{f}_s(x_s, x_c^{(i)})$$

où n est le nombre d'observations, x_s les valeurs du prédicteur pour lequel on souhaite calculer la dépendance partielle et x_c les valeurs des autres prédicteurs. La fonction partielle nous donne ainsi l'effet marginal moyen du prédicteur S , pour une valeur donnée, sur les prédictions. Dans le cas de modèles multi-classes, cela se traduit par la probabilité d'appartenir à une certaine classe pour les valeurs du prédicteur S .

Les PDP évaluent l'influence des prédicteurs en posant l'hypothèse d'indépendance entre eux (Molnar, 2022). Donc, un PDP "plat" n'indique pas l'absence d'influence du prédicteur sur les classes d'abondance, seulement l'absence apparente d'effet marginal de ce prédicteur. De plus, il convient d'identifier les gammes de valeurs observées pour les prédicteurs, afin de ne pas interpréter les gammes de valeurs pour lesquelles un faible nombre de valeurs est disponible.

Les réponses détectées par les PDP présentent des formes diverses (Figure 48). L'influence des prédicteurs sur la probabilité d'appartenir à la classe 0 (**absence**) est similaire aux courbes réponses qui sont présentées dans les études de la distribution spatiale d'espèces basées sur des modèles binaires de présence/absence. Les probabilités d'appartenir aux classes > 0 apportent un niveau d'information supplémentaire sur les niveaux d'abondance potentiels en fonction des valeurs des prédicteurs environnementaux. Ainsi, les classes d'abondance de la truite et de la perche répondent presque inversement à la tem-

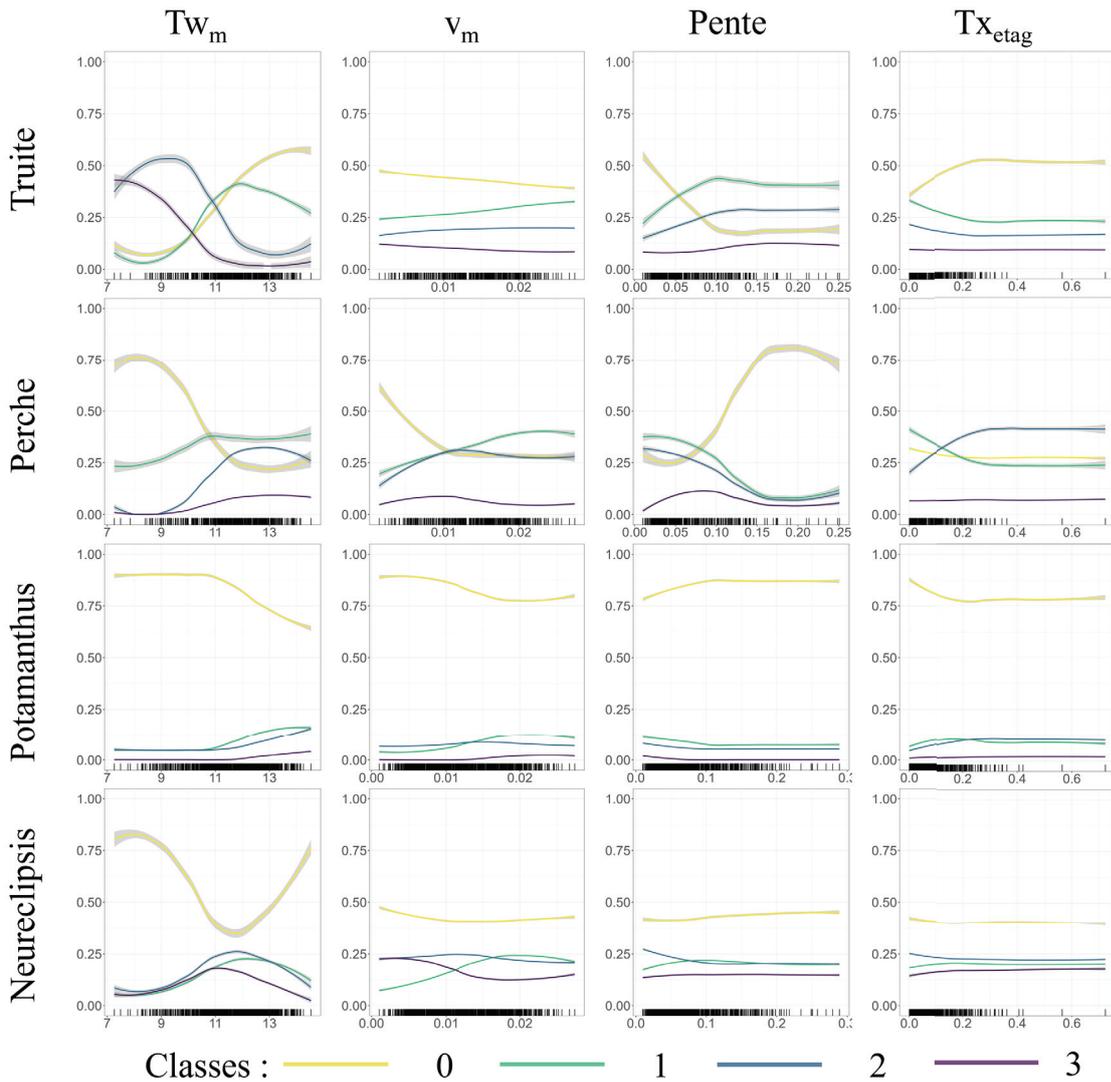


FIGURE 48

Partial Dependance Plot des quatre prédicteurs majeurs (Tw_m : température de l'eau moyenne, v_m : vitesse de courant moyenne, pente et tx_{etag} : taux d'étagement) pour deux modèles poissons (truite commune et perche fluviatile) et les deux modèles macroinvertébrés (Potamanthus et Neureclipsis). Les probabilités d'appartenir à chacune des classes d'abondance sont issues des 20 répliqués de modèles et ont été lissées par une méthode de *loess*. Les marques le long des abscisses de chacun des graphes représentent les valeurs observées pour les prédicteurs.

pérature moyenne de l'eau, avec un effet seuil qui semble se situer autour de 11 °C. Elles sont également sensibles aux trois autres prédicteurs représentés ici, qui retranscrivent plutôt le positionnement sur le gradient amont-aval ; en tête de bassin pour la truite avec de plus fortes vitesses de courant, pentes et de plus faibles taux d'étagement, et inversement pour la perche.

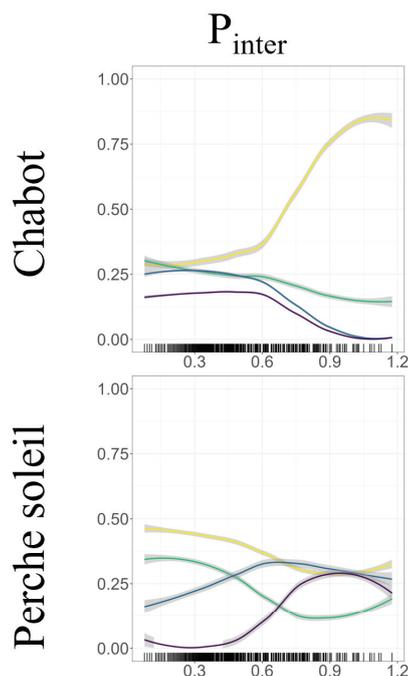


FIGURE 49
Exemple de *Partial Dependence Plot* de P_{inter} . Les codes couleurs sont les mêmes qu'en Figure 48

déterminent, d'une part, les tronçons de présence et d'absence des macroinvertébrés, et d'autre part, les niveaux d'abondance locale des macroinvertébrés (Chesson, 2000 ; Pulliam, 2000). Enfin, les deux compartiments biologiques ne semblent pas se structurer aux mêmes échelles spatiales probablement parce que les processus plus locaux (inférieure à l'échelle du bassin) ont une plus grande influence dans la distribution des macroinvertébrés.

À l'exception du taux d'étagement, les autres prédicteurs des pressions anthropiques (*i.e.* densité de population, gradient agricole et densité d'ouvrage) ne semblent pas être aussi influents que la thermie ou l'hydrologie. Des relations sont quand même observées, notamment sur la classe absence, en réponse au gradient agricole, à la densité d'ouvrage et à la densité de population aux abords des cours d'eau (Annexe E). Cela illustre bien que les taxons se répartissent sur le réseau hydrographique en fonction de leur préférence écologique, et que leur distribution « naturelle » est perturbée par les pressions présentes sur les tronçons.

Bien que le jeu de données macroinvertébrés soit deux fois plus important que celui des poissons (1 077 tronçons de rivière pour les macroinvertébrés contre 482 tronçons

Le même type de réponse est observé chez les autres espèces de poissons, et dépend de leur préferendum écologique pour chacun des prédicteurs. Pour certaines espèces (*Epitheca*, *Isonychia*, van-doise, *Neureclipsis*, *Trichostegia*, chabot et *Leptophlebia*), une réponse marquée à la probabilité d'intermittence des cours d'eau est observée (exemples de réponses en Figure 49). Globalement, l'ensemble des macroinvertébrés montre des réponses moins marquées pour les classes d'abondance > 0, hormis pour les prédicteurs de la thermie. Cela peut indiquer que les effets d'interaction entre prédicteurs sont plus importants chez les macroinvertébrés que chez les poissons. En effet, les PDP ne permettent pas d'illustrer ces interactions entre prédicteurs. Il est aussi très probable que des processus distincts

pour les poissons), les valeurs observées pour les prédicteurs environnementaux sont dans les mêmes gammes (illustrées par les marques au bas de chaque PDP ; voir aussi Tableau 9). Les modèles des deux compartiments biologiques sont donc calibrés dans les mêmes gammes environnementales et leurs prédictions peuvent être considérées ensemble pour retranscrire la composition prédite des communautés de chaque tronçon du réseau H.

D. Intégration des interactions biotiques dans la modélisation de la distribution des abondances

Les interactions biotiques regroupent l'ensemble des relations qui peuvent exister entre les espèces au sein d'une communauté. Elles sont de plusieurs types : elles peuvent profiter aux espèces concernées (*e.g.* mutualisme), seulement à certaines d'entre elles (*e.g.* prédation) ou toutes les desservir (*e.g.* compétition) (Jackson *et al.*, 2001). Ces interactions contribuent à façonner la distribution spatiale des espèces dans l'espace géographique, en restreignant leur niche fondamentale qui est formée par toutes les conditions abiotiques favorables, à leur niche réalisée (Pulliam, 2000 ; Soberón, 2007). Elles ont donc un rôle fondamental dans l'explication des patrons de distribution spatiale des espèces et leur intégration dans les SDM peut permettre d'en accroître les performances (voir par exemple Boulangeat *et al.* 2012 ; Pletterbauer *et al.* 2016 ; Gerhardt Braz *et al.* 2020 ; Lucas *et al.* 2023 ; Woods *et al.* 2023). Bien qu'il soit reconnu comme important d'intégrer ces interactions dans les SDM, comme en atteste la multiplication récente des exemples dans la littérature, il n'existe pas de cadre précis quant à la manière de les intégrer (Anderson, 2017). D'autre part, la contribution relative des interactions biotiques par rapport aux facteurs abiotiques et à la capacité de dispersion des espèces reste inconnue (Galiana *et al.*, 2023), ainsi que l'échelle spatiale à laquelle intervient chacune (Soberon & Nakamura, 2009). Enfin, les interactions biotiques sont susceptibles d'évoluer au cours du temps, en fonction des changements de distribution des espèces qui peuvent impliquer la perte ou le gain de certaines espèces dans les communautés et ainsi la perte ou le gain de certaines interactions. Dans le cadre du changement climatique, la co-occurrence des espèces en interaction est susceptible d'évoluer, et d'impacter les distributions spatiales futures (Lucas *et al.*, 2023), d'où l'importance de mieux comprendre ces processus.

Les tentatives d'intégration des interactions biotiques dans des SDM trouvées dans la littérature concernent majoritairement des interactions trophiques, c'est-à-dire les re-

lations alimentaires qui lient les espèces (proie - prédateur, ressources disponibles ; Pletterbauer *et al.* 2016 ; Lucas *et al.* 2023 ; Woods *et al.* 2023). Les exemples qui intègrent les effets de la compétition intra ou inter-spécifique sont plus rares car cette interaction est plus complexe à mettre en évidence (Gerhardt Braz *et al.*, 2020 ; Stephenson *et al.*, 2022). C'est pourquoi l'étude de l'influence de la compétition sur la distribution spatiale des espèces ne concerne généralement que quelques espèces particulières. Enfin, chez les poissons, il semblerait que ce soient les relations trophiques qui soient les plus structurantes (Matthews, 1998 ; Jackson *et al.*, 2001), ce qui justifie de commencer par intégrer un indicateur de leur ressource en première approche.

Les modèles développés précédemment dans ce chapitre montrent qu'une grande part de la distribution spatiale des abondances peut être expliquée au moyen de variables abiotiques représentatives des conditions hydroclimatiques, du gradient amont-aval et des pressions anthropiques. L'objectif de cette section est d'évaluer l'apport des macroinvertébrés en tant que ressource trophique dans les modèles de distribution de classes d'abondance des espèces de poissons.

D.1. Méthode d'intégration des relations trophiques dans les modèles de distribution

D.1.1. Sélection du jeu de données

Une sous-sélection du jeu de données utilisé pour la modélisation décrite précédemment a été réalisée afin de ne conserver que les tronçons portant à la fois au moins un échantillonnage sur les deux compartiments poissons et macroinvertébrés, soit 408 tronçons. De nouveaux modèles de la distribution des classes d'abondance, basés uniquement sur les prédictors abiotiques présentés précédemment, ont été réalisés sur les données de ces 408 tronçons. Ces modèles, dits naïfs, n'intègrent pas de paramètre biotique et seront considérés comme la référence pour évaluer les modèles intégrant des prédictors d'interaction biotique. Une seconde modélisation intégrant des indicateurs de la quantité et de la diversité de la ressource macroinvertébrés disponible ont été réalisées. Les espèces de poissons à modéliser ont été restreintes aux espèces ayant un régime alimentaire basé sur les macroinvertébrés, au moins en partie et à une étape de leur cycle de vie (tableau 11). L'écologie alimentaire des poissons est complexe car elle varie en fonction de nombreux facteurs pour une même population (*e.g.* ontogénie, sexe, localisation, saison ; Futia *et al.*

2021 ; Ribeiro *et al.* 2021) voire un même individu (stade de vie) et elle est peu connue pour certaines espèces (Gerking, 1994 ; Braga *et al.*, 2012). C'est pourquoi la liste des espèces retenues est assez large.

TABLEAU 11

Régime trophique des espèces de poissons retenues pour la construction de SDM intégrant leur ressource trophique en prédicteur. D'après le site freshwaterecology.org.

Espèces	Régime trophique
Ablette	Omnivore
Barbeau fluviatile	Omnivore
Bouvière	Omnivore
Brème	Omnivore
Chabot fluviatile	Invertivore
Chevaine	Omnivore
Gardon	Omnivore
Goujon	Invertivore
Grémille	Invertivore
Loche franche	Invertivore
Poisson chat	Omnivore
Perche	Invertivore
Perche soleil	Omnivore
Spirilin	Invertivore
Tanche	Omnivore
Truite commune	Invertivore
Vairon ligérien	Invertivore
Vandoise	Omnivore

D.1.2. Calcul des indices de ressource macroinvertébrés

Afin d'approcher la quantité de macroinvertébrés qui constitue la ressource alimentaire des poissons à l'échelle du tronçon de rivière, les différents taxons de macroinvertébrés à disposition (n = 188 genres) ont été regroupés en fonction de leur susceptibilité à la prédation (Floury *et al.*, 2019). L'hypothèse est que plus un groupe est sujet à la prédation, plus il a de chance d'être consommé par les poissons et donc d'avoir de l'influence sur le niveau d'abondance des espèces de poissons. Pour tenter d'approcher cette susceptibilité à la prédation, sept traits autécologiques (Usseglio-Polatera *et al.*, 2000) ont été considérés : taille potentielle (4 modalités), mobilité (5 modalités), substrat (3 modalités), préférence vitesse de courant (4 modalités), distribution transversale (4 modalités), respiration (4 mo-

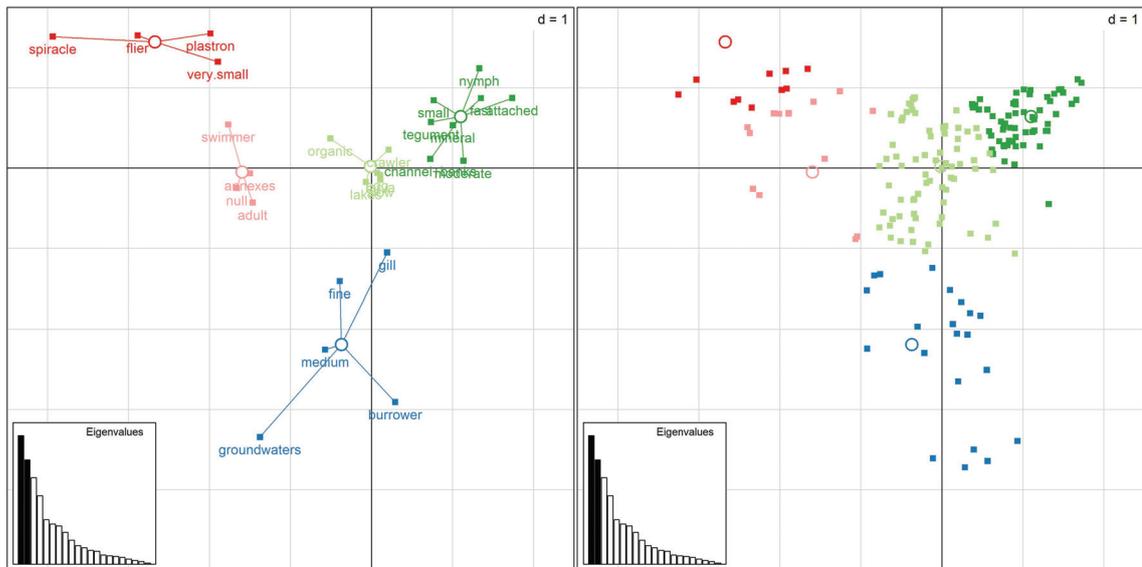


FIGURE 50

Fuzzy Coding Analysis (FCA) à 6 groupes sur les 7 traits considérés pour définir les groupes de macroinvertébrés en fonction de leur susceptibilité à la prédation. Les modalités sont représentées sur le graphe de gauche et les taxons sur le graphe de droite.

dalités) et stade aquatique (4 modalités). Six groupes de susceptibilité à la prédation ont été définis par classification hiérarchique sur la base d'une *Fuzzy Coding Analysis* (FCA) (Chevenet *et al.*, 1994), à partir des 78 traits de 188 genres de macroinvertébrés (Figure 50). Le groupe bleu, situé en bas, est considéré peu sensible à la prédation car il est composé d'organismes très liés au substrat voire enfouis. Les groupes rouge et vert sont plus sujets à la prédation car composés de taxons plus petits, nageant en pleine eau. Ils sont subdivisés en deux groupes chacun, les groupes foncés étant encore plus exposés. Ce sont ces groupes rouge foncé et vert foncé qui ont été utilisés pour identifier les taxons les plus vulnérables à la prédation. Ils ont donc été nommés groupes « vulnérables lentiques » et « vulnérables lotiques » (resp. rouge foncé et vert foncé) du fait du type d'habitat préférentiel des genres qui les composent.

Les abondances cumulées des genres appartenant aux groupes « vulnérables lentiques » et « vulnérables lotiques » ont été utilisées comme indicateurs de la ressource macroinvertébrés prioritairement consommée par les poissons, et donc comme approximation de la ressource trophique « macroinvertébrés » dans les modèles de distribution de classes d'abondance des espèces de poisson. En complément, l'indice d'équitabilité de Piélou a été calculé sur la base des 6 groupes comme proxy de la diversité de la ressource trophique disponible. L'absence de corrélation forte entre les différents prédicteurs, qu'ils soient indicateurs de la ressource macroinvertébré ou prédicteurs environnementaux, a été vérifiée

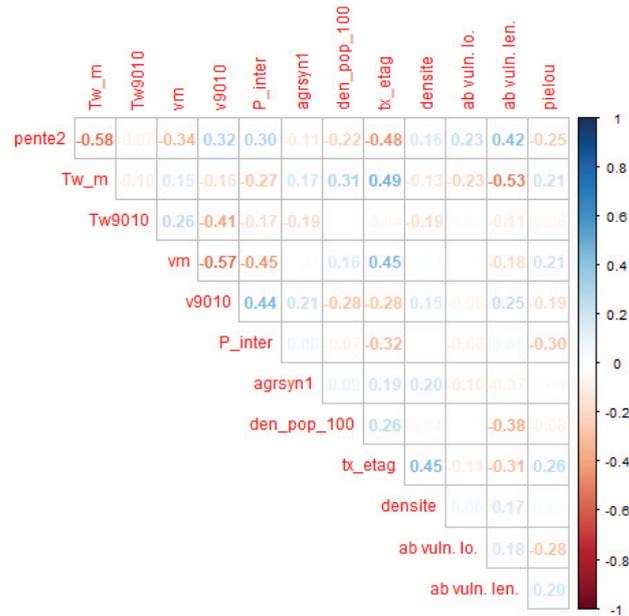


FIGURE 51

Corrélogramme des prédicteurs utilisés dans les SDM avec intégration de la ressource macroinvertébrés. Les valeurs présentées sont les coefficients de corrélation de Spearman (ρ). Les prédicteurs de la ressource macroinvertébrés sont annotés comme suit : abondance du groupe vulnérable lentique = « ab vuln. len. », abondance du groupe vulnérable lotique = « ab vuln. lo. », indice de Piélou = « pielou » ; le reste des correspondances avec les codes prédicteurs peut être consultée au tableau 9.

(coefficient de corrélation de spearman $\rho < 0,7$) avant leur intégration dans les modèles (Figure 51).

Ainsi les modèles suivants ont été construits pour 18 espèces de poissons : *Classes d'abondance* \sim *Pente* + *Tw_m* + *Tw₉₀₁₀* + *v_m* + *v₉₀₁₀* + *densité de population* + *agsyn* + *densité d'ouvrages* + *taux d'étagement* + *probabilité d'intermittence* + *abondance groupe vulnérable lentique* + *abondance groupe vulnérable lotique* + *Piélou*

D.2. Évaluation de l'apport des relations trophiques à la modélisation de la distribution des classes d'abondance

Les modèles intégrant un indicateur du niveau de ressource macroinvertébrés ont été évalués de la même manière que précédemment : calcul de l'AUC multi-classes, du score de Brier et interprétation des relations trouvées par les modèles au moyen de l'ordre d'importance des prédicteurs et de PDP. Les modèles intégrant les indicateurs de ressource macroinvertébrés ont été comparés aux modèles ne les intégrant pas (dits modèles naïfs).

D'après la figure 52, l'intégration d'indicateurs de ressources macroinvertébrés ne permet pas une amélioration notable des performances (basées sur l'AUC et le score de Brier) des modèles de classes d'abondance des 18 espèces de poissons, hormis pour le chabot (gain d'AUC de 0,07).

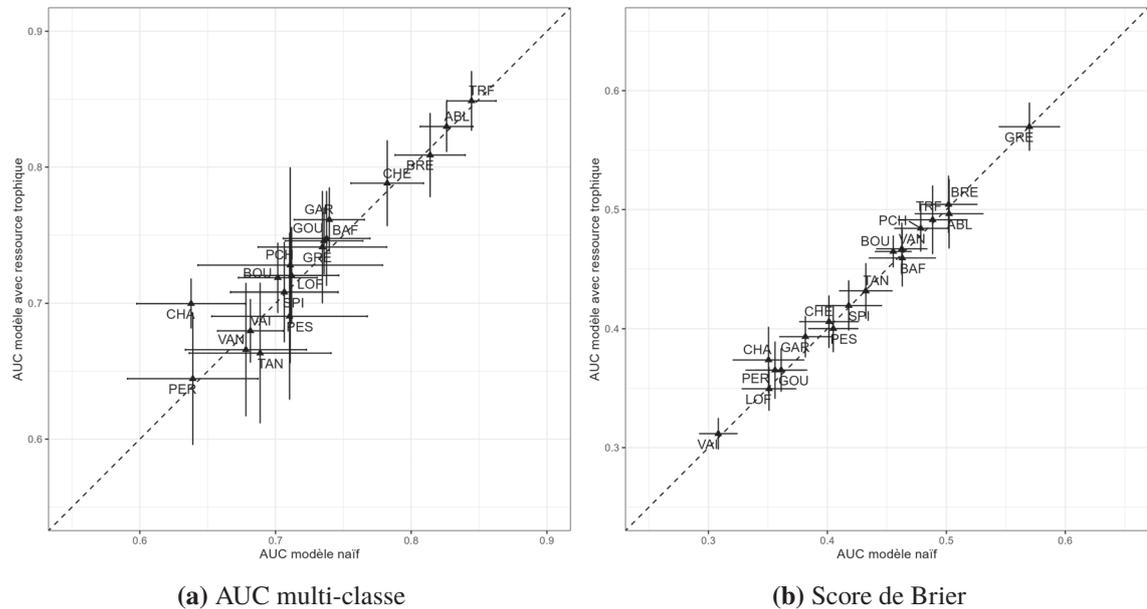


FIGURE 52

Comparaison des performances des modèles (AUC multi-classe et score de Brier) avant et après intégration des indicateurs de ressource macroinvertébrés. Les valeurs présentées sont la moyenne des métriques sur les 20 réplicats, et les écarts-type associés. La correspondance avec les codes espèces poissons peut être consultée au tableau 1.

Comme précédemment (section C.2.2), l'ordre d'importance des prédicteurs dans les modèles poissons (Figure 53) et quelques PDP (Figure 54) sont présentés afin d'interpréter les modèles OF.

Bien que l'ajout des indicateurs de ressource macroinvertébrés n'augmente pas les performances des modèles poissons, ils peuvent « supplanter » certains prédicteurs abiotiques s'ils sont plus influents dans les modèles. Globalement, l'ordre d'importance des prédicteurs des modèles naïfs sont retrouvés globalement aux mêmes rangs que dans les modèles précédents basés sur les données de 484 tronçons (section C.2; Annexe F), avec une légère variation certainement liée à la restriction du jeu de données aux tronçons présentant conjointement des échantillonnages poissons et macroinvertébrés (N = 408 tronçons; Figure 53). Concernant les modèles avec intégration de la ressource macroinvertébrés, l'indice de Piélou n'est retrouvé dans les premiers rangs que pour le SDM du chabot. L'indicateur d'abondance des taxons lotiques était systématiquement retrouvé

aux derniers rangs (*i.e.* intervention tardive dans la structuration des branches de l'OF), ce qui semble indiquer qu'il n'aurait pas ou peu d'influence dans les modèles. L'indicateur d'abondance des taxons lenticques en revanche intervient dans les premiers rangs des modèles ablette, barbeau, bouvière, chabot, gardon et viron, soit 6 espèces sur les 18 modélisées. Il se retrouve même au premier rang pour chacun des 20 réplicats du modèle chabot (d'où l'absence de barre d'erreur sur la figure 53).

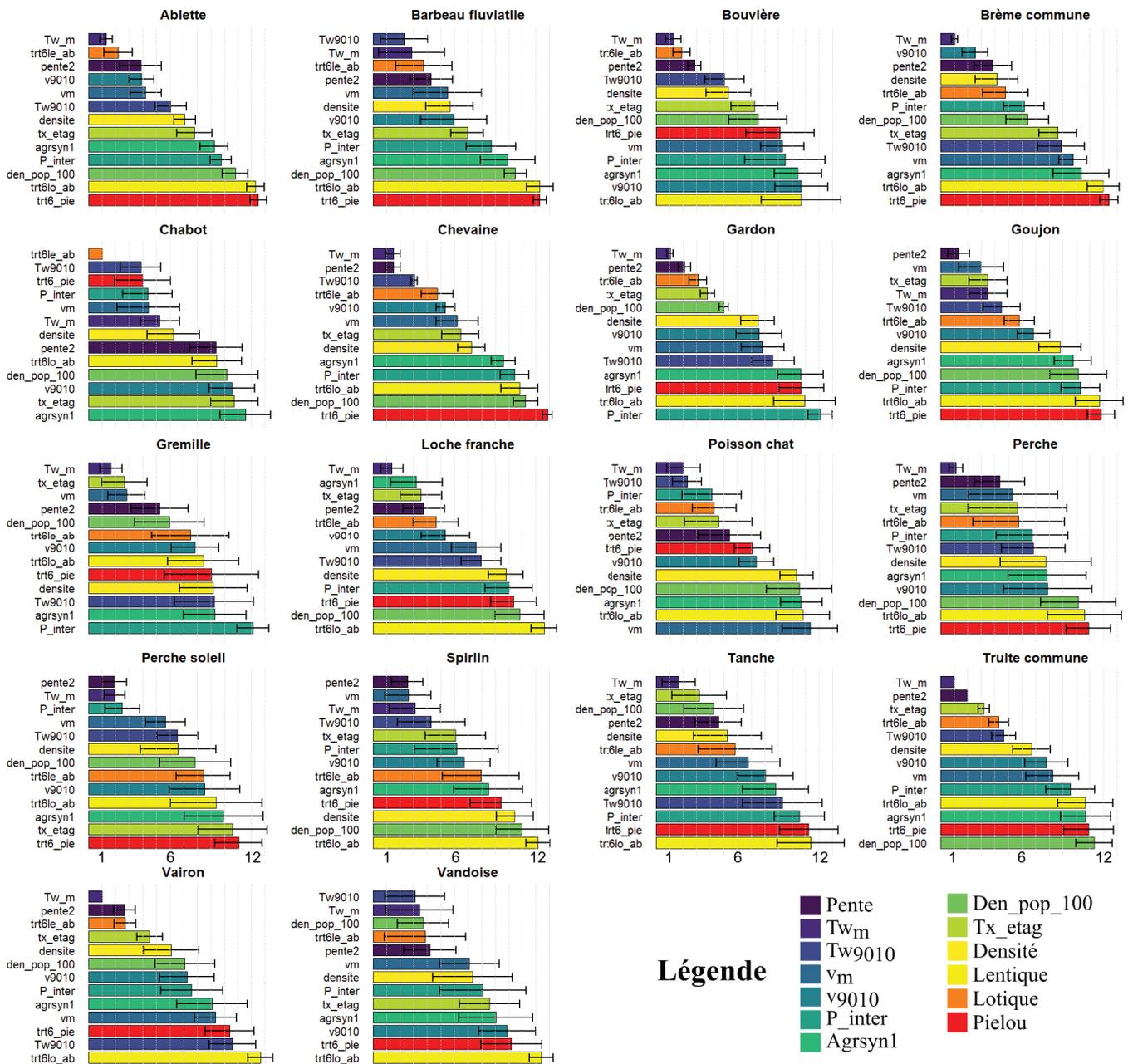


FIGURE 53

Rang des prédicteurs dans les modèles poissons après intégration des indicateurs de ressource macroinvertébrés. Les valeurs présentées sont la moyenne du rang des prédicteurs dans les modèles d'abondance (sur 20 réplicats). Les prédicteurs liés à la ressource macroinvertébrés sont annotés comme suit : abondance du groupe vulnérable lentique = « trt6len_ab », abondance du groupe vulnérable lotique = « trt6lo_ab », indice de Pielou = « trt6_pie » ; le reste des correspondances avec les noms des prédicteurs peut être consultée au tableau 9

Les PDP pour les espèces pour lesquelles au moins l'un des indicateurs de la ressource macroinvertébrés semble être important sont présentés en figure 54.

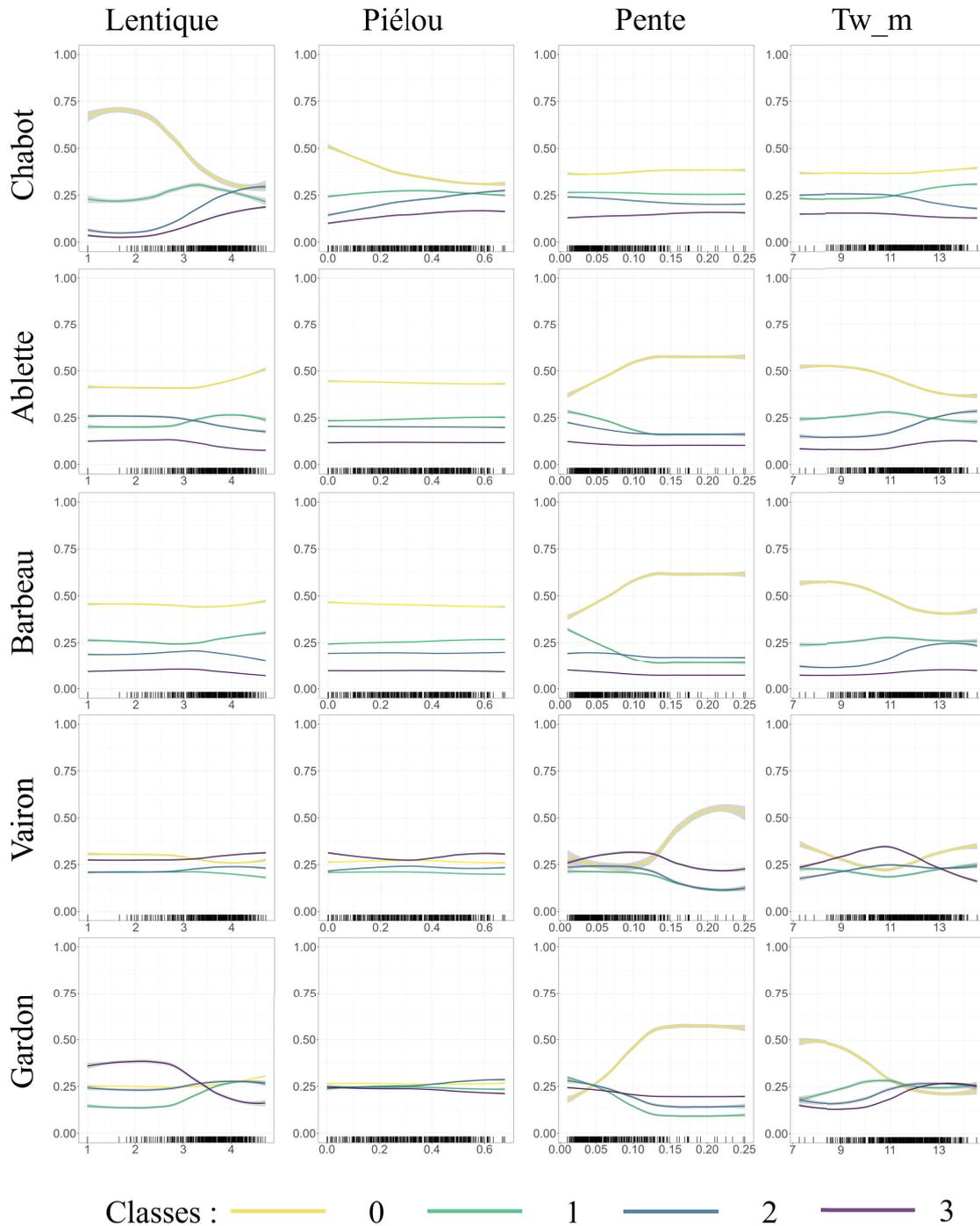


FIGURE 54

Partial Dependence Plot des indicateurs de ressource macroinvertébrés dans les modèles poissons du chabot, barbeau fluviatile, vairon et de l'ablette.

Les PDP des indicateurs de ressources macroinvertébrés ne montrent pas un effet marginal aussi marqué que les prédicteurs environnementaux vus précédemment (Figure 54).

Les classes d'abondance non nulles du chabot montrent une relation positive avec l'indicateur de ressource macroinvertébrés lentières et l'indice de Pielou. En revanche, les réponses observées dans les modèles précédents vis-à-vis des prédictors abiotiques n'apparaissent plus dans ce modèle. Soit l'indicateur de ressource trophique porte une information qui suffit à expliquer la distribution des classes d'abondance d'après le modèle, soit cet indicateur est en interaction avec les autres prédictors, ce qui ne peut pas être détecté au moyen de PDP. Concernant les autres espèces, les PDP ne permettent pas de mettre en évidence une relation claire entre les classes d'abondance et les indicateurs de ressource macroinvertébrés utilisés ici.

La méthode proposée pour l'intégration d'indicateurs de la ressource en macroinvertébrés susceptibles à la prédation dans 18 modèles poisson montre donc des résultats mitigés. D'après les rangs moyens, les indicateurs semblent avoir une certaine importance dans quelques modèles bien que cela n'apparaisse pas aussi nettement dans les PDP. L'intégration de prédictors basés sur le régime trophique de l'espèce modélisée peut donner des résultats de prédictions bien différents, notamment pour les prédictions sous scénarios futurs, dans un contexte de changement climatique (Ohlmann *et al.*, 2023 ; Lucas *et al.*, 2023). D'autres types d'interactions biotiques peuvent également influencer la distribution spatiale des abondances des espèces, comme la compétition interspécifique (Boulangéat *et al.*, 2012). Ces articles ont réalisé des analyses préliminaires bien plus poussées sur les relations inter et intra-spécifiques pour les taxons d'intérêt avant de procéder à la modélisation par SDM. Cependant les connaissances en matière de régime trophique des poissons (Gerking, 1994), mais aussi d'interactions biologiques de manière plus générale, sont encore limitées (Jackson *et al.*, 2001). La revue sur les connaissances de l'écologie trophique des poissons de Braga *et al.* (2012) a montré que, même si le nombre d'espèces à étudier est plus faible en Europe, des lacunes restent à combler. Floury *et al.* (2019) ont, par exemple, mis en évidence le caractère opportuniste du comportement alimentaire de plusieurs espèces sur l'Allier, affluent principal de la Loire, (*e.g.* juvéniles de saumon, loche, barbeau). Ce caractère opportuniste complexifie l'établissement d'un réseau trophique clair pour les espèces de poissons et ce n'est peut-être pas la bonne solution. Jackson *et al.* (2001) ont mis en évidence l'importance de la piscivorie dans la structuration des communautés de poissons et montrent aussi le rôle prépondérant de l'échelle spatiale dans la capacité des

interactions biotiques à expliquer les patrons spatiaux des communautés. La compétition spécifique semble ainsi avoir un rôle majeur à échelle locale, et son influence se réduit par rapport à celle des facteurs abiotiques à mesure qu'on augmente l'échelle spatiale, en accord avec l'hypothèse de bruit Eltonien. Pletterbauer *et al.* (2016) ont montré une amélioration de leur SDM de présence-absence sur l'ombre commun (*Thymallus thymallus*) avec l'intégration des occurrences d'un Trichoptère (*Allogamus auricollis*). Leur unité géographique est constituée de tronçons de longueur moyenne $3,6 \pm 2$ km compris dans un bassin versant de plus 10 km² de surface. Ceci suggère que le grain spatial plus fin des tronçons que nous avons considérés ($1,8 \pm 1,63$ km) pourrait être suffisant pour mettre en évidence le rôle du niveau de ressource trophique macroinvertébrés dans la distribution des espèces de poissons.

En résumé ...

- ★ Le couplage des méthodes *Synthetic Minority Oversampling TEchnique and Cluster-based Undersampling Technique* (SCUT) permet de limiter l'influence de l'asymétrie sur les modèles, et cela réduit notamment le phénomène de sur-apprentissage des modèles sur la classe majoritaire. En accord avec la littérature, l'application de la méthode SCUT permet d'augmenter les performances prédictives des modèles ;
- ★ La modélisation de la distribution des classes d'abondance par *ordinal forest* est performante sur la plupart des taxons de poissons et de macroinvertébrés modélisés ($n = 26$ et $n = 31$, respectivement ; 38 modèles sur 57 avec une AUC multi-classes $> 0,65$; 42 modèles sur 57 avec un score de Brier $< 0,6$) ;
- ★ Les principaux déterminants environnementaux de la distribution spatiale des abondances des taxons étudiés sont la température de l'eau, les variables hydrauliques, la pente, et les pressions sur la continuité écologique. Le taux d'étagement, qui traduit la hauteur supplémentaire induite par les ouvrages le long des cours d'eau, ressort comme particulièrement influent pour plusieurs taxons ;
- ★ La modélisation des classes d'abondance couplée à des méthodes d'interprétation adaptées aux modèles de *machine learning* permet d'apporter des informations plus fines quant aux facteurs explicatifs de la distribution spatiale des abondances des taxons étudiés ;
- ★ La tentative d'intégration d'indicateurs de la quantité de ressource de macroinvertébrés susceptibles à la prédation ne semble pas améliorer les prédictions des modèles pour les 18 espèces de poissons testées. Néanmoins, l'échelle spatiale de l'étude semble adaptée, et l'importance de certains prédicteurs dans les SDM appelle à poursuivre l'intégration des interactions biotiques dans la modélisation de la distribution des abondances.

Chapitre V

**Projections futures des
communautés aquatiques
ligériennes sous scénario de
changement climatique et
conséquences sur leur
structuration spatiale**

A. Introduction

Le changement climatique induit des modifications de la distribution spatiale des espèces aquatiques qui se répercutent sur l'organisation des communautés (voir chapitre I). Les objectifs de ce chapitre sont d'apporter (i) une évaluation des changements potentiellement induits par le changement climatique sur la distribution future des abondances des espèces, et (ii) leurs conséquences en termes de réorganisation des communautés de poissons et de macroinvertébrés. Les modèles de distribution des classes d'abondance développés au chapitre IV ont été utilisés pour projeter les distributions potentielles futures des 57 taxons modélisés (26 espèces de poissons et 31 genres de macroinvertébrés) sous un scénario de changement climatique (RCP 8.5) aux horizons temporels 2050 et 2080.

Les capacités de projection des SDM dans des conditions non incluses dans les données de calibration sont souvent limitées, mais la construction de projections de l'évolution des communautés est nécessaire (Thuiller, 2007). Pour anticiper les effets du changement climatique, par l'adaptation des activités anthropiques ou encore la mise en place de mesures de conservation (Thuiller, 2007 ; Peterson *et al.*, 2011). L'utilisation de modèles consensuels (*i.e.* modèles d'ensemble) a été proposée pour pallier cette limite (Guisan *et al.*, 2017), mais en l'absence de données futures (vérité terrain) il n'est pas possible de déterminer quel est le meilleur modèle pour projeter les distributions des espèces sous scénarios de changement climatique (Elith *et al.*, 2010). Certaines chroniques de suivis long-terme ont permis de confronter les tendances observées à celles prédites par des modèles. Par exemple, Comte *et al.* (2013) ont montré que les changements observés dans les communautés de poissons du territoire français métropolitain sont plus importants que ceux prédits, vraisemblablement du fait de l'intervention de facteurs de pressions non intégrés lors de la modélisation. A contrario, certaines espèces dont les modèles prévoient l'extinction à un horizon donné peuvent persister au-delà de ce dernier dans l'écosystème (*dead species* ; Kuczynski *et al.*, 2023). Mais l'indisponibilité de données futures empêche d'évaluer les capacités d'extrapolation d'un modèle dans la plupart des cas. Des alternatives ont été proposées par le biais (i) d'une évaluation de la capacité d'extrapolation spatiale (ii) ou un partitionnement du jeu de données. Valavi *et al.* (2023) ont ainsi démontré les capacités d'extrapolation supérieures de méthodes de *machine learning*, considérées plus flexibles pour la projection en-dehors de la zone géographique de calibration du modèle (cas a).

La seconde méthode consiste à partitionner les données d'entraînement en fonction d'une variable amenée à évoluer dans le futur, pour placer le modèle dans des conditions de projections similaires (cas b). La capacité du modèle à extrapoler peut ainsi être testée, par exemple, dans des conditions de température plus élevée, en entraînant le modèle sur sites où les températures rencontrées sont basses et en les projetant dans les conditions de températures hautes (voir par exemple Waldock *et al.*, 2021). Cependant, la structuration spatiale de la variable choisie (*e.g.* gradient amont-aval) pour le partitionnement peut faire perdre en représentativité spatiale ou environnementale dans les données d'entraînement du modèle (*e.g.* sélection majoritaire des cours d'eau amont aux températures fraîches) et limiter sa capacité de généralisation. Dans ce chapitre, la méthode *Multivariate Environmental Similarity Surface* (MESS) a été utilisée afin d'évaluer le niveau d'extrapolation du modèle sur chaque tronçon du réseau de projection (Elith *et al.*, 2010).

D'après la théorie de la niche écologique et la *niche breadth hypothesis*, les taxons avec une niche écologique large présentent une distribution large (voir chapitre I; Marino *et al.* 2020 ; Yañez-Arenas *et al.* 2020). Il est donc très probable que l'ensemble des conditions qui leur sont favorables ne soient pas rencontrées dans l'espace géographique et que les modifications entraînées par le changement climatique leur permettent de disposer de plus d'habitats (expansion; Comte & Grenouillet 2013a). À l'inverse, les taxons dont la niche est moins large ont une distribution plus restreinte, et leur distribution est supposée régresser face à une perte des conditions environnementales favorables (contraction). Les taxons dont la niche réalisée se rapproche de la niche fondamentale sont amenés à suivre leurs conditions environnementales favorables et donc glisser le long du gradient longitudinal pour suivre leur niche (glissement). Enfin, l'évolution de la distribution d'une espèce dépend également de si elle est ou non en limite de sa niche dans l'espace géographique. Par exemple, une espèce dont les conditions environnementales lui restent favorables peut simplement se maintenir dans sa distribution actuelle. L'ensemble de ces modifications de distribution ont déjà été décrites et étudiés sur les communautés de poissons françaises par Comte *et al.* (2013). En considérant l'*abundant-center hypothesis* et en supposant que certaines espèces sont en limite de niche, je suppose une corrélation entre la régression de l'aire de répartition d'une espèce et de ses abondances et inversement pour les taxons en expansion (Figure 55). D'autre part, les espèces sont réparties le long du gradient longitudinal en fonction de leur *preferendum* environnemental (*e.g.* tempé-

rature, conditions hydrauliques ; voir chapitre III et IV). Les espèces des zones aval sont supposées y être contraintes par leur limite basse de température, et les espèces des zones amont contraintes vers l'aval par leur limite haute de température (Comte *et al.*, 2013). Par conséquent, avec l'augmentation progressive de la température de l'eau, les espèces sont supposées se déplacer vers l'amont pour suivre leur optimum thermique, avec pour conséquence une expansion ou une réduction de leur aire de répartition selon leur limite supérieure de température respective (Daufresne & Boët, 2007 ; Comte *et al.*, 2013).

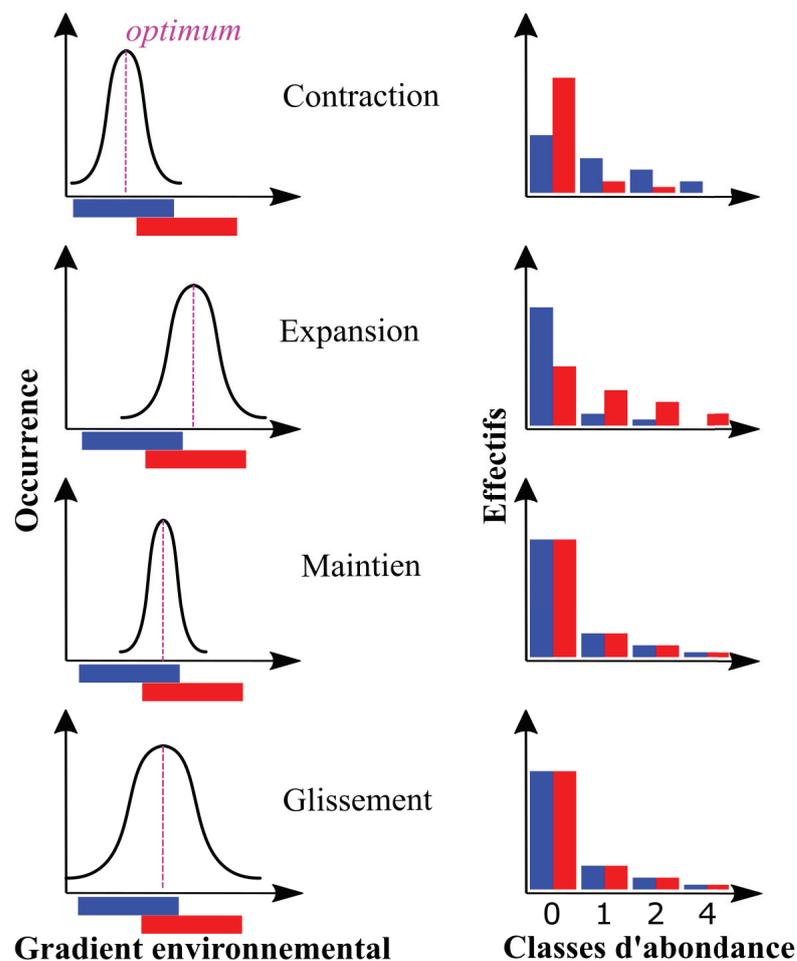


FIGURE 55

Modifications théoriques attendues des abondances d'une espèce en fonction de l'adéquation entre la niche d'une espèce et les conditions environnementales disponibles entre deux horizons temporels sous modifications du gradient environnemental (*e.g.* changement climatique). Le panneau de gauche illustre la niche théorique, supposée normale, et les conditions environnementales, illustrées par une variable, disponibles aux deux horizons temporels actuel (bleu) et futur (rouge ; inspiré de Comte *et al.* 2013). La niche illustrée répond à l'hypothèse de l'abondance-centrée, où les abondances sont supposées maximales à l'optimum (ligne pointillée rose) et décroître de façon monotone à mesure que l'on s'éloigne de l'optimum. Le panneau de droite illustre l'hypothèse selon laquelle les niveaux d'abondance évoluent conjointement à la distribution (*i.e.* contraction, expansion, maintien ou glissement).

Les modifications de la distribution spatiale et des niveaux d'abondances à l'échelle spécifique entraînent des changements à l'échelle de la communauté. En supposant une progression vers l'amont des espèces actuellement inféodées aux zones aval, pour suivre leur niche (que ce soit en glissement ou en expansion), et une régression des espèces des zones amont, du fait de la perte de leurs conditions favorables, je suppose une restructuration plus importante des communautés situées entre le milieu et l'amont du gradient longitudinal que celles des zones situées en aval. En effet, sans connaissance des espèces qui vont coloniser les parties aval du réseau, les potentielles nouvelles arrivées n'ont pas pu être modélisées. D'autre part, avec la perte attendue des taxons spécialistes et rares, une homogénéisation biotique est attendue à l'échelle du bassin versant de la Loire.

Ainsi, les objectifs de ce chapitre sont (i) d'évaluer les modifications de distribution des abondances à l'échelle des taxons (espèces de poissons et genres de macroinvertébrés) en qualifiant le type de changement projeté pour chaque taxon modélisé (*i.e.* régression, expansion, glissement ou maintien) et en évaluant la corrélation entre la modification de l'aire de répartition et du niveaux des abondances et (ii) d'évaluer l'impact des changements de distribution des espèces sur la réorganisation des communautés en quantifiant l'intensité du changement entre la période contemporaine (2007-2017) et deux horizons temporels futurs (2050 et 2080) sous scénario de changement climatique.

B. Données de scénarios de changement climatique

Les chapitres précédents ont permis de montrer le rôle majeur de la température de l'eau (*i.e.* moyenne interannuelle et amplitude de la température de l'eau) et des variables hydrauliques (*i.e.* moyenne interannuelle et amplitude de la vitesse de courant), en association avec la topographie (*i.e.* pente du tronçon) dans la distribution spatiale des taxons par rapport aux autres facteurs étudiés (*e.g.* indicateurs de pressions anthropiques à la continuité écologique, occupation des sols). La projection de la favorabilité potentielle des habitats futurs dans un contexte de changement climatique nécessite d'abord d'établir les conditions environnementales potentielles futures sous un ou plusieurs scénarios de changement climatique. Grâce aux modélisations à base physique de l'hydrologie et de la température de l'eau, via respectivement les modèles EROS et T-NET, des scénarios d'évolutions potentielles de ces variables environnementales sont disponibles (Seyedhashemi *et al.*, 2023). Nous avons donc choisi d'évaluer l'impact que pourraient avoir des

modifications des déterminants environnementaux majeurs sur les communautés aquatiques, en considérant les autres variables incluses dans les modèles comme invariants (*i.e.* pente, distance à l'estuaire, taux d'étagement, densité d'ouvrages, densité de population, gradient d'intensification agricole). En effet, il n'existe pas à notre connaissance de scénarios d'évolution future pour toutes les variables intégrées dans nos modèles (*e.g.* évolutions des ouvrages en travers des cours d'eau et impact sur l'indicateur du taux d'étagement) à l'échelle de tout le réseau hydrographique ligérien considéré.

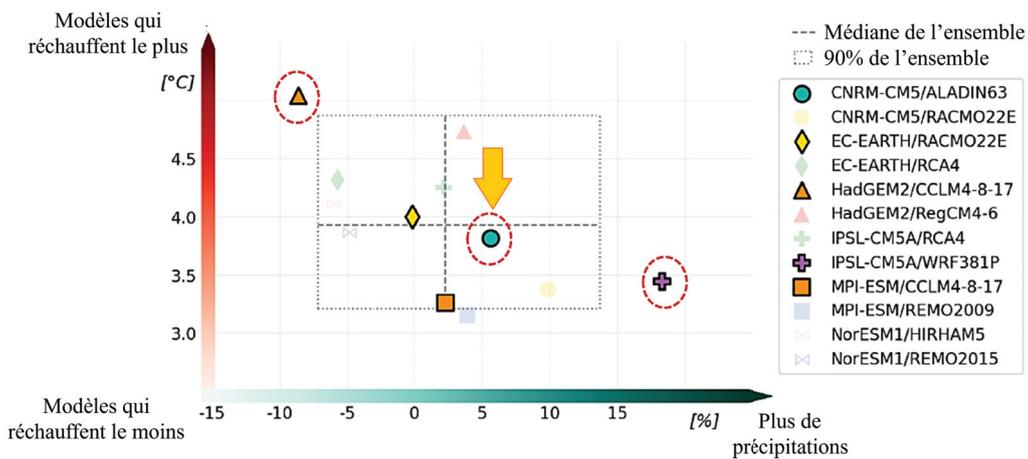


FIGURE 56

Changements de précipitations (en abscisses) et de température de l'air (en ordonnées) en France d'ici à la fin du siècle par rapport à la période historique 1976-2005 sous le scénario RCP 8.5. Les points de couleurs vives représentent les modèles climatiques globaux et régionaux proposés par Météo-France ; les cercles pointillés rouges représentent les modèles sélectionnés pour les projections de la température de l'eau et de l'hydrologie par Seyedhashemi *et al.* (2023).

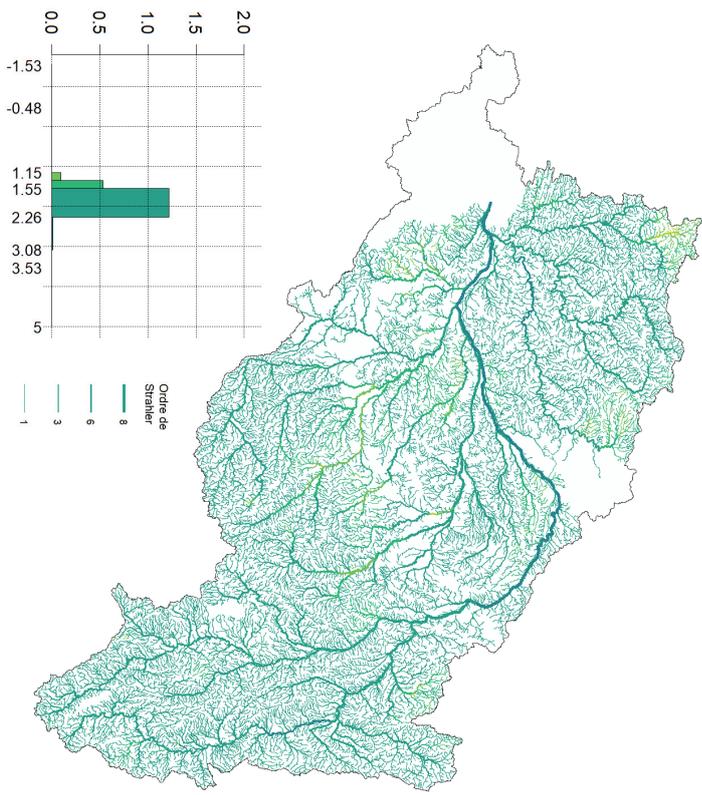
La flèche orange identifie le modèle considéré dans ce chapitre. D'après

<https://www.drias-climat.fr/>.

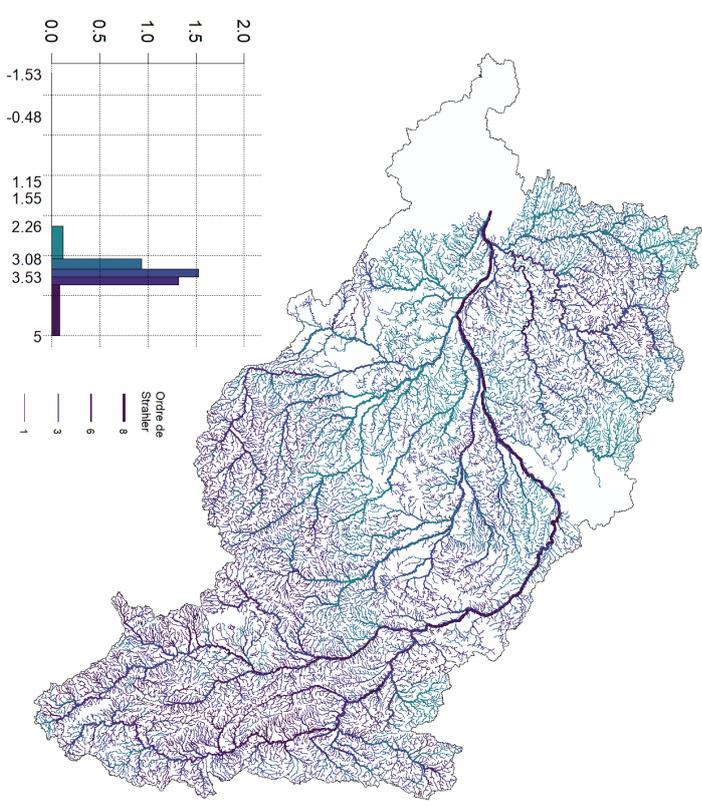
Différents modèles climatiques ont servi à alimenter les données d'entrées des modèles EROS et T-NET pour simuler les projections de l'hydrologie et de la thermie sous plusieurs scénarios RCP et à différents horizons temporels (voir Seyedhashemi *et al.* (2023) pour plus de détails). D'après l'analyse des projections des conditions de thermie et d'hydrologie sous le scénario RCP 8.5 à l'horizon 2071-2100, les modèles climatiques globaux HadGEM2/CCLM4-8-17, IPSL-CM5A/WRF381P et CNRM-CM5/ALADIN63 ont été qualifiés respectivement de chaud et sec, chaud et humide et intermédiaire (Figure 56). Ici, le modèle climatique « intermédiaire », CNRM-CM5/ALADIN63 a été choisi car il représente une trajectoire intermédiaire d'évolution des conditions thermiques et hydrologiques des cours d'eau entre les modèles IPSL-CM5A/WRF381P et HadGEM2/CCLM4-

8-17. Seul le scénario RCP 8.5 a été considéré dans ce chapitre. Les conditions environnementales ont été projetées sur deux périodes temporelles : 2040-2069 et 2070-2099, nommées par la suite respectivement H2050 et H2080. Elles sont comparées à la période actuelle, 2007-2017, nommée ensuite H2010. Comme dans le chapitre II, pour chacune des métriques, les moyennes annuelles ont été calculées à partir des simulations journalières issues des modèles EROS et T-NET puis moyennées à l'échelle de la période considérée pour obtenir les moyennes interannuelles. Ainsi, les métriques moyennes, les écarts-types, les percentiles 10 et 90, et une mesure de l'amplitude (ratio percentile 90 sur 10) ont été utilisés comme prédicteurs dans les modèles (voir équation de modélisation en chapitre IV).

Les Figures 57 à 60 présentent les cartographies des écarts entre les périodes de projections (H2050 et H2080) et la période actuelle (H2010) pour les variables thermiques et hydrauliques. Les projections des températures de l'eau (Figures 57 et 58) montrent un réchauffement global sur l'ensemble du bassin versant de la Loire (écarts > 0), qui s'intensifie dans le temps (écarts moyens de $1,76\text{ °C}$ à l'horizon H2050 et de $3,43\text{ °C}$ à l'horizon H2080). Ce réchauffement est associé à une augmentation de l'amplitude entre les percentiles 90 et 10, indiquant des écarts de température de l'eau encore plus marqués entre l'hiver et l'été, dus à une augmentation plus intense des températures estivales, en particulier pour les moyennes et grandes rivières (Seyedhashemi *et al.*, 2022). Les vitesses moyennes de courant sont également amenées à augmenter (écarts moyen de $0,21\text{ m}\cdot\text{s}^{-1}$ aux horizons H2050 et H2080 ; Figure 59 et 60), en association avec une diminution de l'amplitude entre percentile 90 et percentile 10 due à une accentuation des périodes d'étiages sévères et de l'intensité des phénomènes de crues (Seyedhashemi *et al.*, 2023). Pour plus de détails sur l'évolution projetée des conditions environnementales sous changement climatique voir Seyedhashemi *et al.* (2023).

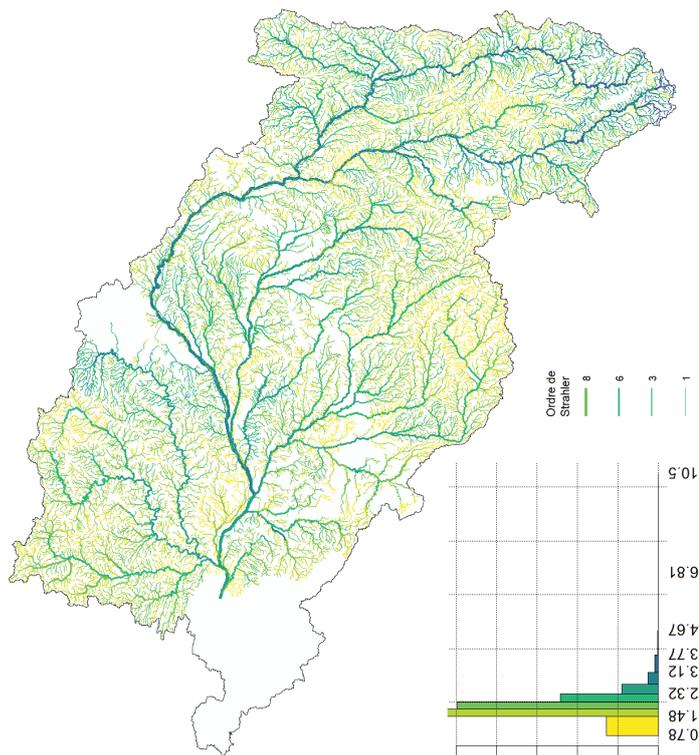


(a)

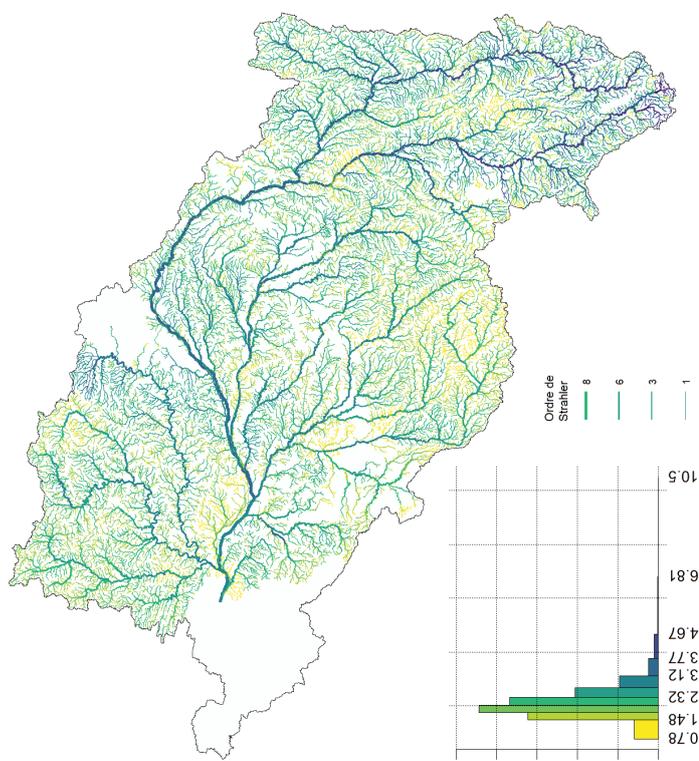


(b)

FIGURE 57 Écarts de la température moyenne interannuelle de l'eau (en °C) entre la période de référence (2007-2017) et les horizons temporels H2050 (a) et H2080 (b) sous le scénario RCP 8.5.

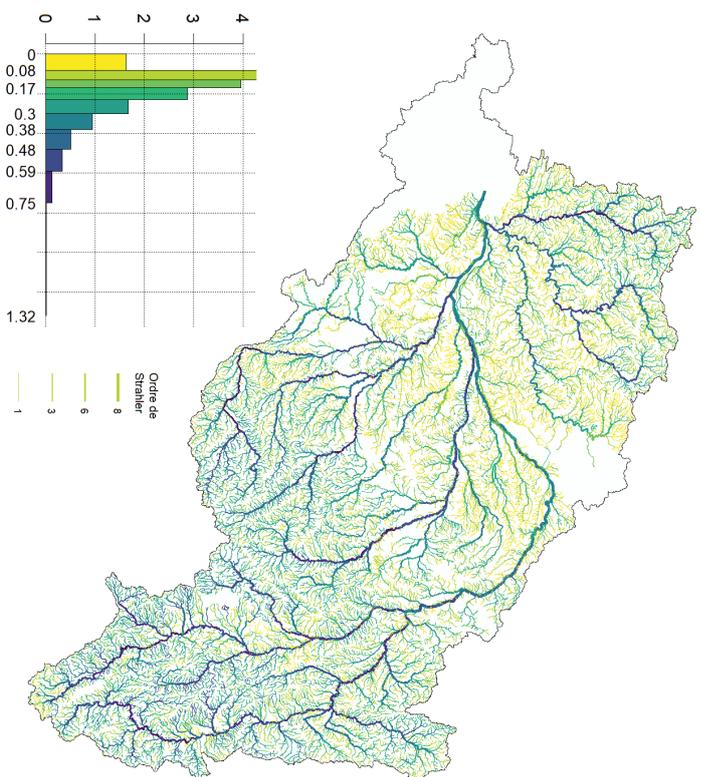


(a)

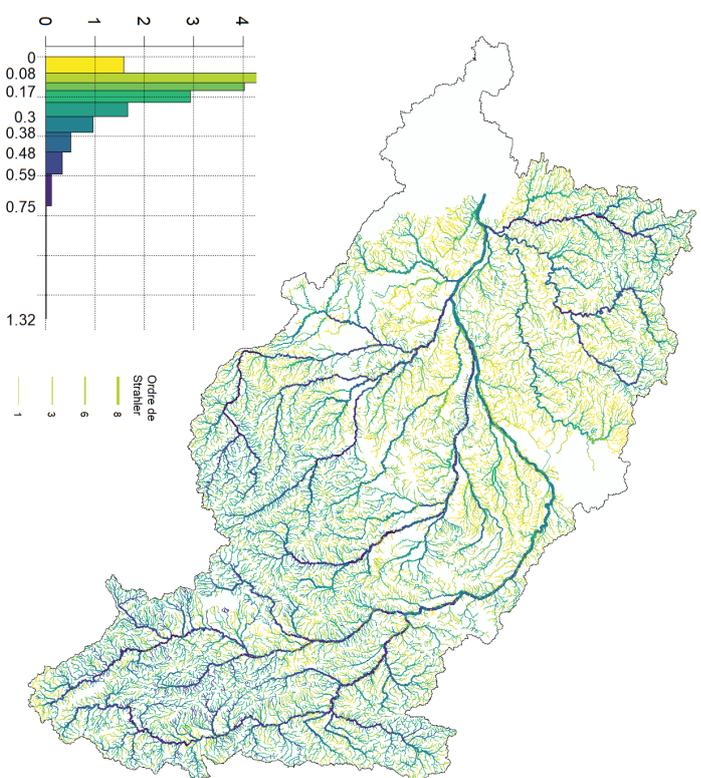


(b)

FIGURE 58
Écart du ratio percentile 90 sur la température de l'eau entre la période de référence (2007-2017) et les horizons temporels H2050 (a)
et H2080 (b) sous le scénario RCP 8.5.

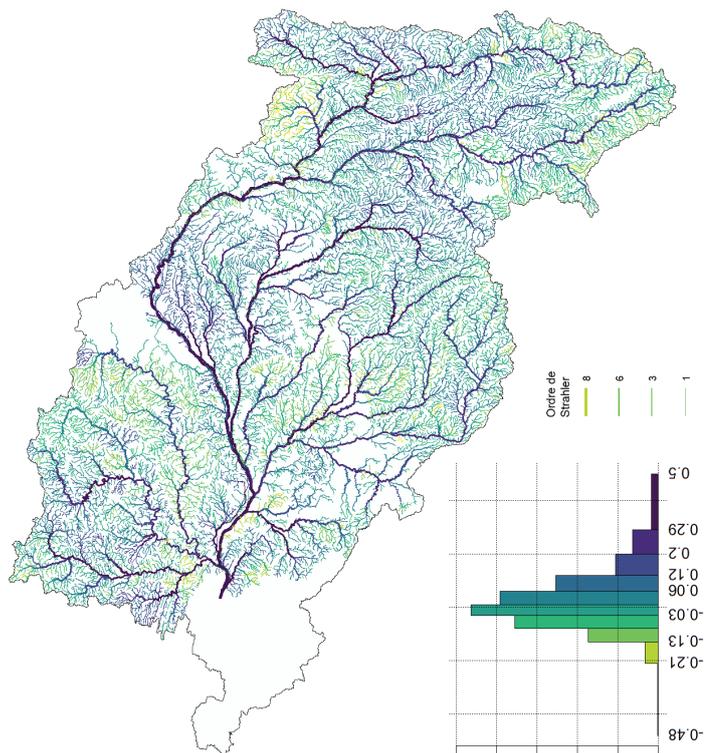


(a)

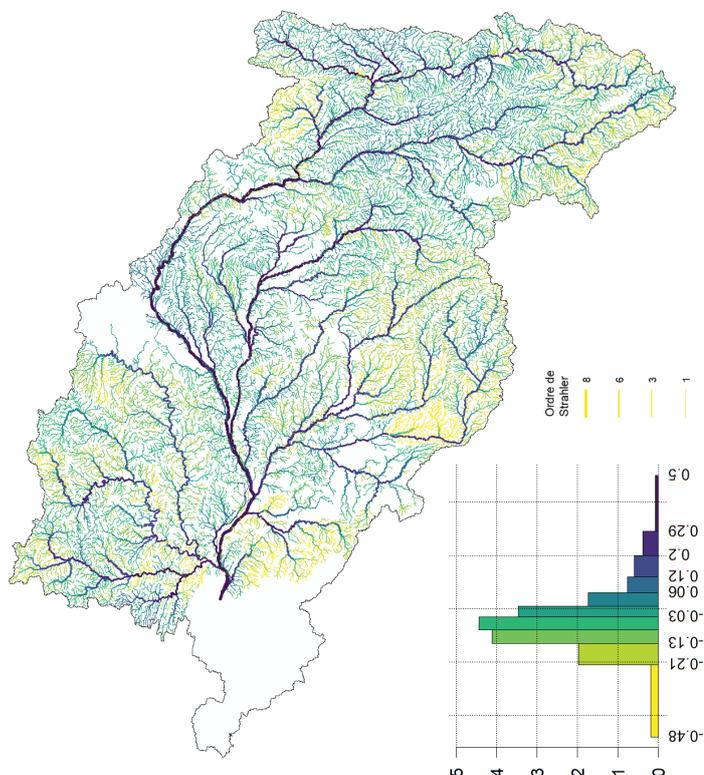


(b)

FIGURE 59 Écart de la vitesse de courant moyenne interannuelle (en $m \cdot s^{-1}$) entre la période de référence (2007-2017) et les horizons temporels H2050 (a) et H2080 (b) sous le scénario RCP 8.5.



(a)



(b)

FIGURE 60

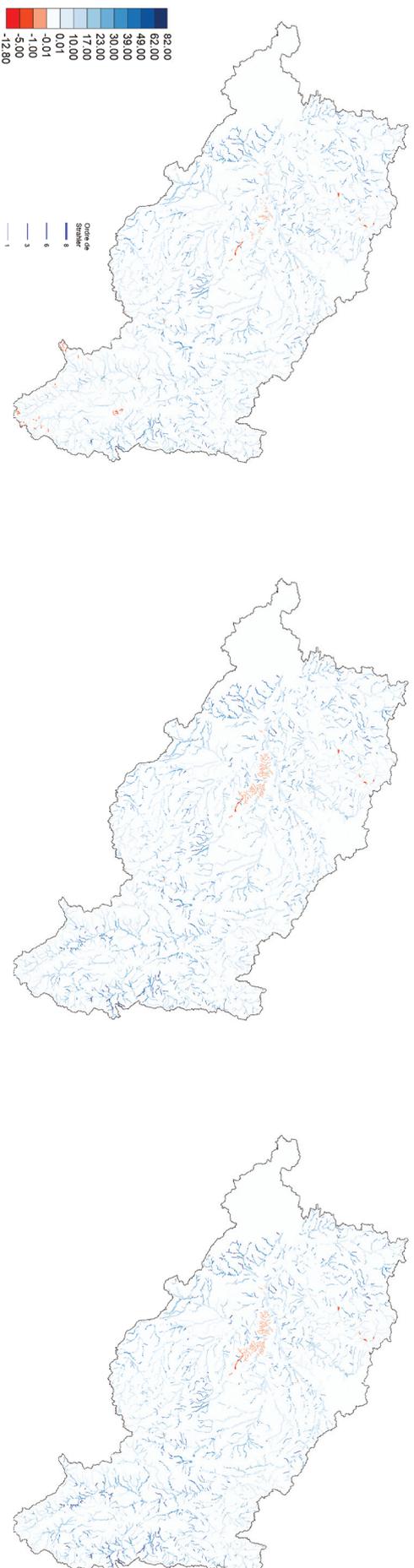
Écart du ratio percentile 90 sur percentile 10 de la vitesse de courant entre la période de référence (2007-2017) et les horizons temporels H2050 (a) et H2080 (b) sous le scénario RCP 8.5.

C. Comparaison des conditions environnementales entre les horizons temporels

La projection de SDM en-dehors des conditions spatiales et temporelles rencontrées dans le jeu d'entraînement peut avoir des implications importantes sur l'interprétation des distributions résultantes (Peterson *et al.*, 2011 ; Guisan *et al.*, 2017). En effet, le modèle est entraîné sur certaines gammes de variables environnementales qui peuvent être très différentes entre les périodes temporelles ; lorsque les conditions de projection n'ont pas été rencontrées par le modèle lors de la phase d'entraînement, celui-ci est en extrapolation pure, et il n'y a aucun moyen de déterminer la précision de la prédiction (Elith & Leathwick, 2009). Il est donc recommandé en premier lieu d'entraîner le modèle sur le plus grand nombre possible de données (Peterson *et al.*, 2011). Afin d'évaluer le niveau d'extrapolation des projections Elith *et al.* (2010) ont proposé la méthode MESS pour comparer les conditions environnementales entre différents jeux de données. Cette méthode calcule pour chaque unité géographique (ici le tronçon de rivière ; $N = 33\ 711$) la similarité entre les conditions environnementales rencontrées dans le jeu de projection et dans le jeu de référence, ici les données utilisées lors de la phase d'entraînement, en tenant compte des combinaisons entre variables. Cette méthode permet de distinguer les unités où le modèle est en extrapolation pure et où l'interprétation devrait être évitée, des unités où les conditions sont relativement proches des conditions connues du modèle (voir Elith *et al.*, 2010 pour les détails de calcul). Les données utilisées pour la « référence » sont l'ensemble des tronçons portant une station d'échantillonnage poissons et macroinvertébrés, soit 1 153 tronçons.

L'analyse MESS indique une proportion très limitée de tronçons en extrapolation pure (1,15 % du réseau à H2010 et 0,96 % à H2050 et H2080) à l'échelle du réseau H (tronçons en rouge sur la Figure 61). Le degré de similarité est globalement faible sur les grands cours d'eau et plus élevé sur les tronçons de têtes de bassins quelque soit l'horizon temporel considéré. À l'horizon H2010, les données comparées sont celles des tronçons utilisés pour constituer le jeu d'entraînement avec l'ensemble des conditions rencontrées sur le reste des tronçons du réseau H. La faible similarité constatée dans ce cas est vraisemblablement due à une plus grande représentativité de ces derniers dans les données d'échantillonnage. Concernant les horizons H2050 et H2080, l'augmentation des températures de l'eau engendre l'apparition de nouvelles conditions, non rencontrées dans les données

d'entraînement, d'où la faible similarité. D'après la méthode MESS, le faible nombre de tronçons en extrapolation pure ($N = 390$ au maximum) indique que les projections issues des modèles sont interprétables sur la majeure partie du réseau. Les faibles valeurs de similarité appellent tout de même à la prudence en fonction des capacités d'extrapolation des modèles, notamment sur les tronçons dans les parties aval. L'utilisation de méthodes de *machine learning*, plus flexibles et reconnues pour leur bonne capacités d'extrapolation, est donc d'autant plus justifiée (Waldock *et al.*, 2021 ; Valavi *et al.*, 2023).



(a)

(b)

(c)

FIGURE 61

Similarité des conditions environnementales rencontrées dans le jeu d'entraînement (N = 1 153 tronçons) par rapport aux conditions de projections à l'échelle du réseau H (n = 33 711 tronçons) aux différents horizons temporels, H2010 (a), H2050 (b) et H2080 (c). Les tronçons en rouge indiquent les zones d'extrapolation pure (à H2010 N = 390, à H2050 N = 324, à H2080 N = 324), les tronçons en bleu (à H2010 N = 33 321, à H2050 N = 33 387, à H2080 N = 33 387) indiquent le degré de similarité entre jeu d'entraînement issus des tronçons échantillonnés et l'ensemble du réseau H.

D. Modifications de la distribution des espèces

D.1. Les différents types de modification de la distribution des espèces

Les projections des modèles d'abondance aux deux horizons temporels H2050 et H2080 sous le scénario RCP 8.5 permettent d'estimer les distributions potentielles futures des espèces à l'échelle du bassin versant de la Loire. Différents exemples de modification potentielle de la distribution future des espèces sont illustrés en Figure 62. L'ensemble des projections obtenues pour les 57 taxons modélisés sont consultables à l'Annexe G. Les cartographies permettent de montrer que les taxons répondront à l'évolution des conditions environnementales induites par le changement climatique. L'ablette (*Alburnus alburnus*) et, dans une moindre mesure, le genre *Choroterpes* montrent par exemple une expansion de leur distribution spatiale vers l'amont, indiquant que les conditions environnementales prédites dans les cours d'eau appartenant aux parties aval du bassin leur restent favorables. La truite (*Salmo trutta*) montre elle une contraction de son aire de répartition ainsi qu'une diminution marquée des effectifs des classes d'abondance moyenne et élevée mais semble se maintenir en tête de bassin avec de faibles abondances. *Trichostegia*, qui occupe actuellement une bonne part du gradient longitudinal, semble quant à lui être amené à quasiment disparaître du bassin de la Loire dès l'horizon H2050, ce qui témoigne peut-être d'une spécialisation pour des conditions environnementales particulières, notamment en terme de thermie.

D.2. Mesure de la modification de la distribution des espèces

Le *Species Range Change* (SRC) est un indice qui permet de mesurer le taux de changement de l'occupation d'un taxon entre deux périodes temporelles en comparant le nombre d'unités géographiques occupées, ici le tronçon du réseau H. Il peut être calculé avec des données d'abondances d'un taxon donné et pondéré par la longueur des tronçons pour intégrer une mesure de la portion du réseau occupée grâce à la formule suivante (Floury *et al.*, 2021) :

$$SRC = \frac{\sum_r d_{f,r} \cdot l_r - \sum_r d_{a,r} \cdot l_r}{\sum_r d_{a,r} \cdot l_r}$$

où $d_{f,r}$ et $d_{a,r}$ correspondent respectivement aux abondances futures et actuelles (ici les CPUE) pour le tronçon r de longueur l_r .

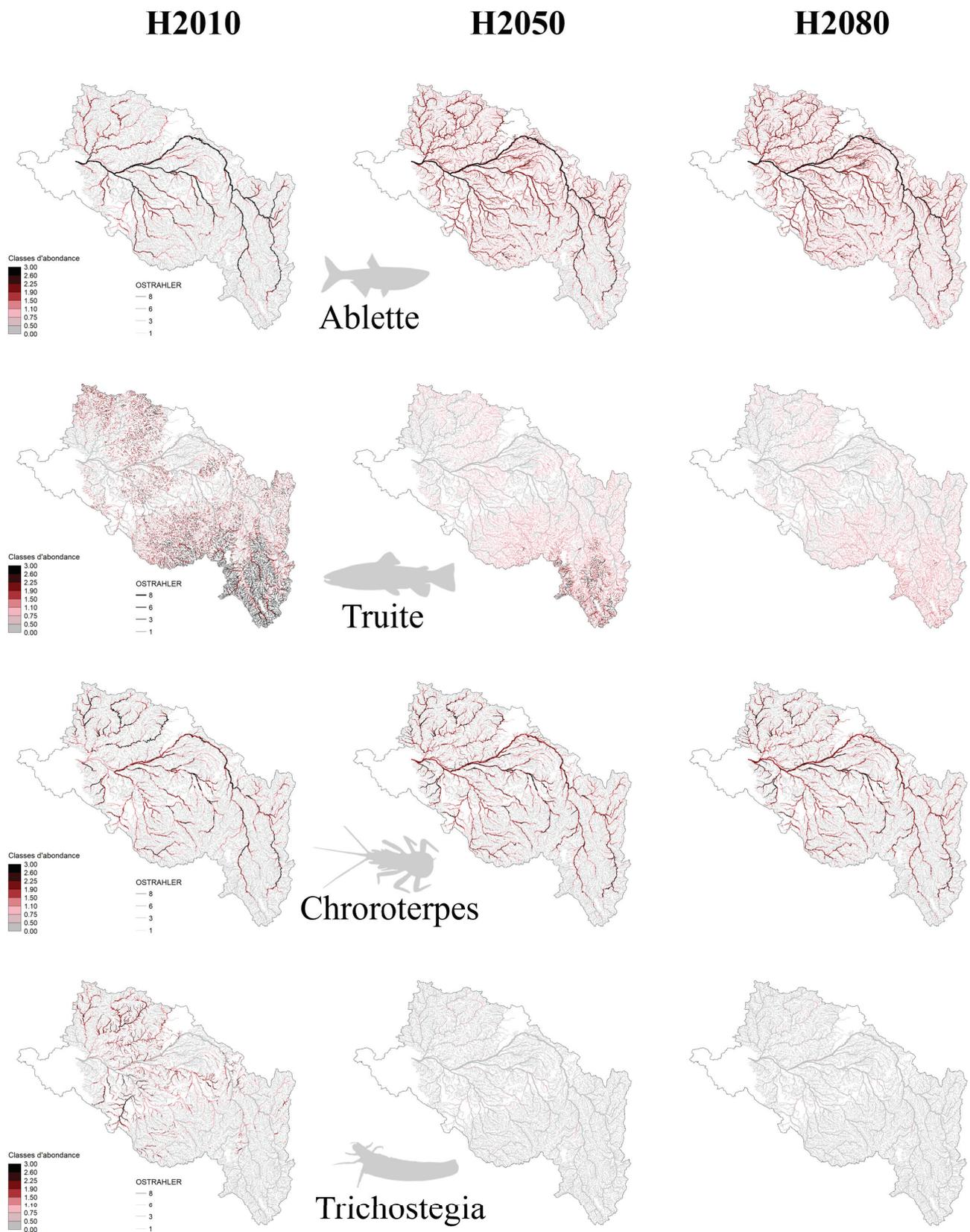


FIGURE 62

Distributions potentielles projetées pour l'ablette (*Alburnus alburnus*), la truite (*Salmo trutta*), *Choroterpes* et *Trichostegia* aux trois horizons temporels H2010 (gauche), H2050 (milieu) et H2080 (droite). Les classes d'abondance moyenne ont été calculées sur les 20 réplicats de chaque modèle pour chaque tronçon du réseau H (N = 33 711).

Les SRC aux deux horizons temporels H2050 et H2080 sont représentés pour chaque taxon en figure 63 et l'ensemble des projections peut être consulté à l'annexe G. Ils montrent une proportion similaire de taxons en expansion (N = 14 espèces de poissons et N = 11 genres de macroinvertébrés) et les taxons en régression (N = 12 espèces de poissons et N = 20 genres macroinvertébrés). Globalement, les valeurs de SRC entre les deux horizons temporels sont assez similaires. Certains taxons montrent néanmoins une légère tendance à poursuivre leur expansion (*e.g.* le silure, la brème ou *Cordulia*) ou leur régression (*e.g.* *Xanthoperla*) tandis que d'autres se stabilisent ou semblent inverser leur dynamique entre l'horizon H2050 et l'horizon H2080 (*e.g.* *Goera*). Les genres de macroinvertébrés *Psychomyia*, *Oligoplectrum*, *Mystacides* et *Euleuctra* semblent conserver des niveaux d'occupation stables à l'horizon H2050 par rapport à l'horizon H2010, mais amorcent une régression à l'horizon H2080. Les trois premiers se situent majoritairement en tête de bassin et régressent vraisemblablement du fait du dépassement de leur limite thermique tandis que les deux autres sont des espèces situées au milieu du gradient amont-aval et voient leurs fortes abondances diminuer. Flourey *et al.* (2021) ont proposé des projections de la distribution de la truite commune sur le bassin versant du Rhône aux périodes temporelles 2029-2038, 2054-2063 et 2079-2088. Ils ont également montré une régression de l'aire de répartition de l'espèce, avec des valeurs de SRC similaires à celles trouvées ici pour cette espèce. Globalement, les taxons en régression perdent à la fois en taux d'occupation et en effectif dans les classes d'abondance moyenne et forte (Figures 64 et 65), et inversement pour les taxons en expansion, vérifiant ainsi notre hypothèse initiale.

Afin d'avoir une vision plus détaillée de l'association entre changement d'occupation et modification des effectifs des abondances par taxon, les Figures 64 et 65 ont été tracées aux trois horizons temporels. Ils permettent de constater que certains taxons en expansion (*e.g.* ablette ou *Potamanthus*) peuvent gagner en taux d'occupation du réseau sans changer drastiquement la distribution des effectifs entre les classes d'abondance, contrairement à notre hypothèse qui supposait une corrélation entre l'évolution de la distribution et des effectifs entre les classes. Certains taxons en expansion (*e.g.* goujon, chevesne ou *Cordulia*) voient leurs classes d'abondance moyenne et surtout forte augmenter en effectifs par rapport à la classe faible, en accord avec notre hypothèse. Quant aux taxons en régression, la plupart voit leurs classes d'abondance moyenne et forte diminuer plus fortement que la classe d'abondance faible. Ces constats montrent que les changements de distribution des

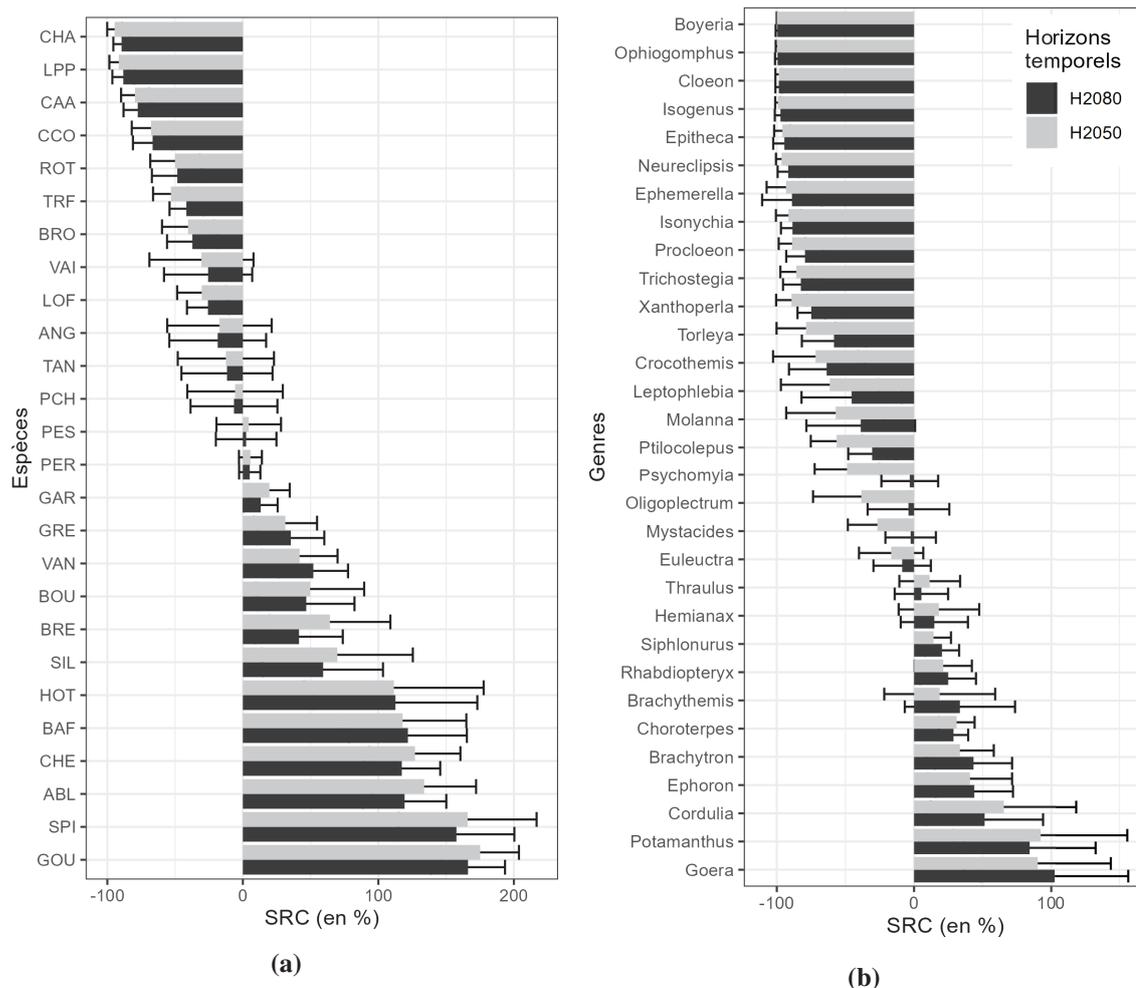


FIGURE 63
Species Range Change pour chaque espèce de poissons (a ; n = 26) et genres de macroinvertébrés (b ; n = 31) entre les horizons H2050 et H2010 (en gris) et H2080 et H2010 (en noir) sous le scénario RCP 8.5. Les taxons sont ordonnés en fonction de leur SRC. Les moyennes et écart-types ont été calculés avec les valeurs obtenues pour les 20 répliqués de modèles. La correspondance avec les codes espèces poissons peut être consultées au tableau 1.

taxons sont accompagnés de modifications dans la répartition des classes d'abondance qui ne sont pas aisément prévisibles sur la base des données de présence-absence, et appuient la nécessité d'utiliser des données d'abondance pour l'étude des effets du changement climatique sur les communautés aquatiques.

E. Modifications à l'échelle des communautés

Il est assez classique de rassembler les prédictions issues des modèles construits pour chaque espèce afin d'évaluer les modifications à l'échelle de la communauté (Thuiller *et al.*, 2005 ; Buisson & Grenouillet, 2009 ; Marmion *et al.*, 2009 ; Maire *et al.*, 2017). Les projections des modèles des 26 espèces de poissons et 31 genres de macroinvertébrés

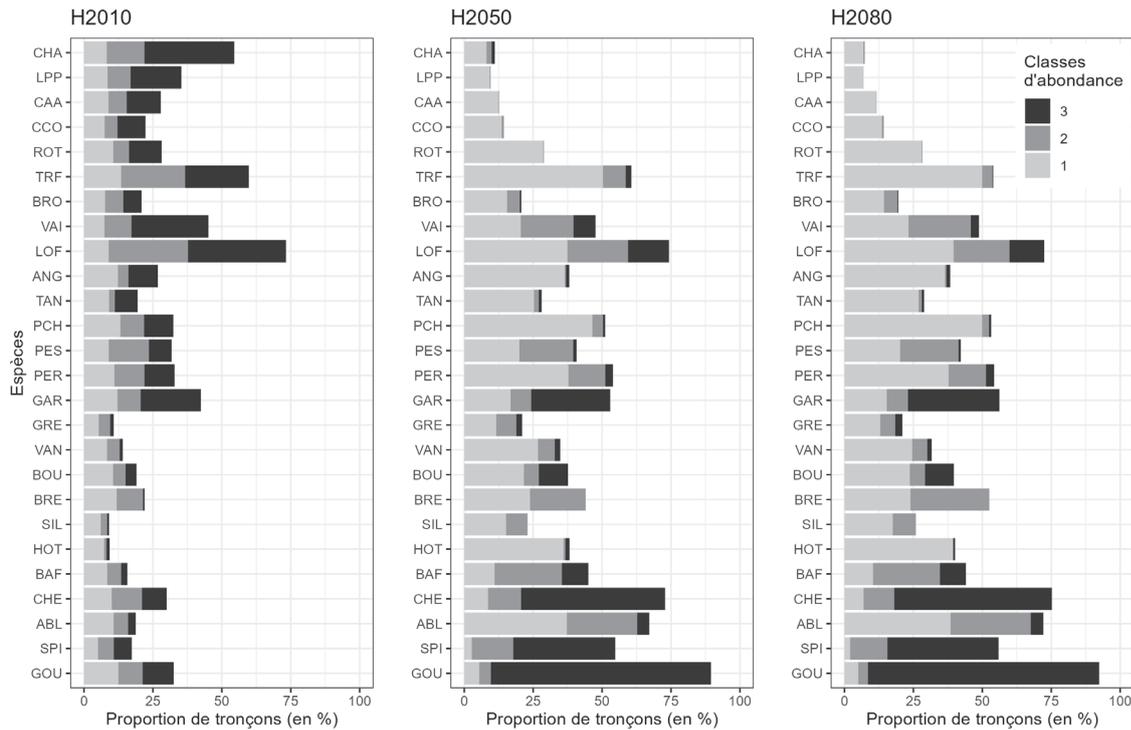


FIGURE 64

Occurrence des classes d'abondance projetées pour les 26 espèces de poissons aux trois horizons temporels H2010, H2050 et H2080 sur l'ensemble du réseau H (N = 33 711 tronçons). Les espèces de poissons sont ordonnées en fonction de leur SRC, comme sur la Figure 63. La correspondance avec les codes espèces poissons peut être consultée au tableau 1.

ont ainsi été rassemblées pour former les communautés prédites à chacun des horizons considérés.

E.1. Modification de la diversité des communautés

La richesse spécifique, soit le nombre d'espèce par tronçon, est une mesure de la diversité d'une communauté. Elle a été calculée à partir des projections rassemblées pour les 57 taxons sur l'ensemble des tronçons du réseau H, et les écarts entre les horizons temporels H2010 d'une part, H2050 et H2080 d'autre part, ont été calculés.

Les modifications de richesse spécifique sont la résultante des changements de distribution des taxons de poissons et de macroinvertébrés. La modification de la richesse spécifique est spatialisée par sous-bassins (Figure ??) avec des bassins où les pertes où gains semblent majoritaires. Hormis pour la zone appartenant au Massif Central, des modifications fortes (gain ou perte > 5 taxons) sont attendues dans tout le bassin. L'augmentation de la richesse est plus élevée dans les zones amont (jusqu'à + 23 taxons, poissons et macroinvertébrés confondus) en lien avec l'augmentation de la favorabilité de certaines

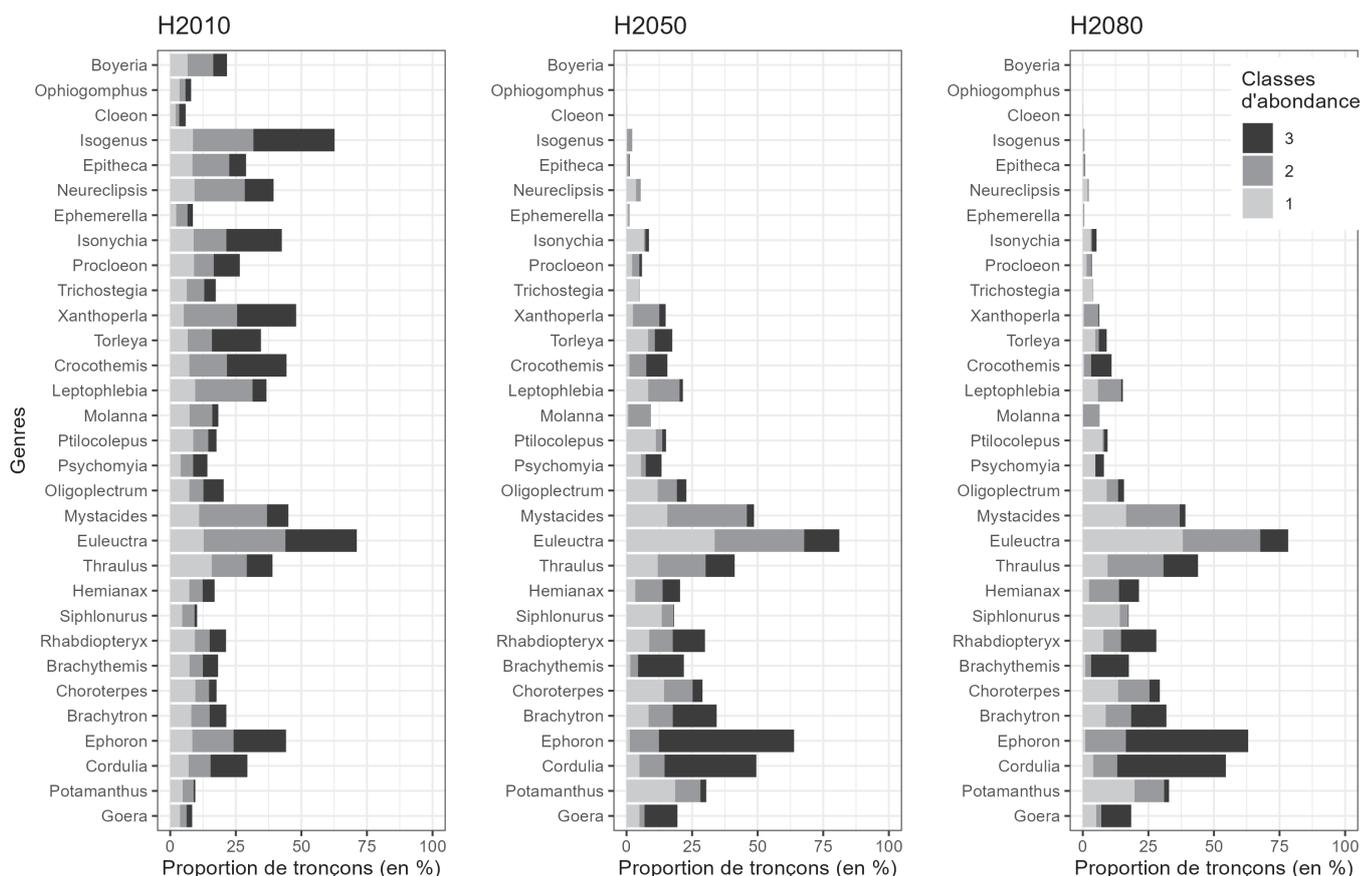
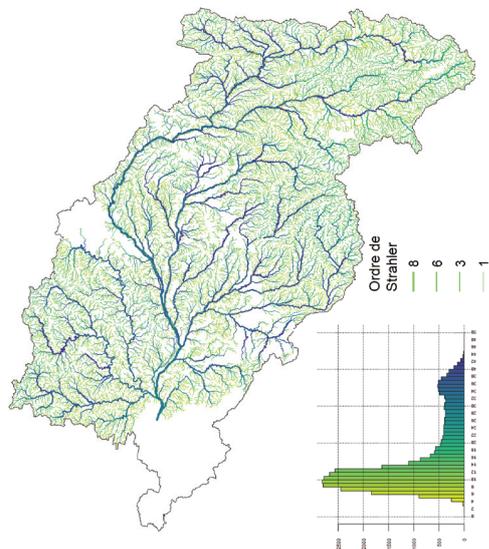


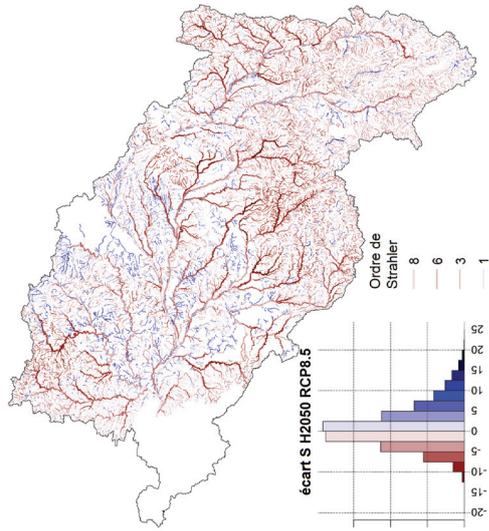
FIGURE 65

Occurrence des classes d'abondance projetées pour les 31 genres de macroinvertébrés aux trois horizons temporels H2010, H2050 et H2080 sur l'ensemble du réseau H (N = 33 711 tronçons). Les genres de macroinvertébrés sont ordonnés en fonction de leur SRC, comme sur la Figure 63.

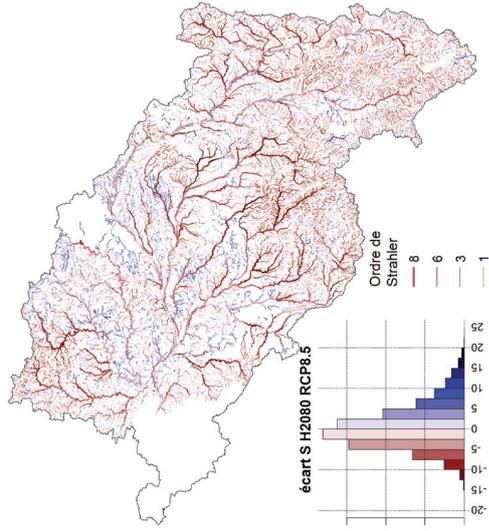
zones amont pour les espèces des zones aval. Certaines zones présentent majoritairement des tronçons avec une perte moyenne de richesse (jusqu'à - 15 taxons, majorité de tronçon rouges), notamment les parties centrales, très à l'Est et au Nord du bassin. Les modifications de la richesse spécifique dépendent également de l'ordre de Strahler et de l'horizon temporel (Figure 69 haut) : pour les ordres 1 à 3, le nombre d'espèces moyen par tronçon augmente à H2050 puis diminue à H2080 ; pour les ordres supérieurs à 3 la richesse diminue dans le temps sauf pour l'ordre de Strahler 8 où elle reste relativement constante. Cette stabilité apparente n'est certainement pas réaliste, car l'arrivée d'espèces exogènes étant inconnue, ces apports n'ont pas été considérés dans les projections. Les modifications de la richesse taxonomique ont une tendance à la hausse pour les poissons et une tendance à la baisse pour les macroinvertébrés (Figures 67 et 68, resp.), ce qui pourrait indiquer une plus grande sensibilité des macroinvertébrés au changement climatique par rapport aux poissons.



(a)



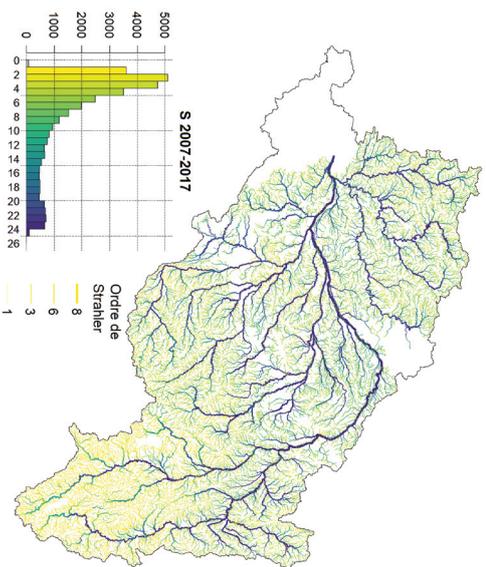
(b)



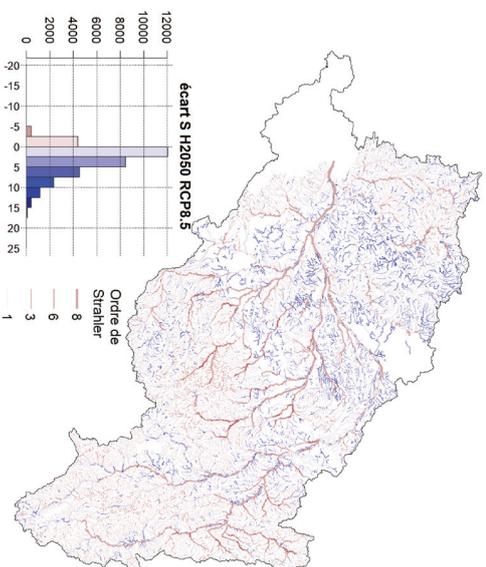
(c)

FIGURE 66

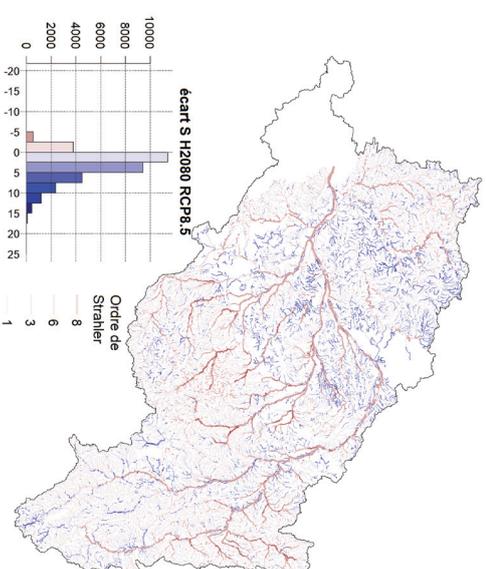
Modification de la richesse spécifique entre les trois horizons temporel. Richesse spécifique de la période actuelle H2010 (a) et écarts entre la richesse spécifique aux horizons H2050 (b) et H2080 (c) par rapport à la période actuelle.



(a)

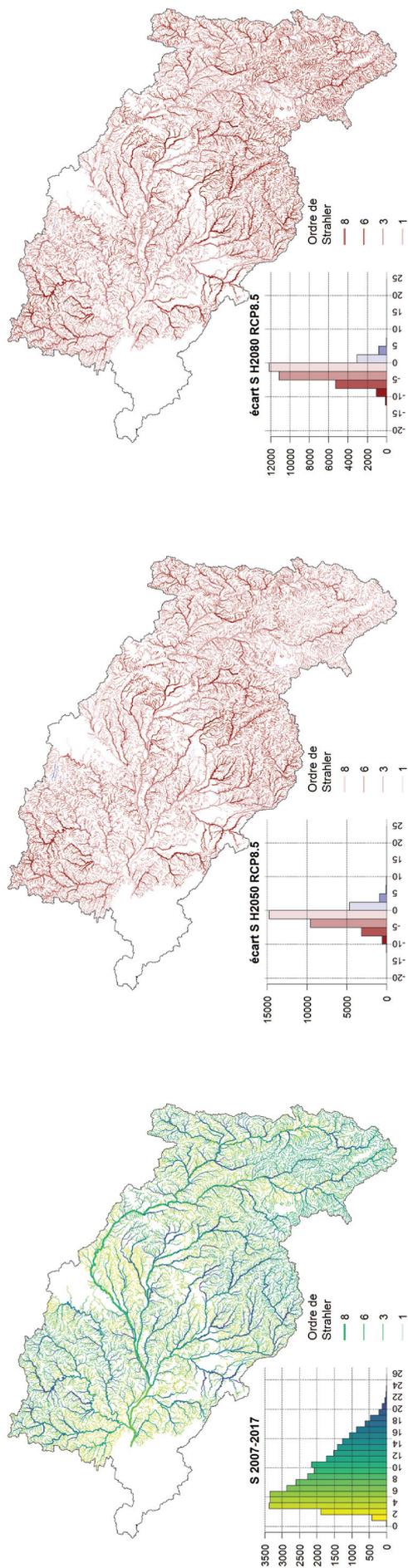


(b)



(c)

FIGURE 67
Modification de la richesse spécifique pour les poissons entre les trois horizons temporel. Richesse spécifique de la période actuelle H2010 (a) et écarts entre la richesse spécifique aux horizons H2050 (b) et H2080 (c) par rapport à la période actuelle.



(a)

(b)

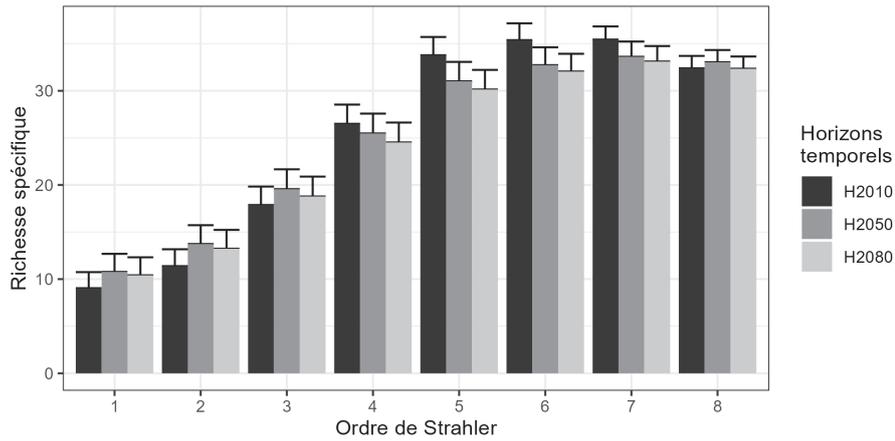
(c)

FIGURE 68

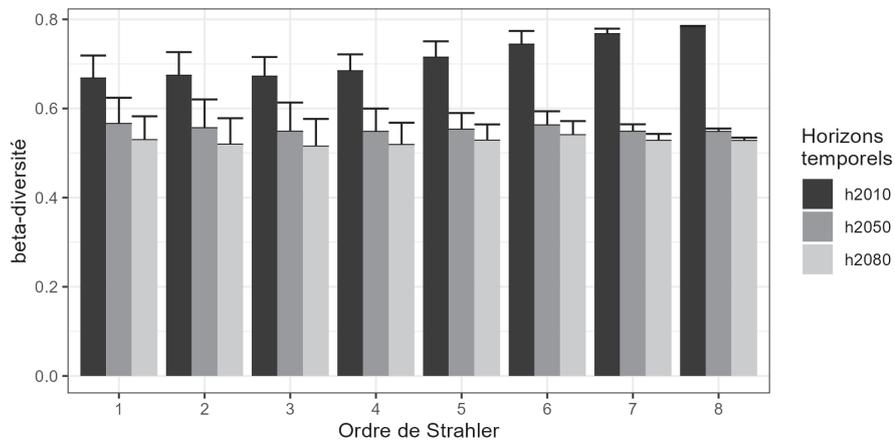
Modification de la richesse spécifique pour les macroinvertébrés entre les trois horizons temporel. Richesse spécifique de la période actuelle H2010 (a) et écarts entre la richesse spécifique aux horizons H2050 (b) et H2080 (c) par rapport à la période actuelle.

Afin de tester l'hypothèse d'homogénéisation biotique, la comparaison de la β -diversité moyenne entre les horizons temporels a été utilisée (Olden, 2006). La structuration des communautés aquatiques dépend fortement de la position sur le gradient amont-aval, avec un accroissement de la richesse spécifique vers l'aval (Matthews, 1998 ; Buisson & Grenouillet, 2009). Les analyses suivantes ont donc été réalisées à l'échelle du bassin versant et par ordre de Strahler afin de comparer les modifications en fonction du positionnement sur le gradient longitudinal. Pour tous les tronçons appartenant à un même ordre de Strahler, une matrice de distance de Bray-Curtis des tronçons pris deux à deux a été calculée afin d'évaluer la β -diversité. L'opération est répétée sur les 20 réplicats de modèles, soit 20 matrices de distance, dont les valeurs sont ensuite moyennées par ordre de Strahler.

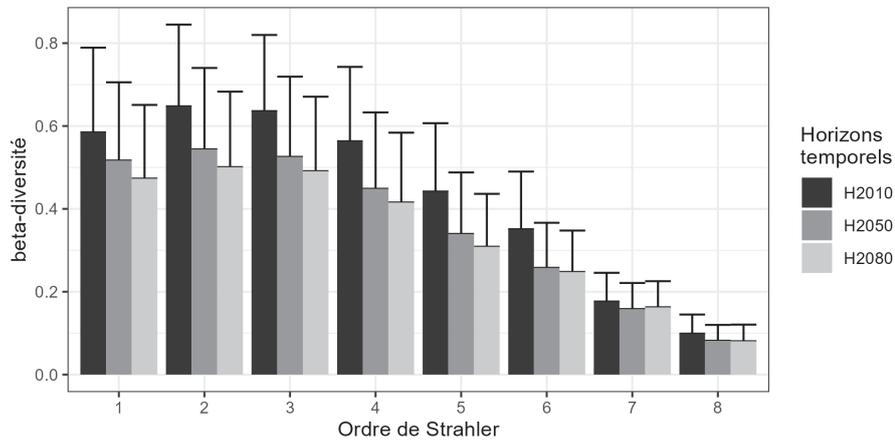
La β -diversité moyenne calculée à partir de tous les tronçons du bassin versant est de $0,44 \pm 0,15$ à l'horizon H2010, $0,36 \pm 0,14$ à H2050 et $0,33 \pm 0,13$ à H2080. La β -diversité globale du bassin diminue donc au cours du temps. Une diminution de la β -diversité intra-ordre de Strahler est également observée pour les ordres de Strahler jusqu'à l'ordre 6 inclus (Figure 69 bas). Il ressort que cette diminution est plus importante dans les milieux amont.



(a)



(b)



(c)

FIGURE 69

Changements des patrons de richesse spécifique (a) et de β -diversité (b et c) des communautés formées par les espèces de poissons et de genres de macroinvertébrés par ordre de Strahler aux trois horizons temporels H2010, H2050 et H2080. Pour la richesse spécifique, les valeurs moyennes par ordre de Strahler et leur écart-type sont présentés. Pour la β -diversité, les valeurs présentées sont les valeurs moyennes par tronçon en fonction de l'ordre de Strahler, à partir de la β -diversité calculée entre tous les tronçons (N = 33 711) (b), et les valeurs moyennes β -diversité calculée entre les tronçons d'un même ordre de Strahler (c).

E.2. Taux de changement des communautés

Le *Temporal Beta-diversity Index* (TBI) est un indicateur qui mesure le changement de composition d'une communauté entre deux périodes de temps pour un site donné (Legendre, 2019). On notera y_1 et y_2 les observations multi-spécifiques aux périodes de temps T1 et T2. Le TBI est calculé à partir de trois métriques basées sur les abondances des taxons et de la formule suivante :

$$TBI = \frac{B + C}{2A + B + C}$$

- ★ A_j est la part d'abondance du taxon j qui est commune aux vecteurs des deux périodes de temps : $A_j = \min(y_{1j}, y_{2j})$. A est la somme des valeurs A_j pour tous les taxons. C 'est la similarité non normalisée entre les deux vecteurs.
- ★ B_j est la part d'abondance du taxon j perdue entre les vecteurs des deux périodes de temps : $B_j = (y_{1j} - y_{2j})$ si $y_{1j} > y_{2j}$; $B_j = 0$ sinon. B est la somme des valeurs B_j pour tous les taxons. C 'est la somme non normalisée des pertes d'abondance de taxons entre T1 et T2.
- ★ C_j est la part d'abondance du taxon j gagnée entre les vecteurs des deux périodes de temps : $C_j = (y_{2j} - y_{1j})$ si $y_{2j} > y_{1j}$; $C_j = 0$ sinon. C est la somme des C_j pour tous les taxons. C 'est la somme non normalisée des gains d'abondance de taxons entre T1 et T2.

La formule du TBI lui permet d'être décomposé en B et C , soit la perte et le gain en abondance par espèce. $(C - B)$ donne la direction du changement entre les deux périodes de temps, soit le processus dominant entre perte et gain d'espèce sur chaque site. La significativité du TBI entre deux périodes de temps est évaluée par un test de permutation.

D'après le TBI, toutes les communautés du réseau hydrographique de la Loire sont amenées à être modifiées (Figures 72). Les modifications les plus fortes sont attendues sur les tronçons appartenant aux zones de tête de bassin (Figures 70 et 71), probablement du fait d'une plus grande représentation des tronçons de têtes de bassin sur le réseau hydrographique. L'intensité de modification semble structuré le long du gradient amont-aval mais aussi par zones hydrographiques. Par exemple, les tronçons appartenant au Massif Central, sur la zone sud du bassin, montrent un taux de changement moindre à l'horizon H2050 que les tronçons de même ordre de Strahler situés ailleurs sur le bassin. La répartition des valeurs de TBI (histogrammes en figures 72 et 70) montre une intensification du

changement entre les horizons H2050 et H2080 à l'échelle du bassin de la Loire, avec une augmentation de la fréquence des valeurs de TBI plus fortes ($TBI > 0.6$). Le partitionnement du TBI montre la coexistence des phénomènes de pertes et gains de taxons, avec une répartition spatialisée sur le bassin, et selon une agrégation spatiale par tronçons proches (Figure 73). La part de pertes est plus élevée (proportion de tronçons où la perte d'abondance de taxon est dominante aux horizons H2050 et H2080, respectivement : 60,34 % et 64,96 %) et s'intensifie entre les deux horizons temporels, notamment sur la zone du Massif Central.

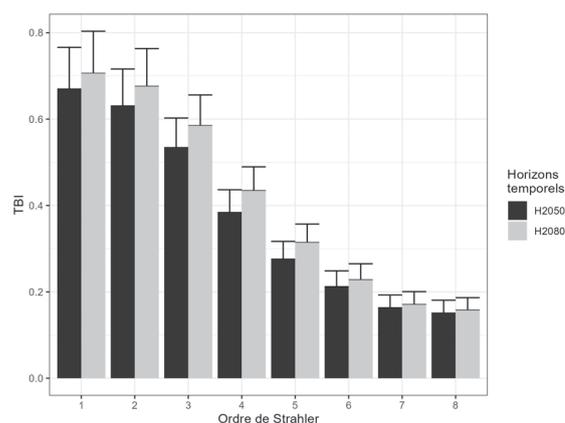


FIGURE 70
Temporal Beta Diversity Index (TBI) moyen pour les deux horizons temporels H2050 et H2080 par ordre de Strahler sous le scénario RCP 8.5.

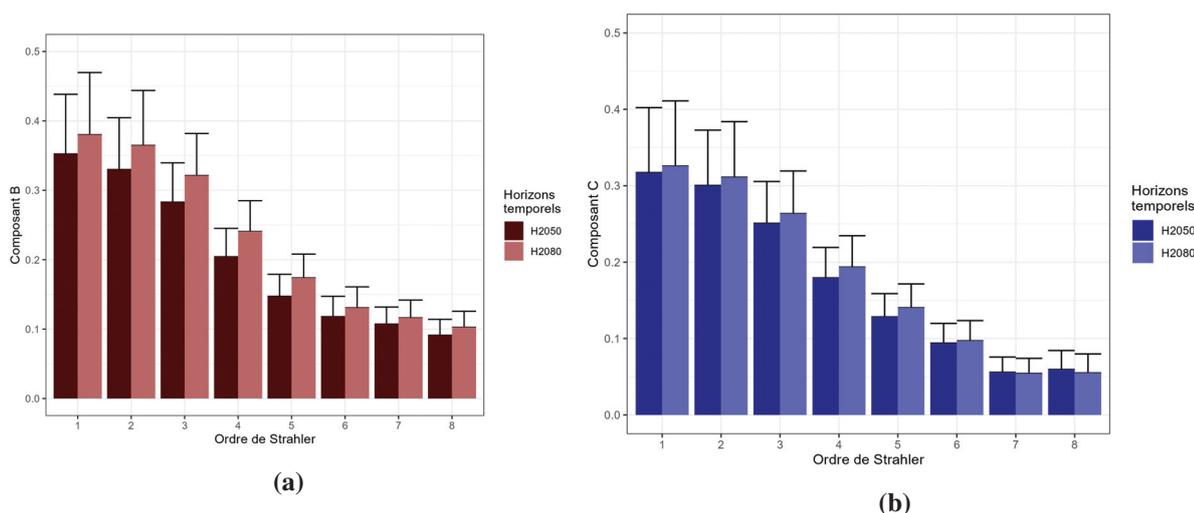


FIGURE 71
Composants B (pertes ; panneau a en rouge) et C (gains ; panneau b en bleu) du Temporal Beta Diversity Index (TBI) entre l'horizon H2010 et les horizons H2050 (barres foncées) et H2080 (barres claires) par ordre de Strahler sous le scénario RCP 8.5.

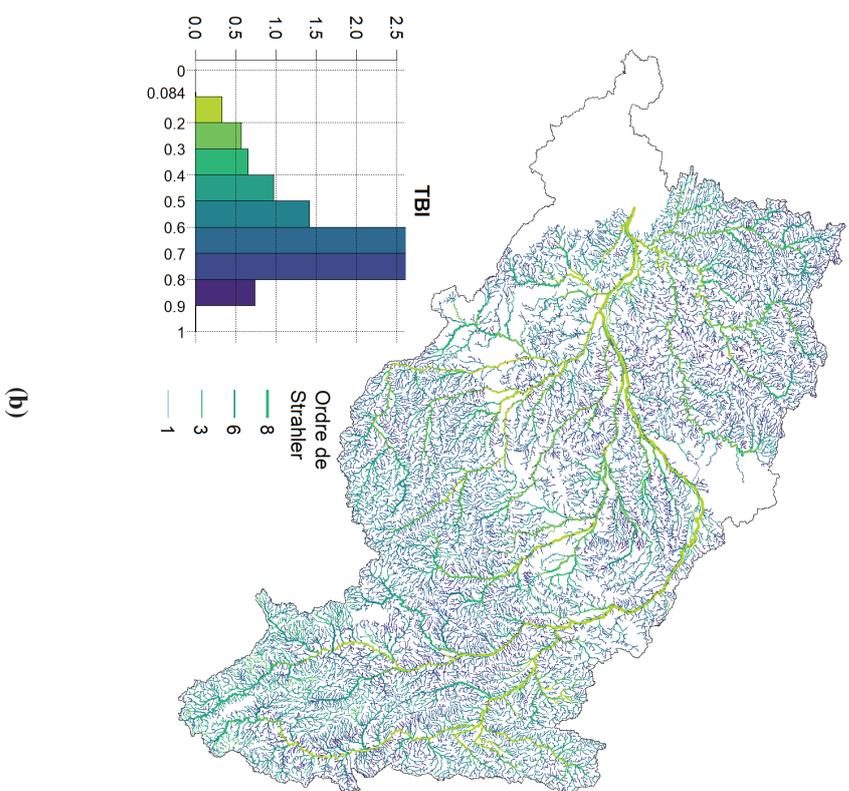
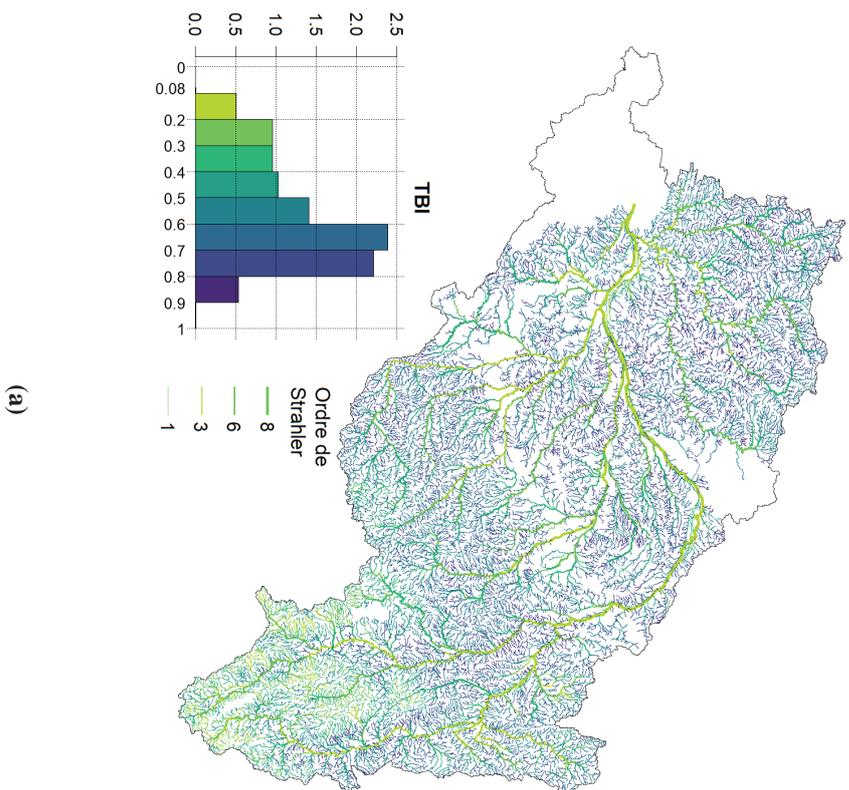
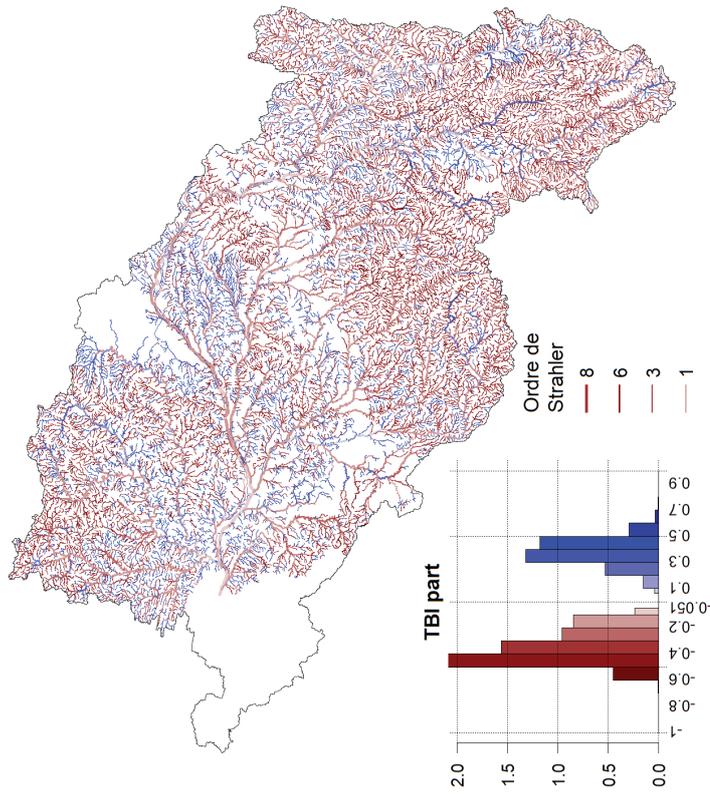
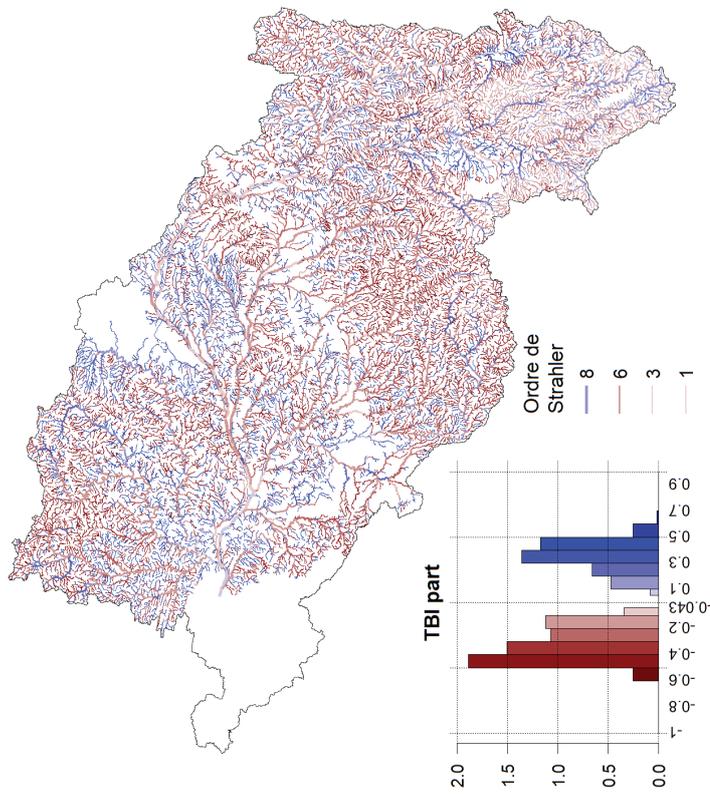


FIGURE 72 *Temporal Beta Diversity Index (TBI) calculés entre les horizons H2050 et H2010 (a), et entre H2080 et H2010 (b). D'après le test de permutation, toutes les valeurs de TBI sont significatives.*



(a)



(b)

FIGURE 73
Partitionnement du Temporal Beta Diversity Index (TBI) en gains (composant C ; dégradé de bleu) et pertes (composant B ; dégradé de rouge) pour les deux horizons temporels H2050 (a) et H2080 (b).

F. Conclusion et perspectives

Les projections des distributions potentielles de 57 taxons aquatiques sous le scénario RCP 8.5 à deux horizons temporels ont permis de montrer que les communautés aquatiques sont amenées à être profondément transformées sous l'effet du changement climatique d'ici la fin du 21^{ème} siècle. À l'échelle spécifique, différents patrons de réponses sont observés sur un gradient entre contraction et expansion de l'occupation du réseau hydrographique. L'hypothèse de corrélation entre les changements d'aire de répartition et d'effectifs des classes d'abondance n'a pas toujours été vérifiée. La complexité des patrons de réponses montre bien l'importance d'intégrer les abondances dans l'analyse des effets du changement climatique sur les communautés. Celles-ci sont amenées à se réorganiser suite à la modification des aires de répartition des taxons, et certains assemblages de taxons particuliers pourraient disparaître tandis que de nouveaux pourraient apparaître. D'après les projections, les changements les plus importants se produiront vraisemblablement dans les cours d'eau de tête de bassins, notamment au travers d'une homogénéisation biotique.

Les projections présentées ici comportent certaines limites :

- (i) Les données de température de l'eau et des variables hydrauliques proviennent de simulations et comportent donc des incertitudes. Bien que la gamme de ces incertitudes soient faibles (*e.g.* le biais moyen est de $0,2 \pm 0,82$ °C pour les simulations de T-NET ; Beaufort *et al.* 2016), il conviendrait d'évaluer la sensibilité des projections de distribution spatiale des abondances à ces incertitudes ;
- (ii) Les SDM produits au chapitre précédent et utilisés ici pour projeter les distributions spatiales des abondances présentent différents niveaux de performance (voir Figure 43). Bien que seuls les taxons pour lesquels les performances ont été jugés suffisantes aient été conservés, certains présentent des niveaux de performance plus faibles (AUC < 0,6 pour 6 taxons) et donc une plus forte incertitude quant à leur projection. Ces taxons sont donc à considérer avec plus d'attention, et l'impact de leur intégration à l'échelle de la communauté pourrait être évalué. Par exemple, le SRC (Figure 63) de 5 des 6 taxons aux plus faibles AUC est proche de -100. Il serait donc intéressant d'évaluer si ces valeurs sont bien liées à une sensibilité aux modifications de conditions environnementales ou à une incertitude de modélisation. Il s'agit néanmoins de taxons dont les abondances sont généralement faibles ;

- (iii) Les images obtenues via la modélisation ne sont que des représentations des distributions potentielles des abondances en fonction de la favorabilité de l'habitat. Par conséquent, rien ne garantit que certains processus ne vont pas entraver l'occupation de tous les habitats favorables par les taxons. Par exemple, aucune prise en compte des capacités des espèces à atteindre les zones favorables, que ce soit leur capacité de dispersion ou les barrières physiques à leur dispersion (*i.e.* obstacles à l'écoulement), n'a été faite au cours de la modélisation. Il apparaît donc à l'avenir nécessaire d'intégrer cette composante dans la modélisation, en identifiant les tronçons atteignables par chaque taxon respectivement (Boulangeat *et al.*, 2012 ; Caradima *et al.*, 2021) ;
- (iv) Les données des espèces utilisées pour la modélisation sont déjà soumises aux effets du changement climatique et montrent déjà des patrons de réponses dans leur distribution spatiale et leurs niveaux d'abondances (Comte *et al.*, 2013 ; Kuczynski *et al.*, 2018). Il n'est donc pas impossible que les taxons observés n'occupent pas les conditions environnementales les plus favorables parmi les conditions disponibles sur le bassin (*mismatch*). Par exemple, Erickson & Smith (2023) ont montré l'inadéquation entre la niche fondamentale de six espèces marines, obtenues à partir d'expérimentations en laboratoire, et la niche réalisée à partir de données sur la période 1900 à nos jours ;
- (v) Les espèces sont susceptibles d'adapter leur réponse sur les plans phénologiques et physiologiques (Baptist *et al.*, 2014) ou encore de développer des stratégies d'évitement, par exemple par des adaptations comportementales (Carbonell *et al.*, 2021). Il conviendrait donc d'évaluer les capacités d'adaptation des espèces en évaluant les stratégies déjà mises en place.

Le regroupement de l'ensemble des projections issues des modèles construits pour chaque espèce afin de dresser un portrait des communautés comprend aussi certaines limites. Les interactions intra et inter-spécifiques qui influencent les dynamiques de populations et donc la structure de la communauté ne sont pas prises en compte. L'évolution de ces interactions pourrait entraîner un changement des fréquences d'abondance de certaines espèces du fait d'une augmentation de la compétition ou de la perte de proies. C'est dans cette idée qu'il a été tenté d'intégrer un indicateur de la ressource macroinvertébrés dans

les modèles poissons au chapitre IV. De plus, de nouveaux assemblages sont susceptibles de se former, avec la mise en contact d'espèces auparavant séparées, dont les potentielles interactions sont inconnues.

Malgré une évolution conjointe des variables environnementales aux deux horizons temporels H2050 et H2080, les analyses ne montrent pas une accélération des modifications des communautés très marquée entre ces deux périodes. Bien que les variables portant sur la température de l'eau et l'hydraulique des tronçons soient les plus structurantes, le taux de modification observé sous l'effet de ces conditions environnementales seules n'engendre pas une intensification très marquée des phénomènes de modifications des communautés, ce qui témoigne probablement de l'importance des autres prédicteurs dans la structuration spatiale des communautés aquatiques. C'est pourquoi il est important de proposer des scénarios d'évolutions de ces prédicteurs (*i.e.* occupation des sols, urbanisation et obstacles à la continuité écologique) et de les intégrer dans les projections pour évaluer l'effet conjoint des modifications environnementales à venir sur les distributions futures des espèces.

En résumé ...

- * Les modèles de distribution de classes d'abondance développés au chapitre IV ont permis de projeter la distribution future potentielle de 57 taxons (26 espèces de poissons et 31 genres de macroinvertébrés) sous le scénario RCP 8.5 de changement climatique, aux horizons temporels 2050 et 2080 ;
- * Les changements de distribution des taxons se répartissent de manière équilibrée entre expansion et régression, avec une modification des effectifs des classes d'abondance qui ne dépend pas du type de modification de la distribution du taxon. Ainsi, l'hypothèse d'une corrélation entre évolution de l'aire de répartition et des effectifs des classes d'abondance n'est pas toujours vérifiée. Ce constat appuie l'importance de coupler l'étude des changements de distribution des espèces à l'évolution potentielle des abondances dans le cadre du changement climatique ;
- * Les modifications de distribution des espèces entraînent des changements importants à l'échelle des communautés : la richesse spécifique est amenée à augmenter en direction des milieux amont, avec la progression des espèces en expansion et celles qui suivent leur niche écologique. En parallèle, les communautés tendent à s'homogénéiser du fait de l'expansion des taxons ubiquistes et de la régression des taxons de tête de bassins ;
- * L'intégration de scénarios d'évolution des prédictors considérés comme fixes dans le présent travail (*i.e.* gradient de pression agricole, densité de population et obstacles à l'écoulement) et des interactions biotiques pourrait permettre d'affiner les projections futures des modèles.

Chapitre VI

Comparaison des patrons de structuration des communautés entre grands bassins versants français

A. Introduction

Les modèles de distribution des abondances des communautés de poissons et de macroinvertébrés ont permis d'identifier les déterminants environnementaux majeurs de la distribution spatiale des abondances (chapitre IV) et de dresser des scénarios d'évolutions potentielles futures sous l'influence du changement climatique (chapitre V). Les analyses réalisées à l'échelle du bassin versant de la Loire (chapitres III à V) ont été permises grâce à la disponibilité des données environnementales à haute résolution spatiale (*i.e.* température de l'eau, variables hydrauliques) issues des modèles à bases physiques (*i.e.* T-NET et EROS). En reflétant plus directement l'habitat des organismes aquatiques (chapitre III), ces données ont permis la construction de modèles performants (chapitre IV). Cependant, elles ne sont pas (encore) disponibles sur les autres grands bassins versants français d'intérêts, à savoir la Seine, la Garonne et le Rhône. Afin d'évaluer l'extension des résultats obtenus sur le bassin ligérien à ces autres contextes hydrographiques, ce chapitre s'intéresse au degré de similarité des patrons de structuration des communautés et des déterminants environnementaux sous-jacents entre ces bassins versants. Le degré de similarité entre les patrons de différents bassins versants est considéré comme un proxy du niveau de transférabilité des résultats obtenus sur la structuration spatiale des communautés entre ces bassins versants.

La question de la transférabilité des patrons de structuration spatiaux des communautés est assez prégnante en écologie puisqu'elle s'inscrit dans la recherche des patrons de structuration et processus sous-jacents généraux. En écologie aquatique, cela s'est traduit par la formalisation des typologies et zonations de cours d'eau (*e.g.* Huet 1959 ; Verneaux 1973 ; Vannote *et al.* 1980, voir chapitre I), qui présentent différentes limites d'applications systématiques si aucune précaution préalable n'est prise (Wasson 1989 ; chapitre I). Par exemple, la composition des communautés biologiques de chaque bassin versant présente ses propres spécificités, du fait d'histoires biogéographiques différentes (chapitre I section D.2), avec son lot d'espèce endémique et/ou rares, qui limitent l'utilisation d'espèces indicatrices « universelles » dans la détermination du niveau typologique. Plus récemment, la généralisation de l'utilisation des modèles de distribution des espèces a aussi posé la question de leur transférabilité des SDM entre différents contextes spatiaux (*e.g.* entre bassins versants) et temporels (*e.g.* projections futures sous scénarios de changement climatique ; Elith *et al.* 2010 ; Valavi *et al.* 2023). Bien que certaines tentatives d'extra-

polation se soient avérées fructueuses (Gies *et al.*, 2015 ; Valavi *et al.*, 2023), la qualité des projections est difficilement vérifiable, et les succès restent rares (Elith *et al.*, 2010 ; Guisan *et al.*, 2017). En effet, même lorsque les conditions environnementales considérées sont similaires, un SDM calibré sur les données issues d'un certain bassin versant est rarement performant lorsqu'il est appliqué à un autre bassin (*i.e.* contexte spatial différent ; Gies *et al.* 2015 ; Guisan *et al.* 2017). Ceci peut s'expliquer par certaines particularités du bassin de projection qui n'ont pas été prise en compte lors de la calibration du modèle, comme les différences d'occupation des sols (Gies *et al.*, 2015), soit par omission d'au moins une variable influente ou encore par manque de représentativité de toutes les gammes de valeurs de certaines variables dans le jeu de calibration (Guisan *et al.*, 2017). Cette limite est néanmoins propre à la modélisation et n'empêche pas de chercher à identifier les déterminants communs à la structuration des communautés entre plusieurs bassins versants, à l'instar des classifications de cours d'eau (Matthews, 1998).

D'après la théorie de filtres, les espèces trouvées localement sont définies comme celles dont les caractéristiques écologiques leur ont permis de passer les différents filtres définis aux échelles spatiale supérieures (Poff, 1997). Vehanen *et al.* (2020) ont ainsi montré le rôle prédominant de l'écorégion, et notamment de la géologie, dans la définition des différents pools d'espèces de poissons à l'échelle locale, de même pour Heino (2002) sur les communautés de macroinvertébrés. Ces deux études ne sont pas sans rappelés les hydroécorégions françaises (Figure 74 ; Wasson *et al.* 2002) qui définissent les conditions environnementales des cours d'eau en fonction de l'hydrologie et de la géologie. Au vu des hydroécorégions très différentes qui composent les conditions environnementales des grands bassins versants étudiés dans ce chapitre, la question du degré de similarité des patrons de structuration de leur communauté est complexe.

Ce chapitre est ainsi dédié à l'étude de la similarité des patrons spatiaux de communautés aquatiques à l'échelle du bassin versant, en particulier les successions longitudinales de communautés, et de leurs relations avec les gradients environnementaux. La section B présente les données biologiques et environnementales mobilisées pour cette analyse sur les quatre bassins versants considérés, à savoir la Loire, la Seine, le Rhône et la Garonne. Ces quatre bassins ont été choisis pour leur taille, de même ordre de grandeur que celui de la Loire, et parce qu'ils font partis des systèmes à enjeux concernés par le cadre de cette thèse. La section C est ainsi consacrée à la comparaison des communautés aquatiques et

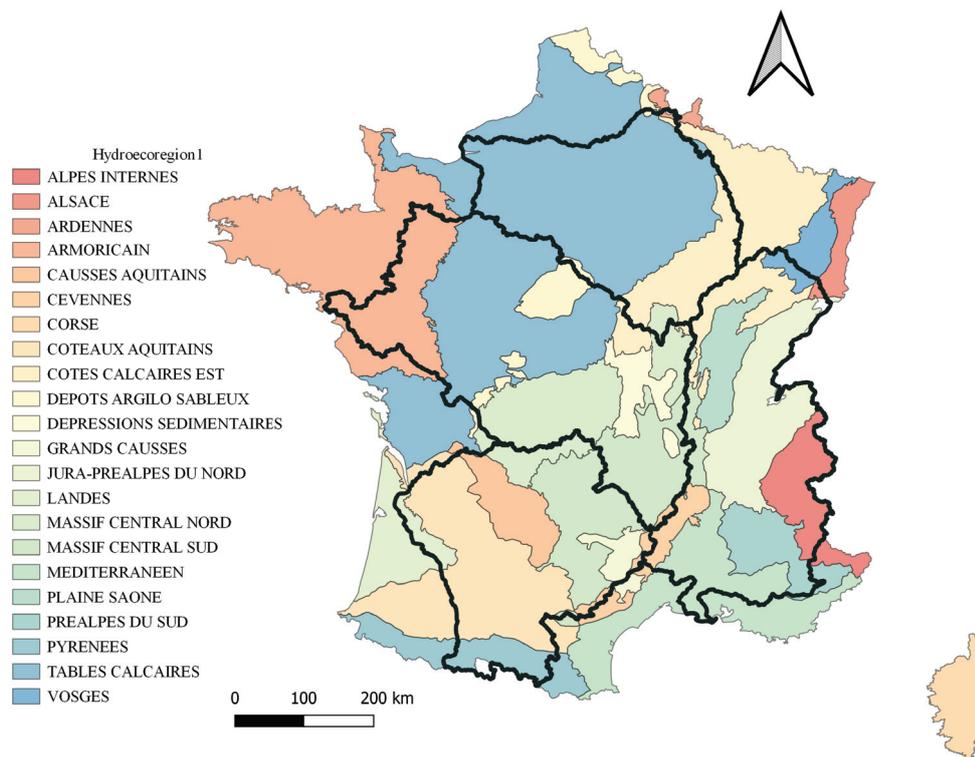


FIGURE 74

Hydroécocorégions de niveau 1 à l'échelle de la France. Les bassins versants d'intérêts, la Loire, la Seine, la Garonne et le Rhône sont encadrés en trait plein noir. D'après Wasson *et al.* (2002).

de leur lien avec les variables environnementales structurantes le long du gradient amont-aval entre les quatre bassins versants étudiés. Enfin, la composition des communautés, et notamment la proportion d'espèces non partagées entre les bassins, fait partie des facteurs potentiellement limitant pour la transférabilité de résultats entre bassins. La section D évalue donc l'influence de la proportion d'espèces communes aux différents bassins sur le degré de similarité des patrons détectés.

B. Données biologiques et environnementales à l'échelle des quatre bassins versants

Les données biologiques mobilisées pour cet axe de travail proviennent de la base Nâïades, déjà présentée au chapitre II. Les données ont été extraites pour les quatre bassins versants d'intérêt, à savoir la Loire, la Seine, le Rhône et la Garonne. Certaines données environnementales mobilisées jusque là pour l'étude du bassin ligérien (*e.g.* modélisations hydrologiques issues de EROS et modélisations de la température de l'eau issues de T-NET) n'étant disponibles à l'échelle de ces quatre bassins, des proxies des conditions

environnementales ont été collectés à une résolution spatiale plus grossière.

B.1. Données biologiques

Le jeu de données poissons extrait à l'échelle nationale est constitué par 12 103 439 individus appartenant à 109 espèces échantillonnées sur 4 014 stations au cours de 22 739 opérations sur la période 1994-2019. Le jeu de données macroinvertébrés extrait à l'échelle nationale comporte lui 65 138 631 individus de 343 taxons appartenant à 12 ordres différents échantillonnés sur 3 195 stations au cours de 22 885 opérations sur la période 1992 - 2017. Ces deux jeux de données ont fait l'objet du même tri qualité que celui présenté au chapitre II (Figure 75) et d'une sélection des données sur la période temporelle 2007-2017 ainsi que des opérations sur les saisons de juin à octobre afin de limiter un éventuel biais temporel. L'ensemble des genres macroinvertébrés présents dans la base Naïades a été conservé dans la suite des analyses.

Le réseau T-NET n'étant disponible qu'à l'échelle du bassin versant de la Loire, le Réseau Hydrographique Théorique (RHT) a été choisi comme unité géographique. Le RHT est un réseau théorique construit à partir de la BD CARTHAGE et constitué de 114 601 tronçons et 283 639 km de linéaire de réseau de cours d'eau à l'échelle nationale (Pella *et al.*, 2012). Cela représente 76 461 tronçons sur les quatre bassins cibles. Le même travail de jointure des données biologiques au réseau RHT a été accompli, avec l'ensemble des stations de suivis hydrobiologiques disponibles sur les quatre bassins cibles. Les stations biologiques ont été projetées sur le tronçon le plus proche dans un rayon de 200 m, les stations non projetées ont été raccordées à la main quand cela était possible. Enfin, les abondances de chaque taxon ont été standardisées en nombre d'individus pour 100 m² (voir chapitre II B.2.1 pour la formule). Au final, 1 512 stations poissons et 2 888 stations macroinvertébrés ont été raccordées au réseau RHT, respectivement sur 1 484 et 2 848 tronçons (Figure 75). Le nombre de tronçons varie en fonction du bassin versant (Tableau 12) en revanche le nombre de taxons est proche entre les bassins.

TABEAU 12

Informations concernant les jeux de données biologiques sur les quatre bassins versants considérés pour l'analyse : Seine, Loire, Rhône et Garonne, sur la période 2007 - 2017. Le nombre de tronçons RHT portant une station d'échantillonnage montre que l'effort d'échantillonnage est plus important sur le bassin de la Loire, notamment pour les macroinvertébrés.

	Poissons			Macroinvertébrés	
	Taille (en km ²)	Nombre de tronçons échantillonnés	Nombre d'espèces	Nombre de tronçons échantillonnés	Nombre de taxons
Garonne	55 846	312	41	688	184
Loire	117 800	535	42	1 122	184
Rhône	97 800	416	45	584	185
Seine	78 650	221	36	454	184

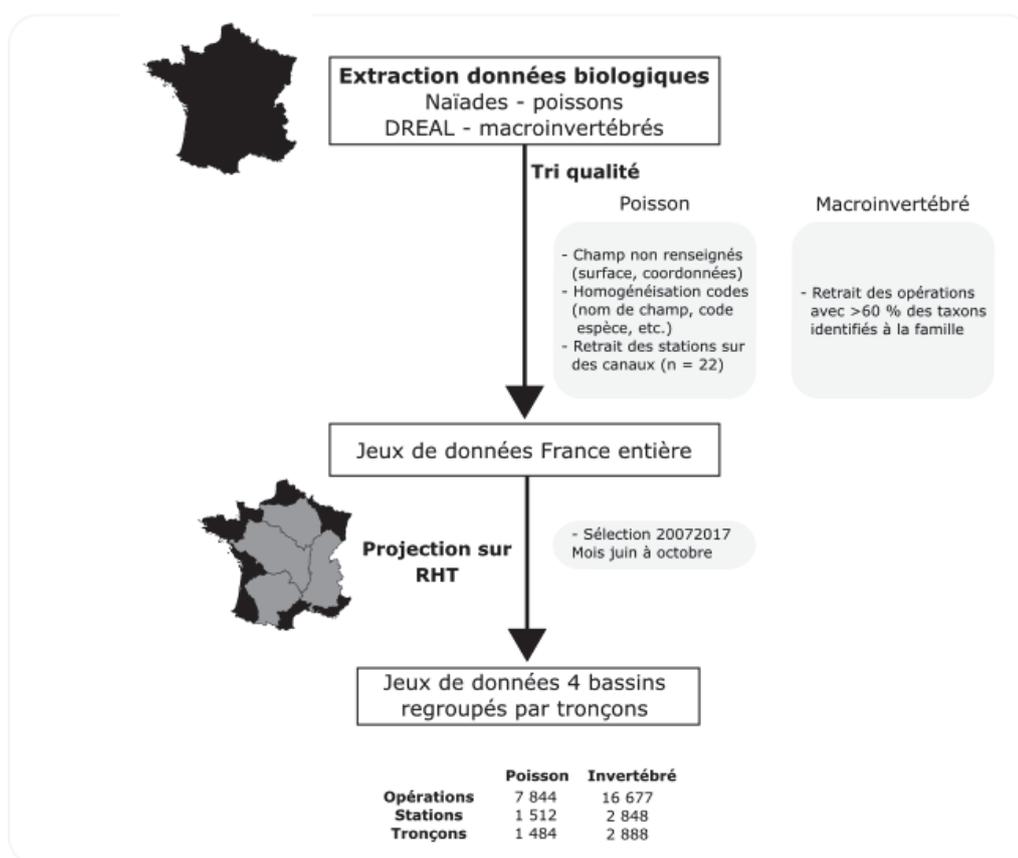


FIGURE 75

Schéma de l'extraction et de la constitution du jeu de données biologiques.

B.2. Données environnementales à l'échelle nationale

B.2.1. Les données de température de l'eau

Le projet TIGRE a permis de reconstituer, par des modèles géostatistiques, deux métriques des températures de l'eau à partir des caractéristiques météorologiques et physiographiques pour l'ensemble du RHT (Beaufort *et al.*, 2022) : la moyenne interannuelle de la température de l'eau des 30 jours consécutifs (MT_w30j) les plus chauds de l'année et celle des 7 jours consécutifs les plus chauds de l'année (MT_w7j) (voir www.thermie-rivieres.inrae.fr). Les données de thermie observées ont été collectées de 2009 à 2017 sur 2 701 stations issues de différents réseaux de suivis (*i.e.* Réseau National Température de l'OFB, suivis des fédérations de pêches). Les métriques MT_w30j et MT_w7j ont été modélisées à l'échelle de la France en utilisant les données climatiques issues de SAFRAN (Quintana-Segui *et al.*, 2008 ; Vidal *et al.*, 2010), les variables morphométriques et hydrologiques du RHT (Pella *et al.*, 2012) et les variables de pressions anthropiques sur la morphologie des cours d'eau provenant de la base SYstèmes Relationnel d'Audit de l'Hydromorphologie des cours d'eau (SYRAH) (Valette *et al.* 2012 ; Figure 76). Les métriques utilisées résultent de la combinaison de plusieurs modèles : réseaux de neurones, *random forest* et régression multiple linéaire. Ces métriques MT_w30j et MT_w7j ont été pensées pour une utilisation dans un cadre écologique. En effet, leur utilisation est retrouvée dans la littérature scientifique qui a souligné le rôle de ces métriques dans la structuration des communautés aquatiques (Verneaux, 1977b ; Caissie, 2006 ; Beaufort *et al.*, 2020a, 2022).

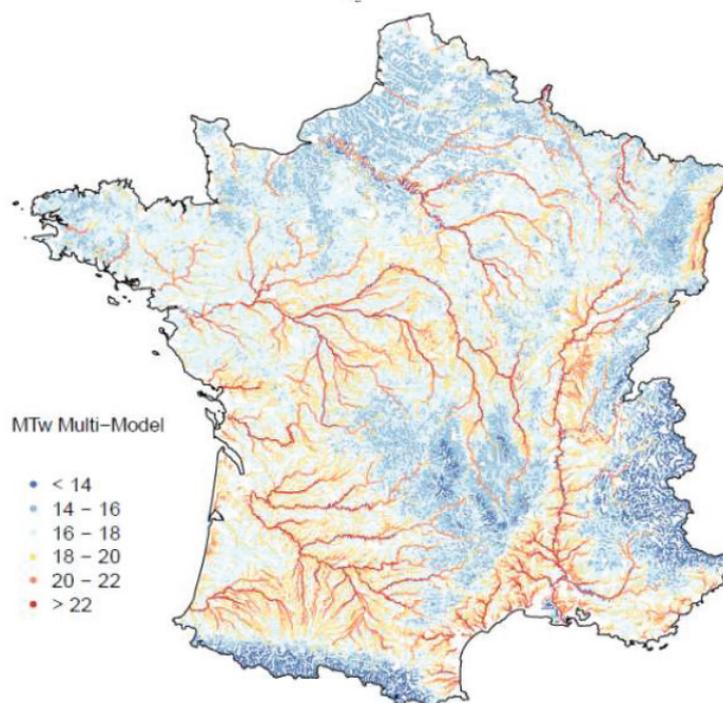


FIGURE 76

Moyenne interannuelle de la température de l'eau sur les 30 jours consécutifs les plus chauds (MT_w30j) calculée sur la période 2009-2017. D'après le projet TIGRE et Beaufort *et al.* (2022).

B.2.2. Les données d'hydrologie et de topographie

Les variables sur la topographie sont fournies avec le RHT et les variables sur l'hydrologie proviennent du modèle EROS. L'ensemble des variables environnementales utilisées ainsi que leur provenance sont présentées dans le tableau 13.

C. Similarité de la structuration des communautés et de ses déterminants environnementaux entre bassins versants

C.1. Comparaison de la structuration des communautés entre bassins versants

Afin d'étudier la structuration spatiale des communautés de poissons et de macroinvertébrés entre les bassins versants, la méthode *Non-metric Multidimensional Scaling* (NMDS) a été utilisée (Legendre & Legendre, 2012). Cette méthode d'ordination se base sur une matrice de distance et est adaptée à l'ordination de données d'abondance, notamment en présence de taxons rares et/ou abondants car elle y est moins sensible que d'autres méthodes d'ordinations multivariées (Minchin, 1987). La distance de Bray-Curtis a été cal-

TABLEAU 13

Liste des variables environnementales mobilisées pour l'analyse de la similarité entre les quatre grands bassins versants Seine, Loire, Rhône et Garonne.

Variable	Origine	Référence
MT _w 7j	Projet TIGRE	Beaufort <i>et al.</i> , 2020, 2022
MT _w 30j	Projet TIGRE	Beaufort <i>et al.</i> , 2020, 2022
Distance à la source	RHT (MNT 50 m IGN BD Alti)	Pella <i>et al.</i> , 2012
Altitude	RHT	Pella <i>et al.</i> , 2012
Pente	RHT	Pella <i>et al.</i> , 2012
Hauteur au module	RHT	Lamouroux <i>et al.</i> , 2010
QMNA5	RHT	Catalogne, 2012

culée sur les données standardisées par la méthode *wisconsin* et 100 itérations ont été utilisées. Les analyses ont été réalisées sur les données poissons et macroinvertébrés séparément. L'ensemble des analyses présentées ici ont été réalisées avec le package *vegan* (Oksanen *et al.*, 2019).

Une *PERmutational Multivariate ANalysis Of VAriance* (PERMANOVA) a été appliquée pour tester l'existence d'une différence significative entre les groupes formés par chacun des bassins versants, illustré par les ellipses sur les Figures 80 et 81 (Anderson, 2001). La PERMANOVA est réalisée avec 999 permutations à partir de la matrice de distance de Bray-Curtis calculée sur les données biologiques à l'échelle des tronçons. Dans le cas où une différence significative est trouvée, un test de dispersion est appliqué pour déterminer si cette différence provient d'une différence de dispersion du nuage de points entre les groupes ou bien d'une différence de centroïde.

Les Figures 77 et 78 présentent les *spiderplot* de chaque bassin versant. Les centroïdes de chaque bassin sont proches dans l'espace factoriel, néanmoins la PERMANOVA détecte une différence significative entre ces groupes de bassins (poissons p-value = 0.001 ; macroinvertébrés p-value = 0.001). Le test d'homogénéité de la dispersion entre les groupes est significatif, indiquant une différence de dispersion plutôt qu'une réelle différence entre les centroïdes (poissons p-value = 0.001 ; macroinvertébrés p-value = 0.001).

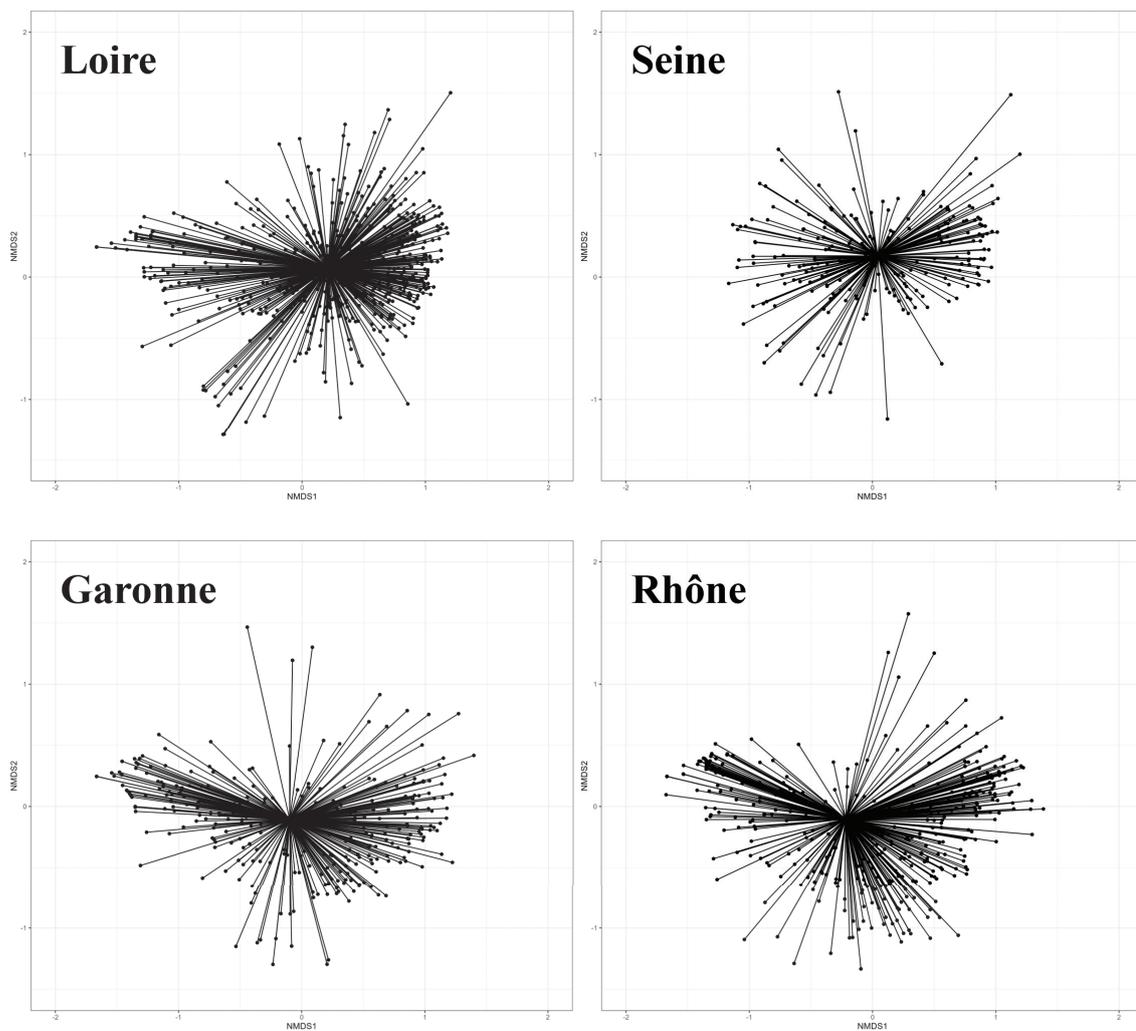


FIGURE 77

Spiderplot de l'ordination basée sur les données poissons des quatre bassins versants. Les *spiderplot* permettent de représenter, pour chaque point ordonné dans l'espace factoriel, la distance (trait plein) au centroïde formé par le nuage des points compris dans le même bassin versant. Les coordonnées de chaque tronçon issues de l'espace factoriel ont été extraites pour chaque bassin et représentées ici séparément à des fins de lisibilité.

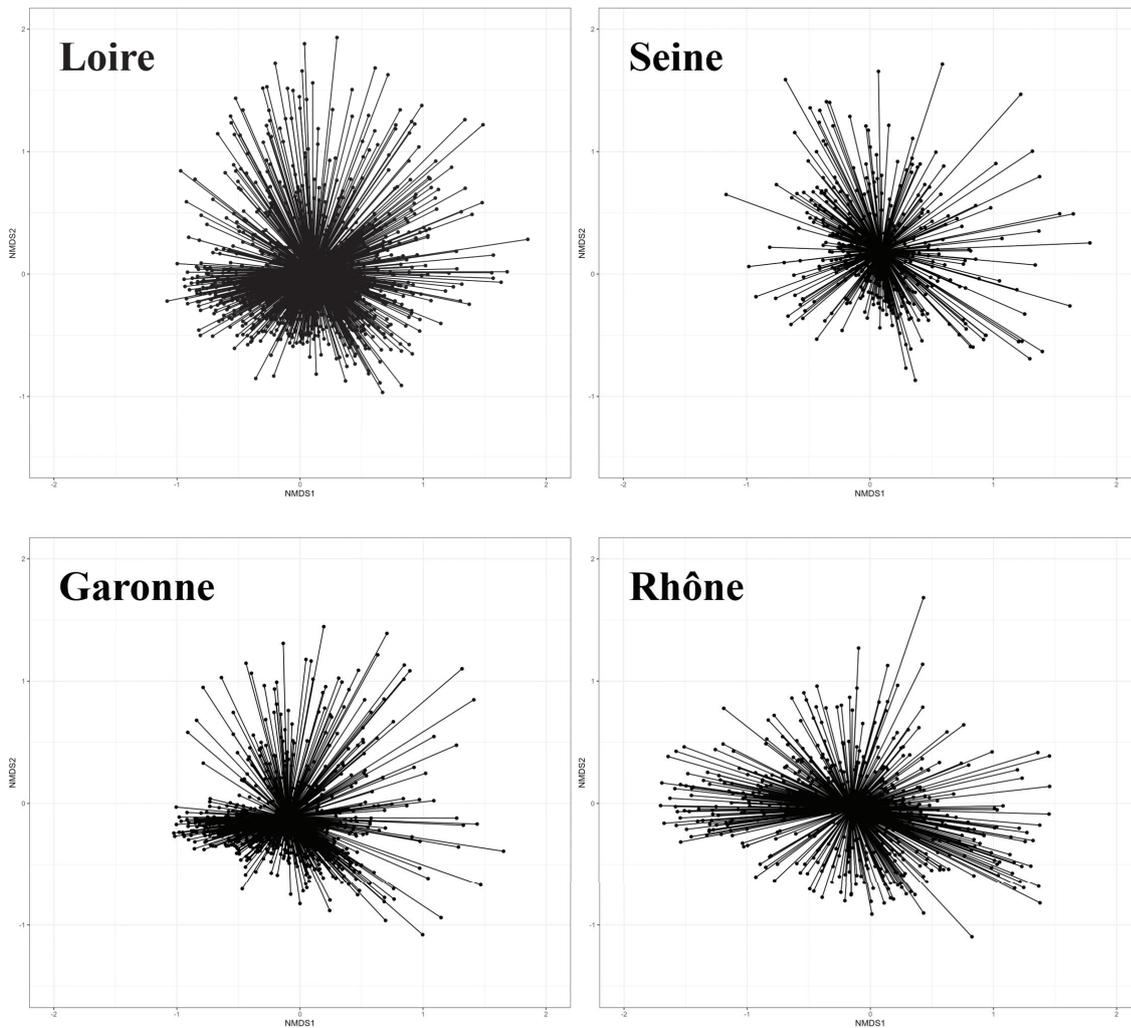


FIGURE 78

Spiderplot de l'ordination basée sur les données macroinvertébrés des quatre bassins versants. Les *spiderplot* permettent de représenter, pour chaque point ordonné dans l'espace factoriel, la distance (trait plein) au centroïde formé par le nuage des points compris dans le même bassin versant. Les coordonnées de chaque tronçon issues de l'espace factoriel ont été extraites pour chaque bassin et représentées ici séparément à des fins de lisibilité.

C. Similarité de la structuration des communautés et de ses déterminants environnementaux entre bassins versants

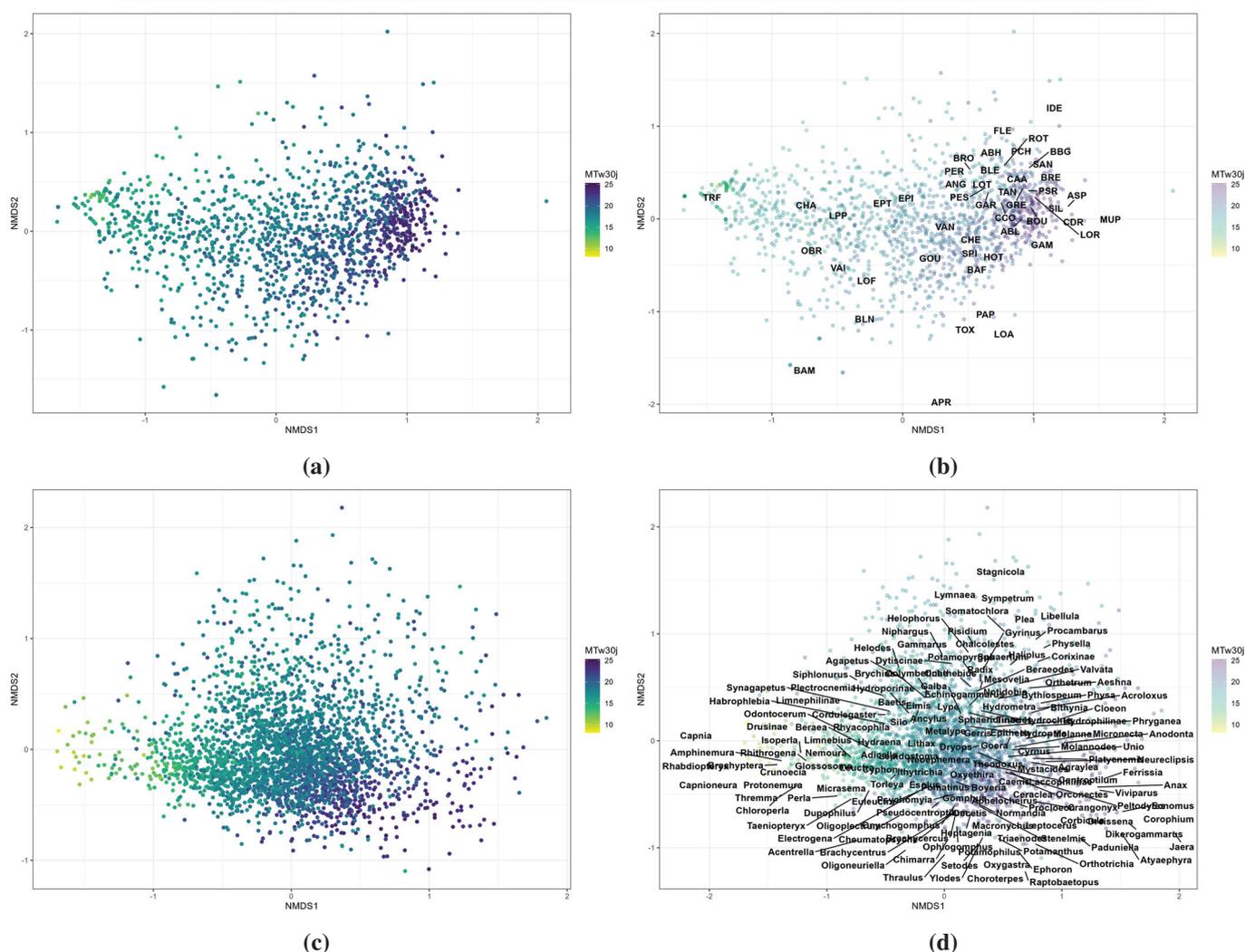


FIGURE 79

Lien entre les données biologiques ordonnées par NMDS et la température de l'eau (MT_{w30j}) pour les poissons (a et b) et les macroinvertébrés (c et d). La température de l'eau sert uniquement de variable illustrative et montre bien une relation avec l'ordination biologique de par sa structuration de gauche à droite. Les figures b et d montrent les centroïdes des espèces/genres le long du gradient thermique. La correspondance avec les codes espèces poissons peut être consultées au tableau 1.

C.2. Déterminants environnementaux de la structuration des communautés

Les valeurs de MT_{w30j} des tronçons projetés sur le plan factoriel de la NMDS (Figure 79. a) ainsi que la position des centroïdes des espèces (Figure 79. b) permettent d'illustrer la corrélation du gradient longitudinal avec la structuration des communautés piscicoles, s'organisant de l'amont (à gauche ; valeurs faibles de MT_{w30j} ; espèces rhéophiles d'eau froide, *e.g.* truite commune, *Rhabdiopteryx*) vers l'aval (à droite ; valeurs fortes de MT_{w30j} ; espèces limnophiles d'eau chaude, *e.g.* brème commune, *Dreissena*). Le signal

semble plus diffus pour l'ordination des macroinvertébrés (Figure 79. c et d), dont la structuration est certainement davantage influencée par d'autres déterminants (voir chapitre III).

Un partitionnement de variance par RDA a été utilisé pour quantifier l'influence des variables environnementales sur la structuration des communautés de poissons et de macroinvertébrés dégagée par les NMDS (Legendre & Legendre, 2012). Les coordonnées de chaque tronçon dans l'espace factoriel à deux dimensions ont été utilisées comme variables réponses. Les groupes de variables explicatives utilisées sont la thermie (MT_w30j et MT_w7j), l'hydrologie (QMNA5 et Q50) et les variables de topographie (distance à la source, altitude et pente du tronçon).

TABLEAU 14
Part de variance expliquée par chaque variable dans la RDA entre les axes des NMDS poissons et macroinvertébrés et les variables environnementales. La part relative de chaque variable est exprimée en %.

Variables environnementales	Poisson - R² (en %)	Macroinvertébrés - R² (en %)
Température moyenne de l'eau sur les 7 jours les plus chauds	23,67	18,83
Température moyenne de l'eau sur les 30 jours les plus chauds	22,90	17,83
Altitude	20,72	25,50
Pente	11,6	10,53
Distance à la source	8,96	12,52
Q50	2,57	3,21
QMNA5	0,85	1,32
Variance totale expliquée	0,5560	0,4664

L'ordre d'importance des variables environnementales est le même pour l'ordination des données poissons et macroinvertébrés (Tableau 14). Les plus explicatives de la structuration dégagée par la NMDS sont MT_w30j , MT_w7j et l'altitude (Tableau 14).

Les régimes hydrologiques sont peu explicatifs de la structuration spatiale des poissons et des macroinvertébrés. Le manque de pouvoir explicatif des variables hydrologiques utilisées ici est peut-être dû à la qualité des variables elle-mêmes. En effet, comme cela a été évoqué dans les précédents chapitres (chapitres III et IV), les débits sont seulement un proxy des habitats hydrauliques, et ne retranscrivent pas toute la variabilité de ces derniers.

La part totale de variance expliquée est plus faible pour les macroinvertébrés (0,4664) que pour les poissons (0,5560). Comme discuté dans les précédents chapitres, bien qu'ils

semblent se structurer en fonction des mêmes déterminants environnementaux que les poissons, d'autres facteurs non pris en compte ici interviennent certainement dans leur structuration (*i.e.* processus locaux, histoire biogéographique ou dynamique de dispersion ; voir chapitre III).

D. Influence du groupe d'espèces sur la transférabilité entre bassins versants

Afin de tester l'influence du groupe d'espèces (ou de taxons pour les macroinvertébrés) sur la similarité des patrons de structuration spatiaux dégagés par analyse multivariée, deux groupes composés de différents taxons ont été construits pour chacun des jeux de données poissons et macroinvertébrés. Nous posons l'hypothèse que la proportion de taxons similaires entre les bassins influe sur le niveau de similarité des patrons de structuration spatiale. Donc, la similarité devrait augmenter avec l'utilisation de taxons communs aux bassins. Cette analyse peut être considérée comme une analyse de sensibilité de la méthode NMDS aux données des espèces rares. Ainsi, le premier groupe comprend la totalité des taxons recensés sur l'ensemble des quatre bassins (groupe 1 ; soit 48 espèces de poissons et 186 genres de macroinvertébrés). Le second comprend seulement les taxons communs aux quatre bassins versants et dont l'occurrence est strictement supérieure à 10% sur chacun des bassins afin de limiter l'effet des espèces rares (groupe 2 ; soit 25 espèces de poissons et 77 genres de macroinvertébrés).

La méthode NMDS a de nouveau été utilisée, avec la même méthodologie que dans la section précédente (*i.e.* matrice de distance de Bray-Curtis, 100 itérations). Une NMDS a été appliquée à chaque jeu de données ne contenant que les abondances des taxons inclus dans les groupes 1 ou 2, et sur les données des quatre bassins simultanément. Les analyses ont été réalisées sur les données poissons et macroinvertébrés séparément.

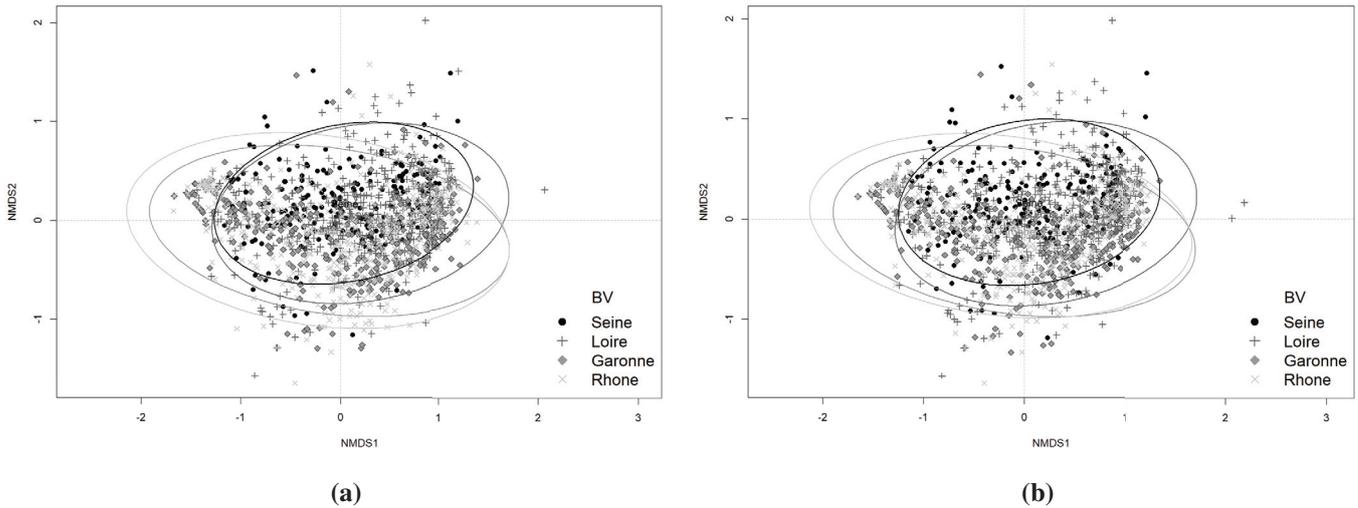


FIGURE 80

Ordinations par NMDS des données poissons en fonction des groupes d'espèces utilisés : (a) groupe 1 complet et (b) groupe 2 constitué des espèces communes aux quatre bassins et restreintes aux espèces dont la fréquence d'occurrence est supérieure à 10% au sein de chaque bassin. Les ellipses regroupent 90% des tronçons de chacun des bassins.

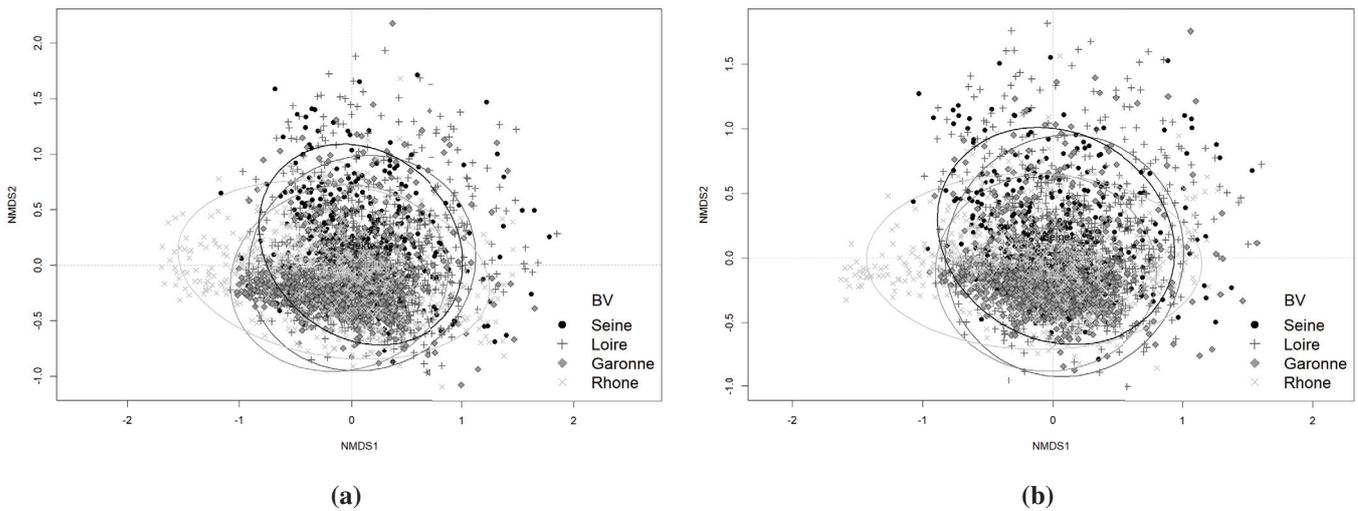


FIGURE 81

Ordinations par NMDS des données macroinvertébrés en fonction des groupes d'espèces utilisés : (a) groupe 1 complet et (b) groupe 2 constitué des genres communs restreints aux taxons dont l'occurrence > 10%. Les ellipses regroupent 90% des tronçons de chacun des bassins.

Les ordinations par NMDS des données d'abondance poissons et macroinvertébrés sur les bassins versants de la Loire, de la Seine, du Rhône et de la Garonne sont présentées aux figures 80 et 81, respectivement. La forme globale des nuages de points formés par les tronçons de chacun des bassins dans la NMDS est très similaire entre les deux groupes de taxons utilisés, que ce soit pour les poissons ou les macroinvertébrés (a vs. b sur les

figures 80 et 81). Ainsi, comme attendu d'après la littérature (Minchin, 1987 ; Vehanen *et al.*, 2020), l'ordination par NMDS ne semble pas ou peu sensible aux espèces rares et/ou peu abondantes. Dans ce cadre, l'hypothèse que le groupe d'espèces considérées influe sur la structuration générale qui peut être dégagée par analyse multivariée est invalidée. La recherche du meilleur groupe d'espèces pour comparer les patrons spatiaux des communautés aquatiques entre plusieurs bassins versants des obtenus à partir de NMDS n'est donc pas pertinente.

Par ailleurs, les tronçons des quatre bassins se superposent en grande partie dans l'espace factoriel, quelque soit le groupe d'espèces considérés, indiquant une relative similarité de leurs communautés. Les ordinations des données macroinvertébrés (Figure 81) montrent des patrons de structuration plus éclatés, ce qui retranscrit la particularité des assemblages de nombreux tronçons, et ce, quel que soit le bassin considéré.

E. Conclusion et perspectives

Les patrons de structuration des quatre grands bassins versants français que sont la Loire, la Seine, le Rhône et la Garonne montrent des similarités vis-à-vis de la structuration le long du gradient amont-aval. Les communautés de poissons et de macroinvertébrés sont également régis par les mêmes grands déterminants environnementaux, soit les températures estivales les plus chaudes (sur 30 jours et 7 jours consécutifs), connus pour être limitantes pour un grand nombre d'organismes aquatiques (Verneaux, 1977b ; Wagner *et al.*, 2023), et la topographie, qui capture l'évolution graduelle des conditions environnementales le long du gradient amont-aval (Huet, 1949 ; Vannote *et al.*, 1980). Dans le cadre des méthodes multivariées et de la résolution des données mobilisées ici, les grands patrons spatiaux étudiés sur la période 2007-2017 sont semblables. Ainsi, les processus structurants les communautés aquatiques des grands bassins versants français sont certainement les mêmes, pour les poissons et les macroinvertébrés, et il n'est pas exclu que les processus régionaux soient dominants dans la structuration des communautés à l'échelle locale, comme l'ont montré Heino (2002) et Vehanen *et al.* (2020). Les patrons de structuration des communautés ligériennes peuvent donc approximer les communautés des autres grands bassins versants, du moins à la large échelle que constitue la taille de ces bassins versants.

L'échelle spatiale de l'étude a une grande influence sur les patrons de structurations qui peuvent être dégagés (Paavola *et al.*, 2006 ; Vehanen *et al.*, 2020). Néanmoins, la plupart des études s'intéressent à la taille de la zone d'étude, mais nous pensons que c'est surtout la résolution spatiale (*i.e.* taille de l'unité géographique par rapport à l'étendue spatiale) qui importe. L'étude de l'influence de différentes résolutions spatiales sur les résultats de structuration qui peuvent être dégagés permettraient de fournir des recommandations pratiques sur la résolution idéale en fonction de l'objectif d'étude. Pour tester cela à partir de nos données, la résolution actuelle pourrait être dégradée en groupant les données de tronçons proches et en relançant le même protocole d'analyses afin de comparer les résultats aux différentes résolutions spatiales.

Le changement climatique opère des changements drastiques sur les communautés aquatiques actuelles (Comte *et al.*, 2013 ; Dudgeon, 2019), et les communautés aquatiques futures vont être amenés à se réorganiser (Daufresne & Boët, 2007 ; Woods *et al.*, 2023 ; voir aussi chapitre V). Les analyses menées dans ce chapitre portent sur les communautés récentes, sur la période 2007-2017, qui pour l'instant sont suffisamment similaires, d'après nos analyses, pour permettre une certaine transférabilité des résultats du bassin à la Loire aux autres bassins. Néanmoins, avec les changements globaux, aucune certitude ne garantit que cette similarité sera maintenue. Les données nécessaires à la modélisation de la distribution des abondances ne sont pas (encore) disponibles sur les autres grands bassins, mais dès que des variables environnementales directes, à haute résolution spatiale seront disponibles, l'évaluation d'éventuels changements dans cette similarité seront à réaliser. Les événements qui ont induits des modifications fortes par le passé ont montré l'importance des différences structurelles de chacun des bassins versants (Dias *et al.*, 2014 ; Keith *et al.*, 2020). Au vu de la littérature et des résultats présentés au chapitre V, une importante réorganisation des communautés aquatiques française est attendue, avec la survenue des processus de la colonisation des communautés des zones avales vers l'amont et d'homogénéisation biotique (Daufresne & Boët, 2007 ; Buisson *et al.*, 2008b ; Maire *et al.*, 2019).

En résumé ...

- ★ Un certain niveau de similarité des patrons de structuration des communautés est supposé traduire une certaine transférabilité de ces patrons entre bassins versants. Afin d'évaluer la portée des résultats obtenus sur le bassin versant de la Loire aux précédents chapitres, la similarité des patrons spatiaux des communautés de poissons et de macroinvertébrés est examinée entre quatre grands bassins versants français : la Loire, la Seine, la Garonne et le Rhône ;
- ★ Les *Non-metric Multidimensional Scaling* (NMDS) couplées à des *PERmutational Multivariate ANalysis Of VAriance* (PERMANOVA) et du partitionnement de variance par *Redundancy Analysis* (RDA), permettent de montrer une certaine similarité entre les patrons de structuration spatiaux des communautés aquatiques de poissons et de macroinvertébrés des quatre grands bassins versants français, avec une structuration graduelle des communautés le long du gradient amont-aval. Les déterminants environnementaux majeurs sont partagés entre les communautés des différents bassins et sont les températures de l'eau sur les 30 jours consécutifs les plus chauds et celles des 7 jours consécutifs les plus chauds, ainsi que la topographie. Ceci, indique que ce sont vraisemblablement les mêmes processus environnementaux qui interviennent dans la structuration des communautés à notre résolution spatiale (*i.e.* grand bassin versant décrits par des tronçons de cours d'eau) ;
- ★ Une analyse de sensibilité de la NMDS à la proportion d'espèces rares considérée a été menée. Cette méthode multivariée s'est révélée très peu sensible aux espèces rares et/ou peu abondantes, comme attendu d'après la littérature. Nous recommandons donc l'utilisation de cette technique d'ordination en présence de nombreux taxons rares ;
- ★ Les changements projetées de la structuration des communautés au chapitre V, à savoir une réorganisation des assemblages d'espèces et une homogénéisation biotique, sont plausible sur les autres grands bassins versants français, mais rien ne garantit que le niveau de similarité observé ici sera maintenu.

Chapitre VII

Conclusion et perspectives de recherche

A. Bilan des travaux

L'objectif général de cette thèse consistait à analyser la structuration spatiale actuelle des communautés aquatiques à large échelle (bassin versant de la Loire) et évaluer sa potentielle évolution dans un contexte de changement climatique. L'état de l'art, de la théorie de la niche écologique à la modélisation de la distribution spatiale des abondances, ont été présentés au chapitre I. La zone d'étude principale, le bassin versant de la Loire, ainsi que la collecte des jeux de données et leur mise en forme pour les analyses sont présentés au chapitre II. Le choix des variables environnementales pour expliquer cette structuration ainsi que l'apport d'une approche multi-compartiments biologiques ont été discutés au chapitre III. La modélisation de la distribution des abondances des espèces a été utilisée pour dresser les modifications futures potentielles des communautés sous scénarios de changement climatique. En effet, les abondances portent des informations essentielles quant à l'état de santé des populations, leurs dynamiques, leur capacité de résilience face aux perturbations, etc. (Potts & Elith, 2006 ; Waldock *et al.*, 2021). Cependant, leur modélisation présente des verrous méthodologiques importants. Une méthode de modélisation de la distribution des abondances des espèces aquatiques a donc été proposée au chapitre IV. À partir de ces SDM d'abondance, l'évaluation de l'influence potentielle du changement climatique sur la structuration en abondance des communautés futures a été réalisée au chapitre V. Enfin, une analyse de la transférabilité des patrons spatiaux à l'échelle des communautés a également été menée pour évaluer le degré de transposition des structurations des communautés ligériennes et de leur évolution potentielle future à d'autres grands bassins versants français (chapitre VI). Les méthodologies et les résultats principaux de ce travail de thèse sont rappelés succinctement dans cette section.

Structuration spatiale des communautés à large échelle

L'analyse de la structuration spatiale des communautés de poissons et de macroinvertébrés du bassin ligérien (chapitre III), et des autres grands bassins français (Seine, Garonne et Rhône ; chapitre VI), a confirmé le rôle majeur du gradient environnemental amont-aval à large échelle, soit l'échelle du bassin versant. Les déterminants majeurs de la structuration de ces communautés sont les régimes thermiques et hydrauliques (valeurs moyennes interannuelles et amplitude) associés à la topographie. L'ensemble des données mobilisées sur le bassin versant de la Loire sont disponibles à une haute résolution spatiale, c'est-à-

dire avec une unité géographique suffisamment petite pour obtenir un grain spatial fin, qui a été permise par la mobilisation d'un réseau hydrographique composé de 33 711 tronçons de longueur moyenne $1,85 \pm 1,63$ km. L'utilisation de variables environnementales qui constituent les conditions de vie directement endurées par les organismes, comme la température de l'eau et les conditions hydrauliques (*i.e.* vitesse du courant, hauteur et largeur en eau), apportent plus d'explicativité de la structuration des communautés à large échelle par rapport à des variables considérées représentatives de ces conditions (*i.e.* température de l'air et débits). Ces variables directes ont été obtenues grâce à des simulations issues de modèles à base physique (EROS et T-NET). Leur utilisation dans un cadre écologique et aux étendues et résolutions spatiales utilisées ici font l'originalité de ce travail.

Ainsi, les communautés des zones amonts présentent de faibles richesses taxonomiques et abondances totales, avec une dominance de taxons psychrophiles et rhéophiles, en adéquation avec les faibles températures de l'eau et les fortes valeurs des paramètres hydrauliques qui s'y trouvent. Les communautés qui leurs succèdent vers l'aval s'enrichissent en diversité et en abondance totale, et une augmentation du nombre de taxons aux préférences pour des températures de l'eau plus chaudes et des vitesses de courant plus faibles. Ces résultats permettent d'étendre les connaissances sur le sujet à une plus grande échelle spatiale (Matthews, 1998 ; Keith *et al.*, 2020) et notamment sur les structurations des communautés de poissons et de macroinvertébrés identifiées par les classifications historiques des cours d'eau (Huet, 1949 ; Illies & Botosaneanu, 1963 ; Verneaux, 1973 ; Vannote *et al.*, 1980). L'approche multi-compartiments a également permis d'invalider le postulat de concordance spatiale entre les communautés de poissons et de macroinvertébrés à large échelle. Cette absence de concordance entre compartiments biologiques a déjà été démontrée, et ce à des étendues et résolutions spatiales inférieures avec des communautés de composition spécifiques différentes (Heino, 2010 ; Vehanen *et al.*, 2020).

Modélisation de la distribution des abondances de taxons issus de multiples compartiments

L'*Ordinal Forest* (OF) est une méthode de *machine learning* qui a été utilisée pour modéliser la distribution des abondances de multiples taxons issus des compartiments poissons (N = 26 espèces) et macroinvertébrés (N = 31 genres) sur l'ensemble du réseau hydrographique du bassin de la Loire. Les comparaisons des performances des différents

algorithmes permettant de modéliser des données d'abondance ont orienté ce choix pour les méthodes de *machine learning*, et en particulier pour les algorithmes *Random Forest* (RF) (Potts & Elith, 2006 ; Waldock *et al.*, 2021 ; Jeliaskov *et al.*, 2022). L'objectif final étant de projeter la distribution potentielle des espèces sous scénario futur de changement climatique, les meilleures performances prédictives des RF semblaient prometteuses. De plus, la flexibilité de ces méthodes leur permet d'appréhender à la fois des relations simples et complexes entre la variable réponse et les prédicteurs (Valavi *et al.*, 2023), ce qui la rend applicable à un grand nombre de taxons quelle que soit la relation attendue. Les données d'abondance étaient disponibles sous la forme de *Captures Par Unités d'Effort* (CPUE, ind/100m²). Elles ont été transformées en quatre classes d'abondance, absence, faible, moyenne et forte, grâce à un algorithme de découpage automatique (*Jenks*). Ce passage en classes a permis de limiter l'influence des incertitudes liées aux échantillonnages (Comte & Grenouillet, 2013b ; Vaudor *et al.*, 2015) et de lever certains verrous de modélisation liés à la sur-dispersion des données d'abondance (Potts & Elith, 2006 ; Waldock *et al.*, 2021). Le manque de performance induit par la sur-inflation de la classe absence a été en partie levé par application de la méthode *SMOTE and Cluster-based Undersampling Technique* (SCUT). Cette technique permet de limiter l'influence de l'asymétrie lors de la phase de calibration du modèle, ce qui accroît les performances prédictives et limite le phénomène de sur-apprentissage (Agrawal *et al.*, 2015). Son efficacité avait déjà été démontrée pour des SDM de présence/absence (Johnson *et al.*, 2012 ; Robinson *et al.*, 2018 ; Steen *et al.*, 2021), mais son application à des modèles de la distribution des abondances des espèces est une première.

L'aspect boîte noire et le manque de capacité des RF à expliquer les relations des organismes avec l'environnement qui sont parfois évoqués comme des limites à leur utilisation ne sont qu'apparentes (Olden & Jackson, 2002 ; Zhao & Hastie, 2021). Parmi les méthodes d'interprétation de ce type de modèles (Molnar, 2022), l'importance basée sur le rang a permis de montrer le rôle de la température de l'eau (moyenne interannuelle et amplitude), des conditions hydrauliques (vitesse de courant moyenne interannuelle et amplitude, probabilité d'intermittence), de la topographie (pente) et des indicateurs de pressions anthropiques (densité de population aux abords des cours d'eau, gradient de pression agricole, taux de fractionnement). Les *Partial Dependence Plots* (PDP) ont permis de montrer la réponse partielle de chaque classe d'abondance de chaque taxon à ces variables environne-

mentales. Les relations entre les niveaux d'abondance et les variables environnementales illustrent ainsi les préférences écologiques des taxons étudiés, et donc leur niche réalisée sur le bassin versant de la Loire. Le rôle du taux d'étagement, qui représente le rapport de la hauteur de chute artificielle par rapport à la pente naturelle, montre l'importance des obstacles à la continuité écologique dans la distribution spatiale des organismes aquatiques. Les prédictifs représentatifs des limites à la dispersion des espèces ne démontre pas toujours leur importance lorsqu'ils sont intégrés dans les SDM (Maire *et al.*, 2017). Cela souligne l'intérêt de considérer un indicateur basé sur les capacités de dispersion des espèces (Vierron & Steinbach, 2020).

Influence du changement climatique sur la distribution spatiale des espèces et conséquences à l'échelle des communautés

Les modèles de distribution des classes d'abondance développés ont permis de projeter les distributions potentielles des 57 taxons modélisés en fonction des projections hydrologiques et thermiques simulées pour le bassin versant de la Loire jusqu'en 2100. Parmi la combinaison des sept scénarios des GES et de modèles climatiques (Seyedhashemi *et al.*, 2023), nous avons choisi d'utiliser le scénario de GES le plus pessimiste (RCP 8.5) et un modèle climatique (CNRM-CM5/ALADIN63), dit « intermédiaire », dans le faisceau des évolutions probables de précipitations et de températures. Du fait d'un manque de données, l'utilisation d'une telle résolution spatiale et de projections de la température de l'eau ainsi que de variables hydrauliques sont rares dans les projections de SDM sous changement climatique. Les projections de chaque SDM des 57 taxons modélisés ont été rassemblées, afin d'établir un devenir possible des communautés aquatiques aux horizons temporels 2050 et 2080. Ces résultats confirment l'existence de plusieurs types de réponses aux changements de conditions environnementales à l'échelle du taxon : l'extension de l'aire de répartition, sa contraction, son glissement vers des milieux aux conditions plus favorables, ou le maintien dans l'aire actuelle (Comte & Grenouillet, 2013a). L'intégration des abondances dans les analyses ont montré que les réponses observées en terme de structuration des niveaux d'abondance ne sont pas aisément prévisibles : contrairement à ce qui était attendu, les niveaux d'abondance n'augmentent pas nécessairement avec l'agrandissement de l'aire de répartition d'un taxon et l'inverse n'est pas toujours vrai non plus. Ces hypothèses étaient basées sur celle de l'abondance-centrée appliquée à l'espace géogra-

phique (Marino *et al.*, 2020), qui prévoit l'existence d'un optimum environnemental où l'abondance serait maximal accompagné d'une diminution progressive de ces abondances à mesure que l'on s'éloigne de cet optimum (Yañez-Arenas *et al.*, 2020). Il existe peu de preuves empiriques de l'hypothèse d'abondance-centrée dans la littérature (Waldock *et al.*, 2019 ; Marino *et al.*, 2020 ; Yañez-Arenas *et al.*, 2020), et il semblerait qu'elle soit aussi dépendante d'autres processus comme la dispersion des espèces (Panter *et al.*, 2023) ou les interactions biotiques (Dallas *et al.*, 2020). L'utilisation des SDM pour déterminer la niche des espèces et vérifier l'hypothèse d'abondance-centrée semble appropriée mais elle souffre trop souvent de quelques limites majeures, dont la violation de l'hypothèse d'équilibre entre les organismes et leur environnement (Dallas *et al.*, 2020). L'utilisation des *Partial Dependence Plots* (PDP) au chapitre IV section C.2.2 a permis d'approcher la niche réalisée selon un, voire deux prédicteurs. Cette approche reste néanmoins limitée à la 2D tandis que la niche écologique est multi-dimensionnelle. Les difficultés à fournir une définition universelle de la niche écologique, et en particulier de sa forme attendue dans l'espace multidimensionnel que forment les conditions environnementales, complexifient également la question de l'hypothèse d'abondance-centrée (Dallas *et al.*, 2020).

B. Perspectives de recherche

Poursuivre le développement des modèles de distribution des abondances

Affiner les prédicteurs abiotiques pour mieux expliquer la distribution des abondances

Les conditions environnementales qui peuvent influencer la distribution spatiale des organismes aquatiques sont nombreuses et les variables a priori les plus influentes sur la distribution spatiale des espèces ont été intégrées dans les SDM présentés au chapitre IV (Verneaux, 1973 ; Vannote *et al.*, 1980 ; Matthews, 1998). Une piste d'amélioration notable est l'intégration d'indicateurs plus fins et représentatifs des pressions anthropiques s'exerçant sur les communautés aquatiques.

Les variables de température de l'eau et d'hydrologie utilisées à l'échelle du bassin versant de la Loire sont issues de modèles à base physique (resp. *Temperature NETWORK* (T-NET) et Ensemble de Rivières Organisé en Sous bassins (EROS)). Ceux-ci ne tiennent pas compte de l'influence anthropique sur les variables environnementales qu'ils modé-

lisent, ni des autres influences comme les résurgences souterraines d'eaux plus fraîches (Seyedhashemi *et al.*, 2021b). La prise en compte de l'influence des barrages et autres retenues d'eau pourrait améliorer les prédictions des variables environnementales (Seyedhashemi *et al.*, 2021a) et probablement celles des modèles de la distribution des abondances.

Les faibles contributions explicatives de certaines variables de pressions anthropiques dans les SDM, dont l'effet est avéré d'après la littérature (Mondy & Usseglio-Polatera, 2014 ; Villeneuve *et al.*, 2018 ; Rumschlag *et al.*, 2023), montrent que les variables choisies ne sont peut-être pas les plus adaptées pour représenter ces pressions. Durant la thèse, nous avons été tenté de construire un modèle de la quantité d'oxygène dissous dans l'eau à l'échelle du réseau hydrographique ligérien, car l'oxygène est une variable qui peut être très limitante pour la survie de nombreux organismes aquatiques (Landman *et al.*, 2005). De plus, la saturation en oxygène dépend de la température de l'eau (Moatar *et al.*, 2001), et le réchauffement global engendré par le changement climatique va impacter fortement ce paramètre de façon directe ou indirecte (Diamond *et al.*, 2021 ; Minaudo *et al.*, 2021 ; Diamond *et al.*, 2023). Mais les données mensuelles disponibles dans les réseaux des Agences de l'Eau n'ont pas permis d'obtenir des modèles géostatistiques performants pour être intégrés en tant que prédicteurs dans les SDM. Afin d'être plus complet, d'autres pressions physico-chimiques pourraient être intégrées, comme les pressions organiques, métalliques, etc. Les variables utilisées pour approcher la qualité de l'eau (*i.e.* densité de population aux abords des cours d'eau, occupation des sols) utilisées dans les SDM proposés à chapitre IV restent grossières. Certaines bases recensent les surplus agricoles d'azote ou de phosphore à l'échelle cantonale (NOPOLLU ; de l'observation et des statistiques 2013) mais les données se sont avérées trop morcelées pour une intégration à l'échelle du réseau hydrographique considéré. Un recensement plus complet de certaines pressions permettrait de gagner en finesse dans la représentation des pressions anthropiques, notamment sur la qualité de l'eau, et des SDM calibrés sur ces données plus précises gagneraient certainement en explicativité.

Finalement, la modélisation a été construite sur chaque tronçon indépendamment des conditions des tronçons à proximité. Bien que l'importance du gradient amont-aval ait été soulignée au travers de l'évolution graduelle des conditions environnementales et donc de la structuration des communautés, la prise en compte de l'influence de l'amont vers l'aval pourrait être améliorée. Cet aspect est particulièrement essentiel pour la considération de

la diffusion de certaines pressions le long du gradient longitudinal, comme l'apport de polluants ou la succession d'obstacles à la continuité écologique (Henriques-Silva *et al.*, 2019).

Explorations de nouveaux paramétrages et analyse de sensibilité

Au cours des développements méthodologiques des modèles de la distribution des classes d'abondance proposés dans ce travail, j'ai identifié différents aspects qui requièrent d'être approfondie : (i) le choix du nombre de classes à modéliser, (ii) la méthode d'évaluation des performances prédictives, et (iii) la méthode de correction de l'asymétrie dans la distribution des classes d'abondance.

Le choix du nombre de classes (ici, $N = 4$) est le résultat d'un compromis entre un nombre de classes suffisamment petit pour que chaque classe contienne assez d'observations, en particulier pour les espèces rares ou peu abondantes, et suffisamment grand pour conserver les informations apportées par la donnée d'abondance, en comparaison avec les approches binaires. L'optimisation du nombre de classes en fonction des caractéristiques des données d'abondance de chaque taxon s'est révélée limitée par le fait que les performances des modèles sont influencées par le nombre de classes, empêchant la comparaison de plusieurs modèles basés sur différents nombre de classes. Et l'objectif étant de combiner les projections des différents modèles pour obtenir une vision de la communauté, l'utilisation de différents nombre de classes semble proscrit.

L'évaluation des performances d'un modèle de *machine learning*, et notamment de modèles multi-classes, est un champ disciplinaire à part entière et le choix de la métrique de performance n'est pas simple (Molnar, 2022). La plupart des métriques ne sont sensibles qu'à un aspect particulier de la performance prédictive (Delgado & Tibau, 2019 ; Konowalik & Nosol, 2021), comme le nombre exact de bonnes prédictions (*i.e.* accuracy). Or, dans le cas d'un modèle multi-classes construit sur des données asymétriques (*i.e.* sur-représentation d'une classe, très fréquent en zéro avec des données d'abondance), l'utilisation de métriques de performance basées sur l'*accuracy* peut pousser le modèle à optimiser les bonnes prédictions de la classe majoritaire, au détriment des performances prédictives sur les autres classes (Chawla *et al.*, 2011). Cela peut également conduire à produire un modèle qui a sur-appris. Le sur-apprentissage est un phénomène difficile à détecter, bien qu'il existe des techniques dédiées, et qui limite fortement la capacité de gé-

néralisation d'un modèle. C'est un phénomène à éviter lorsque l'objectif de la modélisation est de réaliser des projections sous changement climatique (Elith & Leathwick, 2009 ; Valavi *et al.*, 2023). La construction de répliqués de modèle à partir de jeux d'entraînement tirés aléatoirement peut aider à limiter ce phénomène (Guisan *et al.*, 2017). Finalement, l'utilisation de plusieurs métriques de performances non redondantes est à privilégier car cela permet de bien comprendre le fonctionnement d'un modèle, en identifiant ses forces et ses faiblesses (Konowalik & Nosol, 2021).

La forte asymétrie des données d'abondance a été conservée lors du passage en classes. La méthode *Synthetic Minority Oversampling Technique* (SMOTE) (Chawla *et al.*, 2011) a été mobilisée pour limiter l'influence de la classe sur-représentée dans chaque modèle. Elle s'est montrée efficace pour la plupart des taxons, et elle a permis d'augmenter les performances des SDM. Cependant, le phénomène de sur-représentation étant accentué par la rareté d'une espèce, l'application de la technique SMOTE n'a pas permis la modélisation des abondances des espèces rares. Il existe d'autres méthodes pour limiter l'influence de l'asymétrie dans les données d'entraînement (Robinson *et al.*, 2018 ; Steen *et al.*, 2021) qui n'ont pas été explorées dans le cadre de ce travail. Par exemple, le sous-échantillonnage spatialisé permet de mieux répartir les données sélectionnées et en même temps de tenir compte de l'auto-corrélation spatiale entre les données (Guisan *et al.*, 2017 ; Steen *et al.*, 2021). Néanmoins, cette méthode peut conduire à un nombre insuffisant de données pour les espèces rares (Steen *et al.*, 2021).

Comment parvenir à intégrer les interactions biotiques dans les modèles de distribution d'espèces ?

Parmi les interactions biotiques à ajouter dans la modélisation, l'exploration des relations trophiques entre taxons apparaît comme la plus prometteuse et la moins complexe à intégrer en première approche. Par exemple, Lucas *et al.* (2023) ont dressé le régime trophique de 19 populations d'ours brun (*Ursus arctos*) à travers l'Europe, en intégrant la qualité énergétique de la ressource. Ils ont ensuite comparé les distributions potentielles futures sous changement climatique issues d'un modèle basé purement sur des prédicteurs abiotiques et d'un modèle intégrant la ressource trophique. Ils ont ainsi montré que les réductions d'aire de répartition de l'ours brun étaient bien plus drastiques lorsque ses ressources trophiques étaient considérées. Des exemples de SDM intégrant aussi la compétition interspécifique existent (*e.g.* Boulangeat *et al.* 2012 ; Gerhardt Braz *et al.* 2020 ;

Stephenson *et al.* 2022). Néanmoins, des connaissances fines sur les régimes alimentaires des espèces modélisées sont nécessaires, comme la composition du régime, sa variabilité en fonction des populations et des différents stade de vie (Floury *et al.*, 2019). Chez les poissons, ces connaissances ne sont pas acquises pour toutes les espèces d'intérêt (Gerking, 1994 ; Braga *et al.*, 2012). Pour pallier ce manque, le recensement des connaissances pré-existantes et manquantes sur les espèces concernées devrait être mené, suivi de l'acquisition de nouvelles données pour les espèces dont le régime alimentaire reste méconnu. Ces deux étapes peuvent se révéler assez ardues et demander d'importantes ressources (humaines, financières et temporelles). J'ai identifié deux approches possibles pour combler les lacunes : (i) l'acquisition de nouvelles données empiriques, (ii) la recherche de relations statistiques via les données pré-existantes, en se basant par exemple sur la co-occurrence d'espèces. L'analyse des contenus stomachaux apporte des informations précieuses sur la composition et la quantité de chaque proie ingérée par un poisson mais cela demande des moyens assez conséquents et une expertise à l'identification des proies (Floury *et al.*, 2019 ; Tirelli *et al.*, 2023). Les développements récents des méthodes de *meta-barcoding* et des bases de données génétiques de macroinvertébrés (*e.g.* NCBI.nlm.nih.gov ou bold-systems.org) offrent la possibilité d'identifier plus rapidement la composition du régime alimentaire d'un poisson via des prélèvements non létaux du contenu stomachal (Campbell *et al.*, 2023), voire par l'analyse des fèces (Guillerault *et al.*, 2017). Cependant, la réalisation de cette étude pour plusieurs dizaines d'espèces à l'échelle d'un bassin versant de la taille de celui de la Loire ($\sim 10^5$ km²) demanderait des moyens conséquents. La recherche de relations statistiques entre les fréquences d'abondance d'un prédateur et de ses proies peut être une solution de contournement. Cela peut également permettre de détecter le caractère plus ou moins opportuniste d'une espèce. Par ailleurs, le régime alimentaire entre populations ou à l'échelle de la vie d'un individu peut être variable (Futia *et al.*, 2021 ; Ribeiro *et al.*, 2021). Il convient alors d'établir la représentativité des différents stades de vie dans les données d'échantillonnage et de dresser leur régime alimentaire en vue de la modélisation de la distribution des abondances représentées.

Vers une vision plus complète des communautés aquatiques : modéliser les espèces rares

La modélisation de la distribution des espèces rares présente un certain nombre de difficultés (Jeliaskov *et al.*, 2022). Du fait de leur rareté, la distribution des abondances présente une sur-représentation des absences, qui posent des difficultés techniques quel que soit le modèle utilisé (Potts & Elith, 2006 ; Waldock *et al.*, 2021). De plus, des difficultés se posent dès la phase d'échantillonnage, car la plupart des protocoles ont des efficacités limitées à détecter les espèces rares (Jeliaskov *et al.*, 2022) ; d'où la nécessité des études sur l'influence de la probabilité de détection de chaque taxon sur sa modélisation (Comte & Grenouillet, 2013b ; Vaudor *et al.*, 2015). À ce jour, ces analyses n'ont été conduites que sur des modèles de présence/absence. Il conviendrait donc de les étendre aux modèles de distribution des abondances.

Actuellement, modéliser les espèces rares constitue un véritable défi technique alors que cette part de la biodiversité est la plus menacée par les changements globaux et appelle à des mesures de conservation rapides et efficaces (Jeliaskov *et al.*, 2022). Le rôle majeur de certaines espèces rares dans le fonctionnement des écosystèmes n'est plus à prouver, et l'absence de redondance entraîne la perte de ces fonctions à l'extinction de l'espèce (Violle *et al.*, 2017 ; Auber *et al.*, 2022). Dans ce travail, il a été montré que de faibles niveaux d'abondance réduisent les performances des modèles, en accord avec d'autres études (Waldock *et al.*, 2021 ; Erickson & Smith, 2023). L'application de méthodes de correction dédiées aux distributions asymétriques (SMOTE ; Chawla *et al.* 2011) s'est révélée utile pour les SDM présentés ici, mais est resté insuffisant pour certains taxons, en accord avec d'autres résultats trouvés dans la littérature (Robinson *et al.*, 2018 ; Steen *et al.*, 2021). Trente-sept taxons ont ainsi dû être écartés de la modélisation en raison de leur rareté trop importante ($\leq 5\%$ occurrence ; 11 espèces de poisson et 26 genres de macroinvertébrés). Des méthodologies ont été proposées récemment, comme un regroupement des différents tirages de jeux tests dans la lignée des techniques de *bootstrap* (Collart & Guisan, 2023). Cependant, elles sont le résultat de développement qui concerne principalement les SDM de présence-absence et mériteraient d'être testées sur des SDM d'abondance. Les SDM d'abondances des espèces rares ont fait l'objet d'autres propositions mais qui ne paraissent pas encore satisfaisantes pour être exploitées (Jeliaskov *et al.*, 2022 ; Erickson & Smith, 2023).

Les modèles de distribution construits à l'échelle de la communauté ont l'avantage de pouvoir intégrer les espèces rares sans être soumis aux difficultés techniques d'un modèle à l'échelle de l'espèce, notamment au trop faible nombre de données (Ferrier & Guisan, 2006 ; Collart & Guisan, 2023). Les *Joint-Species Distribution Models* (JSDM) sont basés sur l'hypothèse que les patrons de réponse à l'environnement sont partagés à l'échelle de la communauté. Les réponses d'espèces trop rares pour être modélisées individuellement pourrait ainsi être approchées par les réponses d'espèces plus communes et fonctionnellement proches (Ovaskainen, 2020 ; pour un exemple d'approximation de la distribution d'espèces rares par SDM d'abondance d'espèces communes voir Woods *et al.*, 2023). Ces modèles ont fait l'objet de développements méthodologiques récents qui devrait bientôt leur permettre de modéliser la structuration des abondances à l'échelle de la communauté (Pollock *et al.*, 2014 ; Ovaskainen, 2020). En particulier, Erickson & Smith (2023) ont comparé différents types de modèles, dont des JSDM, confrontés à différents types de données rares. Ils ont montré que les performances dépendaient des caractéristiques des données d'abondance et que ces performances restent insuffisantes en l'état. Néanmoins, les efforts récents sur la modélisation des espèces rares se multiplient (Jeliaskov *et al.*, 2022 ; Collart & Guisan, 2023 ; Erickson & Smith, 2023) et devraient aboutir à une méthode qui permettrait une utilisation appliquée des prédictions, par exemple dans un cadre de biologie de la conservation (*e.g.* identification de zones prioritaires pour la conservation d'une espèce cible).

Adopter une approche fonctionnelle pour évaluer l'impact du changement climatique sur les communautés aquatiques

Au cours de la thèse, une approche purement taxonomique a été adoptée pour étudier la distribution spatiale des abondances et leurs potentielles évolutions sous scénario de changement climatique. Néanmoins, la biodiversité présente plusieurs facettes (*i.e.* taxonomiques, fonctionnelles et phylogénétique), dont les réponses au changement climatique peuvent différer (Buisson *et al.*, 2008b ; Olden, 2006). L'approche fonctionnelle permet d'identifier directement quelles caractéristiques des organismes (*e.g.* *preferendum* thermique, spécialisation) sont sélectionnées sous l'effet d'une contrainte, comme le régime des perturbations d'un habitat (Statzner & Bêche, 2010) ou le changement climatique (Buisson *et al.*, 2013). La combinaison des approches taxonomiques et fonctionnelles offre

donc une vision plus complète des processus ayant cours dans les écosystèmes, c'est pourquoi cette approche multi-facettes est très plébiscitée (Maire *et al.*, 2013 ; Comte *et al.*, 2021b). Par exemple, Woods *et al.* (2023) ont pu mettre en évidence que les poissons prédateurs sont favorisés au contraire des taxons rhéophiles d'eaux fraîches. Ils montrent aussi une perte de diversité fonctionnelle dans les zones aval du réseau hydrographique, et l'inverse dans les zones amont. L'application d'une approche fonctionnelle aux SDM développés ici pour le bassin de la Loire permettrait de compléter les conclusions dressées au chapitre V. L'analyse des traits basés sur les préférences écologiques en terme de température de l'eau et des conditions hydrauliques pourrait permettre de déterminer quels taxons sont favorisés ou non, et surtout pourquoi. Il est également important de tenir compte de l'homogénéisation fonctionnelle, qui n'est pas toujours corrélée à une homogénéisation taxonomique (Olden, 2006). Enfin, au-delà des espaces rares en tant que telle, la rareté fonctionnelle est une composante biologique essentielle au bon fonctionnement des écosystèmes et sa perte a de lourdes conséquences pour les écosystèmes (Violle *et al.*, 2017 ; Auber *et al.*, 2022). Il conviendrait donc d'identifier les traits fonctionnels rares présents sur le bassin de la Loire et d'évaluer leur devenir possible afin de mettre en place des mesures de gestion plus ciblées.

Perspectives d'applications à la gestion des milieux aquatiques dans le contexte des changements globaux

Le présent travail a été réalisé pour apporter des éléments théoriques sur la distribution spatiale des abondances des espèces dans un contexte de changement climatique, voire de changements globaux, en vue de répondre à des problématiques très appliquées. L'objectif général était de mieux comprendre quels sont les déterminants principaux de la structuration des communautés afin d'en anticiper les potentielles modifications futures et de commencer dès à présent à réfléchir à des stratégies d'adaptation au changement climatique (Baptist *et al.*, 2014).

D'importantes modifications des conditions environnementales et dans la structuration spatiale des communautés aquatiques vont se produire, et peuvent avoir des conséquences pour la biodiversité ainsi que pour les activités humaines (Baptist *et al.*, 2014). Du point de vue de la biodiversité, l'anticipation des changements passe par l'identification des espèces les plus menacées et l'application de mesures de conservation spécifiques (*e.g.* protection de leurs aires de distribution présentes et futures, avec maintien de la continuité écologique si ces deux aires diffèrent). Concernant les activités humaines, les modifications climatiques demandent déjà une certaine adaptation des usages de l'eau (*e.g.* prélèvements pour l'irrigation, gestion hydro-électrique), et elles sont rejointes par les enjeux de conservation de la biodiversité. En effet, les usages de l'eau sont susceptibles d'avoir un effet sur les organismes aquatiques dans le milieu concerné. Ces premiers sont donc réfléchis et réglementés pour concilier usages et maintien d'un écosystème fonctionnel, parfois à l'échelle locale. Ainsi, les changements locaux peuvent rendre les pratiques de la période actuelle inadaptée aux conditions environnementales et biologiques futures. Il convient donc de réfléchir à la mise en place de mesures d'adaptation dès à présent, afin de suivre ces évolutions temporelles.

Bien que l'ensemble des composantes (*i.e.* biotiques, abiotiques, dispersion) impliquées dans la structuration spatiale des communautés ait été considéré dans ce travail, les méthodes utilisées et les différents choix qui ont été fait posent des limites à l'utilisation des résultats actuels. Par exemple, les cartographies proposées en chapitre V et en Annexe G ne sont que des distributions *potentielles* futures, et bien que notre modélisation tente de s'approcher au maximum de la réalité, ces résultats n'en sont pas moins qu'une représentation simplifiée (Peterson *et al.*, 2011 ; Guisan *et al.*, 2017). D'autre part, l'approche

à l'échelle du bassin versant permet d'étudier les processus écologiques à large échelle, mais peuvent faire perdre en précision à l'échelle locale (Alabia *et al.*, 2018 ; Fernandes *et al.*, 2020). Pour pallier cela, les projections ont été réalisées à haute résolution spatiale, à l'échelle du tronçon de cours d'eau d'une longueur moyenne de l'ordre du kilomètre. Les SDM de classes d'abondance ont montré une certaine instabilité des prédictions dans les zones en limite d'aire de répartition de l'espèce modélisée et les niveaux de performance de modélisation pour certains taxons restent assez bas. Enfin, les communautés des grands ordres de Strahler (> 5) sont les moins représentées dans les données utilisées ici (Figure 28). Par conséquent, l'utilisation des projections en limite d'aire de répartition ou dans les zones aval devraient être précautionneuses car plus incertaines.

L'importance de la température de l'eau dans la structuration des communautés et de leur abondance a été une nouvelle fois montrée ici, à l'instar de nombreux autres travaux (Illies & Botosaneanu, 1963 ; Verneaux, 1973 ; Matthews, 1998). Nous avons aussi souligné le rôle des conditions hydrauliques mais aussi d'autres facteurs comme les pressions anthropiques représentées par la densité de population aux abords des cours d'eau, l'occupation des sols, etc. Les analyses multivariées ont permis de montrer la succession graduelle des conditions environnementales et de la structuration des communautés le long du réseau hydrographique, ce que ne permettent pas de représenter les classifications de cours d'eau ou se succèdent des « boîtes » dépourvues de zones de transitions (Matthews, 1998 ; Wasson, 1989). Wehrly *et al.* (2003) ont par exemple reconnu qu'il est difficile d'attribuer une zone de cours d'eau à une classe typologique en se basant uniquement sur les espèces, sachant que ces dernières se distribuent le long d'un continuum et peuvent être trouvées au-delà de la limite de leurs préférences thermiques. Ce constat complexifie la question de la ré-évaluation des conditions environnementales impliquées dans le passage d'une communauté à une autre (Verneaux, 1976). Ainsi, il apparaît difficile d'établir une réglementation générique en matière de rejets thermiques en définissant une valeur fixe de température de l'eau (*i.e.* delta de température entre le prélèvement et le rejet, ou valeur maximale de température de rejet). De plus, le déplacement des communautés et surtout la transformation des assemblages compliquent le raisonnement à partir d'une typologie qui est établie sur la période actuelle. (Wehrly *et al.*, 2003) ont par exemple reconnu qu'il est difficile d'attribuer une zone de cours d'eau à une classe typologique en se basant uniquement sur les espèces, sachant que ces dernières se distribuent le long d'un continuum

et peuvent être trouvées au-delà de la limite de leurs préférences thermiques. Certaines méthodes d'analyse relativement récentes, comme les *gradient forest*, pourrait aider à déterminer l'existence d'un tel seuil entre le passage d'une communauté à une autre, et ce pas uniquement pour la température de l'eau mais aussi pour les autres déterminants environnementaux. Il semble d'ailleurs plus vraisemblable que ce soit la combinaison d'un ensemble de règles sur les différents déterminants environnementaux qui induit les changements de communautés qu'un seuil sur un paramètre unique. De plus, ces règles dépendent aussi du positionnement sur le gradient amont-aval et de la composition de la communauté considérée (Verneaux, 1976). Par exemple, les parties situées en aval d'un bassin versant comportent plus d'espèces tolérantes à de grandes amplitudes thermiques que les milieux amont, ce qui implique que les seuils de température provoquant une modification des assemblages des espèces ne sont pas les mêmes le long d'un cours d'eau. Enfin, au vu des modifications attendues sous l'effet du changement climatique, une ré-actualisation de ces seuils réglementaires en tenant compte des réorganisations futures des communautés aquatiques semblent indispensables. Il semble également important de tenir compte de l'impact des différents processus de réorganisation (*e.g.* homogénéisation biotique) sur la capacité de résilience des écosystèmes face aux perturbations. En effet, ces modifications rapides et drastiques fragilisent de plus en plus les écosystèmes et appellent à la mise en place de mesures de soutien pour aider ces communautés à faire face aux changements (Olden, 2006 ; Dudgeon, 2019).

Références bibliographiques

- Agence française pour la biodiversité - AFB, Centre de ressources Cours d'Eau (2021) *Recueil d'expériences sur l'hydromorphologie*. Recueils d'expériences. Office français pour la biodiversité édition.
- Agrawal, A., Viktor, H.L. & Paquet, E. (2015) SCUT : Multi-Class Imbalanced Data Classification using SMOTE and Cluster-based Undersampling :. *Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, pp. 226–234. SCITEPRESS - Science and Technology Publications, Lisbon, Portugal.
- Alabia, I.D., García Molinos, J., Saitoh, S.I., Hirawake, T., Hirata, T. & Mueter, F.J. (2018) Distribution shifts of marine taxa in the Pacific Arctic under contemporary climate changes. *Diversity and Distributions*, **24**, 1583–1597.
- Allen, D.C., Datry, T., Boersma, K.S., Bogan, M.T., Boulton, A.J., Bruno, D., Busch, M.H., Costigan, K.H., Dodds, W.K., Fritz, K.M., Godsey, S.E., Jones, J.B., Kaletova, T., Kampf, S.K., Mims, M.C., Neeson, T.M., Olden, J.D., Pastor, A.V., Poff, N.L., Ruddell, B.L., Ruhi, A., Singer, G., Vezza, P., Ward, A.S. & Zimmer, M. (2020) River ecosystem conceptual models and non-perennial rivers : A critical review. *WIREs Water*, **7**.
- Alric, B., Dézerald, O., Meyer, A., Billoir, E., Coulaud, R., Larras, F., Mondy, C.P. & Usseglio-Polatera, P. (2020) How diatom-, invertebrate- and fish-based diagnostic tools can support the ecological assessment of rivers in a multi-pressure context : temporal trends over the past two decades in France. *Science of The Total Environment*, p. 143915.
- Anderson, M.J. (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, **26**, 32–46.
- Anderson, R.P. (2017) When and how should biotic interactions be considered in models of species niches and distributions ? *Journal of Biogeography*, **44**, 8–17.
- Auber, A., Waldock, C., Maire, A., Goberville, E., Albouy, C., Algar, A.C., McLean, M., Brind'Amour, A., Green, A.L., Tupper, M., Vigliola, L., Kaschner, K., Kesner-Reyes, K., Beger, M., Tjiputra, J., Toussaint, A., Violle, C., Mouquet, N., Thuiller, W. & Mouillot, D. (2022) A functional vulnerability framework for biodiversity conservation. *Nature Communications*, **13**, 4774.
- Backus-Freer, J. & Pyron, M. (2015) Concordance among fish and macroinvertebrate assemblages in streams of Indiana, USA. *Hydrobiologia*, **758**, 141–150.
- Bae, M.J., Li, F., Kwon, Y.S., Chung, N., Choi, H., Hwang, S.J. & Park, Y.S. (2014) Concordance of diatom, macroinvertebrate and fish assemblages in streams at nested spatial scales : Implications for ecological integrity. *Ecological Indicators*, **47**, 89–101.
- Baptist, F., Poulet, N. & Séon-Massin, N. (2014) *Freshwater fish and climate change. Current situation and adaptation strategies*. Knowledge for action series. Onema édition.
- Baudoin, J.M., Burgun, V., Chanseau, M., Larinier, M., Ovidio, M., Sremski, W., Steinbach, P. & Voetgle, B. (2014) Évaluer le franchissement des obstacles par les poissons. Principes et méthodes. Informations sur la continuité écologique - ICE. Technical report, ONEMA, Université de Liège, Ecogea.

- Beaufort, A., Curie, F., Moatar, F., Ducharne, A., Melin, E. & Thiery, D. (2016) T-NET, a dynamic model for simulating daily stream temperature at the regional scale based on a network topology : T-NET model for simulating stream temperature at a regional scale. *Hydrological Processes*, **30**, 2196–2210.
- Beaufort, A., Diamond, J.S., Sauquet, E. & Moatar, F. (2022) Spatial extrapolation of stream thermal peaks using heterogeneous time series at a national scale. *Hydrology and Earth System Sciences*, **26**, 3477–3495.
- Beaufort, A., Moatar, F. & Sauquet, E. (2020a) Thermie en rivière : Analyse géostatistique et description de régime : Application à l'échelle de la France. Technical report, INRAE UR RiverLy, Université de Tours GÉHCO.
- Beaufort, A., Moatar, F., Sauquet, E., Loicq, P. & Hannah, D.M. (2020b) Influence of landscape and hydrological factors on stream–air temperature relationships at regional scale. *Hydrological Processes*, **34**, 583–597.
- Bivand, R. (2022) classInt : Choose Univariate Class Intervals.
- Bivand, R. & Lewin-Koh, N. (2022) maptools : Tools for handling spatial objects.
- Boulangéat, I., Gravel, D. & Thuiller, W. (2012) Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances : The role of dispersal and biotic interactions in explaining species distributions and abundances. *Ecology Letters*, **15**, 584–593.
- Braga, R.R., Bornatowski, H. & Vitule, J.R.S. (2012) Feeding ecology of fishes : an overview of worldwide publications. *Reviews in Fish Biology and Fisheries*, **22**, 915–929.
- Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5–32.
- Buisson, L., Blanc, L. & Grenouillet, G. (2008a) Modelling stream fish species distribution in a river network : the relative effects of temperature versus physical factors. *Ecology of Freshwater Fish*, **17**, 244–257.
- Buisson, L. & Grenouillet, G. (2009) Contrasted impacts of climate change on stream fish assemblages along an environmental gradient. *Diversity and Distributions*, **15**, 613–626.
- Buisson, L., Grenouillet, G., Villéger, S., Canal, J. & Laffaille, P. (2013) Toward a loss of functional diversity in stream fish assemblages under climate change. *Global Change Biology*, **19**, 387–400.
- Buisson, L., Thuiller, W., Casajus, N., Lek, S. & Grenouillet, G. (2010) Uncertainty in ensemble forecasting of species distribution. *Global Change Biology*, **16**, 1145–1157.
- Buisson, L., Thuiller, W., Lek, S., Lim, P. & Grenouillet, G. (2008b) Climate change hastens the turnover of stream fish assemblages. *Global Change Biology*, **14**, 2232–2248.
- Caissie, D. (2006) The thermal regime of rivers : a review. *Freshwater Biology*, **51**, 1389–1406.
- Campbell, L.J., Castillo, N.A., Shenker, J., Owens, L.A., Santos, R.O., Adams, A.J., Rehage, J.S., Denton, K.E. & Goldberg, T.L. (2023) Bone appétit : DNA metabarcoding as a non-lethal alternative to morphological dietary assessment in Atlantic bonefish (*Albula vulpes*). *Environmental Biology of Fishes*, **106**, 337–348.
- Caradima, B., Scheidegger, A., Brodersen, J. & Schuwirth, N. (2021) Bridging mechanistic conceptual models and statistical species distribution models of riverine fish. *Ecological Modelling*, **457**, 109680.
- Carbonell, J.A., Wang, Y. & Stoks, R. (2021) Evolution of cold tolerance and thermal plasticity in life history, behaviour and physiology during a poleward range expansion. *Journal of Animal Ecology*, pp. 1365–2656.13482.
- Cha, Y., Shin, J., Go, B., Lee, D.S., Kim, Y., Kim, T. & Park, Y.S. (2021) An interpretable machine learning method for supporting ecosystem management : Application to species distribution models of freshwater macroinvertebrates. *Journal of Environmental Management*, **291**, 112719.

- Chase, J.M. (2011) Ecological niche theory. *The theory of ecology*, pp. 93–107. University of Chicago Press edition.
- Chase, J.M. & Leibold, M.A. (2003) *Ecological niches : linking classical and contemporary approaches*. Interspecific interactions. University of Chicago Press, Chicago.
- Chase, J.M., Liebergesell, M., Sagouis, A., May, F., Blowes, S.A., Berg, A., Bernard, E., Brosi, B.J., Cadotte, M.W., Cayuela, L., Chiarello, A.G., Cosson, J., Cresswell, W., Dami, F.D., Dauber, J., Dickman, C.R., Didham, R.K., Edwards, D.P., Farneda, F.Z., Gavish, Y., Gonçalves-Souza, T., Guadagnin, D.L., Henry, M., López-Baucells, A., Kappes, H., Mac Nally, R., Manu, S., Martensen, A.C., McCollin, D., Meyer, C.F.J., Neckel-Oliveira, S., Nogueira, A., Pons, J., Raheem, D.C., Ramos, F.N., Rocha, R., Sam, K., Slade, E., Stireman, J.O., Struebig, M.J., Vasconcelos, H. & Ziv, Y. (2019) Frag SAD : A database of diversity and species abundance distributions from habitat fragments. *Ecology*, **100**.
- Chawla, N.V., Bowyer, K.W., Hall, L.O. & Kegelmeyer, W.P. (2011) SMOTE : Synthetic Minority Over-sampling Technique.
- Chesson, P. (2000) Mechanisms of Maintenance of Species Diversity. *Annual Review of Ecology and Systematics*, **31**, 343–366.
- Chevenet, F., Doleadec, S. & Chessel, D. (1994) A fuzzy coding approach for the analysis of long-term ecological data. *Freshwater Biology*, **31**, 295–309.
- Clarke, K.R. (1993) Non-parametric multivariate analyses of changes in community structure. *Austral Ecology*, **18**, 117–143.
- Collart, F., Broennimann, O., Guisan, A. & Vanderpoorten, A. (2023) Ecological and biological indicators of the accuracy of species distribution models : lessons from European bryophytes. *Ecography*, p. e06721.
- Collart, F. & Guisan, A. (2023) Small to train, small to test : Dealing with low sample size in model evaluation. *Ecological Informatics*, p. 102106.
- Colwell, R.K. & Rangel, T.F. (2009) Hutchinson's duality : The once and future niche. *Proceedings of the National Academy of Sciences*, **106**, 19651–19658.
- Comte, L., Buisson, L., Daufresne, M. & Grenouillet, G. (2013) Climate-induced changes in the distribution of freshwater fish : observed and predicted trends : Climate change and freshwater fish. *Freshwater Biology*, **58**, 625–639.
- Comte, L., Carvajal-Quintero, J., Tedesco, P.A., Giam, X., Brose, U., Erős, T., Filipe, A.F., Fortin, M., Irving, K., Jacquet, C., Larsen, S., Sharma, S., Ruhi, A., Becker, F.G., Casatti, L., Castaldelli, G., Dala-Corte, R.B., Davenport, S.R., Franssen, N.R., García-Berthou, E., Gavioli, A., Gido, K.B., Jimenez-Segura, L., Leitão, R.P., McLarney, B., Meador, J., Milardi, M., Moffatt, D.B., Occhi, T.V.T., Pompeu, P.S., Propst, D.L., Pyron, M., Salvador, G.N., Stefferud, J.A., Sutela, T., Taylor, C., Terui, A., Urabe, H., Vehanen, T., Vitule, J.R.S., Zeni, J.O. & Olden, J.D. (2021a) RivFishTIME : A global database of fish time-series to study global change ecology in riverine systems. *Global Ecology and Biogeography*, **30**, 38–50.
- Comte, L. & Grenouillet, G. (2013a) Do stream fish track climate change ? Assessing distribution shifts in recent decades. *Ecography*, **36**, 1236–1246.
- Comte, L. & Grenouillet, G. (2013b) Species distribution modelling and imperfect detection : comparing occupancy versus consensus methods. *Diversity and Distributions*, **19**, 996–1007.
- Comte, L. & Grenouillet, G. (2015) Distribution shifts of freshwater fish under a variable climate : comparing climatic, bioclimatic and biotic velocities. *Diversity and Distributions*, **21**, 1014–1026.
- Comte, L., Olden, J.D., Tedesco, P.A., Ruhi, A. & Giam, X. (2021b) Climate and land-use changes interact to drive long-term reorganization of riverine fish communities globally. *Proceedings of the National Academy of Sciences*, **118**, e2011639118.
- Culler, L.E., Wood, Z.T., Diaz, J., Fey, S.B., Timmins, D. & Ayres, M.P. (2018) Streams in an uninhabited watershed have predictably different thermal sensitivities to variable summer air temperatures. *Freshwater Biology*, **63**, 676–686.

- Dallas, T.A., Santini, L., Decker, R. & Hastings, A. (2020) Weighing the evidence for the abundant-center hypothesis. *Biodiversity Informatics*, p. 11.
- Daufresne, M. & Boët, P. (2007) Climate change impacts on structure and diversity of fish communities in rivers. *Global Change Biology*, **13**, 2467–2478.
- de l'observation et des statistiques, S. (2013) NOPOLU-Agri. Outils de spatialisation des pressions de l'agriculture. Méthodologie et résultats pour les surplus d'azote et les émissions des gaz à effet de serre. Campagne 2010-2011. Technical report, Service de l'observation et des statistiques.
- Delgado, R. & Tibau, X.A. (2019) Why Cohen's Kappa should be avoided as performance measure in classification. *PLOS ONE*, **14**, e0222916.
- Descy, J.P. (2009) Continental Atlantic Rivers. *Rivers of Europe*, pp. 151–198. Elsevier.
- Diamond, J.S., Bernal, S., Boukra, A., Cohen, M.J., Lewis, D., Masson, M., Moatar, F. & Pinay, G. (2021) Stream network variation in dissolved oxygen : Metabolism proxies and biogeochemical controls. *Ecological Indicators*, **131**, 108233.
- Diamond, J.S., Moatar, F., Cohen, M.J., Poirel, A., Martinet, C., Maire, A. & Pinay, G. (2022) Metabolic regime shifts and ecosystem state changes are decoupled in a large river. *Limnology and Oceanography*, **67**.
- Diamond, J.S., Moatar, F., Recoura-Massaquant, R., Chaumot, A., Zarnetske, J., Valette, L. & Pinay, G. (2023) Hypoxia is common in temperate headwaters and driven by hydrological extremes. *Ecological Indicators*, **147**, 109987.
- Dias, M.S., Oberdorff, T., Hugueny, B., Leprieur, F., Jézéquel, C., Cornu, J.F., Brosse, S., Grenouillet, G. & Tedesco, P.A. (2014) Global imprint of historical connectivity on freshwater fish biodiversity. *Ecology Letters*, **17**, 1130–1140.
- Dolph, C.L., Huff, D.D., Chizinski, C.J. & Vondracek, B. (2011) Implications of community concordance for assessing stream integrity at three nested spatial scales in Minnesota, U.S.A. : Concordance and scale in stream assessment. *Freshwater Biology*, **56**, 1652–1669.
- Dormann, C.F., Bobrowski, M., Dehling, D.M., Harris, D.J., Hartig, F., Lischke, H., Moretti, M.D., Pagel, J., Pinkert, S., Schleuning, M., Schmidt, S.I., Sheppard, C.S., Steinbauer, M.J., Zeuss, D. & Kraan, C. (2018) Biotic interactions in species distribution modelling : 10 questions to guide interpretation and avoid false conclusions. *Global Ecology and Biogeography*, **27**, 1004–1016.
- Dovciak, A.L. & Perry, J.A. (2002) In Search of Effective Scales for Stream Management : Does Agroecoregion, Watershed, or Their Intersection Best Explain the Variance in Stream Macroinvertebrate Communities? *Environmental Management*, **30**, 365–377.
- Dudgeon, D. (2019) Multiple threats imperil freshwater biodiversity in the Anthropocene. *Current Biology*, **29**, R960–R967.
- Elith, J., Leathwick, J.R. & Hastie, T. (2008) A working guide to boosted regression trees. *Journal of Animal Ecology*, **77**, 802–813.
- Elith, J., Kearney, M. & Phillips, S. (2010) The art of modelling range-shifting species : The art of modelling range-shifting species. *Methods in Ecology and Evolution*, **1**, 330–342.
- Elith, J. & Leathwick, J.R. (2009) Species Distribution Models : Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677–697.
- Elton, C.S. (1927) *Animal ecology*. University of Chicago Press, Chicago, the macmillan company edition.
- Erickson, K.D. & Smith, A.B. (2023) Modeling the rarest of the rare : a comparison between multi-species distribution models, ensembles of small models, and single-species models at extremely low sample sizes. *Ecography*, p. e06500.
- European Commission (2000) Directive 2000/60/EC of the European Parliament and of the Council establishing a framework for Community action in the field of water policy.

- European Environment Agency (2012) Corine Land Cover (CLC). Technical report, European Environment Agency.
- Fernandes, J.A., Rutterford, L., Simpson, S.D., Butenschön, M., Frölicher, T.L., Yool, A., Cheung, W.W.L. & Grant, A. (2020) Can we project changes in fish abundance and distribution in response to climate? *Global Change Biology*, p. gcb.15081.
- Ferrier, S. & Guisan, A. (2006) Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, **43**, 393–404.
- Ficklin, D.L., Luo, Y., Stewart, I.T. & Maurer, E.P. (2012) Development and application of a hydroclimatological stream temperature model within the Soil and Water Assessment Tool. *Water Resources Research*, **48**.
- Filz, K.J., Schmitt, T. & Engler, J.O. (2013) How fine is fine-scale? Questioning the use of fine-scale bioclimatic data in species distribution models used for forecasting abundance patterns in butterflies. *European Journal of Entomology*, **110**, 311–317.
- Floury, M., Colombet, J. & Desvilettes, C. (2019) Interspecific competition between restocked 0+ salmon parr (*Salmo salar*) and native fish species in a large European river. *Ecology of Freshwater Fish*, **28**, 69–84.
- Floury, M., Pollock, L.J., Buisson, L., Thuiller, W., Chandesris, A. & Souchon, Y. (2021) Combining expert-based and computational approaches to design protected river networks under climate change. *Diversity and Distributions*, p. ddi.13411.
- Floury, M., Souchon, Y. & Looy, K.V. (2018) Climatic and trophic processes drive long-term changes in functional diversity of freshwater invertebrate communities. *Ecography*, **41**, 209–218.
- Floury, M., Usseglio-Polatera, P., Ferreol, M., Delattre, C. & Souchon, Y. (2013) Global climate change in large European rivers : long-term effects on macroinvertebrate communities and potential local confounding factors. *Global Change Biology*, **19**, 1085–1099.
- Futia, M.H., Colborne, S.F., Fisk, A.T., Gorsky, D., Johnson, T.B., Lantry, B.F., Lantry, J.R. & Rinchar, J. (2021) Comparisons among three diet analyses demonstrate multiple patterns in the estimated adult diet of a freshwater piscivore, *Salvelinus namaycush*. *Ecological Indicators*, **127**, 107728.
- Galiana, N., Lurgi, M., Montoya, J.M., Araújo, M.B. & Galbraith, E.D. (2023) Climate or diet? The importance of biotic interactions in determining species range size. *Global Ecology and Biogeography*, p. geb.13686.
- Ganz, K. (2021) scutr : Balancing multiclass datasets for classification tasks. R package version 0.1.2.
- Gao, J., Peng, Z., Zang, H., Wang, Y., Ding, N., He, S., Datry, T. & Wang, B. (2023) Species sensitivity and functional uniqueness determine the response of macroinvertebrate functional diversity to species loss in urban streams. *Freshwater Biology*, p. 14055.
- Gause, G. (1934) The struggle for existence Williams and Wilkins. *Baltimore, Maryland*.
- Gergs, R. & Rothhaupt, K.O. (2015) Invasive species as driving factors for the structure of benthic communities in Lake Constance, Germany. *Hydrobiologia*, **746**, 245–254.
- Gerhardt Braz, A., Eduardo de Viveiros Grelle, C., de Souza Lima Figueiredo, M. & de Moraes Weber, M. (2020) Interspecific competition constrains local abundance in highly suitable areas. *Ecography*, **43**, 1560–1570.
- Gerking, S.D. (1994) *Feeding ecology of fish*. Elsevier.
- GIEC (2023) AR6 Synthesis report : climate change 2023. Technical report.
- Gies, M., Sondermann, M., Hering, D. & Feld, C.K. (2015) Are species distribution models based on broad-scale environmental variables transferable across adjacent watersheds? A case study with eleven macroinvertebrate species. *Fundamental and Applied Limnology / Archiv für Hydrobiologie*, **186**, 63–97.

- Grenouillet, G., Brosse, S., Tudesque, L., Lek, S., Baraillé, Y. & Loot, G. (2008) Concordance among stream assemblages and spatial autocorrelation along a fragmented gradient. *Diversity and Distributions*, **14**, 592–603.
- Griffiths, D. (2015) Connectivity and vagility determine spatial richness gradients and diversification of freshwater fish in North America and Europe. *Biological Journal of the Linnean Society*, **116**, 773–786.
- Grinnell, J. (1917) The niche-relationships of the California Thrasher. *The Auk*, **34**, 427–433. Publisher : JSTOR.
- Guillerault, N., Bouletreau, S., Iribar, A., Valentini, A. & Santoul, F. (2017) Application of DNA metabarcoding on faeces to identify European catfish *Silurus glanis* diet : dna metabarcoding of *s. glanis* faeces. *Journal of Fish Biology*, **90**, 2214–2219.
- Guisan, A., Thuiller, W. & Zimmermann, N.E. (2017) *Habitat suitability and distribution models with applications in R*. Ecology, biodiversity and conservation. Cambridge University Press, Cambridge, United Kingdom ; New York, NY.
- Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.
- Haesen, S., Lenoir, J., Gril, E., De Frenne, P., Lembrechts, J., Kopecký, M., Macek, M., Man, M., Wild, J. & Van Meerbeek, K. (2023) Uncovering the hidden niche : incorporating microclimate temperature into species distribution models. preprint, Life Sciences.
- Hand, D.J. & Till, R.J. (2001) A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. *Machine Learning*, **45**, 171–186.
- Hawkins, C.P. & Norris, R.H. (2000) Performance of different landscape classifications for aquatic bioassessments : introduction to the series. *Journal of the North American Benthological Society*, **19**, 367–369.
- Heino, J. (2002) Concordance of species richness patterns among multiple freshwater taxa : a regional perspective. *Biodiversity and Conservation*, **11**, 137–147.
- Heino, J. (2010) Are indicator groups and cross-taxon congruence useful for predicting biodiversity in aquatic ecosystems ? *Ecological Indicators*, **10**, 112–117.
- Heino, J. & Soininen, J. (2007) Are higher taxa adequate surrogates for species-level assemblage patterns and species richness in stream organisms ? *Biological Conservation*, **137**, 78–89.
- Henriques-Silva, R., Logez, M., Reynaud, N., Tedesco, P.A., Brosse, S., Januchowski-Hartley, S.R., Oberdorff, T. & Argillier, C. (2019) A comprehensive examination of the network position hypothesis across multiple river metacommunities. *Ecography*, **42**, 284–294.
- Hornung, R. (2020) Ordinal Forests. *Journal of Classification*, **37**, 4–17.
- Hornung, R. (2022) ordinalForest : Ordinal Forests : Prediction and Variable Ranking with Ordinal Target Variables.
- Hubbell, S. (2001) *The unified neutral theory of biodiversity and biogeography*.
- Huet, M. (1949) Aperçu des relations entre la pente et les populations piscicoles des eaux courantes. *Schweizerische Zeitschrift für Hydrologie*, **11**, 332–351.
- Huet, M. (1959) Profiles and Biology of Western European Streams as Related to Fish Management. *Transactions of the American Fisheries Society*, **88**, 155–163.
- Hutchinson, G.E. (1957) Concluding remarks. *Cold spring harbor symposium on quantitative biology*, **22**, 415–427.
- Illies, J. & Botosaneanu, L. (1963) Problèmes et méthodes de la classification et de la zonation écologique des eaux courantes, considérées surtout du point de vue faunistique. *SIL Communications, 1953-1996*, **12**, 1–57.

- Infante, D.M., David Allan, J., Linke, S. & Norris, R.H. (2009) Relationship of fish and macroinvertebrate assemblages to environmental factors : implications for community concordance. *Hydrobiologia*, **623**, 87–103.
- INSEE (2021) Recensement de la population. Technical report, INSEE.
- Irving, K., Jähnig, S.C. & Kuemmerlen, M. (2022) Disentangling the effect of climatic and hydrological predictor variables on benthic macroinvertebrate distributions from predictive models. *Hydrobiologia*, **849**, 1021–1040.
- Jackson, D.A., Peres-Neto, P.R. & Olden, J.D. (2001) What controls who is where in freshwater fish communities — the roles of biotic, abiotic, and spatial factors. *Canadian Journal of Fisheries and Aquatic Sciences*, **58**, 14.
- Jeliazkov, A., Gavish, Y., Marsh, C.J., Geschke, J., Brummitt, N., Rocchini, D., Haase, P., Kunin, W.E. & Henle, K. (2022) Sampling and modelling rare species : Conceptual guidelines for the neglected majority. *Global Change Biology*, p. gcb.16114.
- Johnson, R.A., Chawla, N.V. & Hellmann, J.J. (2012) Species distribution modeling and prediction : A class imbalance problem. *2012 Conference on Intelligent Data Understanding*, pp. 9–16. IEEE, Boulder, CO, USA.
- Johnson, R.K., Goedkoop, W. & Sandin, L. (2004) Spatial scale and ecological relationships between the macroinvertebrate communities of stony habitats of streams and lakes. *Freshwater Biology*, **49**, 1179–1194.
- Johnson, R.K. & Hering, D. (2010) Spatial congruency of benthic diatom, invertebrate, macrophyte, and fish assemblages in European streams. *Ecological Applications*, **20**, 978–992.
- Johnson, R.H. (1910) *Determinate evolution in the color-pattern of the lady-beetles*. Carnegie Inst, Washington.
- Junk, W.J., Bayley, P.B. & Sparks, R.E. (1989) The Flood Pulse Concept in river-floodplain systems. *Canadian Special Publications of Fisheries and Aquatic Science*, **106**, 110–127.
- Kearney, M. (2006) Habitat, environment and niche : what are we modelling ? *Oikos*, **115**, 186–191.
- Keith, P., Poulet, N., Denys, G., Changeux, T., Feunteun, E. & Persat, H. (2020) *Les poissons d'eau douce de France*. Inventaires et biodiversité. Biotope Éditions, Mèze ; Muséum National d'Histoire naturelle, Paris, Paris, 2 edition.
- Kilgour, B.W. & Barton, D.R. (1999) Associations between stream fish and benthos across environmental gradients in southern Ontario, Canada. *Freshwater Biology*, **41**, 553–566.
- Konowalik, K. & Nosol, A. (2021) Evaluation metrics and validation of presence-only species distribution models based on distributional maps with varying coverage. *Scientific Reports*, **11**, 1482.
- Kuczynski, L., Legendre, P. & Grenouillet, G. (2018) Concomitant impacts of climate change, fragmentation and non-native species have led to reorganization of fish communities since the 1980s. *Global Ecology and Biogeography*, **27**, 213–222.
- Kuczynski, L., Ontiveros, V.J. & Hillebrand, H. (2023) Biodiversity time series are biased towards increasing species richness in changing environments. *Nature Ecology & Evolution*.
- Laffaille, P., Brosse, S., Ombredane, D. & Feunteun, E. (2020) Richesse et organisation spatiale des communautés de poissons d'eau douce. *Les poissons d'eau douce de France*, Inventaires et biodiversité, pp. 128–159. Muséum national d'Histoire naturelle, Paris, biotope editions, mèze edition.
- Lamouroux, N., Dolédec, S. & Gayraud, S. (2004) Biological traits of stream macroinvertebrate communities : effects of microhabitat, reach, and basin filters. *Journal of the North American Benthological Society*, **23**, 449–466.
- Landman, M.J., Van Den Heuvel, M.R. & Ling, N. (2005) Relative sensitivities of common freshwater fish and invertebrates to acute hypoxia. *New Zealand Journal of Marine and Freshwater Research*, **39**, 1061–1067.

- Lasne, E., Bergerot, B., Lek, S. & Laffaille, P. (2007) Fish zonation and indicator species for the evaluation of the ecological status of rivers : example of the Loire basin (France). *River Research and Applications*, **23**, 877–890.
- Leach, J.A. & Moore, R.D. (2019) Empirical Stream Thermal Sensitivities May Underestimate Stream Temperature Response to Climate Warming. *Water Resources Research*, **55**, 5453–5467.
- Legendre, P. (2019) A temporal beta-diversity index to identify sites that have changed in exceptional ways in space–time surveys. *Ecology and Evolution*, **9**, 3500–3514.
- Legendre, P. & Legendre, L. (2012) *Numerical ecology*. Number 24 in Developments in environmental modelling. Elsevier, Amsterdam, third english edition edition.
- Loicq, P., Moatar, F., Jullian, Y., Dugdale, S.J. & Hannah, D.M. (2018) Improving representation of riparian vegetation shading in a regional stream temperature model using LiDAR data. *Science of the total environment*, **624**, 480–490.
- Lucas, P.M., Thuiller, W., Talluto, M.V., Polaina, E., Albrecht, J., Selva, N., De Barba, M., Maiorano, L., Penteriani, V., Guégen, M., Balkenhol, N., Dutta, T., Fedororca, A., Frank, S.C., Zedrosser, A., Afonso-Jordana, I., Ambarlı, H., Ballesteros, F., Bashta, A.T., Bilgin, C.C., Bogdanović, N., Bojars, E., Bojarska, K., Bragalanti, N., Brøseth, H., Chynoweth, M.W., Ćirović, D., Ciucci, P., Corradini, A., De Angelis, D., de Gabriel Hernando, M., Domokos, C., Dutsov, A., Ertürk, A., Filacorda, S., Frangini, L., Groff, C., Heikkinen, S., Hoxha, B., Huber, D., Huitu, O., Ionescu, G., Ionescu, O., Jerina, K., Jurj, R., Karamanlidis, A.A., Kindberg, J., Kojola, I., López-Bao, J.V., Männil, P., Melovski, D., Mertzanis, Y., Molinari, P., Molinari-Jobin, A., Mustoni, A., Naves, J., Ogurtsov, S., Özüt, D., Palazón, S., Pedrotti, L., Perovi ?, A., Piminov, V.N., Pop, I.M., Popa, M., Psaralexi, M., Quenette, P.Y., Rauer, G., Reljic, S., Revilla, E., Saarma, U., Saveljev, A.P., Sayar, A.O., Şekercioğlu, C.H., Sergiel, A., Sîrbu, G., Skrbinšek, T., Skuban, M., Soyumert, A., Stojanov, A., Tammeleht, E., Tirronen, K., Trajçe, A., Trbojević, I., Trbojević, T., Zieba, F., Zlatanova, D., Zwijacz-Kozica, T. & Pollock, L.J. (2023) Including biotic interactions in species distribution models improves the understanding of species niche : a case of study with the brown bear in Europe. preprint, Ecology.
- Léger (1909) Principes de la méthode rationnelle du peuplement des cours d'eau à salmonidés. *Trav Lab Piscicult Grenoble*, **1**, 533 :568.
- MacArthur, R. & Wilson, E. (1967) The theory of island biogeography.
- Maire, A., Laffaille, P., Maire, J.F. & Buisson, L. (2017) Identification of Priority Areas for the Conservation of Stream Fish Assemblages : Implications for River Management in France : Priority Areas for the Conservation of Fish Assemblages in France. *River Research and Applications*, **33**, 524–537.
- Maire, A., Buisson, L., Biau, S., Canal, J. & Laffaille, P. (2013) A multi-faceted framework of diversity for prioritizing the conservation of fish assemblages. *Ecological Indicators*, **34**, 450–459.
- Maire, A., Thierry, E., Viechtbauer, W. & Daufresne, M. (2019) Poleward shift in large-river fish communities detected with a novel meta-analysis framework. *Freshwater Biology*.
- Malmqvist, B. & Rundle, S. (2002) Threats to the running water ecosystems of the world. *Environmental Conservation*, **29**, 134–153.
- Marino, N.A.C., Céréghino, R., Gilbert, B., Petermann, J.S., Srivastava, D.S., Omena, P.M., Bautista, F.O., Guzman, L.M., Romero, G.Q., Trzcinski, M.K., Barberis, I.M., Corbara, B., Debastiani, V.J., Dézerald, O., Kratina, P., Leroy, C., MacDonald, A.A.M., Montero, G., Pillar, V.D., Richardson, B.A., Richardson, M.J., Talaga, S., Gonçalves, A.Z., Piccoli, G.C.O., Jocqué, M. & Farjalla, V.F. (2020) Species niches, not traits, determine abundance and occupancy patterns : A multi-site synthesis. *Global Ecology and Biogeography*, **29**, 295–308.
- Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K. & Thuiller, W. (2009) Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, **15**, 59–69.
- Matthews, W.J. (1998) *Patterns in Freshwater Fish Ecology*. Springer US : Imprint : Springer, Boston, MA.

- Melles, S.J., Jones, N.E. & Schmidt, B.J. (2014) Evaluation of Current Approaches to Stream Classification and a Heuristic Guide to Developing Classifications of Integrated Aquatic Networks. *Environmental Management*, **53**, 549–566.
- Minaudo, C., Abonyi, A., Leitão, M., Lançon, A.M., Floury, M., Descy, J.P. & Moatar, F. (2021) Long-term impacts of nutrient control, climate change, and invasive clams on phytoplankton and cyanobacteria biomass in a large temperate river. *Science of The Total Environment*, **756**, 144074.
- Minchin, P.R. (1987) An evaluation of the relative robustness of techniques for ecological ordination. I.C. Prentice & E. van der Maarel, eds., *Theory and models in vegetation science*, pp. 89–107. Springer Netherlands, Dordrecht.
- Moatar, F., Miquel, J. & Poirel, A. (2001) A quality-control method for physical and chemical monitoring data. Application to dissolved oxygen levels in the river Loire (France). *Journal of Hydrology*, **252**, 25–36.
- Moatar, F., Descy, J.P., Rodrigues, S., Souchon, Y., Floury, M., Grosbois, C., Minaudo, C., Leitao, M., Wantzen, K.M. & Bertrand, F. (2022) The Loire River basin. *Rivers of Europe*, pp. 245–271. Elsevier.
- Moatar, F. & Dupont, N., eds. (2016) *La Loire fluviale et estuarienne : un milieu en évolution*. Quae, Versailles.
- Molnar, C. (2022) *Interpretable Machine Learning : A Guide for Making Black Box Models Explainable*. 2 edition.
- Mondy, C.P. & Usseglio-Polatera, P. (2014) Using fuzzy-coded traits to elucidate the non-random role of anthropogenic stress in the functional homogenisation of invertebrate assemblages. *Freshwater Biology*, **59**, 584–600.
- Morel, M., Booker, D.J., Gob, F. & Lamouroux, N. (2020) Intercontinental predictions of river hydraulic geometry from catchment physical characteristics. *Journal of Hydrology*, p. 124292.
- Oberdorff, T., Tedesco, P.A., Hugueny, B., Leprieur, F., Beauchard, O., Brosse, S. & Dürr, H.H. (2011) Global and Regional Patterns in Riverine Fish Species Richness : A Review. *International Journal of Ecology*, **2011**, 1–12.
- Office Français pour la Biodiversité (mars 2020) Nâïades.
- Ohlmann, M., Poggiato, G., Dray, S., Thuiller, W., Matias, C. & Miele, V. (2023) Quantifying the overall effect of biotic interactions on species distributions along environmental gradients.
- Oksanen, J., Blanchet, G., Friendly, M., Kindt, R. & Legendre, P. (2019) vegan : Community Ecology Package.
- Olden, J.D. (2006) Biotic homogenization : a new research agenda for conservation biogeography. *Journal of Biogeography*, **33**, 2027–2039.
- Olden, J.D. & Jackson, D.A. (2002) Illuminating the “black box” : a randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*, **154**, 135–150.
- Olden, J.D., LeRoy Poff, N., Douglas, M.R., Douglas, M.E. & Fausch, K.D. (2004) Ecological and evolutionary consequences of biotic homogenization. *Trends in Ecology & Evolution*, **19**, 18–24.
- Ovaskainen, O. (2020) *Joint species distribution modelling : with applications in R*. Ecology, biodiversity and conservation. Cambridge University Press, Cambridge, United Kingdom ; New York, NY.
- Paavola, R., Muotka, T., Virtanen, R., Heino, J., Jackson, D. & Mäki-Petäys, A. (2006) Spatial Scale Affects Community Concordance Among Fishes, Benthic Macroinvertebrates, And Bryophytes In Streams. *Ecological Applications*, **16**, 368–379.

- Panter, C.T., Bachman, S.P., Baines, O., Bruelheide, H., Kambach, S., Sporberr, M., Field, R. & Schrodt, F. (2023) Species abundances often conform to ‘abundant-centre’ patterns depending on dispersal capabilities. *bioRxiv*. Publisher : Cold Spring Harbor Laboratory _eprint : <https://www.biorxiv.org/content/early/2023/04/03/2023.03.31.535106.full.pdf>.
- Park, Y.S., Grenouillet, G., Esperance, B. & Lek, S. (2006) Stream fish assemblages and basin land cover in a river network. *Science of The Total Environment*, **365**, 140–153.
- Parmesan, C., Gaines, S., Gonzalez, L., Kaufman, D.M., Kingsolver, J., Townsend Peterson, A. & Sagarin, R. (2005) Empirical perspectives on species borders : from traditional biogeography to global change. *Oikos*, **108**, 58–75.
- Pebesma, E. (2018) Simple Features for R : Standardized Support for Spatial Vector Data.
- Pella, H., Lejot, J., Lamouroux, N. & Snelder, T. (2012) Le réseau hydrographique théorique (RHT) français et ses attributs environnementaux. *Géomorphologie : relief, processus, environnement*, **18**, 317–336.
- Peres-Neto, P.R. & Jackson, D.A. (2001) How well do multivariate data sets match ? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia*, **129**, 169–178.
- Peres-Neto, P.R., Legendre, P., Dray, S. & Borcard, D. (2006) Variation partitioning of species data matrices : estimation and comparison of fractions. *Ecology*, **87**, 2614–2625.
- Persat, H. & Keith, P. (2020) Biogéographie et historique de la mise en place des peuplements ichtyologiques de France métropolitaine. *Les poissons d’eau douce de France*, Inventaires et biodiversité, pp. 43–98. Biotope Éditions, Mèze ; Muséum National d’Histoire naturelle, Paris, Paris, 2nde édition.
- Peterson, A.T., Papeş, M. & Soberón, J. (2015) Mechanistic and Correlative Models of Ecological Niches. *European Journal of Ecology*, **1**, 28–38.
- Peterson, A.T. & Soberón, J. (2012) Species Distribution Modeling and Ecological Niche Modeling : Getting the Concepts Right. *Natureza & Conservação*, **10**, 102–107.
- Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M. & Araújo, M.B. (2011) *Ecological niches and geographic distributions*. Number 49 in Monographs in population biology. Princeton University Press, Princeton, N.J.
- Petts, G.E. & Amoros, C. (1996) *Fluvial hydrosystems*. Chapman et Hall.
- Picard, C., Floury, M., Seyedhashemi, H., Morel, M., Pella, H., Lamouroux, N., Buisson, L., Moatar, F. & Maire, A. (2022) Direct habitat descriptors improve the understanding of the organization of fish and macroinvertebrate communities across a large catchment. *PLOS ONE*, **17**, e0274167.
- Pletterbauer, F., Graf, W. & Schmutz, S. (2016) Effect of biotic dependencies in species distribution models : The future distribution of *Thymallus thymallus* under consideration of *Allogamus auricollis*. *Ecological Modelling*, **327**, 95–104.
- Poff, N.L. (1997) Landscape Filters and Species Traits : Towards Mechanistic Understanding and Prediction in Stream Ecology. *Journal of the North American Benthological Society*, **16**, 391–409.
- Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O’Hara, R.B., Parris, K.M., Vesk, P.A. & McCarthy, M.A. (2014) Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution*, **5**, 397–406.
- Potts, J.M. & Elith, J. (2006) Comparing species abundance models. *Ecological Modelling*, **199**, 153–163.
- Poulet, N., Beaulaton, L. & Dembski, S. (2011) Time trends in fish populations in metropolitan France : insights from national monitoring data. *Journal of Fish Biology*, **79**, 1436–1452.
- Prié, V., Molina, Q. & Gamboa, B. (2014) French naiad (Bivalvia : Margaritiferidae, Unionidae) species distribution models : prediction maps as tools for conservation. *Hydrobiologia*, **735**, 81–94.

- Pulliam, H. (2000) On the relationship between niche and distribution. *Ecology Letters*, **3**, 349–361.
- Quintana-Segui, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L. & Morel, S. (2008) Analysis of near-surface atmospheric variables : Validation of the SAFRAN analysis over France. *Journal of applied meteorology and climatology*, **47**, 92–107.
- R Core Team (2022) *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ribeiro, D., Gkenas, C., Gago, J. & Ribeiro, F. (2021) Variation in Diet Patterns of the Invasive Top Predator Sander lucioperca (Linnaeus, 1758) across Portuguese Basins. *Water*, **13**, 2053.
- Ricker, W.E. (1934) *An ecological classification of certain Ontario streams*, volume 49. University of Toronto Press, Toronto.
- Robinson, O.J., Ruiz-Gutierrez, V. & Fink, D. (2018) Correcting for bias in distribution modelling for rare species using citizen science data. *Diversity and Distributions*, **24**, 460–472.
- Rumschlag, S.L., Mahon, M.B., Jones, D.K., Battaglin, W., Behrens, J., Bernhardt, E.S., Bradley, P., Brown, E., De Laender, F., Hill, R., Kunz, S., Lee, S., Rosi, E., Schäfer, R., Schmidt, T.S., Simonin, M., Smalling, K., Voss, K. & Rohr, J.R. (2023) Density declines, richness increases, and composition shifts in stream macroinvertebrates. *Science Advances*, **9**, eadf4896.
- Sales, L.P., Hayward, M.W. & Loyola, R. (2021) What do you mean by “niche”? Modern ecological theories are not coherent on rhetoric about the niche concept. *Acta Oecologica*, **110**, 103701.
- Sauquet, E., Gottschalk, L. & Leblois, E. (2000) Mapping average annual runoff : a hierarchical approach applying a stochastic interpolation scheme. *Hydrological Sciences Journal*, **45**, 799–815.
- Scheiner, S.M. & Willig, M.R., eds. (2011) *The theory of ecology*. University of Chicago Press, Chicago, Ill.
- Segal, R.D., Massaro, M., Carlile, N. & Whitsed, R. (2021) Small-scale species distribution model identifies restricted breeding habitat for an endemic island bird. *Animal Conservation*, **24**, 959–969.
- Syedhashemi, H., Moatar, F., Vidal, J.P., Diamond, J.S., Beaufort, A., Chandesris, A. & Valette, L. (2021a) Thermal signatures identify the influence of dams and ponds on stream temperature at the regional scale. *Science of The Total Environment*, p. 142667.
- Syedhashemi, H., Moatar, F., Vidal, J.P. & Thiéry, D. (2023) Past and future discharge and stream temperature at high spatial resolution in a large European basin (Loire basin, France). preprint, ESSD – Land/Hydrology.
- Syedhashemi, H., Vidal, J.P., Diamond, J.S., Thiéry, D., Monteil, C., Hendrickx, F., Maire, A. & Moatar, F. (2021b) Regional, multi-decadal analysis reveals that stream temperature increases faster than air temperature. *Hydrology and Earth System Sciences Discussions*, pp. 1–31.
- Syedhashemi, H., Vidal, J.P., Diamond, J.S., Thiéry, D., Monteil, C., Hendrickx, F., Maire, A. & Moatar, F. (2022) Regional, multi-decadal analysis on the Loire River basin reveals that stream temperature increases faster than air temperature. *Hydrology and Earth System Sciences*, **26**, 2583–2603.
- Sillero, N., Campos, J.C., Arenas-Castro, S. & Barbosa, A.M. (2023) A curated list of R packages for ecological niche modelling. *Ecological Modelling*, **476**, 110242.
- Snelder, T.H., Datry, T., Lamouroux, N., Larned, S.T., Sauquet, E., Pella, H. & Catalogne, C. (2013) Regionalization of patterns of flow intermittence from gauging station records. *Hydrology and Earth System Sciences*, **17**, 2685–2699.
- Soberon, J. & Nakamura, M. (2009) Niches and distributional areas : Concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences*, **106**, 19644–19650.
- Soberon, J. & Peterson, A.T. (2005) Interpretation of Models of Fundamental Ecological Niches and Species’ Distributional Areas. *Biodiversity Informatics*, **2**.

- Soberón, J. (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, **10**, 1115–1123.
- Soubeyroux, J.M., Bernus, S., Corre, L., Drouin, A., Dubuisson, B., Etchevers, P., Gouget, V., Josse, P., Kerdoncuff, M., Samacoits, R. & Tocquer, F. (2020) Les nouvelles projections climatiques de référence DRIAS 2020 pour la métropole. Technical report, Météo-France.
- Souchon, Y. & Tissot, L. (2012) Synthesis of thermal tolerances of the common freshwater fish species in large Western Europe rivers. *Knowledge and Management of Aquatic Ecosystems*, p. 03.
- Southwood, T.R.E. (1977) Habitat, the Templet for Ecological Strategies? *The Journal of Animal Ecology*, **46**, 336.
- Statzner, B. & Bêche, L.A. (2010) Can biological invertebrate traits resolve effects of multiple stressors on running water ecosystems? *Freshwater Biology*, **55**, 80–119.
- Statzner, B. & Higler, B. (1986) Stream hydraulics as a major determinant of benthic invertebrate zonation patterns. *Freshwater Biology*, **16**, 127–139.
- Steen, V.A., Tingley, M.W., Paton, P.W.C. & Elphick, C.S. (2021) Spatial thinning and class balancing : Key choices lead to variation in the performance of species distribution models with citizen science data. *Methods in Ecology and Evolution*, **12**, 216–226.
- Stephenson, F., Gladstone-Gallagher, R.V., Bulmer, R.H., Thrush, S.F. & Hewitt, J.E. (2022) Inclusion of biotic variables improves predictions of environmental niche models. *Diversity and Distributions*, **28**, 1373–1390.
- Stewart, D.R., Underwood, Z.E., Rahel, F.J. & Walters, A.W. (2018) The effectiveness of surrogate taxa to conserve freshwater biodiversity : Freshwater Conservation Planning. *Conservation Biology*, **32**, 183–194.
- Tachet, H., Richoux, P., Bournaud, M. & Usseglio-Polatera, P. (2010) *Invertébrés d'eau douce : systématique, biologie, écologie*, volume 15. CNRS éditions Paris.
- Thuiller, W. (2007) Climate change and the ecologist. *Nature*, **448**, 550–552.
- Thuiller, W., Brotons, L., Araújo, M.B. & Lavorel, S. (2004) Effects of restricting environmental range of data to project current and future species distributions. *Ecography*, **27**, 165–172.
- Thuiller, W., Lavorel, S., Araújo, M.B., Sykes, M.T. & Prentice, I.C. (2005) Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences*, **102**, 8245–8250.
- Tirelli, F.P., Migliorini, R.P., Appel, M.E., Trigo, T.C., Kasper, C.B., Peters, F.B., Figueiró, H.V. & Eizirik, E. (2023) Using DNA barcoding and species distribution models to characterize the trophic ecology of Geoffroy's cat in the Brazilian Pampa. *Mammal Research*.
- Tison-Rosebery, J., Leboucher, T., Archaimbault, V., Belliard, J., Carayon, D., Ferréol, M., Floury, M., Jeliakov, A., Tales, E., Villeneuve, B. & Passy, S. (2022) Decadal biodiversity trends in rivers reveal recent community rearrangements. *Science of The Total Environment*, **823**, 153431.
- Tourinho, L. & Vale, M.M. (2023) Choosing among correlative, mechanistic, and hybrid models of species' niche and distribution. *Integrative Zoology*, **18**, 93–109.
- Townsend, C.R., Doledec, S., Norris, R., Peacock, K. & Arbuckle, C. (2003) The influence of scale and geography on relationships between stream community composition and landscape variables : description and prediction. *Freshwater Biology*, **48**, 768–785.
- Townsend, C.R. & Hildrew, A.G. (1976) Field Experiments on the Drifting, Colonization and Continuous Redistribution of Stream Benthos. *The Journal of Animal Ecology*, **45**, 759.
- Usseglio-Polatera, P., Bournaud, M., Richoux, P. & Tachet, H. (2000) Biological and ecological traits of benthic freshwater macroinvertebrates : relationships and definition of groups with similar traits. *Freshwater Biology*, **43**, 175–205.
- Valavi, R., Elith, J., Lahoz-Monfort, J.J. & Guillera-Arroita, G. (2021) Modelling species presence-only data with random forests. *Ecography*, **44**, 1731–1742.

- Valavi, R., Elith, J., Lahoz-Monfort, J.J. & Guillera-Arroita, G. (2023) Flexible species distribution modelling methods perform well on spatially separated testing data. *Global Ecology and Biogeography*, p. geb.13639.
- Valette, L., Piffady, J., Chandesris, A. & Souchon, Y. (2012) SYRAH-CE : description des données et modélisation du risque d'altération hydromorphologique des cours d'eau pour l'état des lieux DCE. *Rapport Technique Onema-Irstea*.
- Vannote, R.L., Minshall, G.W., Cummins, K.W., Sedell, J.R. & Cushing, C.E. (1980) The River Continuum Concept. *Canadian Journal of Fisheries and Aquatic Sciences*, **37**, 130–137.
- Vaudor, L., Lamouroux, N., Olivier, J.M., Forcellini, M. & Maly, U. (2015) How sampling influences the statistical power to detect changes in abundance : an application to river restoration. *Freshwater Biology*, p. 16.
- Vehanen, T., Sutela, T. & Harjunpää, A. (2020) The effects of ecoregions and local environmental characteristics on spatial patterns in boreal riverine fish assemblages. *Ecology of Freshwater Fish*.
- Verneaux, J. (1973) *Cours d'eau de Franche-Comté (massif du Jura) : recherches écologiques sur le réseau hydrographique du Doubs – essai de biotypologie*. PhD Thesis, Université de Franche-Comté.
- Verneaux, J. (1976) Biotypologie de l'écosystème « eau courante ». La structure biotypologique. *Comptes Rendus de l'Académie des Sciences de Paris*.
- Verneaux, J. (1977a) Biotypologie de l'écosystème « eau courante ». Détermination approchée de l'appartenance typologique d'un peuplement ichtyologique. *Comptes Rendus de l'Académie des Sciences de Paris*, **284**, 675–678.
- Verneaux, J. (1977b) Biotypologie de l'écosystème « eau courante ». Déterminisme approché de la structure biotypologique. *Comptes Rendus de l'Académie des Sciences de Paris*, **284**, 77–79.
- Vidal, J.P., Martin, E., Franchistéguy, L., Baillon, M. & Soubeyrou, J.M. (2010) A 50-year high-resolution atmospheric reanalysis over France with the Safran system. *International Journal of Climatology*, **30**, 1627–1644.
- Vierron, A. & Steinbach, P. (2020) Valorisation de la connaissance des obstacles à l'écoulement - Calcul de trois indicateurs hydromorphologiques complémentaires de pression sur les cours d'eau. Technical report, Office Français de la Biodiversité.
- Villeneuve, B., Piffady, J., Valette, L., Souchon, Y. & Usseglio-Polatera, P. (2018) Direct and indirect effects of multiple stressors on stream invertebrates across watershed, reach and site scales : A structural equation modelling better informing on hydromorphological impacts. *Science of The Total Environment*, **612**, 660–671.
- Villéger, S., Blanchet, S., Beauchard, O., Oberdorff, T. & Brosse, S. (2011) Homogenization patterns of the world's freshwater fish faunas. *Proceedings of the National Academy of Sciences*, **108**, 18003–18008.
- Violle, C., Thuiller, W., Mouquet, N., Munoz, F., Kraft, N.J., Cadotte, M.W., Livingstone, S.W. & Mouillot, D. (2017) Functional Rarity : The Ecology of Outliers. *Trends in Ecology & Evolution*, **32**, 356–367.
- Wagner, T., Hansen, G.J.A., Schliep, E.M., Bethke, B.J., Honsey, A.E., Jacobson, P.C., Kline, B.C. & White, S.L. (2020) Improved understanding and prediction of freshwater fish communities through the use of joint species distribution models. *Canadian Journal of Fisheries and Aquatic Sciences*, pp. cjfas-2019-0348.
- Wagner, T., Schliep, E.M., North, J.S., Kundel, H., Custer, C.A., Ruzich, J.K. & Hansen, G.J.A. (2023) Predicting climate change impacts on poikilotherms using physiologically guided species abundance models. *Proceedings of the National Academy of Sciences*, **120**, e2214199120.
_eprint : <https://www.pnas.org/doi/pdf/10.1073/pnas.2214199120>.
- Waldock, C., Stuart-Smith, R.D., Albouy, C., Cheung, W.W.L., Edgar, G.J., Mouillot, D., Tjiputra, J. & Pellissier, L. (2021) A quantitative review of abundance-based species distribution models. *Ecology*.

- Waldock, C., Stuart-Smith, R.D., Edgar, G.J., Bird, T.J. & Bates, A.E. (2019) The shape of abundance distributions across temperature gradients in reef fishes. *Ecology Letters*, **22**, 685–696.
- Ward, J.V. (1989) The Four-Dimensional Nature of Lotic Ecosystems. *Journal of the North American Benthological Society*, **8**, 2–8.
- Wasson, J.G. (1989) Éléments pour une typologie fonctionnelle des eaux courants : 1. Revue critique de quelques approches existantes. *Bull Ecol*, **20**, 109–127.
- Wasson, J., Chandesris, A., Pella, H. & Blanc, L. (2002) Les hydro-écorégions de France métropolitaine : Approche régionale de la typologie des eaux courantes et éléments pour la définition des peuplements de référence d'invertébrés. Technical report, Cemagref, Lyon.
- Wehrly, K.E., Wiley, M.J. & Seelbach, P.W. (2003) Classifying Regional Variation in Thermal Regime Based on Stream Fish Community Patterns. *Transactions of the American Fisheries Society*, **132**, 18–38.
- Whittaker, R.J., Araújo, M.B., Jepson, P., Ladle, R.J., Watson, J.E.M. & Willis, K.J. (2005) Conservation Biogeography : assessment and prospect : Conservation Biogeography. *Diversity and Distributions*, **11**, 3–23.
- Wiens, J.A. (2002) Riverine landscapes : taking landscape ecology into the water. *Freshwater Biology*, **47**, 501–515.
- Woods, T., Freeman, M.C., Krause, K.P. & Maloney, K.O. (2023) Observed and projected functional reorganization of riverine fish assemblages from global change. *Global Change Biology*, p. gcb.16707.
- Yañez-Arenas, C., Martín, G., Osorio-Olvera, L., Escobar-Luján, J., Castaño-Quintero, S., Chiappa-Carrara, X. & Martínez-Meyer, E. (2020) The abundant niche-centroid hypothesis : key points about unfilled niches and the potential use of supraspecific modeling units. *Biodiversity Informatics*, p. 11.
- Zhao, Q. & Hastie, T. (2021) Causal Interpretations of Black-Box Models. *Journal of Business & Economic Statistics*, **39**, 272–281.

Annexes

Table des annexes

Annexe A.	Liste des communications réalisées pendant la thèse	239
Annexe B.	Article publié dans Plos One	239
Annexe C.	Procédure de jointure des réseaux T-NET et RHT	284
Annexe D.	Boxplot des performances des modèles de la distribution des classes d'abondance pour les espèces de poissons et les genres de macroinvertébrés	285
Annexe E.	Exemples supplémentaires de <i>Partial Dependance Plot</i>	286
Annexe F.	Rang des prédicteurs avant intégration des indicateurs de ressources macroinvertébrés dans les modèles basés sur le jeu de données communs	287
Annexe G.	Cartographies des projections futures pour les poissons (n = 26 espèces) et les macroinvertébrés (n = 31 genres)	288

A. Liste des communications réalisées pendant la thèse

Article publié :

- Picard, C., Flourey, M., Seyedhashemi, H., Morel, M., Pella, H., Lamouroux, N., Buisson, L., Moatar, F., & Maire, A. (2022). Direct habitat descriptors improve the understanding of the organization of fish and macroinvertebrate communities across a large catchment. PLOS ONE, 17(9), e0274167. <https://doi.org/10.1371/journal.pone.0274167>

Liste des communications orales :

- Journées des thèses INRAE - 05/2021 à INRAE Villeurbanne.

Présentation orale : Spatial organization of aquatic communities at large scale : environmental drivers under climate change.

- Symposium for European Freshwater Sciences (SEFS12) - 07/2021.

Présentation orale : Direct habitat descriptors to better understand the spatial organization of aquatic communities.

- Congress of the International Society of Limnology (SIL) - 08/2022 à Berlin.

Présentation orale : Species abundance models in freshwaters : developments for fish and macroinvertebrates in a large catchment.

- Colloque thermie-hydrobiologie - 11/2022 à Saclay.

Présentation orale : Structuration spatiale des communautés aquatiques à large échelle : influence des facteurs environnementaux dans un contexte de changement climatique.

Présentation disponible en ligne au lien suivant :

https://www.youtube.com/watch?v=m7Ax7NPQsxE&list=PL1jC-T5uY_CVNvAm2vepVZouuR03RjHx9&index=6

- DECRYPThèse, journées des thèses de l'école doctorale E2M2 - 05/2023 à Villeurbanne.

Poster : Modélisation de la distribution des abondances des communautés aquatiques du bassin versant de la Loire.

RESEARCH ARTICLE

Direct habitat descriptors improve the understanding of the organization of fish and macroinvertebrate communities across a large catchment

Coline Picard^{1,2,*}, Mathieu Floury³, Hanieh Seyedhashemi^{2,4}, Maxime Morel², Hervé Pella², Nicolas Lamouroux², Laëticia Buisson⁵, Florentina Moatar², Anthony Maire¹

1 EDF R&D LNHE—Laboratoire National d'Hydraulique et Environnement, Chatou, France, **2** INRAE, UR RiverLy, Centre de Lyon-Villeurbanne, Villeurbanne, France, **3** Univ Lyon, Université Claude Bernard Lyon 1, CNRS, ENTPE, UMR 5023 LEHNA, Villeurbanne, France, **4** EA 6293 GéHCO, Université de Tours, Tours, France, **5** Laboratoire Écologie Fonctionnelle et Environnement, Université de Toulouse, CNRS, Toulouse INP, Université Toulouse 3 –Paul Sabatier (UPS), Toulouse, France

* coline.picard3@gmail.com



OPEN ACCESS

Citation: Picard C, Floury M, Seyedhashemi H, Morel M, Pella H, Lamouroux N, et al. (2022) Direct habitat descriptors improve the understanding of the organization of fish and macroinvertebrate communities across a large catchment. *PLoS ONE* 17(9): e0274167. <https://doi.org/10.1371/journal.pone.0274167>

Editor: Frank O. Masese, University of Eldoret, KENYA

Received: July 21, 2021

Accepted: August 23, 2022

Published: September 22, 2022

Copyright: © 2022 Picard et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The biological data underlying the results presented in the study are available from <http://www.naiades.eaufrance.fr/acces-donnees#/hydrobiologie> The environmental data underlying the results presented in the study are available from Beaufort et al. (2016) [DOI: [10.1002/hyp.10787](https://doi.org/10.1002/hyp.10787)], Pella et al. (2012) [DOI: [10.4000/geomorphologie.9933](https://doi.org/10.4000/geomorphologie.9933)] and Morel et al. (2020) [DOI: [10.1016/j.jhydrol.2019.124292](https://doi.org/10.1016/j.jhydrol.2019.124292)].

Abstract

In large-scale aquatic ecological studies, direct habitat descriptors (e.g. water temperature, hydraulics in river reaches) are often approximated by coarse-grain surrogates (e.g. air temperature, discharge respectively) since they are easier to measure or model. However, as biological variability can be very strong at the habitat scale, surrogate variables may have a limited ability to capture all of this variability, which may lead to a lesser understanding of the ecological processes or patterns of interest. In this study, we aimed to compare the capacity of direct habitat descriptors vs. surrogate environmental variables to explain the organization of fish and macroinvertebrate communities across the Loire catchment in France (10⁵ km²). For this purpose, we relied on high-resolution environmental data, extensive biological monitoring data (>1000 sampling stations) and multivariate analyses. Fish and macroinvertebrate abundance datasets were considered both separately and combined to assess the value of a cross-taxa approach. We found that fish and macroinvertebrate communities exhibited weak concordance in their organization and responded differently to the main ecological gradients. Such variations are probably due to fundamental differences in their life-history traits and mobility. Regardless of the biological group considered, direct habitat descriptors (water temperature and local hydraulic variables) consistently explained the organization of fish and macroinvertebrate communities better than surrogate descriptors (air temperature and river discharge). Furthermore, the organization of fish and macroinvertebrate communities was slightly better explained by the combination of direct or surrogate environmental variables when the two biological groups were considered together than when considered separately. Tied together, these results emphasize the importance of using a cross-taxa approach in association with high-resolution direct habitat variables to more accurately explain the organization of aquatic communities.

Funding: This study was funded by Electricité de France as part of the research team Hynes built within the EDF-INRAE Framework Agreement on Water Engineering and Aquatic Systems. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Freshwater ecosystems are amongst the most vulnerable and threatened ecosystems worldwide [1–3]. Developing efficient conservation plans to counteract the erosion of their biodiversity requires a better understanding of the environmental drivers and ecological processes that shape aquatic communities in space and time [4–6]. Nonetheless, characterizing organism–environment relationships in rivers is challenging because of the dendritic nature of hydrographic networks that constrains their organization [7, 8]. Historically, freshwater ecologists have attempted to measure the degree of alteration of a system by comparing its state to biological reference conditions [9, 10]. This has notably led to the formulation of river typologies and zonations to describe the natural (i.e. with a minimal degree of alteration) organization of aquatic organisms and identify their environmental determinants [e.g. 11–14]. The main environmental determinants that naturally shape aquatic communities are river thermal and hydrological regimes, topography, organic matter inputs and local (i.e. at reach scale) habitat characteristics [15–18]. Most of these drivers are structured longitudinally (e.g. increases in water temperature and discharge from the source to the mouth), forming the upstream–downstream gradient along which aquatic communities are distributed [19, 20].

A prevailing assumption states that at large scale, the different groups of freshwater organisms (e.g. fish, macroinvertebrates, aquatic plants) are assembled and structured by similar environmental filters [21], which is expected to result in shared spatial patterns [22, 23]. Consequently, most studies have focused on a single group, such as fish or macroinvertebrates, and have assumed its representativeness with respect to the other groups [6, 12, 18]. However, recent studies have questioned this assumption and investigated the spatial concordance between biological groups in freshwater systems at different spatial scales [24, 25]. For instance, Backus-Freer and Pyron [26] showed that the spatial distribution of macroinvertebrates is better predicted by local environmental variations (e.g. in-stream cover, turbidity) than the distribution of fish, resulting in a low concordance between the organization of fish and macroinvertebrate communities. Other studies have suggested that the concordance of spatial patterns among aquatic organisms may increase with spatial scale, probably due to the weaker contribution of local specificities in coarser regional patterns [22, 23]. In addition, spatial concordance among communities may result from mechanisms other than environmental filtering, including biotic interactions, dispersal dynamics, assemblage homogenization due to the loss of sensitive taxa and scale effects [16, 23, 27]. Therefore, considering different biological groups when investigating broad ecological patterns and processes appears relevant, if not necessary, to ensure that the results are integrative and generalizable. This also allows the identification of patterns at ecosystem level, not just at population or community level.

In this study, we aimed at comparing the capacity of surrogate *vs.* direct environmental variables to explain the organization of aquatic communities in a large catchment. Surrogate variables refer to variables classically used as indirect proxies for the environmental conditions experienced by aquatic organisms (e.g. air temperature to approximate water temperature) or that measure general characteristics of the river reach not directly related to the habitat of the organisms (e.g. discharge to approximate current velocity or water depth) [28]. Many studies have had to rely on surrogate variables in the absence of finer or more precise data available at large scale or over a sufficiently long period [e.g. 5, 29–31]. In contrast, direct variables are those that directly translate the environmental conditions experienced by aquatic organisms. For this purpose, we used an extensively sampled dataset of fish and macroinvertebrate communities combined with high-resolution environmental variables (including both direct and surrogate variables) over the Loire catchment in France. Fish and macroinvertebrate communities were considered separately and together, making it possible to investigate potential

differences between biological groups particularly regarding spatial concordance. First, we hypothesized that direct variables would better explain the organization of fish and macroinvertebrate communities than surrogate variables since they more accurately described the conditions experienced by organisms and their ecological preferences [28]. In addition, we hypothesized that direct variables would better explain the organization of macroinvertebrate communities than that of fish communities due to their higher sensibility to local factors, in relation to the smaller home-range size and lower dispersal capacity of macroinvertebrates [24–26]. Last, given the complexity of the relationships between aquatic communities and environmental gradients, we hypothesized the organization of the whole community (i.e. fish and macroinvertebrates combined) to reflect more accurately the biological gradient and to be better explained by environmental predictors than that of fish and macroinvertebrates taken separately [16, 18, 27].

Materials and methods

Study area

The Loire catchment is the largest in France, with a drainage area of about 117,000 km², covering 20% of the French metropolitan territory. The Loire River is 1,012 km long and runs through a variety of landscapes and climates from its source at Gerbier-de-Jonc to its estuary at Saint-Nazaire where it flows into the Atlantic Ocean. The upper part of the catchment is located within a forested mountain range (the Massif Central) while agriculture dominates in the downstream areas. The sedimentary basin is mostly occupied by cereal farming and vineyards and the western and south-eastern parts of the catchment being mainly dedicated to livestock (mostly cattle). Annual precipitation varies between 600 and 1,300 mm per year over the catchment and average air temperature ranges from 6 to 12.5°C. The Loire catchment shows a variety of hydrological regimes with eight regimes out of the 12 present in France [32, 33]. Specific discharge ranges from 3 l/s/km² (in the Maine catchment) to 40 l/s/km² (in the Cévennes). Seasonal variation in discharge can be high for some catchments due to very little groundwater supply [33]. The mean interannual discharge of the Loire River in its downstream part (at Montjean-sur-Loire) is 840 m³/s, with flood peaks 5–10 times higher (Banque Hydro; www.hydro.eaufrance.fr). The study area includes the entire Loire catchment except the part under marine influence, downstream of Montjean-sur-Loire (110 kilometers upstream from the estuary) (Fig 1).

Environmental and biological datasets

Direct vs. surrogate environmental variables. All environmental variables were selected on the basis of knowledge of their potential influence on fish and macroinvertebrate communities (i.e. literature, expertise). A set of 23 environmental variables were considered, including descriptors of water and air temperatures, discharge, local hydraulic conditions and topography (see Table 1 for the full list of variables). They were described at the scale of the river reach (mean length \pm standard deviation = 1.7 \pm 1.4 kilometers), which is the spatial unit of the physical-based model Temperature-NETwork (T-NET). This model is based on a hydrographic network of the Loire catchment through 52,278 reaches and use the equilibrium temperature concept [34].

Discharge data were obtained from the semi-distributed hydrological EROS model [35], which simulates daily discharge at the outlet of 368 sub-basins within the Loire catchment (sub-catchments area range from 40 to 1600 km² with mean = 300 km²). Within a sub-basin, daily discharge is extrapolated depending on drainage area. The EROS model performed well

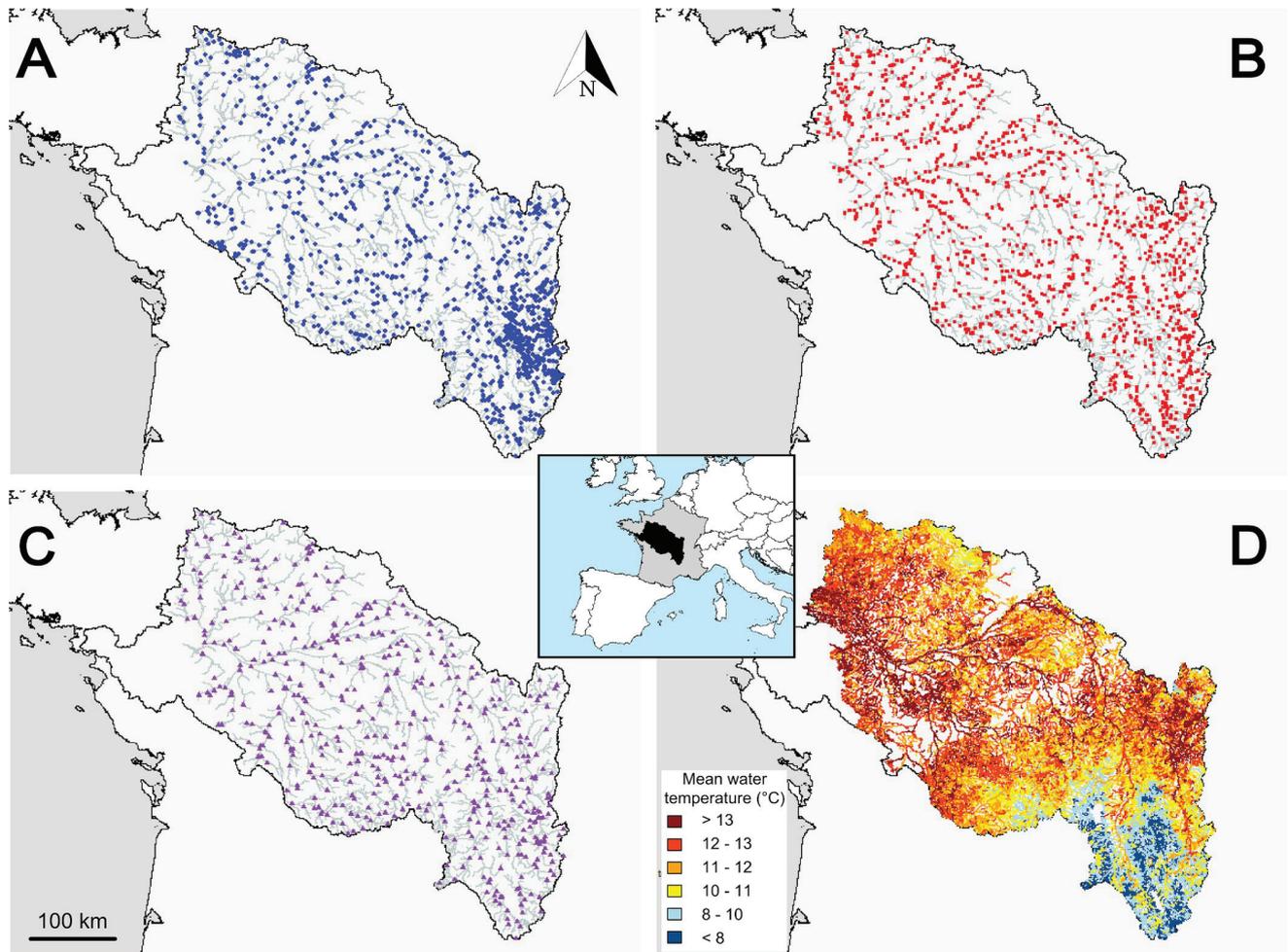


Fig 1. Study area and location of the sampled reaches for (A) fish, (B) macroinvertebrates and (C) both (i.e. “common” dataset) and (D) T-NET hydrographic network. In panels A, B and C, a simplified version of the T-NET hydrographic network is shown. In panel D, the 52,278 T-NET reaches are colored according to their mean annual water temperature (°C). The inserted map in the center locates the Loire catchment within Europe (source: Natural Earth).

<https://doi.org/10.1371/journal.pone.0274167.g001>

on 44 stations at the annual scale with a mean relative bias \pm standard deviation of $-0.12 \pm 10.62\%$.

Air temperature was derived from SAFRAN, which is a mesoscale atmospheric analysis system for surface variables at the hourly time step using ground data observations at the spatial resolution of a grid of 8 x 8 km [36, 37]. At annual scale the RMSE between observations and corresponding SAFRAN outputs was of 1°C [37].

Discharge statistics were translated into direct hydraulic statistics (i.e. reach width, water depth and current velocity) using the hydraulic geometry models from Morel et al. [38]. These models improve the classical hydraulic geometry equations of Leopold and Maddock [39] and relate hydraulics to discharge using power laws. To date, the models of Morel et al. [38] provide the best estimates of reach hydraulics at a given river discharge in the absence of detailed field hydraulic measurements. These models were available for a hydrographic network of the Loire catchment (Réseau Hydrographique Théorique, RHT) [40] coarser from that of the T-NET model. Therefore, we made a correspondence between RHT and T-NET river reaches, so that 49% of T-NET reaches had corresponding RHT reaches. For T-NET reaches without

Table 1. Range of environmental values covered by each dataset. Average value and range (i.e. [minimum: maximum]) of the 23 environmental descriptors computed on: (1) the entire T-NET network (N = 52,278 reaches), (2) reaches with fish data (N = 1,191), (3) reaches with macroinvertebrate data (N = 1,195), and (4) reaches with fish and macroinvertebrate data (“common” dataset; N = 532). *Ta* and *Tw* refer to air and water temperature, respectively. Suffixes *m*, *sd*, *10* and *90* refer to the mean, standard deviation, lower 10% quantile and upper 90% quantile, respectively.

	Environmental variables	T-NET network	Fish	Macroinvertebrates	Common
Surrogate descriptors	Discharge m (m ³ /s)	2.9 [<0.01: 820.3]	10.2 [<0.01: 675.4]	8.5 [0.01: 675.4]	14.7 [0.01: 675.4]
	Discharge 10 (m ³ /s)	0.7 [<0.01: 212.1]	2.5 [<0.01: 171.1]	2.1 [<0.01: 171.1]	3.6 [<0.01: 171.1]
	Discharge 90 (m ³ /s)	6.2 [<0.01: 1704.9]	21.6 [<0.01: 1408.7]	18 [0.02: 1408.7]	31.2 [0.02: 1408.7]
	Discharge sd (m ³ /s)	2.7 [<0.01: 667.4]	9.2 [<0.01: 547.4]	7.9 [0.01: 547.4]	13.4 [0.01: 547.4]
	Ta m (°C)	10.7 [6.1: 12.5]	10.3 [6.2: 12.5]	10.9 [6.2: 12.5]	10.8 [6.2: 12.5]
	Ta 10 (°C)	1.9 [-2.5: 4.2]	1.4 [-2.5: 4.2]	2.0 [-2.5: 4.2]	2.0 [-2.1: 4.2]
	Ta 90 (°C)	19.4 [15.3: 21.1]	19.2 [15.3: 21.1]	19.6 [15.3: 21.1]	19.6 [15.3: 21.1]
	Ta sd (°C)	6.6 [5.5: 7.3]	6.7 [5.7: 7.3]	6.6 [5.7: 7.3]	6.7 [5.7: 7.3]
Direct descriptors	Depth m (m)	0.24 [0.02: 2.86]	0.42 [0.05: 2.61]	0.46 [0.07: 2.61]	0.55 [0.07: 2.61]
	Depth 10 (m)	0.15 [0.01: 1.79]	0.26 [0.04: 1.65]	0.28 [0.04: 1.65]	0.34 [0.05: 1.65]
	Depth 90 (m)	0.30 [0.03: 3.70]	0.53 [0.06: 3.34]	0.58 [0.08: 3.34]	0.70 [0.08: 3.34]
	Velocity m (m/s)	0.08 [<0.01: 4.14]	0.20 [<0.01: 3.78]	0.20 [0.01: 3.78]	0.28 [0.01: 3.78]
	Velocity 10 (m/s)	0.02 [<0.01: 1.23]	0.05 [<0.01: 1.08]	0.05 [<0.01: 1.08]	0.08 [<0.01: 1.08]
	Velocity 90 (m/s)	0.15 [<0.01: 7.97]	0.40 [0.01: 7.37]	0.40 [0.01: 7.37]	0.56 [0.01: 7.37]
	Width m (m)	4.6 [0.1: 204.9]	11.2 [0.5: 176.3]	11.3 [0.7: 176.3]	15.6 [0.7: 176.3]
	Width 10 (m)	4.0 [0.08: 181.2]	9.7 [0.4: 156.7]	9.8 [0.6: 156.7]	13.5 [0.6: 156.7]
	Width 90 (m)	5.0 [0.1: 219.0]	12.1 [0.5: 187.8]	12.2 [0.7: 187.8]	16.8 [0.7: 187.8]
	Tw m (°C)	11.2 [5.4: 16.2]	11.0 [5.9: 14.1]	11.5 [6.8: 14.1]	11.5 [6.8: 14.1]
	Tw 10 (°C)	6.3 [0.1: 11.5]	4.3 [0.1: 9.5]	4.8 [0.2: 8.7]	4.6 [0.5: 8.4]
	Tw 90 (°C)	17.2 [8.4: 27.6]	18.5 [10.2: 24.5]	19.1 [11.4: 24.5]	19.3 [11.4: 24.5]
	Tw sd (°C)	4.1 [0.3: 8.0]	5.3 [2.0: 7.7]	5.4 [2.7: 7.7]	5.5 [2.7: 7.7]
Others	Distance to the estuary (km)	514.5 [110.1: 1004.5]	601.0 [114.7: 999.5]	512.0 [120.2: 999.5]	511.4 [138.2: 999.5]
	Slope (‰)	0.02 [<0.01: 0.40]	0.02 [<0.01: 0.17]	0.01 [<0.01: 0.12]	0.01 [<0.01: 0.12]

<https://doi.org/10.1371/journal.pone.0274167.t001>

RHT correspondence, the hydraulic geometry coefficients were extrapolated from the nearest downstream T-NET reach with RHT correspondence.

Water temperature data were obtained from T-NET, which simulates daily water temperature for each reach of the entire river network of the Loire catchment (Fig 1D). Comparison of annual average values (i.e. the time-scale of this study) between simulations and observations at 72 stations (2010–2014) showed good performance of the T-NET model with mean bias \pm standard deviation of $0.2 \pm 0.82^\circ\text{C}$ and mean root-mean square error (RMSE) = 0.7°C .

Based on the daily T-NET, SAFRAN and EROS outputs, we computed the interannual average of four summary statistics (i.e. annual mean, lower 10% quantile, upper 90% quantile and standard deviation) for water temperature, air temperature and discharge, respectively, in order to describe the thermal and hydrological regimes of each river reach over the period 1990–2010. Three summary statistics (i.e. annual mean, lower 10% quantile and upper 90% quantile) were calculated for hydraulic variables. Water temperature and hydraulics descriptors were considered as direct variables whereas discharge and air temperature descriptors were considered as surrogate variables. The mean slope of the river reach and the distance to the estuary, which are known to influence aquatic communities, were also considered. The mean slope of the river reach was extracted from the digital elevation model IGN BD ALTI® 25x25 m [41]. The distance to the estuary was obtained by summing the length of all T-NET reaches located downstream of the reach of interest. Metrics of discharge, hydraulic variables and slope were log-transformed prior to analyses to approach normality.

Fish and macroinvertebrate communities. Biological data were extracted from the French national hydrobiological monitoring database built in the context of the European Water Framework Directive [42] and the regulatory hydrobiological monitoring of five nuclear power plants (NPP) operated by Électricité de France (EDF) in the Loire catchment. Sampling operations were mostly conducted in summer and fall.

Fish data were assembled from the Naïades database [43] of the Office Français de la Biodiversité (OFB) and from the NPP database [44]. All fish samples kept for analysis were collected using standardized electrofishing protocols over the period 1994–2017 [5, 45]. The initial fish dataset consisted of abundance data for 50 species (see [S1 Table](#) for species list) collected during 5,225 fishing operations conducted in 1,393 stations. Fieldwork was conducted with adequate administrative permits for electrofishing in accordance with French laws and ethical rules.

Macroinvertebrate data were assembled from the regional databases of the Directions Régionales de l'Environnement, de l'Aménagement et du Logement (DREAL) and covered the period 1992–2015. All samples were collected following a common normalized protocol (multi-habitat sampling; norm XP T90-333; [46]). According to this protocol, eight samples were collected in different habitat types (i.e. combinations of substrates, water velocity and height) with a Surber net. Taxa identification ranged from order to genus level, but only data available at genus level were kept for consistency. The initial macroinvertebrate dataset consisted of abundance data for 230 genera collected during 4,645 sampling operations conducted at 1,215 stations ([S2 Table](#)). Abundances of fish species and macroinvertebrate genera were converted into catch per unit effort (CPUE), defined as the number of specimens per 100 m² sampled (i.e. densities).

We associated the biological data to their corresponding T-NET reach by projecting sampling stations onto the T-NET hydrographic network. When stations had been sampled more than once over the study period or when two or more stations were located on the same T-NET river reach, CPUE were averaged for each taxon and species. This resulted in a total of 1,191 and 1,195 T-NET reaches sampled at least once for fish and macroinvertebrates, respectively. They were evenly distributed over the Loire catchment ([Fig 1A and 1B](#)). Finally, a third dataset was built with the 532 T-NET river reaches that included at least one sampling operation for fish and at least one for macroinvertebrates (hereafter, “common” dataset; [Fig 1C](#)). To limit the influence of the rarest species and genera, those with a frequency of occurrences of less than 0.5% were excluded from the analyses. Hence, the final dataset included the CPUE of 43 fish species and 186 macroinvertebrate genera.

Data analysis

Non-metric multidimensional scaling (NMDS) was used to examine the organization of aquatic communities at the scale of the entire Loire catchment. This iterative ordination method is based on a ranked dissimilarity matrix between statistical samples (here T-NET reaches) and is adapted to ecological datasets including rare and low abundance taxa, especially when data from different biological groups are considered [47]. In addition, NMDS performs well in highlighting key organizational patterns at the community level, without placing too much emphasis on patterns specific to certain species or taxa. First, for each biological dataset, a dissimilarity matrix based on genera or species Wisconsin-transformed CPUE was built using Bray-Curtis dissimilarity among T-NET reaches. The best ordination was selected after 200 iterations from a random initial ordination. We assessed the quality of the NMDS using the stress value, which is a measure of the goodness-of-fit between the ranks of the observed dissimilarities and ordination distances. When stress value is lower than 0.2, the ordination is considered correct (accurate if stress values are lower than 0.1) [48].

Pairwise ordinations between biological groups were then compared using a Procrustes analysis. This method assesses the similarity of two factorial spaces and thus here the spatial concordance between biological groups. Procrustes analysis was preferred to a Mantel test because of its greater statistical power [49]. The Procrustes sum of squares m^2 , also called concordance value, was used to measure the similarity between two ordinations [22]. It ranges from 0 to 1 with low m^2 values indicating a higher degree of similarity. Because Procrustes analyses must be conducted on ordinations performed with the same samples, NMDS were reproduced considering only the 532 T-NET reaches common to all datasets (i.e. fish, macroinvertebrate and “common”). Finally, a fourth NMDS was computed based on the 23 environmental descriptors (Table 1), for the 532 reaches common to all three biological datasets. This fourth ordination was then compared to each biological ordination using Procrustes analyses to assess the concordance between environmental gradients and the organization of each biological group.

To assess the influence of the previously described environmental variables on the organization of aquatic communities resulting from the NMDS, we perform a redundancy analysis (RDA). RDA were used as an extension of linear regression allowing to handle the multiple response variables and to overcome the collinearity inherent to the calculated metrics (i.e. mean, lower 10th quantile, upper 90th quantile and standard deviation) [50]. First, for each biological group (i.e. fish, macroinvertebrates and common) and each environmental variable (i.e. water temperature, air temperature, hydraulic variables and discharge), a RDA was performed with the three NMDS axes as response variables and groups of 4 metrics of the corresponding environmental variable, as explanatory variables. Second, to compare the effect of direct vs. surrogate variables, a variation partitioning in RDA was performed per biological group with the three NMDS axes as response variables and all metrics of direct or surrogate variables as explanatory variables (i.e. metrics of water temperature and hydraulic variables for direct variables and metrics of air temperature and discharge for surrogate variables). The respective explanatory power of each variable (coefficient of determination R^2) for each biological group was then assessed [51]. For the hydraulic variables, for which we computed a total of nine metrics (for each of the three variables, i.e. current velocity, water depth and reach width), all combinations of four out of nine metrics were tested, to keep the same number of explanatory variables and the same degree of freedom between the RDA. To minimize the influence of spatial autocorrelation inherent to these datasets on the statistical assessments, RDA was combined with a Monte-Carlo procedure. We first selected all reaches contained in 70% of the sub-basins (i.e. among the 368 sub-basins considered in the EROS model) as calibration dataset to perform a RDA. This entire procedure was repeated 999 times. To assess both the model explanatory performance and the significance of the difference in average R^2 values between groups of environmental variables, the overlap between confidence intervals around the R^2 was evaluated. Confidence intervals was calculated as 2×standard deviation. The predictive performance of the model was then also assessed on the remaining 30% of sub-basins (see S1 Appendix for details).

Finally, we tested the sensitivity of the results obtained by RDA to uncertainties in environmental variables derived from physical-based models (i.e. water and air temperatures, discharge and hydraulic variables). To do so, we replicated the cross-validation procedure, except that each value of water temperature, air temperature, discharge and hydraulic variables was modified by adding a randomly sampled error scaled to the uncertainties provided for each environmental model (see S1 Appendix for details).

Data formatting and statistical analyses were performed using the R software [52]. The vegan R package [53] was used for NMDS and RDA computation and the maptools R package [54] was used for GIS data manipulation.

Results

A total of 2,950,799 fish specimens belonging to 43 different species were sampled over the study period in the Loire catchment. This represented an average of 97 [11; 253] (mean [5th and 95th quantiles]) CPUEs (i.e. specimens per 100 m² sampled) and 10 [1; 26] species per T-NET reach. Fish data included 18 fish families with cyprinids being the most represented (20 species). A total of 8,334,023 macroinvertebrates specimens belonging to 186 different genera were sampled over the study period. This represented an average of 3,851 [541; 10,123] CPUEs and 43 [16; 71] genera per T-NET reach. The most represented macroinvertebrate orders were: Trichoptera (55 genera), Ephemeroptera (27 genera), Coleoptera (27 genera), Gastropod (17 genera), Odonata (17 genera) and Plecoptera (14 genera). The range of environmental conditions captured by the T-NET reaches sampled was representative of the overall variety of environmental conditions found in the Loire catchment (Table 1).

The NMDS ordination derived from the fish dataset had a stress value of 0.11, indicating a good representation (Fig 2). Fish communities with positive values along the first axis had a low species richness and were dominated by brown trout, while those with negative values had a high species richness and were composed mainly of cyprinids (S1A Fig). The NMDS ordination derived from the macroinvertebrate dataset had a stress value of 0.16, indicating a good representation (Fig 4). Macroinvertebrate communities showing positive values on the second axis were characterized by assemblages with high genera richness, while assemblages with negative values had mainly low genera richness (S1B Fig). The NMDS ordination derived from the common dataset (i.e. reaches with both fish and macroinvertebrate sampling stations) had a stress value of 0.13, indicating a good representation (Fig 5). Regarding the position of fish species centroids on the first factorial plan of the 'common' NMDS, the pattern was very similar to that found for the ordination based only on fish: the first axis opposed fish communities dominated by brown trout with a low taxa richness to communities composed mainly of cyprinids with a high taxa richness (S1C Fig). The pattern was not as clear for macroinvertebrates, with rich macroinvertebrate communities including abundant and common genera located near the center of the factorial space and more distinct assemblages including rarer genera located relatively far from the center.

Procrustes analyses indicated very weak concordance between each pair of biological ordinations (fish-macroinvertebrates: $m^2 = 0.99$; fish-common: $m^2 = 0.99$; macroinvertebrates-common: $m^2 = 0.99$). The concordance with the environmental ordination was higher for the "common" ordination ($m^2 = 0.61$) than for fish ($m^2 = 0.97$) and macroinvertebrates ($m^2 = 0.95$).

Fish ordination was mainly organized along the upstream-downstream gradient as illustrated by mean water temperature (Fig 2). Based on RDA (Fig 3), the most structuring variables of the fish ordination were, by order of importance, water temperature (average explanatory R^2 across the 999 iterations = 47.1%), hydraulic variables ($R^2 = 43.9$), slope and distance to the estuary ($R^2 = 43.3$), air temperature ($R^2 = 41.0$) and discharge ($R^2 = 28.1$). When considered together, direct variables explained the fish ordination with an $R^2 = 54.7$ and the surrogate variables with an $R^2 = 51.8$.

Considering macroinvertebrate ordination, the structuration by the main environmental gradients was visible (e.g. thermal gradient, Fig 4) but appeared less clearly than for fish. Based on RDA (Fig 3), the most structuring variables of the macroinvertebrate ordination were, by order of importance, which is the same than for fish, water temperature ($R^2 = 31.2$), hydraulic variables ($R^2 = 28.8$), air temperature ($R^2 = 22.2$), slope and distance to the estuary ($R^2 = 22.1$) and discharge ($R^2 = 18.1$). The direct variables taken together explained the macroinvertebrate ordination with an $R^2 = 40.8$ and the surrogate variable with an $R^2 = 37.8$.

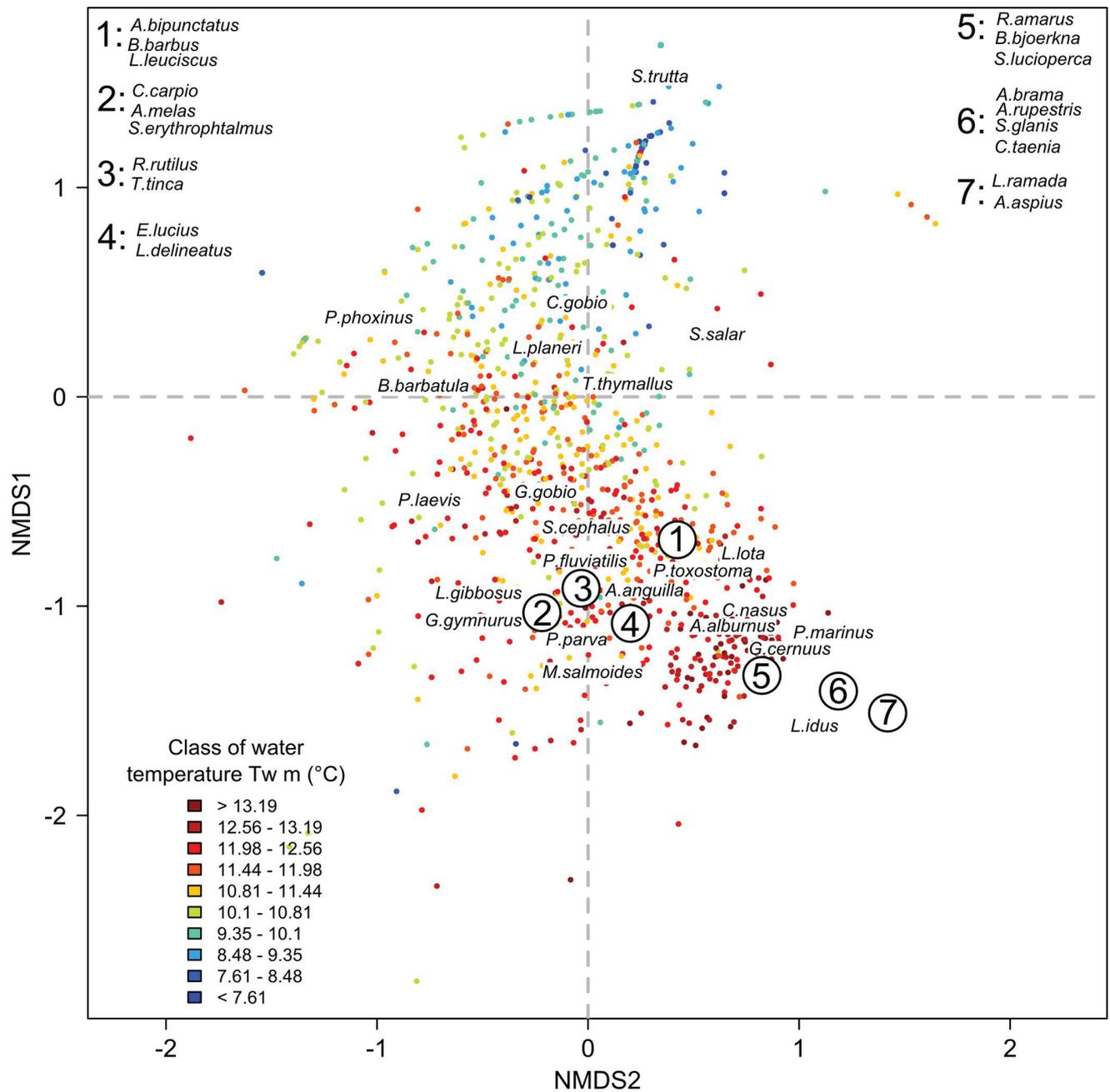


Fig 2. Fish ordination. Position of the T-NET reaches sampled for fish on the first factorial plane. Reaches (points; $N = 1,191$) are colored according to their mean water temperature ($T_w m$; in $^{\circ}C$). Numbers indicate groups of nearby species in the factorial space, whose names have been listed in the upper corners for better readability (See [S1 Table](#) for full species names and [S3 Table](#) for the coordinates of species centroids on the three NMDS axes).

<https://doi.org/10.1371/journal.pone.0274167.g002>

The ordination produced using the common dataset (i.e. with fish and macroinvertebrate considered together, [Fig 5](#)) showed similarities with fish and macroinvertebrate ordinations, both in terms of placement of reaches relative to each other and placement of taxon centroids. Based on RDA ([Fig 3](#)), the most structuring variables of the common ordination, by order of importance, were water temperature ($R^2 = 48.4$), hydraulic variables ($R^2 = 43.7$), slope and

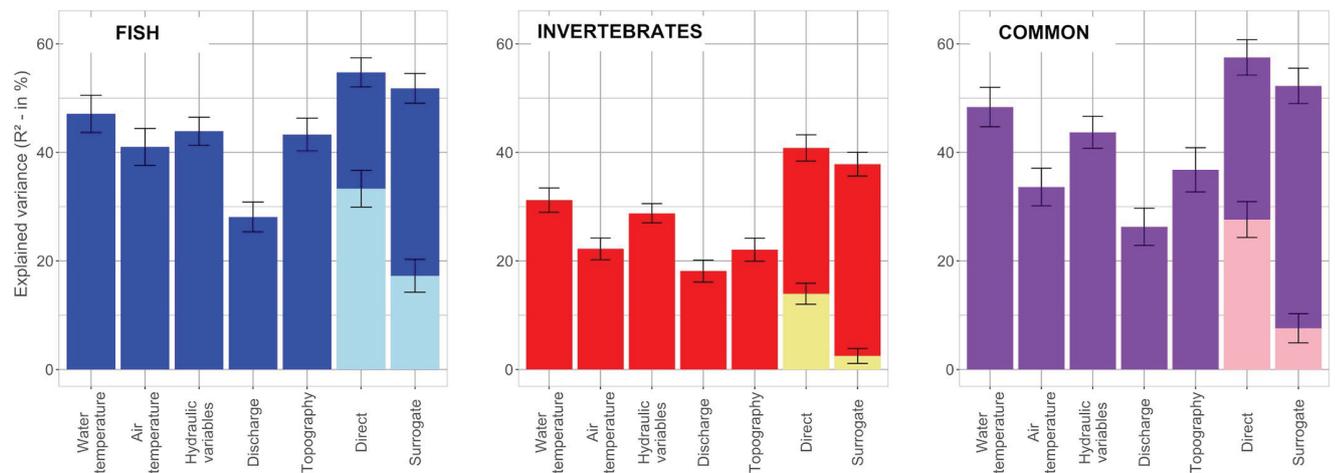


Fig 3. Importance of the different environmental variables in structuring the biological ordinations (i.e. fish, macroinvertebrate and common NMDS): Average coefficients of determination (R^2 ; %) and associated 95% confidence interval derived from RDA (combined with variation partitioning for direct and surrogate models) and obtained through 999 random sub-samples of reaches contained in 70% of sub-basins. Shared fractions of explained variance are shown in lighter color within the bars corresponding to direct and surrogate models.

<https://doi.org/10.1371/journal.pone.0274167.g003>

distance to the estuary ($R^2 = 36.8$), air temperature ($R^2 = 33.6$) and discharge ($R^2 = 26.3$). The direct variables taken together explained the common ordination with an $R^2 = 57.5$ and the surrogates variables with an $R^2 = 52.3$.

Regarding the explanatory power of RDA, the difference in R^2 between each pair of direct and surrogate variables (i.e. water temperature vs. air temperature, hydraulic variables vs. discharge, and direct vs. surrogate combined) were all statistically significant, as shown by the non-overlapping confidence intervals (only with means for the direct and surrogate models), with the direct variables explaining all three biological ordinations better than the surrogate variables. Regarding the predictive power of RDA, most R^2 were also higher for the direct than for the surrogate variables, although the differences were mostly non-significant due to wider confidence intervals (S1 Appendix). Finally, these results were found to be robust to uncertainties in the environmental variables derived from physical-based models (i.e. water and air temperatures, discharge and hydraulic variables), according to the sensitivity analysis (S1 Appendix).

Discussion

Main environmental determinants of the organization of aquatic communities in the Loire catchment

The NMDS performed on the three biological datasets (i.e. fish, macroinvertebrates and common) highlighted the role of the longitudinal gradient in structuring aquatic communities at the extent of a large catchment. The NMDS based on fish densities closely transcribed this longitudinal gradient, the relationships with the environmental variables showing the opposition between thermal and hydraulic-hydrological variables (increasing from upstream to downstream) and topographic variables (decreasing). Fish communities varied along this axis, from salmonid-dominated communities in the upstream reaches to cyprinid-dominated communities in the downstream areas, with a concomitant increase in species richness and total density. The resulting organization of fish communities is consistent with that generally described in the literature [15, 20]. The relationship between the upstream-downstream environmental

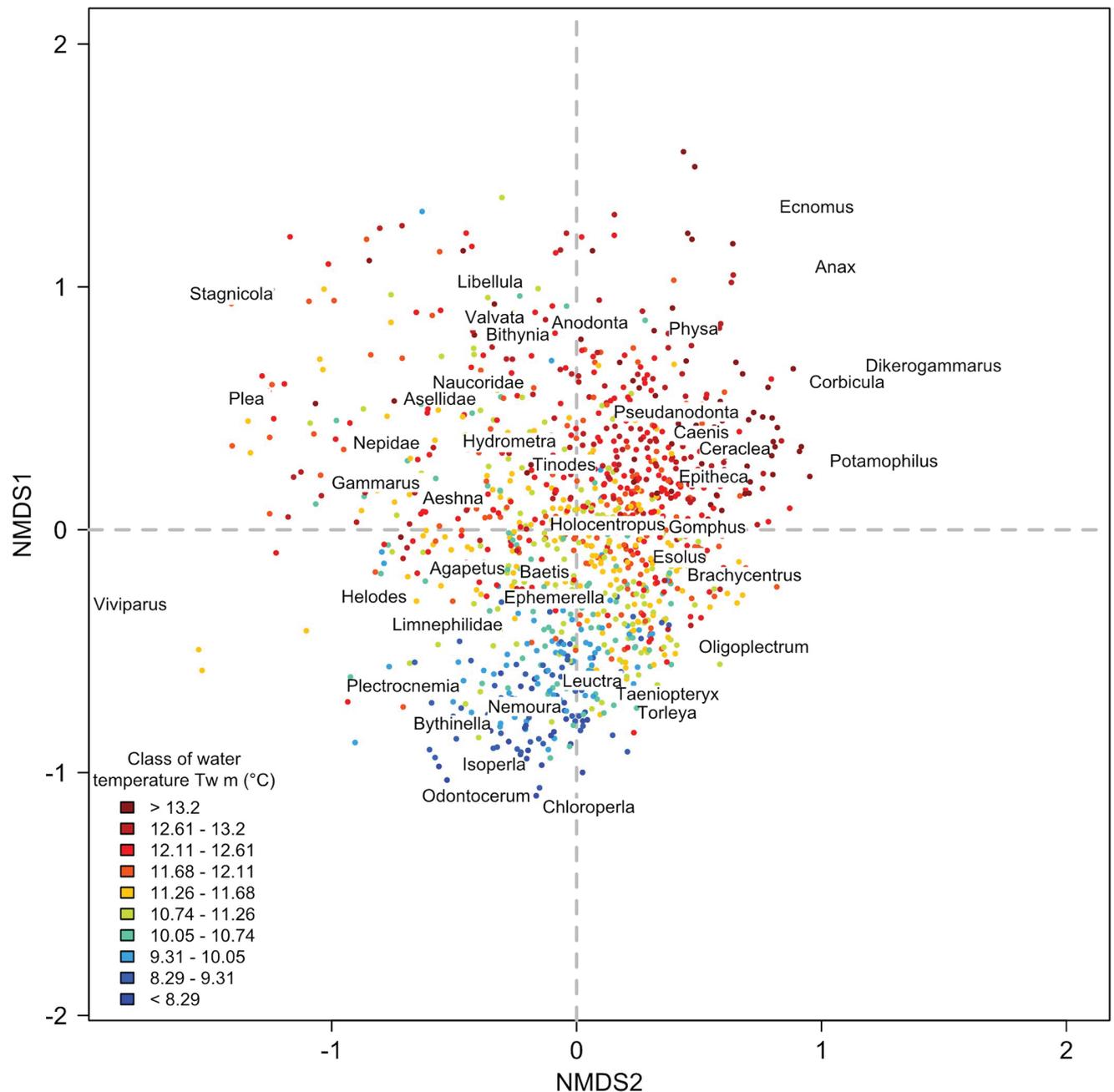


Fig 4. Macroinvertebrate ordination. Position of the T-NET reaches sampled for macroinvertebrates on the first factorial plane. Reaches (points; $N = 1,195$) are colored according to their mean water temperature ($T_w m$; in $^{\circ}C$). For better readability, only centroids of the most represented macroinvertebrate genera are displayed (see [S2 Table](#) for correspondence between code and taxon name and [S4 Table](#) for the coordinates of all the genera centroids on the three NMDS axes).

<https://doi.org/10.1371/journal.pone.0274167.g004>

gradient and macroinvertebrate densities was less clear but still appeared on the second NMDS axis. All R^2 between the macroinvertebrate ordination and the environmental variables were slightly lower than those found for fish, which indicates that the higher variability of the macroinvertebrate community was slightly less explained by our set of environmental variables. This suggests that macroinvertebrate communities are not as strongly structured by

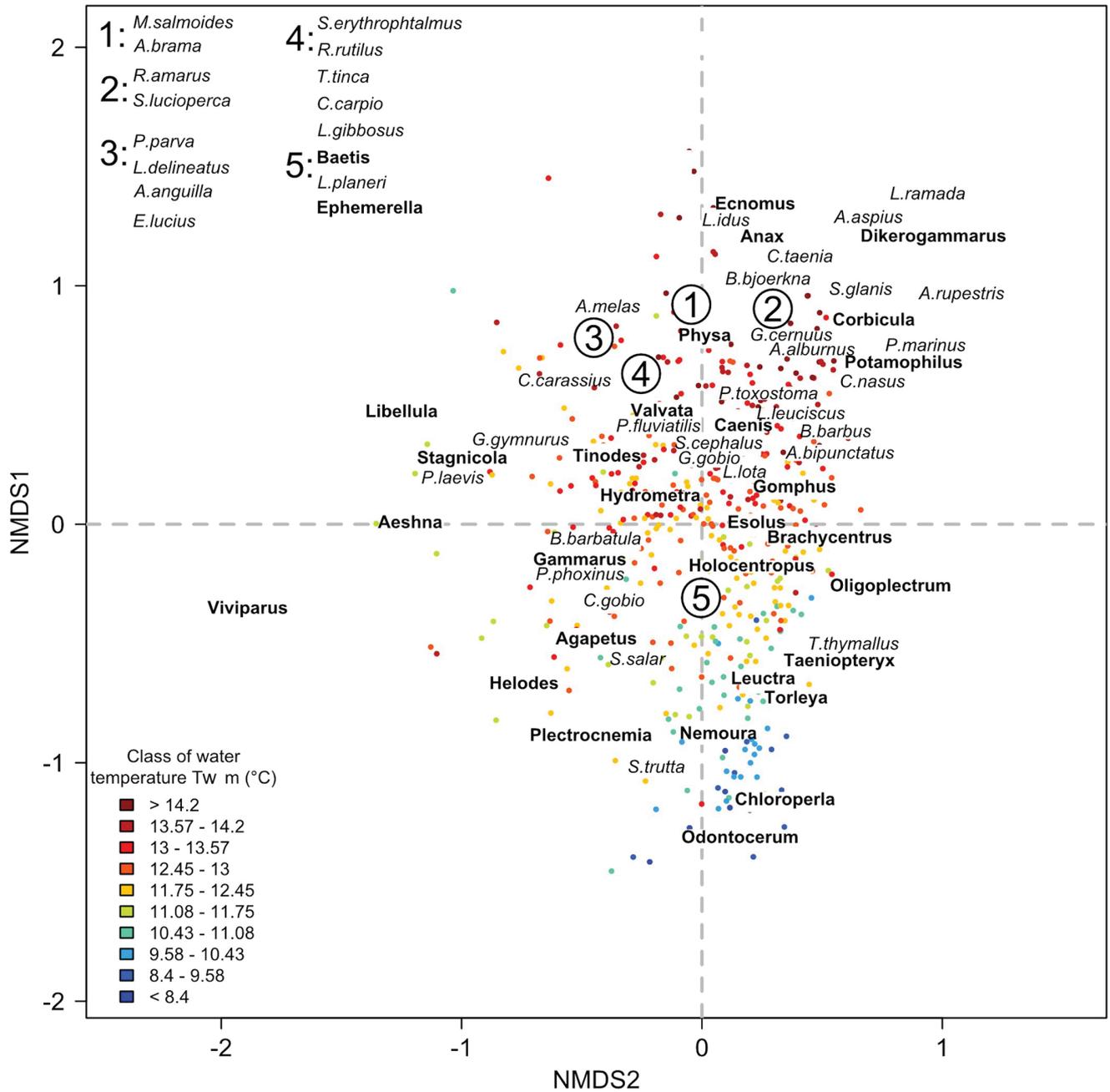


Fig 5. Common ordination. Position of the T-NET reaches sampled for both fish and macroinvertebrates on the first factorial plane. Reaches (points; $N = 532$) were colored according to their mean water temperature ($T_w m$; in $^{\circ}C$). Numbers indicate groups of nearby fish species (in italics) or macroinvertebrate genera (in bold) in the factorial space, whose names have been listed in the upper left corner for better readability. See S1 and S2 Tables for full species and genera names. See S5 Table for the coordinates of all species and genera centroids on the three NMDS axes.

<https://doi.org/10.1371/journal.pone.0274167.g005>

direct variables and the longitudinal gradient at this spatial extent as fish communities are. Key differences in the ecology of fish and macroinvertebrates may explain this result. For example, the greater ability of fish have access to more diverse habitats and can therefore occupy habitats more closely aligned with their ecological requirements [26, 55] Overall, for the three

biological NMDS, we did not observe clustered groups of reaches or any other patchy patterns, although the sampling stations covered very distinctive geographical entities (e.g. Massif Central, Loire valley). This supports the idea that succession between communities is gradual and continuous, with transitions where species with different ecological preferences co-exist, rather than being bounded (as formulated in the River Continuum Concept) [18].

The community organization of the aquatic groups studied was essentially explained by the same set of environmental variables. The most influential determinants were water temperature (especially mean and quantile 90), hydraulic variables (i.e. water depth, reach width and current velocity) and slope associated to the distance to estuary. This finding is consistent with the literature on the topic. First, water temperature, especially warm temperature, limits the spatial distribution of aquatic species according to their thermal tolerance and oxygen requirements [28, 56, 57]. Nonetheless, while being frequently recognized in the literature, few studies have shown the key role of water temperature in structuring aquatic communities at large scale using water rather than air temperature data. Second, hydraulic conditions represent an important local environmental filter that constrains species depending on their hydraulic habitat preferences, also influencing other habitat characteristics (e.g. substrate size, type of vegetation) [58, 59].

Direct vs. surrogate environmental drivers of the organization of aquatic communities

We found that direct environmental variables (i.e. water temperature and hydraulic descriptors) explained significantly better the organization of aquatic communities at the scale of the Loire catchment than surrogate ones (i.e. air temperature and discharge). This corroborates our first hypothesis assuming that direct variables better reflect the actual conditions experienced by aquatic organisms than surrogate ones, therefore better explaining their community organization. We found significant differences between water and air temperature in their ability to explain the organization of fish and macroinvertebrate communities. Studies generally agreed that water temperature is highly correlated with air temperature, especially when the spatial resolution of the study is large [28]. As water temperature measurements are rarely available at large extent, air temperature data are therefore often used as a proxy in large extent ecological studies (e.g. [5, 60, 61]). However, air temperature does not capture all the variability of water temperature. This is highlighted by the ranges of values taken by the two kind of metrics at the scale of the whole T-NET network (see Table 1), with for instance mean air temperature ranging from 6.1 to 12.5°C whereas mean water temperature ranges from 5.4 to 16.2°C (the differences being even higher for the values of lower 10% quantile and upper 90% quantile). Consequently, air temperature does not substitute well for water temperature under certain conditions, such as in the most upstream reaches or near groundwater inputs [28, 34]. For all three biological datasets, the relationships between the NMDS ordinations and the discharge were weaker than with the corresponding hydraulic variables. This supports the idea that the resolution at which discharge metrics have been computed was not the most appropriate to correctly transcribe the relationships between organisms and discharge, unlike hydraulic metrics, which describe habitats at a finer grain. Indeed, hydraulic metrics provide additional dimensions relatively to discharge, because they better reflect local geomorphological variations that influence habitats [38, 62, 63]. Finally, when the direct environmental variables were considered together, they still better explained the organization of fish and macroinvertebrate communities than the full set of surrogate environmental variables. Moreover, the differences between direct and surrogate variables were even greater for the common dataset. These two results reinforce the hypothesis that the combination of several biological groups associated to

the use of direct environmental variables is most suitable for accurately assessing the environmental-organism relationships in lotic hydrosystems. As far as possible, we recommend using several biological groups and direct variables to study spatial patterns of aquatic communities. Nonetheless, surrogate variables remain valuable when direct variables are not available, especially at large scale where the differences in explaining the organization of aquatic communities can be slight as found in the present study. In particular, direct variables are likely to perform better at finer spatial extents and may be more interesting for environmental management as they are more directly linked to actual levers of action. These differences were further reduced with respect to the predictive power of both variables. Therefore, the choice of using direct or surrogate environmental variables may also depend on whether the objective is to explain and understand the organization of aquatic communities and its main drivers, or rather to predict spatial patterns under different scenarios (e.g. climate change).

The environmental variables studied here only partially explained the organization of fish and macroinvertebrate communities. The spatial resolution of the study was the T-NET reach (~1.7 km on average; $N = 52,278$) and we believe that such accuracy, density and coverage at the extent of a large hydrographic network (cumulative length of the river system = 88,058 km) made it highly suitable for precisely investigating the organization of aquatic communities in the Loire basin. Setting the study at a broad spatial extent allows capturing organism responses to major environmental gradients arising from climate, geology or topography [64]. Nevertheless, some processes occurring at a markedly different spatial scale may be blurred at such large spatial extent [65, 66]. For instance, large spatial extent may include distinct biogeographic units, which are known to be driven by other factors such as historical events (e.g. major glaciations) and dispersal capacity of organisms [67–69]. While the catchment scale is now widely recognized as appropriate for examining the organization of fish communities and for management and conservation measures (e.g. Huet [12]; Miranda et al. [70]; Oberdorff et al. [69]), the appropriate extent at which macroinvertebrate communities should be studied remains under debate [71]. For example, Hawkins and Norris [72] and Johnson et al. [73] agreed that the organization of macroinvertebrate communities is better explained by local habitat factors (e.g. substrate, aquatic vegetation) than coarser ones (e.g. land-use in the upstream drainage area, position of the stream in the hydrographic network) whereas Townsend et al. [74] found the opposite. Overall, it seems that the organization of aquatic communities is ruled by the interplay of different spatial scales, with broad-scale processes influencing those occurring at finer scales [17, 58, 73, 75, 76]. One way to capture these scale interactions would be to build hierarchical models able to account for several scales together (e.g. Ville-neuve et al. [17]). Here, the NMDS based on macroinvertebrate densities suggested the role of processes occurring at two levels: catchment scale, governed mainly by thermal and hydrologic regimes, and reach scale at which hydraulic variables show important variations. However, explanatory variables that may be in interaction with variables already considered may have been overlooked, and their inclusion could improve our understanding of the organization of aquatic communities. Furthermore, limitations in the modelling of environmental variables may also influence their contribution in explaining the organization of aquatic communities. The sensitivity analysis performed (S1 Appendix) showed that the results were robust to the uncertainties arising from the physical-based environmental models. Nevertheless, we acknowledge that the T-NET and EROS models, from which the water temperature and discharge data used in this study are derived, do not account for all human-induced alterations of streams, notably physical barriers (e.g. dams, weirs) that are known to alter the hydrological and thermal regimes of rivers [77, 78]. Although the Loire basin is acknowledged as one of the last wild river in Europe because of a limited impact of human activities on its hydrological regime, further developments to integrate anthropogenic uses would be interesting [33].

Concordance between the organization of fish and macroinvertebrate communities and the value of a cross-taxa approach

The concordance between the three NMDS ordinations (*i.e.* fish, macroinvertebrates and “common”) was very weak. This result highlighted that fish and macroinvertebrate communities do not seem to share very similar spatial patterns at the extent of the Loire catchment. Thus, while being driven by the same set of environmental variables, both communities seem to respond differently to environmental gradients. This reinforces the assumption that macroinvertebrate and fish communities are sensitive to environmental gradients but at a different spatial extent. The weak concordance between their community organization confirms the relevance of studying fish and macroinvertebrates together to identify generalizable distribution patterns of aquatic communities instead of considering only one or the other. Most studies that used a similar approach (*i.e.* multivariate ordination and concordance assessment) have also found weak concordances between biological groups at spatial extents similar to the Loire catchment (*e.g.* fish and macroinvertebrates in Kilgour and Barton [79]; many biological groups in the review from Heino [24]; diatoms, macroinvertebrates, macrophytes and fish in Johnson and Hering [80]). In addition, some studies have compared the concordance between taxonomic groups at different spatial scales. Dolph et al. [23] and Paavola et al. [22] found greater concordance in the organization of fish and macroinvertebrate communities at broader spatial extent (*i.e.* similar to the Loire catchment) whereas Bae et al. [81] found the opposite. The first two studies showed that fish and macroinvertebrate communities were concordant at broad geographical scale, over study areas of similar size to that of the Loire catchment (*i.e.* 101 stations within two ecoregions in Finland in Paavola et al. [22] and 670 stations in Minnesota, USA in Dolph et al. [23]). However, there are noteworthy differences between these studies and the present one. Paavola et al. [22] focused on boreal aquatic communities, which live under much colder conditions and have lower richness and abundance. Dolph et al. [23] based their results on concordance between taxonomic groups using species-poor North American communities with a total of 5 fish species and 15 macroinvertebrates taxa. Moreover, in both studies, the broad spatial extent is constituted by grouping small sampling areas from different distant sub-basins, thus not considering a single large catchment that would include the full range of environmental conditions. Furthermore, the weak concordance found between the ordination based on the “common” dataset (*i.e.* combining fish and macroinvertebrate sampling) and the other ordinations indicated that neither fish nor macroinvertebrates were more representative of the organization of the combined aquatic communities. The influence of environmental drivers differed: macroinvertebrates showed weaker correlations with most environmental variables than fish, despite our second hypothesis. The concordance between the environmental ordination and the “common” ordination was higher than with fish and macroinvertebrate ordinations, which is certainly due to a better embedding of the environmental gradient through the biological gradient captured by the NMDS combining fish and macroinvertebrate data (*i.e.* richer and more complex communities). This observation was also confirmed by the slightly higher R^2 found between most explaining environmental variables and the “common” ordination. This supports our third hypothesis that combining different biological groups improves the fit with the main environmental gradients, even though it increases the biological variability and complexity.

Some hypotheses may explain the differences between biological groups found here. First, the macroinvertebrate dataset included 186 genera against 43 species for fish. Consequently, communities were more diverse for macroinvertebrates than for fish (on average 43 genera per reach for macroinvertebrates *vs.* 10 species per reach for fish). The observed CPUE were also more variable for macroinvertebrates (CPUE mean = $3,851 \pm 3,121$ specimens per 100 m^2

per T-NET reach) than for fish (CPUE mean = 97 ± 140 specimens per 100 m^2 per T-NET reach). This could lead to a higher variability to be explained by the same set of environmental variables for macroinvertebrates than for fish. Furthermore, the extent of the spatial distribution of macroinvertebrate genera was on overall more variable than for fish species, with ubiquitous genera occurring along the whole longitudinal gradient (e.g. the genus *Baetis* was present in more than 95% of the reaches sampled; [S2 Table](#)) and other genera showing very particular ecological preferences and restricted occurrences within the Loire catchment (e.g. the genus *Leptophlebia* was present in less than 2% of the reaches sampled, mainly in upstream reaches; [S2 Table](#)). In contrast, fish ordination has highlighted typical 'indicator' species like *S. trutta* or *T. thymallus* whose ecological requirements are well known and which are found in specific parts of the Loire catchment. Considering the high number of genera in the macroinvertebrate dataset and the large extent of the area studied, it seems plausible that the Loire catchment includes distinct sub-basins, each having distinct macroinvertebrate biogeographic patterns. The organization of macroinvertebrate communities would thus be less driven by the upstream-downstream environmental gradient. In contrast, the longitudinal organization of fish assemblages seems to apply to the entire catchment. Second, the identification of macroinvertebrates was limited to the genus level, which may group species with different ecological preferences [72, 82]. Previous studies had shown that family-level data can be sufficient to examine the main spatial patterns in macroinvertebrate communities, although more precise patterns can only be detected with a finer level of taxonomic identification [83]. In addition, the sensitivity of non-metric ordination methods to the identification level of macroinvertebrates is known to be reduced under the family level [55, 79]. According to these studies, the use of the genus level seems to be a good compromise between practical limitations (i.e. field surveys, time allocated to taxa identification) and appropriate identification level to examine ecological patterns. Third, fish typically use larger areas of stream to complete their life cycle than macroinvertebrates: above the kilometer for most fish species vs. a few meters for most macroinvertebrate genera [26, 79]. Both groups can therefore be sensitive to the same environmental variables but not at the same spatial grain [84]. Macroinvertebrates are more sensitive to local determinants (*sensu* habitat variables) and show stronger responses to local perturbations than fish, especially due to their lower dispersal and avoidance abilities in their larval form [58, 85, 86]. This implies that the identification of patterns in the organization of macroinvertebrate communities at large extent might be blurred by local processes and disturbances acting at fine grain. Fish are probably less affected by these fine-grained factors as they are able to move to suitable habitats when habitat alteration is local and temporary. Stochastic processes of colonization dynamic (e.g. organisms' drifts, flying macroinvertebrates) may also play a more important role for macroinvertebrates than for fish [26].

Conclusion and perspectives

Using high-resolution environmental datasets in a large catchment, we demonstrated the ability of direct environmental descriptors to better explain the organization of fish and macroinvertebrate communities. We also addressed the concordance between the organization of fish and macroinvertebrate communities and identified common environmental determinants. We found that fish and macroinvertebrates exhibited weak concordance in their community organization due to their different responses to ecological gradients, which precludes using one biological group as a proxy for the other. Future research should continue to examine together data from different taxonomic groups, paying particular attention to the role of spatial scale (i.e. extent and resolution), and pursue the efforts to make accurate, fine-grained environmental variables available through monitoring and modelling over large spatial extents.

Including anthropogenic influences in environmental models (e.g. T-NET, EROS) is an interesting avenue to grasp more environmental variability and to gain insights into the spatial patterns of aquatic communities, in line with recent meta-community concepts. Thus, improving our understanding of how communities are organized in a changing environment (i.e. climate change) is of primary importance for field managers and biodiversity stakeholders.

Supporting information

S1 Fig. Taxonomic richness of the sampled T-NET reaches plotted on the first factorial plan of the fish, macroinvertebrates and common ordinations (A, B and C, respectively). Reaches are colored according to their taxonomic richness from low (blue) to high (red) number of species or genera.
(PDF)

S1 Table. List of 50 fish species present in the initial dataset. For each species, the number of occurrences (with the frequency of occurrence under brackets) over the 1191 T-NET reaches sampled is provided, as well as the average CPUE per reach (in number of specimens per 100 m² sampled). The species marked with an asterisk correspond to the rarest species (frequency of occurrence < 0.5%) that were excluded prior to analyses.
(PDF)

S2 Table. List of 230 macroinvertebrates genera present in the initial dataset. For each taxa, the number of occurrences (with the frequency of occurrence under brackets) over the 1195 T-NET reaches sampled is provided, as well as the average CPUE per reach (in number of specimens per 100 m² sampled). The taxa marked with an asterisk correspond to the rarest taxa (frequency of occurrence < 0.5%) that were excluded prior to analyses.
(PDF)

S3 Table. Position of the centroids of the 43 fish species on the three factorial axes of the NMDS performed on the fish dataset.
(PDF)

S4 Table. Position of the centroids of the 186 macroinvertebrates taxa on the three factorial axes of the NMDS performed on the macroinvertebrate dataset.
(PDF)

S5 Table. Position of the centroids of the 43 fish species (bold) and 170 macroinvertebrates taxa (non-bold) on the three factorial axes of the NMDS performed on the common dataset.
(PDF)

S1 Appendix. Monte Carlo cross-validation procedure and sensitivity analysis of RDA to uncertainties in environmental variables derived from physical-based models.
(PDF)

Acknowledgments

This research is part of team HYNES built within the EDF-INRAE Framework Agreement on Water Engineering and Aquatic Systems. We thank Dominique Thiery (BRGM), who provided the simulations of discharge with EROS model and Yann Julien (University of Tours) for his help with the calculations of water temperature and discharge descriptors. The SAFRAN database was provided by the French national meteorological service (Météo-France). We thank all the people who have been involved in the monitoring. Fieldwork was

conducted with adequate administrative permits for electrofishing and in accordance with French laws and ethical rules.

Author Contributions

Conceptualization: Laëtitia Buisson, Florentina Moatar, Anthony Maire.

Data curation: Coline Picard, Mathieu Floury, Hanieh Seyedhashemi, Maxime Morel, Hervé Pella.

Formal analysis: Coline Picard, Mathieu Floury.

Methodology: Coline Picard, Mathieu Floury, Nicolas Lamouroux, Laëtitia Buisson, Florentina Moatar, Anthony Maire.

Resources: Hanieh Seyedhashemi, Maxime Morel, Hervé Pella, Nicolas Lamouroux.

Supervision: Mathieu Floury, Florentina Moatar, Anthony Maire.

Writing – original draft: Coline Picard.

Writing – review & editing: Mathieu Floury, Hanieh Seyedhashemi, Maxime Morel, Nicolas Lamouroux, Laëtitia Buisson, Florentina Moatar, Anthony Maire.

References

1. Malmqvist B, Rundle S. Threats to the running water ecosystems of the world. *Envir Conserv.* 2002; 29: 134–153. <https://doi.org/10.1017/S0376892902000097>
2. Reid AJ, Carlson AK, Creed IF, Eliason EJ, Gell PA, Johnson PTJ, et al. Emerging threats and persistent conservation challenges for freshwater biodiversity: Emerging threats to freshwater biodiversity. *Biological Reviews.* 2019 [cited 23 Apr 2019]. <https://doi.org/10.1111/brv.12480> PMID: 30467930
3. Dudgeon D. Multiple threats imperil freshwater biodiversity in the Anthropocene. *Current Biology.* 2019; 29: R960–R967. <https://doi.org/10.1016/j.cub.2019.08.002> PMID: 31593677
4. Closs GP, Krkosek M, Olden JD, editors. *Conservation of Freshwater Fishes.* Cambridge: Cambridge University Press; 2016. <https://doi.org/10.1017/CBO9781139627085>
5. Maire A, Laffaille P, Maire J-F, Buisson L. Identification of Priority Areas for the Conservation of Stream Fish Assemblages: Implications for River Management in France: Priority Areas for the Conservation of Fish Assemblages in France. *River Research and Applications.* 2017; 33: 524–537. <https://doi.org/10.1002/rra.3107>
6. Vehanen T, Sutela T, Harjunpää A. The effects of ecoregions and local environmental characteristics on spatial patterns in boreal riverine fish assemblages. *Ecology of Freshwater Fish.* 2020. <https://doi.org/10.1111/eff.12550>
7. Campbell Grant EH, Lowe WH, Fagan WF. Living in the branches: population dynamics and ecological processes in dendritic networks. *Ecol Letters.* 2007; 10: 165–175. <https://doi.org/10.1111/j.1461-0248.2006.01007.x> PMID: 17257104
8. Erős T, Olden JD, Schick RS, Schmera D, Fortin M-J. Characterizing connectivity relationships in freshwaters using patch-based graphs. *Landscape Ecol.* 2012; 27: 303–317. <https://doi.org/10.1007/s10980-011-9659-2>
9. Reynoldson TB, Norris RH, Resh VH, Day KE, Rosenberg DM. The Reference Condition: A Comparison of Multimetric and Multivariate Approaches to Assess Water-Quality Impairment Using Benthic Macroinvertebrates. *Journal of the North American Benthological Society.* 1997; 16: 833–852. <https://doi.org/10.2307/1468175>
10. Soranno PA, Wagner T, Martin SL, McLean C, Novitski LN, Provence CD, et al. Quantifying regional reference conditions for freshwater ecosystem management: A comparison of approaches and future research needs. *Lake and Reservoir Management.* 2011; 27: 138–148. <https://doi.org/10.1080/07438141.2011.573614>
11. Shelford VE. Ecological succession: 1. Stream fishes and the method of physiographic. *The Biological Bulletin.* 1911; 21: 9–35. <https://doi.org/10.2307/1535983>

12. Huet M. Profiles and Biology of Western European Streams as Related to Fish Management. *Transactions of the American Fisheries Society*. 1959; 88: 155–163. [https://doi.org/10.1577/1548-8659\(1959\)88\[155:PABOWE\]2.0.CO;2](https://doi.org/10.1577/1548-8659(1959)88[155:PABOWE]2.0.CO;2)
13. Verneaux J. Cours d'eau de Franche-Comté (massif du Jura): recherches écologiques sur le réseau hydrographique du Doubs—essai de biotypologie. Université de Franche-Comté. 1973.
14. Wasson J-G. Éléments pour une typologie fonctionnelle des eaux courants: 1. Revue critique de quelques approches existantes. *Bull Ecol*. 1989; 20: 109–127. <https://doi.org/10.1007/s00221-012-3100-1> PMID: [22562588](https://pubmed.ncbi.nlm.nih.gov/22562588/)
15. Matthews WJ. *Patterns in Freshwater Fish Ecology*. Boston, MA: Springer US; Imprint: Springer; 1998. Available: <http://public.eblib.com/choice/publicfullrecord.aspx?p=3081363>
16. Jackson DA, Peres-Neto PR, Olden JD. What controls who is where in freshwater fish communities—the roles of biotic, abiotic, and spatial factors. 2001; 58: 14.
17. Villeneuve B, Piffady J, Valette L, Souchon Y, Usseglio-Polatera P. Direct and indirect effects of multiple stressors on stream invertebrates across watershed, reach and site scales: A structural equation modelling better informing on hydromorphological impacts. *Science of The Total Environment*. 2018; 612: 660–671. <https://doi.org/10.1016/j.scitotenv.2017.08.197> PMID: [28866394](https://pubmed.ncbi.nlm.nih.gov/28866394/)
18. Vannote RL, Minshall GW, Cummins KW, Sedell JR, Cushing CE. The River Continuum Concept. *Can J Fish Aquat Sci*. 1980; 37: 130–137. <https://doi.org/10.1139/f80-017>
19. Ward JV. The Four-Dimensional Nature of Lotic Ecosystems. *Journal of the North American Benthological Society*. 1989; 8: 2–8. <https://doi.org/10.2307/1467397>
20. Laffaille P, Brosse S, Ombredane D, Feunteun É. *Richesse et organisation spatiale des communautés de poissons d'eau douce*. Biotope Editions, Mèze. Les poissons d'eau douce de France. Biotope Editions, Mèze. Muséum national d'Histoire naturelle, Paris; 2020. pp. 128–159.
21. Heino J, Soininen J, Alahuhta J, Lappalainen J, Virtanen R. Metacommunity ecology meets biogeography: effects of geographical region, spatial dynamics and environmental filtering on community structure in aquatic organisms. *Oecologia*. 2017; 183: 121–137. <https://doi.org/10.1007/s00442-016-3750-y> PMID: [27714463](https://pubmed.ncbi.nlm.nih.gov/27714463/)
22. Paavola R, Muotka T, Virtanen R, Heino J, Jackson D, Mäki-Petäys A. Spatial Scale Affects Community Concordance Among Fishes, Benthic Macroinvertebrates, And Bryophytes In Streams. *Ecological Applications*. 2006; 16: 368–379. <https://doi.org/10.1890/03-5410> PMID: [16705986](https://pubmed.ncbi.nlm.nih.gov/16705986/)
23. Dolph CL, Huff DD, Chizinski CJ, Vondracek B. Implications of community concordance for assessing stream integrity at three nested spatial scales in Minnesota, U.S.A.: Concordance and scale in stream assessment. *Freshwater Biology*. 2011; 56: 1652–1669. <https://doi.org/10.1111/j.1365-2427.2011.02589.x>
24. Heino J. Are indicator groups and cross-taxon congruence useful for predicting biodiversity in aquatic ecosystems? *Ecological Indicators*. 2010; 10: 112–117. <https://doi.org/10.1016/j.ecolind.2009.04.013>
25. Stewart DR, Underwood ZE, Rahel FJ, Walters AW. The effectiveness of surrogate taxa to conserve freshwater biodiversity: Freshwater Conservation Planning. *Conservation Biology*. 2018; 32: 183–194. <https://doi.org/10.1111/cobi.12967> PMID: [28612963](https://pubmed.ncbi.nlm.nih.gov/28612963/)
26. Backus-Freer J, Pyron M. Concordance among fish and macroinvertebrate assemblages in streams of Indiana, USA. *Hydrobiologia*. 2015; 758: 141–150. <https://doi.org/10.1007/s10750-015-2281-6>
27. Olden JD. Biotic homogenization: a new research agenda for conservation biogeography. *Journal of Biogeography*. 2006; 33: 2027–2039. <https://doi.org/10.1111/j.1365-2699.2006.01572.x>
28. Caissie D. The thermal regime of rivers: a review. *Freshwater Biology*. 2006; 51: 1389–1406. <https://doi.org/10.1111/j.1365-2427.2006.01597.x>
29. Charbonnel A, Laffaille P, Biffi M, Blanc F, Maire A, Némoz M, et al. Can Recent Global Changes Explain the Dramatic Range Contraction of an Endangered Semi-Aquatic Mammal Species in the French Pyrenees? Courchamp F, editor. *PLoS ONE*. 2016; 11: e0159941. <https://doi.org/10.1371/journal.pone.0159941> PMID: [27467269](https://pubmed.ncbi.nlm.nih.gov/27467269/)
30. Arevalo E. An innovative bivariate approach to detect joint temporal trends in environmental conditions: Application to large French rivers and diadromous fish. *Science of The Total Environment*. 2020; 12. <https://doi.org/10.1016/j.scitotenv.2020.141260> PMID: [32805565](https://pubmed.ncbi.nlm.nih.gov/32805565/)
31. Floury M, Pollock LJ, Buisson L, Thuiller W, Chandèsris A, Souchon Y. Combining expert-based and computational approaches to design protected river networks under climate change. Grech A, editor. *Divers Distrib*. 2021; ddi.13411. <https://doi.org/10.1111/ddi.13411>
32. Sauquet E, Gottschalk L, Krasovskaia I. Estimating mean monthly runoff at ungauged locations: an application to France. *Hydrology Research*. 2008; 39: 403–423.
33. Moatar F, Dupont N, editors. *La Loire fluviale et estuarienne: un milieu en évolution*. Versailles: Quae; 2016.

34. Beaufort A, Curie F, Moatar F, Ducharme A, Melin E, Thiery D. T-NET, a dynamic model for simulating daily stream temperature at the regional scale based on a network topology: T-NET model for simulating stream temperature at a regional scale. *Hydrological Processes*. 2016; 30: 2196–2210. <https://doi.org/10.1002/hyp.10787>
35. Thiéry D, Moutzopoulos C. Un modèle hydrologique spatialisé pour la simulation de très grands bassins: le modèle EROS formé de grappes de modèles globaux élémentaires. VIIIèmes journées hydrologiques de l'ORSTOM " Régionalisation en hydrologie, application au développement", Le Barbé et E Servat (Eds). 1995; 285–295.
36. Loicq P, Moatar F, Jullian Y, Dugdale SJ, Hannah DM. Improving representation of riparian vegetation shading in a regional stream temperature model using LiDAR data. *Science of the total environment*. 2018; 624: 480–490. <https://doi.org/10.1016/j.scitotenv.2017.12.129> PMID: 29268220
37. Vidal J-P, Martin E, Franchistéguy L, Baillon M, Soubeyrou J-M. A 50-year high-resolution atmospheric reanalysis over France with the Safran system. *International Journal of Climatology*. 2010; 30: 1627–1644.
38. Morel M, Booker DJ, Gob F, Lamouroux N. Intercontinental predictions of river hydraulic geometry from catchment physical characteristics. *Journal of Hydrology*. 2020; 124292. <https://doi.org/10.1016/j.jhydrol.2019.124292>
39. Leopold LB, Maddock T. *The hydraulic geometry of stream channels and some physiographic implications*. US Government Printing Office; 1953.
40. Pella H, Lejot J, Lamouroux N, Snelder T. Le réseau hydrographique théorique (RHT) français et ses attributs environnementaux [The French theoretical hydrographic network (RHT) and its environmental attributes]. *Géomorphologie: Relief, Processus, Environnement*. 2012; 18: 317–336. <https://doi.org/10.4000/geomorphologie.9933>
41. Institut national de l'information géographique et forestière (IGN). IGN BD ALTI@ 25x25 m. In: Géoportail [Internet]. Available: <https://geoservices.ign.fr>
42. European Union. Directive 2000/60/EC of the European Parliament and of the Council establishing a framework for Community action in the field of water policy. 2000.
43. Office Français pour la Biodiversité (OFB). Naïades. Available: www.naiades.eaufrance.fr
44. Maire A, Thierry E, Viechtbauer W, Daufresne M. Poleward shift in large-river fish communities detected with a novel meta-analysis framework. *Freshwater Biology*. 2019 [cited 23 Apr 2019]. <https://doi.org/10.1111/fwb.13291>
45. Poulet N, Beaulaton L, Dembski S. Time trends in fish populations in metropolitan France: insights from national monitoring data. *Journal of Fish Biology*. 2011; 79: 1436–1452. <https://doi.org/10.1111/j.1095-8649.2011.03084.x> PMID: 22136235
46. Flourey M, Souchon Y, Looy KV. Climatic and trophic processes drive long-term changes in functional diversity of freshwater invertebrate communities. *Ecography*. 2018; 41: 209–218. <https://doi.org/10.1111/ecog.02701>
47. Minchin PR. An evaluation of the relative robustness of techniques for ecological ordination. In: Prentice IC, van der Maarel E, editors. *Theory and models in vegetation science*. Dordrecht: Springer Netherlands; 1987. pp. 89–107. https://doi.org/10.1007/978-94-009-4061-1_9
48. Clarke KR. Non-parametric multivariate analyses of changes in community structure. *Austral Ecol*. 1993; 18: 117–143. <https://doi.org/10.1111/j.1442-9993.1993.tb00438.x>
49. Peres-Neto PR, Jackson DA. How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia*. 2001; 129: 169–178. <https://doi.org/10.1007/s004420100720> PMID: 28547594
50. Peres-Neto PR, Legendre P, Dray S, Borcard D. Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology*. 2006; 87: 2614–2625. [https://doi.org/10.1890/0012-9658\(2006\)87\[2614:vposdm\]2.0.co;2](https://doi.org/10.1890/0012-9658(2006)87[2614:vposdm]2.0.co;2) PMID: 17089669
51. Legendre P, Legendre L. *Numerical ecology*. Third English edition. Amsterdam: Elsevier; 2012.
52. R Development Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2019. Available: <https://www.R-project.org/>
53. Oksanen J, Blanchet G, Friendly M, Kindt R, Legendre P. *vegan: Community Ecology Package*. R package version 2.5–4; 2019. Available: <https://CRAN.R-project.org/package=vegan>
54. Bivand R, Lewin-Koh N. *maptools: Tools for Handling Spatial Objects*. 2019. Available: <https://CRAN.R-project.org/package=maptools>
55. Grenouillet G, Brosse S, Tudesque L, Lek S, Baraillé Y, Loot G. Concordance among stream assemblages and spatial autocorrelation along a fragmented gradient. *Diversity and Distributions*. 2008; 14: 592–603. <https://doi.org/10.1111/j.1472-4642.2007.00443.x>

56. Souchon Y, Tissot L. Synthesis of thermal tolerances of the common freshwater fish species in large Western Europe rivers. *Knowledge and Management of Aquatic Ecosystems*. 2012; 03. <https://doi.org/10.1051/kmae/2012008>
57. Mee JA, Robins GL, Post JR. Patterns of fish species distributions replicated across three parallel rivers suggest biotic zonation in response to a longitudinal temperature gradient. *Ecology of Freshwater Fish*. 2018; 27: 44–61. <https://doi.org/10.1111/eff.12322>
58. Lamouroux N, Dolédec S, Gayraud S. Biological traits of stream macroinvertebrate communities: effects of microhabitat, reach, and basin filters. *Journal of the North American Benthological Society*. 2004; 23: 449–466. [https://doi.org/10.1899/0887-3593\(2004\)023<0449:BTOSMC>2.0.CO;2](https://doi.org/10.1899/0887-3593(2004)023<0449:BTOSMC>2.0.CO;2)
59. Rolls RJ, Heino J, Ryder DS, Chessman BC, Growns IO, Thompson RM, et al. Scaling biodiversity responses to hydrological regimes: Hydrology, freshwater biodiversity and scale. *Biol Rev*. 2018; 93: 971–995. <https://doi.org/10.1111/brv.12381> PMID: 29115026
60. Buisson L, Blanc L, Grenouillet G. Modelling stream fish species distribution in a river network: the relative effects of temperature versus physical factors. *Ecology of Freshwater Fish*. 2008; 17: 244–257. <https://doi.org/10.1111/j.1600-0633.2007.00276.x>
61. Comte L, Murienne J, Grenouillet G. Species traits and phylogenetic conservatism of climate-induced range shifts in stream fishes. *Nat Commun*. 2014; 5: 5053. <https://doi.org/10.1038/ncomms6053> PMID: 25248802
62. Booker DJ. Predicting wetted width in any river at any discharge. *Earth Surf Process Landforms*. 2010; n/a-n/a. <https://doi.org/10.1002/esp.1981>
63. Stutzner B, Higler B. Stream hydraulics as a major determinant of benthic invertebrate zonation patterns. *Freshwater Biology*. 1986; 16: 127–139. <https://doi.org/10.1111/j.1365-2427.1986.tb00954.x>
64. Infante DM, David Allan J, Linke S, Norris RH. Relationship of fish and macroinvertebrate assemblages to environmental factors: implications for community concordance. *Hydrobiologia*. 2009; 623: 87–103. <https://doi.org/10.1007/s10750-008-9650-3>
65. Wiens JA. Spatial Scaling in Ecology. *Functional Ecology*. 1989; 3: 385. <https://doi.org/10.2307/2389612>
66. Wiens JA. Riverine landscapes: taking landscape ecology into the water. *Freshwater Biol*. 2002; 47: 501–515. <https://doi.org/10.1046/j.1365-2427.2002.00887.x>
67. Dias MS, Cornu J-F, Oberdorff T, Lasso CA, Tedesco PA. Natural fragmentation in river networks as a driver of speciation for freshwater fishes. *Ecography*. 2013; 36: 683–689. <https://doi.org/10.1111/j.1600-0587.2012.07724.x>
68. Dias MS, Oberdorff T, Huguency B, Leprieux F, Jézéquel C, Cornu J-F, et al. Global imprint of historical connectivity on freshwater fish biodiversity. Anderson M, editor. *Ecol Lett*. 2014; 17: 1130–1140. <https://doi.org/10.1111/ele.12319> PMID: 25039890
69. Oberdorff T, Dias MS, Jézéquel C, Albert JS, Arantes CC, Bigorne R, et al. Unexpected fish diversity gradients in the Amazon basin. *Sci Adv*. 2019; 5: eaav8681. <https://doi.org/10.1126/sciadv.aav8681> PMID: 31535018
70. Miranda LE, Killgore KJ, Slack WT. Spatial organization of fish diversity in a species-rich basin. *River Research and Applications*. 2018; 1–9. <https://doi.org/10.1002/rra.3392>
71. Dovciak AL, Perry JA. In Search of Effective Scales for Stream Management: Does Agroecoregion, Watershed, or Their Intersection Best Explain the Variance in Stream Macroinvertebrate Communities? *Environmental Management*. 2002; 30: 365–377. <https://doi.org/10.1007/s00267-002-2529-6> PMID: 12148071
72. Hawkins CP, Norris RH. Performance of different landscape classifications for aquatic bioassessments: introduction to the series. *Journal of the North American Benthological Society*. 2000; 19: 367–369. <https://doi.org/10.1899/0887-3593-19.3.367>
73. Johnson RK, Goedkoop W, Sandin L. Spatial scale and ecological relationships between the macroinvertebrate communities of stony habitats of streams and lakes. *Freshwater Biol*. 2004; 49: 1179–1194. <https://doi.org/10.1111/j.1365-2427.2004.01262.x>
74. Townsend CR, Doledec S, Norris R, Peacock K, Arbuckle C. The influence of scale and geography on relationships between stream community composition and landscape variables: description and prediction. *Freshwater Biol*. 2003; 48: 768–785. <https://doi.org/10.1046/j.1365-2427.2003.01043.x>
75. Poff NL. Landscape Filters and Species Traits: Towards Mechanistic Understanding and Prediction in Stream Ecology. *Journal of the North American Benthological Society*. 1997; 16: 391–409. <https://doi.org/10.2307/1468026>
76. Brosse S, Arbuckle CJ, Townsend CR. Habitat scale and biodiversity: influence of catchment, stream reach and bedform scales on local invertebrate diversity. *Biodiversity & Conservation*. 2003; 12: 2057–2075. <https://doi.org/10.1023/A:1024107915183>

77. Seyedhashemi H, Vidal J-P, Diamond JS, Thiéry D, Monteil C, Hendrickx F, et al. Regional, multi-decadal analysis on the Loire River basin reveals that stream temperature increases faster than air temperature. *Hydrol Earth Syst Sci.* 2022; 26: 2583–2603. <https://doi.org/10.5194/hess-26-2583-2022>
78. Zaidel PA, Roy AH, Houle KM, Lambert B, Letcher BH, Nislow KH, et al. Impacts of small dams on stream temperature. *Ecological Indicators.* 2021; 120: 106878. <https://doi.org/10.1016/j.ecolind.2020.106878>
79. Kilgour BW, Barton DR. Associations between stream fish and benthos across environmental gradients in southern Ontario, Canada. *Freshwater Biol.* 1999; 41: 553–566. <https://doi.org/10.1046/j.1365-2427.1999.00402.x>
80. Johnson RK, Hering D. Spatial congruency of benthic diatom, invertebrate, macrophyte, and fish assemblages in European streams. *Ecological Applications.* 2010; 20: 978–992. <https://doi.org/10.1890/08-1153.1> PMID: 20597284
81. Bae M-J, Li F, Kwon Y-S, Chung N, Choi H, Hwang S-J, et al. Concordance of diatom, macroinvertebrate and fish assemblages in streams at nested spatial scales: Implications for ecological integrity. *Ecological Indicators.* 2014; 47: 89–101. <https://doi.org/10.1016/j.ecolind.2014.07.030>
82. Tachet H, Richoux P, Bournaud M, Usseglio-Polatera P. *Invertébrés d'eau douce: systématique, biologie, écologie.* CNRS éditions Paris; 2010.
83. Heino J, Soininen J. Are higher taxa adequate surrogates for species-level assemblage patterns and species richness in stream organisms? *Biological Conservation.* 2007; 137: 78–89. <https://doi.org/10.1016/j.biocon.2007.01.017>
84. Fausch KD, Torgersen CE, Baxter CV, Li HW. Landscapes to Riverscapes: Bridging the Gap between Research and Conservation of Stream Fishes. *BioScience.* 2002; 52: 483. [https://doi.org/10.1641/0006-3568\(2002\)052\[0483:LTRBTG\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2002)052[0483:LTRBTG]2.0.CO;2)
85. Allen AP, Whittier TR, Larsen DP, Kaufmann PR, O'Connor RJ, Hughes RM, et al. Concordance of taxonomic composition patterns across multiple lake assemblages: effects of scale, body size, and land use. *Can J Fish Aquat Sci.* 1999; 56: 2029–2040. <https://doi.org/10.1139/f99-139>
86. Heino J, Paavola R, Virtanen R, Muotka T. Searching for biodiversity indicators in running waters: do bryophytes, macroinvertebrates, and fish show congruent diversity patterns? *Biodivers Conserv.* 2005; 14: 415–428. <https://doi.org/10.1007/s10531-004-6064-z>

Supporting information for the article:

Direct habitats descriptors improve the understanding of the spatial organization of fish and macroinvertebrate communities across a large catchment

Coline PICARD ^{1,2,*}, Mathieu FLOURY ³, Hanieh SEYEDHASEMI ^{2,4}, Maxime MOREL ²,
Hervé PELLA ², Nicolas LAMOUREUX ², Laëtitia BUISSON ⁵, Florentina MOATAR ² and
Anthony MAIRE ¹

¹ EDF R&D LNHE - Laboratoire National d'Hydraulique et Environnement, Chatou, France

² INRAE, UR RiverLy, centre de Lyon-Villeurbanne, Villeurbanne, France

³ Univ Lyon, Université Claude Bernard Lyon 1, CNRS, ENTPE, UMR 5023 LEHNA, F-69622, Villeurbanne, France

⁴ EA 6293 GÉHCO, Université de Tours, Tours, France

⁵ Laboratoire écologie fonctionnelle et environnement, Université de Toulouse, CNRS, Toulouse INP, Université Toulouse 3 – Paul Sabatier (UPS), Toulouse, France

* Corresponding author - E-mail address: coline.picard3@gmail.com (C. Picard).

S1 Table. List of 50 fish species present in the initial dataset. For each species, the number of occurrences (with the frequency of occurrence under brackets) over the 1191 T-NET reaches sampled is provided, as well as the average CPUE per reach (in number of specimens per 100 m² sampled). The species marked with an asterisk correspond to the rarest species (frequency of occurrence < 0.5%) that were excluded prior to analyses.

Scientific name	Family	Occurrences (frequency)	Average CPUE per reach (ind/100 m ²)
<i>Abramis brama</i>	Cyprinidae	228 (19.1%)	0.12
<i>Alburnoides bipunctatus</i>	Cyprinidae	363 (30.5%)	3.54
<i>Alburnus alburnus</i>	Cyprinidae	348 (29.2%)	2.37
<i>Alosa alosa</i> *	Clupeidae	1 (0.1%)	< 0.01
<i>Alosa fallax fallax</i> *	Clupeidae	1 (0.1%)	< 0.01
<i>Ambloplites rupestris</i>	Centrarchidae	11 (0.9%)	0.01
<i>Ameiurus melas</i>	Ictaluridae	269 (22.6%)	0.64
<i>Ameiurus nebulosus</i> *	Ictaluridae	3 (0.3%)	< 0.01
<i>Anguilla anguilla</i>	Anguillidae	358 (30.1%)	0.23
<i>Aspius aspius</i>	Cyprinidae	21 (1.8%)	< 0.01
<i>Barbatula barbatula</i>	Nemacheilidae	818 (68.7%)	8.72
<i>Barbus barbus</i>	Cyprinidae	349 (29.3%)	1.29
<i>Blicca bjoerkna</i>	Cyprinidae	192 (16.1%)	0.3
<i>Chondrostoma nasus</i>	Cyprinidae	184 (15.4%)	0.22
<i>Cobitis taenia</i>	Cobitidae	25 (2.1%)	0.02
<i>Cottus gobio</i>	Cottidae	579 (48.6%)	9.04
<i>Cyprinus carpio</i>	Cyprinidae	277 (23.3%)	0.12
<i>Esox lucius</i>	Esocidae	304 (25.5%)	0.07
<i>Gasterosteus gymnurus</i>	Gasterosteidae	50 (4.2%)	0.2
<i>Gobio gobio</i>	Cyprinidae	744 (62.5%)	9.62
<i>Gymnocephalus cernuus</i>	Percidae	177 (14.9%)	0.05
<i>Hypophthalmichthys molitrix</i> *	Cyprinidae	4 (0.3%)	< 0.01
<i>Lampetra fluviatilis</i> *	Petromyzontidae	4 (0.3%)	< 0.01
<i>Lampetra planeri</i>	Petromyzontidae	374 (31.4%)	1.34
<i>Lepomis gibbosus</i>	Centrarchidae	479 (40.2%)	1.47
<i>Leucaspis delineatus</i>	Cyprinidae	105 (8.8%)	0.02
<i>Leuciscus idus</i>	Cyprinidae	10 (0.8%)	< 0.01
<i>Leuciscus leuciscus</i>	Cyprinidae	311 (26.1%)	0.3
<i>Liza ramada</i>	Mugilidae	16 (1.3%)	< 0.01
<i>Lota lota</i>	Gadidae	51 (4.3%)	0.01
<i>Micropterus salmoides</i>	Centrarchidae	45 (3.8%)	0.01
<i>Carrassius sp.</i>	Cyprinidae	234 (19.6%)	0.13
<i>Parachondrostoma toxostoma</i>	Cyprinidae	19 (1.6%)	0.01
<i>Perca fluviatilis</i>	Percidae	534 (44.8%)	1.03
<i>Petromyzon marinus</i>	Petromyzontidae	56 (4.7%)	0.01
<i>Phoxinus phoxinus</i>	Cyprinidae	801 (67.3%)	27.52
<i>Platichthys flesus</i> *	Pleuronectidae	1 (0.1%)	< 0.01
<i>Pseudorasbora parva</i>	Cyprinidae	221 (18.6%)	0.56

<i>Pungitius laevis</i>	Gasterosteidae	177 (14.9%)	1.88
<i>Rhodeus amarus</i>	Cyprinidae	285 (23.9%)	1.72
<i>Rutilus rutilus</i>	Cyprinidae	602 (50.5%)	5.22
<i>Salmo salar</i>	Salmonidae	59 (5.0%)	0.29
<i>Salmo trutta</i>	Salmonidae	788 (66.2%)	12.35
<i>Sander lucioperca</i>	Percidae	154 (12.9%)	0.02
<i>Scardinius erythrophthalmus</i>	Cyprinidae	345 (29.0%)	0.43
<i>Silurus glanis</i>	Siluridae	126 (10.6%)	0.05
<i>Squalius cephalus</i>	Cyprinidae	679 (57.0%)	5.42
<i>Telestes souffia</i> *	Cyprinidae	2 (0.2%)	< 0.01
<i>Thymallus thymallus</i>	Salmonidae	53 (4.5%)	0.02
<i>Tinca tinca</i>	Cyprinidae	306 (25.7%)	0.11

S2 Table. List of 230 macroinvertebrates genera present in the initial dataset. For each taxa, the number of occurrences (with the frequency of occurrence under brackets) over the 1195 T-NET reaches sampled is provided, as well as the average CPUE per reach (in number of specimens per 100 m² sampled). The taxa marked with an asterisk correspond to the rarest taxa (frequency of occurrence < 0.5%) that were excluded prior to analyses.

Genus	Order	Family	Occurrences (frequency)	Average CPUE per reach (ind/100 m ²)
<i>Acentrella</i>	Ephemeroptera	Baetidae	30 (2.5%)	0.24
<i>Acroloxus</i>	Gastropoda	Acroloxidae	158 (13.2%)	0.52
<i>Adicella</i>	Trichoptera	Leptoceridae	286 (23.9%)	1.27
<i>Aeshna</i>	Odonata	Aeshnidae	7 (0.6%)	< 0.01
<i>Agapetus</i>	Trichoptera	Glossosomatidae	239 (20.0%)	6.8
<i>Agraylea</i>	Trichoptera	Hydroptilidae	37 (3.1%)	0.16
<i>Allotrichia</i>	Trichoptera	Hydroptilidae	7 (0.6%)	0.02
<i>Ameletus</i> *	Ephemeroptera	Ameletidae	2 (0.2%)	< 0.01
<i>Ametropus</i> *	Ephemeroptera	Ametropodidae	1 (0.1%)	< 0.01
<i>Amphinemura</i>	Plecoptera	Nemouridae	88 (7.4%)	1.23
<i>Anaciaeschna</i> *	Odonata	Aeshnidae	1 (0.1%)	< 0.01
<i>Anax</i>	Odonata	Aeshnidae	29 (2.4%)	0.03
<i>Ancylus</i>	Gastropoda	Planorbidae	999 (83.6%)	40.5
<i>Anodonta</i>	Bivalvia	Unionidae	88 (7.4%)	0.09
<i>Aphelocheirus</i>	Heteroptera	Aphelocheiridae	503 (42.1%)	15.61
<i>Aplexa</i> *	Gastropoda	Physidae	1 (0.1%)	< 0.01
<i>Arcynopteryx</i> *	Plecoptera	Perlodidae	6 (0.5%)	0.01
<i>Arthroplea</i> *	Ephemeroptera	Arthropleidae	1 (0.1%)	< 0.01
<i>Asellidae</i>	Isopoda	Asellidae	831 (69.5%)	63.13
<i>Athripsodes</i>	Trichoptera	Leptoceridae	743 (62.2%)	11.08
<i>Atyaephyra</i>	Decapoda	Atyidae	38 (3.2%)	1.37
<i>Aulonogyrus</i> *	Coleoptera	Gyrinidae	3 (0.3%)	< 0.01
<i>Austropotamobius</i>	Decapoda	Astacidae	11 (0.9%)	0.03
<i>Baetis</i>	Ephemeroptera	Baetidae	1138 (95.2%)	262.05
<i>Belgrandia</i> *	Gastropoda	Hydrobiidae	1 (0.1%)	0.33
<i>Beraea</i>	Trichoptera	Beraeidae	27 (2.3%)	0.02
<i>Beraeamyia</i> *	Trichoptera	Beraeidae	2 (0.2%)	< 0.01
<i>Beraeodes</i>	Trichoptera	Beraeidae	187 (15.6%)	1.71
<i>Bithynia</i>	Gastropoda	Bithyniidae	452 (37.8%)	20.65
<i>Boyeria</i>	Odonata	Aeshnidae	259 (21.7%)	0.4
<i>Brachycentrus</i>	Trichoptera	Brachycentridae	381 (31.9%)	15.03
<i>Brachycercus</i>	Ephemeroptera	Caenidae	104 (8.7%)	0.27
<i>Brachyptera</i>	Plecoptera	Taeniopterygidae	44 (3.7%)	1.53

<i>Brachythemis*</i>	Odonata	Libellulidae	2 (0.2%)	< 0.01
<i>Brachytron*</i>	Odonata	Aeshnidae	2 (0.2%)	< 0.01
<i>Brychius</i>	Coleoptera	Haliplidae	33 (2.8%)	0.19
<i>Bythinella</i>	Gastropoda	Hydrobiidae	110 (9.2%)	0.95
<i>Bythiospeum</i>	Gastropoda	Hydrobiidae	25 (2.1%)	17.36
<i>Caenis</i>	Ephemeroptera	Caenidae	808 (67.6%)	54.29
<i>Calamoceras</i>	Trichoptera	Calamoceratidae	10 (0.8%)	0.05
<i>Calopteryx</i>	Odonata	Calopterygidae	1022 (85.5%)	15.04
<i>Capnia*</i>	Plecoptera	Capniidae	3 (0.3%)	0.48
<i>Capnioneura*</i>	Plecoptera	Capniidae	2 (0.2%)	< 0.01
<i>Capnopsis*</i>	Plecoptera	Capniidae	2 (0.2%)	0.09
<i>Centroptilum</i>	Ephemeroptera	Baetidae	612 (51.2%)	9.48
<i>Ceraclea</i>	Trichoptera	Leptoceridae	313 (26.2%)	1.62
<i>Chalcolestes</i>	Odonata	Lestidae	32 (2.7%)	0.03
<i>Cheumatopsyche</i>	Trichoptera	Hydropsychidae	398 (33.3%)	46.43
<i>Chimarra</i>	Trichoptera	Philopotamidae	84 (7.0%)	1.7
<i>Chloroperla</i>	Plecoptera	Chloroperlidae	54 (4.5%)	0.18
<i>Choroterpes</i>	Ephemeroptera	Leptophlebiidae	42 (3.5%)	0.16
<i>Cloeon</i>	Ephemeroptera	Baetidae	352 (29.5%)	4.02
<i>Corbicula</i>	Bivalvia	Corbiculidae	198 (16.6%)	25.72
<i>Cordulegaster</i>	Odonata	Cordulegasteridae	238 (19.9%)	0.63
<i>Cordulia*</i>	Odonata	Corduliidae	3 (0.3%)	< 0.01
<i>Corophium</i>	Amphipoda	Coropiidae	9 (0.8%)	1.08
<i>Crangonyx</i>	Amphipoda	Crangonyctidae	73 (6.1%)	3.28
<i>Crunoecia</i>	Trichoptera	Lepidostomatidae	18 (1.5%)	0.01
<i>Cymatia*</i>	Heteroptera	Corixidae	2 (0.2%)	< 0.01
<i>Cyphon</i>	Coleoptera	Scirtidae	20 (1.7%)	0.1
<i>Cyrnus</i>	Trichoptera	Polycentropodidae	466 (39.0%)	4.03
<i>Dictyogenus*</i>	Plecoptera	Perlodidae	3 (0.3%)	< 0.01
<i>Dikerogammarus</i>	Amphipoda	Pontogammaridae	46 (3.8%)	8.26
<i>Dinocras</i>	Plecoptera	Perlidae	107 (9.0%)	1.57
<i>Diplacodes*</i>	Odonata	Libellulidae	1 (0.1%)	< 0.01
<i>Diplectronea*</i>	Trichoptera	Hydropsychidae	6 (0.5%)	0.06
<i>Donacia*</i>	Coleoptera	Chrysomelidae	1 (0.1%)	< 0.01
<i>Dreissena</i>	Bivalvia	Dreissenidae	9 (0.8%)	0.6
<i>Dryops</i>	Coleoptera	Dryopidae	275 (23.0%)	0.28
<i>Dupophilus</i>	Coleoptera	Elmidae	458 (38.3%)	20.24
<i>Dytiscidae</i>	Coleoptera	Dytiscidae	705 (59.0%)	5.38
<i>Ecdyonurus</i>	Ephemeroptera	Heptageniidae	527 (44.1%)	10.32
<i>Echinogammarus</i>	Amphipoda	Gammaridae	492 (41.2%)	525.51
<i>Ecnomus</i>	Trichoptera	Ecnomidae	86 (7.2%)	0.91

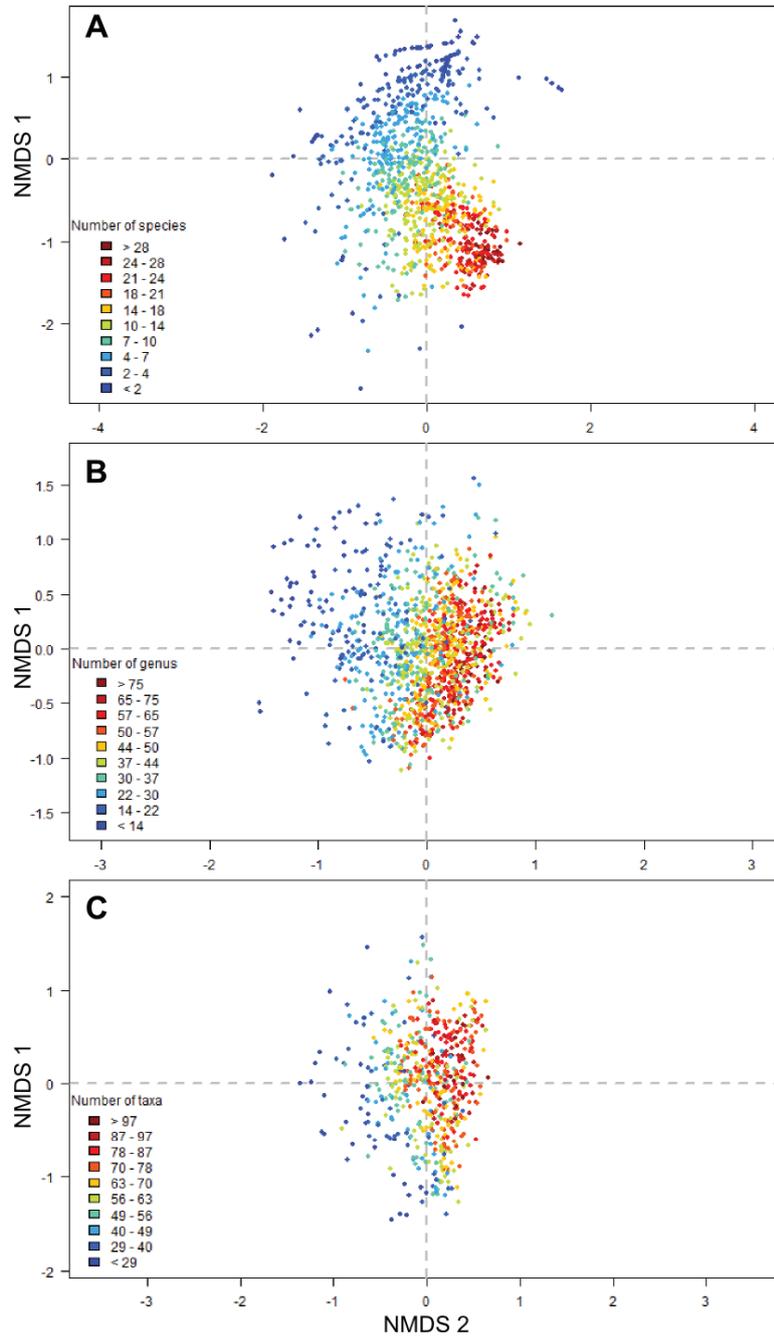
<i>Electrogena</i>	Ephemeroptera	Heptageniidae	155 (13.0%)	0.41
<i>Elmis</i>	Coleoptera	Elmidae	1036 (86.7%)	112.31
<i>Epeorus</i>	Ephemeroptera	Heptageniidae	271 (22.7%)	9.16
<i>Ephemera</i>	Ephemeroptera	Ephemeridae	937 (78.4%)	35.3
<i>Ephemerella</i>	Ephemeroptera	Ephemerellidae	944 (79.0%)	96.81
<i>Ephoron</i>	Ephemeroptera	Polymitarciidae	114 (9.5%)	1.67
<i>Epithea</i>	Odonata	Corduliidae	9 (0.8%)	0.01
<i>Erotesis*</i>	Trichoptera	Leptoceridae	2 (0.2%)	< 0.01
<i>Esolus</i>	Coleoptera	Elmidae	877 (73.4%)	69.69
<i>Euleuctra</i>	Plecoptera	Leuctridae	328 (27.4%)	4.22
<i>Ferrissia</i>	Gastropoda	Planorbidae	207 (17.3%)	2.95
<i>Galba</i>	Gastropoda	Lymnaeidae	46 (3.8%)	0.08
<i>Gammarus</i>	Amphipoda	Gammaridae	1094 (91.5%)	725.43
<i>Gerris</i>	Heteroptera	Gerridae	599 (50.1%)	1.43
<i>Glossosoma</i>	Trichoptera	Glossosomatidae	180 (15.1%)	2.54
<i>Goera</i>	Trichoptera	Goeridae	428 (35.8%)	2.08
<i>Gomphus</i>	Odonata	Gomphidae	507 (42.4%)	2.67
<i>Gyrinus</i>	Coleoptera	Gyrinidae	100 (8.4%)	0.14
<i>Habroleptoides</i>	Ephemeroptera	Leptophlebiidae	176 (14.7%)	3.6
<i>Habrophlebia</i>	Ephemeroptera	Leptophlebiidae	384 (32.1%)	10.81
<i>Haliplus</i>	Coleoptera	Haliplidae	302 (25.3%)	1.47
<i>Helodes</i>	Coleoptera	Scirtidae	249 (20.8%)	0.93
<i>Helophorus</i>	Coleoptera	Helophoridae	261 (21.8%)	0.62
<i>Hemianax*</i>	Odonata	Aeshnidae	1 (0.1%)	< 0.01
<i>Heptagenia</i>	Ephemeroptera	Heptageniidae	410 (34.3%)	5.77
<i>Holocentropus</i>	Trichoptera	Polycentropodidae	157 (13.1%)	0.71
<i>Hydraena</i>	Coleoptera	Hydraenidae	786 (65.8%)	16.31
<i>Hydrochus</i>	Coleoptera	Hydrochidae	88 (7.4%)	0.1
<i>Hydrocyphon</i>	Coleoptera	Scirtidae	96 (8.0%)	0.59
<i>Hydrometra</i>	Heteroptera	Hydrometridae	230 (19.2%)	0.28
<i>Hydrophilidae</i>	Coleoptera	Hydrophilidae	386 (32.3%)	0.6
<i>Hydropsyche</i>	Trichoptera	Hydropsychidae	1054 (88.2%)	203.89
<i>Hydroptila</i>	Trichoptera	Hydroptilidae	708 (59.2%)	16.5
<i>Hydroscapha*</i>	Coleoptera	Hydroscaphidae	2 (0.2%)	< 0.01
<i>Hygrobia*</i>	Coleoptera	Hygrobiidae	2 (0.2%)	0.01
<i>Isogenus*</i>	Plecoptera	Perlodidae	3 (0.3%)	0.01
<i>Isoperla</i>	Plecoptera	Perlodidae	260 (21.8%)	4.38
<i>Ithytrichia</i>	Trichoptera	Hydroptilidae	440 (36.8%)	19.62
<i>Lasiocephala</i>	Trichoptera	Lepidostomatidae	146 (12.2%)	4.58
<i>Lepidostoma</i>	Trichoptera	Lepidostomatidae	456 (38.2%)	12.93
<i>Leptocerus</i>	Trichoptera	Leptoceridae	138 (11.5%)	2.72

<i>Leptophlebia</i>	Ephemeroptera	Leptophlebiidae	16 (1.3%)	0.03
<i>Lestes</i> *	Odonata	Lestidae	3 (0.3%)	< 0.01
<i>Leucorrhinia</i> *	Odonata	Libellulidae	2 (0.2%)	< 0.01
<i>Leuctra</i>	Plecoptera	Leuctridae	721 (60.3%)	60.35
<i>Libellula</i>	Odonata	Libellulidae	42 (3.5%)	0.07
<i>Limnebius</i>	Coleoptera	Hydraenidae	92 (7.7%)	0.21
<i>Limnephilidae</i>	Trichoptera	Limnephilidae	936 (78.3%)	20.39
<i>Limnius</i>	Coleoptera	Elmidae	935 (78.2%)	41.57
<i>Lithax</i>	Trichoptera	Goeridae	31 (2.6%)	0.5
<i>Lithoglyphus</i> *	Gastropoda	Hydrobiidae	3 (0.3%)	< 0.01
<i>Lymnaea</i>	Gastropoda	Lymnaeidae	19 (1.6%)	0.03
<i>Lype</i>	Trichoptera	Psychomyiidae	532 (44.5%)	1.51
<i>Macromia</i> *	Odonata	Corduliidae	1 (0.1%)	< 0.01
<i>Macronychus</i>	Coleoptera	Elmidae	180 (15.1%)	0.4
<i>Macrolea</i> *	Coleoptera	Chrysomelidae	1 (0.1%)	< 0.01
<i>Margaritifera</i> *	Bivalvia	Margaritiferidae	4 (0.3%)	< 0.01
<i>Marthamea</i> *	Plecoptera	Perlidae	1 (0.1%)	< 0.01
<i>Mesovelia</i>	Heteroptera	Mesoveliidae	39 (3.3%)	0.06
<i>Metalype</i>	Trichoptera	Psychomyiidae	15 (1.3%)	0.02
<i>Micrasema</i>	Trichoptera	Brachycentridae	274 (22.9%)	21.5
<i>Micronecta</i>	Heteroptera	Corixidae	684 (57.2%)	42.47
<i>Molanna</i>	Trichoptera	Molannidae	105 (8.8%)	0.32
<i>Molannodes</i>	Trichoptera	Molannidae	13 (1.1%)	0.02
<i>Mystacides</i>	Trichoptera	Leptoceridae	830 (69.5%)	12.34
<i>Myxas</i>	Gastropoda	Lymnaeidae	10 (0.8%)	0.02
<i>Naucoridae</i>	Heteroptera	Naucoridae	54 (4.5%)	1.19
<i>Nemoura</i>	Plecoptera	Nemouridae	353 (29.5%)	10.43
<i>Nemurella</i> *	Plecoptera	Nemouridae	2 (0.2%)	< 0.01
<i>Neophemera</i> *	Ephemeroptera	Neophemeridae	1 (0.1%)	< 0.01
<i>Nepidae</i>	Heteroptera	Nepidae	178 (14.9%)	0.19
<i>Neureclipsis</i>	Trichoptera	Polycentropodidae	130 (10.9%)	3.03
<i>Niphargus</i>	Amphipoda	Niphargidae	46 (3.8%)	0.06
<i>Normandia</i>	Coleoptera	Elmidae	111 (9.3%)	2.19
<i>Noterus</i> *	Coleoptera	Noteridae	4 (0.3%)	< 0.01
<i>Notidobia</i>	Trichoptera	Sericostomatidae	254 (21.3%)	1.07
<i>Notonectidae</i>	Heteroptera	Notonectidae	190 (15.9%)	0.24
<i>Ochthebius</i>	Coleoptera	Hydraenidae	76 (6.4%)	0.09
<i>Odontocerum</i>	Trichoptera	Odontoceridae	148 (12.4%)	2.55
<i>Oecetis</i>	Trichoptera	Leptoceridae	520 (43.5%)	5.02
<i>Oecismus</i>	Trichoptera	Sericostomatidae	10 (0.8%)	0.01
<i>Oligoneuriella</i>	Ephemeroptera	Oligoneuriidae	75 (6.3%)	1.59

<i>Oligoptectrum</i>	Trichoptera	Brachycentridae	124 (10.4%)	46.94
<i>Onychogomphus</i>	Odonata	Gomphidae	424 (35.5%)	1.98
<i>Ophiogomphus</i>	Odonata	Gomphidae	26 (2.2%)	0.05
<i>Orconectes</i>	Decapoda	Cambaridae	275 (23.0%)	0.42
<i>Orectochilus</i>	Coleoptera	Gyrinidae	524 (43.8%)	3.62
<i>Orthetrum</i>	Odonata	Libellulidae	26 (2.2%)	0.04
<i>Orthotrichia</i>	Trichoptera	Hydroptilidae	108 (9.0%)	3.24
<i>Oulimnius</i>	Coleoptera	Elmidae	1003 (83.9%)	70.38
<i>Oxyethira</i>	Trichoptera	Hydroptilidae	17 (1.4%)	0.02
<i>Oxygastra</i>	Odonata	Corduliidae	18 (1.5%)	0.05
<i>Pacifastacus</i>	Decapoda	Astacidae	215 (18.0%)	1.7
<i>Paragomphus*</i>	Odonata	Gomphidae	2 (0.2%)	< 0.01
<i>Paraleptophlebia</i>	Ephemeroptera	Leptophlebiidae	551 (46.1%)	10.03
<i>Peltodytes</i>	Coleoptera	Haliplidae	11 (0.9%)	0.01
<i>Perla</i>	Plecoptera	Perlidae	189 (15.8%)	1.85
<i>Perlodes</i>	Plecoptera	Perlodidae	180 (15.1%)	1.14
<i>Philopotamus</i>	Trichoptera	Philopotamidae	89 (7.4%)	1.13
<i>Phryganea</i>	Trichoptera	Phryganeidae	18 (1.5%)	0.01
<i>Physa</i>	Gastropoda	Physidae	374 (31.3%)	8.3
<i>Physella</i>	Gastropoda	Physidae	247 (20.7%)	4.72
<i>Pisidium</i>	Bivalvia	Sphaeriidae	1058 (88.5%)	151.98
<i>Platycnemis</i>	Odonata	Platycnemididae	624 (52.2%)	9.16
<i>Plea</i>	Heteroptera	Pleidae	35 (2.9%)	0.23
<i>Plectrocnemia</i>	Trichoptera	Polycentropodidae	93 (7.8%)	0.32
<i>Polycentropus</i>	Trichoptera	Polycentropodidae	804 (67.3%)	20.68
<i>Pomatinus</i>	Coleoptera	Dryopidae	169 (14.1%)	0.24
<i>Potamanthus</i>	Ephemeroptera	Potamanthidae	145 (12.1%)	4.95
<i>Potamophilus</i>	Coleoptera	Elmidae	91 (7.6%)	0.3
<i>Potamopyrgus</i>	Gastropoda	Hydrobiidae	811 (67.9%)	491.73
<i>Potomida</i>	Bivalvia	Unionidae	9 (0.8%)	0.01
<i>Procambarus</i>	Decapoda	Cambaridae	10 (0.8%)	0.01
<i>Proclleon</i>	Ephemeroptera	Baetidae	513 (42.9%)	5.18
<i>Protonemura</i>	Plecoptera	Nemouridae	320 (26.8%)	34.48
<i>Pseudanodonta</i>	Bivalvia	Unionidae	15 (1.3%)	0.01
<i>Pseudocentropilum</i>	Ephemeroptera	Baetidae	125 (10.5%)	0.4
<i>Pseudoneureclipsis*</i>	Trichoptera	Ecnomidae	1 (0.1%)	< 0.01
<i>Psychomyia</i>	Trichoptera	Psychomyiidae	459 (38.4%)	13.04
<i>Ptilocolepus*</i>	Trichoptera	Hydroptilidae	5 (0.4%)	0.01
<i>Radix</i>	Gastropoda	Lymnaeidae	562 (47.0%)	7.12
<i>Raptobaetopus</i>	Ephemeroptera	Baetidae	34 (2.8%)	0.12
<i>Rhabdiopteryx*</i>	Plecoptera	Taeniopterygidae	4 (0.3%)	0.03

<i>Rhithrogena</i>	Ephemeroptera	Heptageniidae	325 (27.2%)	9.33
<i>Rhyacophila</i>	Trichoptera	Rhyacophilidae	803 (67.2%)	16.51
<i>Riolus</i>	Coleoptera	Elmidae	271 (22.7%)	9.64
<i>Schizopelex*</i>	Trichoptera	Sericostomatidae	1 (0.1%)	< 0.01
<i>Scirtes*</i>	Coleoptera	Scirtidae	5 (0.4%)	0.01
<i>Sericostoma</i>	Trichoptera	Sericostomatidae	485 (40.6%)	17.14
<i>Setodes</i>	Trichoptera	Leptoceridae	108 (9.0%)	1.01
<i>Silo</i>	Trichoptera	Goeridae	486 (40.7%)	8.96
<i>Siphonurus</i>	Ephemeroptera	Siphonuridae	17 (1.4%)	0.07
<i>Siphonoperla</i>	Plecoptera	Chloroperlidae	164 (13.7%)	1.03
<i>Somatochlora</i>	Odonata	Corduliidae	33 (2.8%)	0.03
<i>Sphaerium</i>	Bivalvia	Sphaeriidae	624 (52.2%)	43.52
<i>Stactobia*</i>	Trichoptera	Hydroptilidae	1 (0.1%)	< 0.01
<i>Stactobiella*</i>	Trichoptera	Hydroptilidae	1 (0.1%)	< 0.01
<i>Stagnicola</i>	Gastropoda	Lymnaeidae	36 (3.0%)	0.08
<i>Stenelmis</i>	Coleoptera	Elmidae	315 (26.4%)	3.3
<i>Sympecma</i>	Odonata	Lestidae	7 (0.6%)	0.01
<i>Sympetrum</i>	Odonata	Libellulidae	8 (0.7%)	< 0.01
<i>Synagapetus</i>	Trichoptera	Glossosomatidae	12 (1.0%)	0.19
<i>Taeniopteryx</i>	Plecoptera	Taeniopterygidae	146 (12.2%)	2.09
<i>Theodoxus</i>	Gastropoda	Neritidae	236 (19.7%)	24.89
<i>Thraululus</i>	Ephemeroptera	Leptophlebiidae	10 (0.8%)	0.02
<i>Thremma</i>	Trichoptera	Uenoidae	13 (1.1%)	0.25
<i>Tinodes</i>	Trichoptera	Psychomyiidae	302 (25.3%)	1.81
<i>Torleya</i>	Ephemeroptera	Ephemerellidae	147 (12.3%)	1.37
<i>Triaenodes</i>	Trichoptera	Leptoceridae	33 (2.8%)	0.1
<i>Trichostegia*</i>	Trichoptera	Phryganeidae	3 (0.3%)	< 0.01
<i>Unio</i>	Bivalvia	Unionidae	97 (8.1%)	0.08
<i>Valvata</i>	Gastropoda	Valvatidae	178 (14.9%)	2.48
<i>Viviparus</i>	Gastropoda	Viviparidae	74 (6.2%)	4.88
<i>Wormaldia</i>	Trichoptera	Philopotamidae	46 (3.8%)	0.17
<i>Xanthoperla</i>	Plecoptera	Chloroperlidae	10 (0.8%)	0.15
<i>Ylodes</i>	Trichoptera	Leptoceridae	34 (2.8%)	0.08

S1 Figure. Taxonomic richness of the sampled T-NET reaches plotted on the first factorial plan of the fish, macroinvertebrates and common ordinations (A, B and C, respectively). Reaches are colored according to their taxonomic richness from low (blue) to high (red) number of species or genera.



S3 Table. Position of the centroids of the 43 fish species on the three factorial axes of the NMDS performed on the fish dataset.

Scientific name	Axis 1	Axis 2	Axis 3
<i>Abramis brama</i>	-1.27	0.66	0.25
<i>Alburnoides bipunctatus</i>	-0.67	0.37	-0.76
<i>Alburnus alburnus</i>	-1.09	0.68	-0.40
<i>Ambloplites rupestris</i>	-1.41	1.23	-0.52
<i>Ameiurus melas</i>	-1.06	-0.27	0.16
<i>Anguilla anguilla</i>	-0.93	0.27	0.29
<i>Aspius aspius</i>	-1.48	1.31	-0.15
<i>Barbatula barbatula</i>	0.05	-0.78	-0.02
<i>Barbus barbus</i>	-0.74	0.42	-0.82
<i>Blicca bjoerkna</i>	-1.35	0.91	0.16
<i>Carassius sp.</i>	-1.06	-0.26	0.43
<i>Chondrostoma nasus</i>	-1.02	0.79	-0.74
<i>Cobitis taenia</i>	-1.44	1.16	0.00
<i>Cottus gobio</i>	0.44	-0.05	1.02
<i>Cyprinus carpio</i>	-1.00	-0.37	0.26
<i>Esox lucius</i>	-1.02	0.18	0.62
<i>Gasterosteus gymnurus</i>	-1.07	-0.51	0.11
<i>Gobio gobio</i>	-0.46	-0.20	-0.46
<i>Gymnocephalus cernuus</i>	-1.18	0.95	-0.13
<i>Lampetra planeri</i>	0.23	-0.20	0.52
<i>Lepomis gibbosus</i>	-0.94	-0.52	-0.04
<i>Leucaspis delineatus</i>	-1.06	0.23	0.35
<i>Leuciscus idus</i>	-1.57	1.07	0.30
<i>Leuciscus leuciscus</i>	-0.69	0.37	-0.47
<i>Liza ramada</i>	-1.44	1.40	-0.29
<i>Lota lota</i>	-0.75	0.73	-0.30
<i>Micropterus salmoides</i>	-1.31	0.02	0.32
<i>Parachondrostoma toxostoma</i>	-0.83	0.54	-1.30
<i>Perca fluviatilis</i>	-0.78	-0.02	0.25
<i>Petromyzon marinus</i>	-1.14	1.16	-0.57
<i>Phoxinus phoxinus</i>	0.37	-1.04	-0.36
<i>Pseudorasbora parva</i>	-1.15	-0.06	-0.21
<i>Pungitius laevis</i>	-0.49	-0.75	1.42
<i>Rhodeus amarus</i>	-1.23	0.63	-0.20
<i>Rutilus rutilus</i>	-0.92	0.00	0.08
<i>Salmo salar</i>	0.30	0.61	-0.02
<i>Salmo trutta</i>	1.53	0.34	-0.10
<i>Sander lucioperca</i>	-1.30	0.74	0.01
<i>Scardinius erythrophthalmus</i>	-1.13	-0.22	0.15

<i>Silurus glanis</i>	-1.38	1.11	-0.17
<i>Squalius cephalus</i>	-0.63	-0.01	-0.36
<i>Thymallus thymallus</i>	0.06	0.18	-0.19
<i>Tinca tinca</i>	-0.97	0.10	0.21

S4 Table. Position of the centroids of the 186 macroinvertebrates taxa on the three factorial axes of the NMDS performed on the macroinvertebrate dataset.

Macroinvertebrates' genus	Axis 1	Axis 2	Axis 3
<i>Acentrella</i>	-0.60	0.40	0.49
<i>Acroloxus</i>	0.73	0.00	-0.45
<i>Adicella</i>	-0.66	-0.05	-0.34
<i>Aeshna</i>	0.13	-0.50	0.20
<i>Agapetus</i>	-0.16	-0.45	-0.74
<i>Agraylea</i>	0.51	0.73	0.16
<i>Allotrichia</i>	0.42	0.72	0.50
<i>Amphinemura</i>	-1.00	-0.40	0.17
<i>Anax</i>	1.08	1.06	0.01
<i>Ancylus</i>	-0.08	-0.15	0.23
<i>Anodonta</i>	0.85	0.05	-0.04
<i>Aphelocheirus</i>	0.07	0.52	0.22
<i>Asellidae</i>	0.55	-0.56	0.40
<i>Athripsodes</i>	0.20	0.11	-0.11
<i>Atyaephyra</i>	1.21	0.91	0.02
<i>Austropotamobius</i>	-0.73	-0.65	-0.28
<i>Baetis</i>	-0.17	-0.13	0.27
<i>Beraea</i>	-0.81	-0.61	0.12
<i>Beraeodes</i>	0.28	-0.30	-0.57
<i>Bithynia</i>	0.80	-0.24	0.10
<i>Boyeria</i>	-0.07	0.34	-0.24
<i>Brachycentrus</i>	-0.19	0.69	0.09
<i>Brachycercus</i>	-0.10	0.57	0.64
<i>Brachyptera</i>	-0.97	-0.31	0.36
<i>Brychius</i>	0.07	0.01	-1.28
<i>Bythinella</i>	-0.80	-0.51	0.00
<i>Bythiospeum</i>	0.49	-0.01	-0.95
<i>Caenis</i>	0.39	0.51	0.12
<i>Calamoceras</i>	-0.08	0.40	-0.83
<i>Calopteryx</i>	0.08	0.00	0.08
<i>Centroptilum</i>	0.25	0.33	-0.17
<i>Ceraclea</i>	0.37	0.65	0.09
<i>Chalcolestes</i>	1.07	-0.11	0.18
<i>Cheumatopsyche</i>	-0.16	0.69	0.23
<i>Chimarra</i>	-0.29	0.88	0.16
<i>Chloroperla</i>	-1.15	0.05	-0.10
<i>Choroterpes</i>	0.18	0.99	0.84
<i>Cloeon</i>	1.10	-0.22	0.34
<i>Corbicula</i>	0.61	1.10	0.30
<i>Cordulegaster</i>	-0.72	-0.55	-0.03
<i>Corophium</i>	0.56	1.68	0.58
<i>Crangonyx</i>	0.59	0.87	0.56

<i>Crunoecia</i>	-1.21	-0.44	-0.31
<i>Cyphon</i>	-0.86	0.26	-0.39
<i>Cyrnus</i>	0.34	0.24	0.11
<i>Dikerogammarus</i>	0.64	1.46	0.56
<i>Dinocras</i>	-1.25	-0.15	0.00
<i>Dreissena</i>	1.75	1.08	0.12
<i>Dryops</i>	-0.05	-0.35	0.09
<i>Dupophilus</i>	-0.82	0.09	0.01
<i>Dytiscidae</i>	-0.06	-0.37	0.08
<i>Ecdyonurus</i>	-0.65	0.03	0.40
<i>Echinogammarus</i>	0.28	0.04	-0.90
<i>Ecnomus</i>	1.33	0.98	0.31
<i>Electrogena</i>	-0.21	0.25	0.48
<i>Elmis</i>	-0.28	-0.14	-0.37
<i>Epeorus</i>	-1.04	-0.03	0.10
<i>Ephemera</i>	-0.09	-0.24	-0.01
<i>Ephemerella</i>	-0.28	-0.09	0.33
<i>Ephoron</i>	0.35	0.92	0.77
<i>Epitheca</i>	0.21	0.56	0.21
<i>Esolus</i>	-0.14	0.42	-0.17
<i>Euleuctra</i>	-0.25	0.44	-0.10
<i>Ferrissia</i>	0.88	0.57	0.58
<i>Galba</i>	0.27	-0.38	0.05
<i>Gammarus</i>	0.19	-0.82	-0.12
<i>Gerris</i>	0.23	-0.07	0.26
<i>Glossosoma</i>	-1.01	-0.07	0.01
<i>Goera</i>	0.18	0.03	-0.07
<i>Gomphus</i>	0.01	0.53	0.39
<i>Gyrinus</i>	0.37	-0.32	-0.35
<i>Habroleptoides</i>	-0.92	-0.02	0.23
<i>Habrophlebia</i>	-0.61	-0.37	0.39
<i>Haliplus</i>	0.85	-0.12	-0.29
<i>Helodes</i>	-0.27	-0.83	-0.61
<i>Helophorus</i>	0.26	-0.69	0.01
<i>Heptagenia</i>	-0.03	0.63	0.56
<i>Holocentropus</i>	0.02	0.14	0.17
<i>Hydraena</i>	-0.56	-0.22	0.05
<i>Hydrochus</i>	0.34	0.06	0.05
<i>Hydrocyphon</i>	-1.01	0.04	0.13
<i>Hydrometra</i>	0.36	-0.27	0.27
<i>Hydrophilidae</i>	0.32	-0.05	0.29
<i>Hydropsyche</i>	-0.06	0.11	0.20
<i>Hydroptila</i>	0.25	0.22	0.10
<i>Isoperla</i>	-0.97	-0.34	0.42
<i>Ithytrichia</i>	-0.12	0.40	-0.43
<i>Lasiocephala</i>	-0.47	0.08	-0.71

<i>Lepidostoma</i>	-0.22	0.28	-0.58
<i>Leptocerus</i>	0.41	0.80	-0.21
<i>Leptophlebia</i>	-0.39	0.18	-0.05
<i>Leuctra</i>	-0.62	0.06	0.25
<i>Libellula</i>	1.02	-0.35	0.45
<i>Limnebius</i>	-0.91	-0.14	-0.03
<i>Limnephilidae</i>	-0.40	-0.53	-0.11
<i>Limnius</i>	-0.24	0.01	-0.45
<i>Lithax</i>	-0.69	0.16	0.01
<i>Lymnaea</i>	0.65	0.28	-0.41
<i>Lype</i>	0.11	-0.04	-0.39
<i>Macronychus</i>	0.17	0.86	-0.27
<i>Mesovelia</i>	0.24	0.09	-0.13
<i>Metalype</i>	0.47	0.18	0.20
<i>Micrasema</i>	-0.99	0.03	0.04
<i>Micronecta</i>	0.43	0.18	0.35
<i>Molanna</i>	0.62	0.30	-0.82
<i>Molannodes</i>	0.62	0.50	-0.87
<i>Mystacides</i>	0.12	0.30	-0.02
<i>Myxas</i>	0.41	-0.59	-0.11
<i>Naucoridae</i>	0.61	-0.40	1.98
<i>Nemoura</i>	-0.73	-0.21	0.60
<i>Nepidae</i>	0.35	-0.78	0.09
<i>Neureclipsis</i>	0.90	0.44	0.54
<i>Niphargus</i>	-0.11	-0.10	-0.31
<i>Normandia</i>	0.22	0.77	-0.63
<i>Notidobia</i>	0.20	-0.27	-0.68
<i>Notonectidae</i>	0.53	-0.64	0.26
<i>Ochthebius</i>	0.37	-0.54	-0.30
<i>Odontocерum</i>	-1.09	-0.41	-0.26
<i>Oecetis</i>	-0.12	0.64	0.02
<i>Oecismus</i>	-1.07	0.32	-0.37
<i>Oligoneuriella</i>	-0.42	0.64	0.68
<i>Oligoplectrum</i>	-0.49	0.72	0.18
<i>Onychogomphus</i>	-0.15	0.63	0.25
<i>Ophiogomphus</i>	-0.02	0.97	0.20
<i>Orconectes</i>	0.52	0.54	0.17
<i>Orectochilus</i>	-0.24	0.34	-0.07
<i>Orthetrum</i>	0.89	0.24	-0.33
<i>Orthotrichia</i>	0.74	1.15	0.11
<i>Oulimnius</i>	0.09	0.08	-0.25
<i>Oxyethira</i>	-0.41	0.36	-0.43
<i>Oxygastra</i>	0.24	1.16	-0.13
<i>Pacifastacus</i>	-0.40	0.03	0.49
<i>Paraleptophlebia</i>	-0.38	0.21	0.29
<i>Peltodytes</i>	0.79	0.04	-0.03

<i>Perla</i>	-1.00	0.02	0.13
<i>Perlodes</i>	-1.01	-0.01	0.15
<i>Philopotamus</i>	-1.27	-0.35	-0.04
<i>Phryganea</i>	0.99	0.03	-0.43
<i>Physa</i>	0.82	0.48	0.17
<i>Physella</i>	0.93	-0.04	0.40
<i>Pisidium</i>	0.43	-0.68	-0.43
<i>Platycnemis</i>	0.63	0.34	0.33
<i>Plea</i>	0.54	-1.35	0.14
<i>Plectrocnemia</i>	-0.64	-0.71	-0.06
<i>Polycentropus</i>	-0.10	0.17	0.01
<i>Pomatinus</i>	-0.08	0.23	-0.21
<i>Potamanthus</i>	0.28	0.61	1.07
<i>Potamophilus</i>	0.28	1.25	0.22
<i>Potamopyrgus</i>	0.37	-0.24	-0.46
<i>Potomida</i>	0.66	1.39	-0.40
<i>Procambarus</i>	0.99	0.40	0.23
<i>Procloeon</i>	0.28	0.61	0.32
<i>Protonemura</i>	-1.12	-0.23	0.10
<i>Pseudanodonta</i>	0.49	0.41	-0.52
<i>Pseudocentropilum</i>	-0.41	0.30	0.25
<i>Psychomyia</i>	-0.11	0.62	0.48
<i>Radix</i>	0.55	-0.21	-0.03
<i>Raptobaetopus</i>	0.48	1.44	0.49
<i>Rhithrogena</i>	-0.93	-0.20	0.15
<i>Rhyacophila</i>	-0.50	-0.12	0.16
<i>Riolus</i>	0.20	0.14	-0.97
<i>Sericostoma</i>	-0.60	-0.20	-0.49
<i>Setodes</i>	-0.03	0.87	0.30
<i>Silo</i>	-0.48	-0.15	-0.33
<i>Siphonurus</i>	-0.57	-0.35	0.95
<i>Siphonoperla</i>	-1.09	-0.16	0.23
<i>Somatochlora</i>	0.55	-0.19	0.06
<i>Sphaerium</i>	0.66	-0.35	-0.30
<i>Stagnicola</i>	0.97	-1.41	0.40
<i>Stenelmis</i>	0.06	0.72	-0.33
<i>Sympecma</i>	0.46	0.05	0.51
<i>Sympetrum</i>	0.92	0.24	0.25
<i>Synagapetus</i>	-0.08	-0.03	1.36
<i>Taeniopteryx</i>	-0.68	0.37	0.06
<i>Theodoxus</i>	0.37	0.49	-0.53
<i>Thraulus</i>	-0.22	0.95	0.12
<i>Thremma</i>	-1.62	-0.42	-0.05
<i>Tinodes</i>	0.27	-0.05	0.14
<i>Torleya</i>	-0.74	0.37	-0.14
<i>Triaenodes</i>	-0.15	0.90	-0.36

<i>Unio</i>	0.35	0.35	-0.43
<i>Valvata</i>	0.86	-0.33	-0.16
<i>Viviparus</i>	-0.31	-1.84	1.57
<i>Wormaldia</i>	-1.03	0.15	-0.13
<i>Xanthoperla</i>	0.44	1.47	0.58
<i>Ylodes</i>	0.05	0.96	0.01

S5 Table. Position of the centroids of the 43 fish species (bold) and 170 macroinvertebrates taxa (non-bold) on the three factorial axes of the NMDS performed on the common dataset.

Taxa name	Axis 1	Axis 2	Axis 3
<i>Abramis brama</i>	0.88	-0.01	-0.13
<i>Alburnoides bipunctatus</i>	0.29	0.57	-0.39
<i>Alburnus alburnus</i>	0.73	0.46	-0.21
<i>Ambloplites rupestris</i>	0.96	1.08	-0.12
<i>Ameiurus melas</i>	0.92	-0.39	-0.55
<i>Anguilla anguilla</i>	0.72	-0.12	0.06
<i>Aspius aspius</i>	1.27	0.69	-0.17
<i>Barbatula barbatula</i>	-0.06	-0.44	-0.23
<i>Barbus barbus</i>	0.41	0.55	-0.22
<i>Blicca bjoerkna</i>	0.99	0.27	-0.01
<i>Carassius sp.</i>	0.61	-0.57	-0.49
<i>Chondrostoma nasus</i>	0.60	0.71	-0.34
<i>Cobitis taenia</i>	1.14	0.41	-0.19
<i>Cottus gobio</i>	-0.32	-0.37	0.33
<i>Cyprinus carpio</i>	0.58	-0.33	-0.45
<i>Esox lucius</i>	0.70	-0.15	0.11
<i>Gasterosteus gymnurur</i>	0.35	-0.76	-1.28
<i>Gobio gobio</i>	0.29	0.03	-0.43
<i>Gymnocephalus cernuus</i>	0.81	0.37	0.05
<i>Lampetra planeri</i>	-0.28	-0.04	0.07
<i>Lepomis gibbosus</i>	0.61	-0.33	-0.56
<i>Leucaspius delineatus</i>	0.74	-0.34	-0.15
<i>Leuciscus idus</i>	1.28	0.10	-0.25
<i>Leuciscus leuciscus</i>	0.44	0.41	-0.21
<i>Liza ramada</i>	1.39	0.94	-0.33
<i>Lota lota</i>	0.22	0.18	-0.02
<i>Micropterus salmoides</i>	0.89	-0.10	-0.01
<i>Parachondrostoma toxostoma</i>	0.55	0.28	-0.33
<i>Perca fluviatilis</i>	0.44	-0.18	-0.21
<i>Petromyzon marinus</i>	0.72	0.93	-0.21
<i>Phoxinus phoxinus</i>	-0.21	-0.51	0.01
<i>Pseudorasbora parva</i>	0.74	-0.19	-0.59
<i>Pungitius laevis</i>	0.20	-1.04	0.65
<i>Rhodeus amarus</i>	0.90	0.18	-0.25
<i>Rutilus rutilus</i>	0.66	-0.25	-0.23
<i>Salmo salar</i>	-0.56	-0.27	1.27
<i>Salmo trutta</i>	-1.02	-0.19	-0.37
<i>Sander lucioperca</i>	0.87	0.27	-0.10
<i>Scardinius erythrophthalmus</i>	0.65	-0.40	-0.54
<i>Silurus glanis</i>	0.98	0.66	0.01

<i>Squalius cephalus</i>	0.34	0.07	-0.33
<i>Thymallus thymallus</i>	-0.52	0.63	-0.14
<i>Tinca tinca</i>	0.63	-0.31	-0.28
<i>Acentrella</i>	-0.81	0.32	-0.27
<i>Acroloxus</i>	0.66	-0.51	0.46
<i>Adicella</i>	-0.77	0.23	0.28
<i>Agapetus</i>	-0.48	-0.44	0.37
<i>Agraylea</i>	0.60	0.45	0.31
<i>Amphinemura</i>	-1.37	0.04	-0.23
<i>Anax</i>	1.19	0.25	0.45
<i>Ancylus</i>	-0.14	-0.08	0.00
<i>Anodonta</i>	0.73	-0.12	0.20
<i>Aphelocheirus</i>	0.25	0.38	-0.01
<i>Athripsodes</i>	0.23	-0.11	0.15
<i>Atyaephyra</i>	1.34	0.18	0.21
<i>Baetis</i>	-0.27	0.01	-0.21
<i>Beraea</i>	-0.75	0.02	-0.20
<i>Beraeodes</i>	0.05	-0.65	0.24
<i>Bithynia</i>	0.64	-0.16	0.26
<i>Boyeria</i>	-0.07	0.14	0.22
<i>Brachycentrus</i>	-0.06	0.53	0.13
<i>Brachycercus</i>	0.10	0.34	-0.48
<i>Brachyptera</i>	-1.12	-0.27	-0.52
<i>Brychius</i>	-0.19	-0.61	0.99
<i>Bythinella</i>	-0.88	-0.08	0.19
<i>Bythiospeum</i>	0.27	-0.65	1.13
<i>Caenis</i>	0.43	0.17	0.17
<i>Calopteryx</i>	0.03	-0.06	-0.11
<i>Centroptilum</i>	0.26	-0.08	0.30
<i>Ceraclea</i>	0.47	0.31	0.18
<i>Chalcolestes</i>	0.73	-0.68	0.83
<i>Cheumatopsyche</i>	-0.02	0.53	0.01
<i>Chimarra</i>	-0.05	0.77	0.22
<i>Chloroperla</i>	-1.16	0.35	-0.03
<i>Choroterpes</i>	0.47	0.82	0.01
<i>Cloeon</i>	0.84	-0.36	-0.16
<i>Corbicula</i>	0.86	0.71	0.16
<i>Cordulegaster</i>	-0.99	-0.32	-0.20
<i>Corophium</i>	1.03	1.09	-0.12
<i>Crangonyx</i>	0.73	0.48	-0.11
<i>Crunoecia</i>	-1.43	0.29	0.11
<i>Cyphon</i>	-1.19	0.49	0.26
<i>Cyrnus</i>	0.32	-0.02	-0.01
<i>Dikerogammarus</i>	1.23	0.96	-0.22
<i>Dinocras</i>	-1.29	0.46	0.18

<i>Dreissena</i>	1.69	0.08	0.26
<i>Dryops</i>	-0.20	-0.20	-0.30
<i>Dupophilus</i>	-0.84	0.22	0.02
<i>Ecdyonurus</i>	-0.69	0.29	-0.25
<i>Echinogammarus</i>	0.25	-0.35	0.70
<i>Ecnomus</i>	1.35	0.22	0.14
<i>Electrogena</i>	-0.26	0.21	-0.44
<i>Elmis</i>	-0.42	-0.18	0.26
<i>Epeorus</i>	-1.15	0.41	-0.01
<i>Ephemera</i>	-0.17	-0.22	0.15
<i>Ephemerella</i>	-0.32	0.05	-0.03
<i>Ephoron</i>	0.59	0.78	-0.02
<i>Epitheca</i>	0.39	0.08	-0.48
<i>Esolus</i>	0.00	0.23	0.31
<i>Euleuctra</i>	-0.16	0.35	0.24
<i>Ferrissia</i>	0.94	0.07	0.02
<i>Galba</i>	0.21	-0.47	0.30
<i>Gammarus</i>	-0.15	-0.51	-0.26
<i>Gerris</i>	0.02	-0.22	-0.16
<i>Glossosoma</i>	-1.16	0.29	0.18
<i>Goera</i>	0.07	-0.23	0.09
<i>Gomphus</i>	0.15	0.39	-0.18
<i>Gyrinus</i>	0.21	-0.55	0.17
<i>Habroleptoides</i>	-0.97	0.29	-0.19
<i>Habrophlebia</i>	-0.84	-0.10	-0.42
<i>Haliplus</i>	0.64	-0.55	0.27
<i>Helodes</i>	-0.66	-0.74	-0.13
<i>Helophorus</i>	-0.05	-0.54	-0.31
<i>Heptagenia</i>	0.15	0.52	-0.28
<i>Holocentropus</i>	-0.18	0.21	0.09
<i>Hydraena</i>	-0.71	0.00	-0.15
<i>Hydrochus</i>	0.04	-0.04	-0.24
<i>Hydrocyphon</i>	-1.02	0.32	-0.08
<i>Hydrometra</i>	0.12	-0.21	-0.31
<i>Hydropsyche</i>	-0.09	0.13	-0.11
<i>Hydroptila</i>	0.23	-0.08	-0.16
<i>Isoperla</i>	-1.14	0.15	-0.15
<i>Ithytrichia</i>	-0.08	0.11	0.32
<i>Lasiocephala</i>	-0.62	-0.04	0.53
<i>Lepidostoma</i>	-0.18	0.03	0.62
<i>Leptocerus</i>	0.57	0.29	0.55
<i>Leptophlebia</i>	-0.63	0.09	0.22
<i>Leuctra</i>	-0.66	0.25	-0.17
<i>Libellula</i>	0.48	-1.25	0.14
<i>Limnebius</i>	-1.01	0.26	0.12

<i>Limnius</i>	-0.30	-0.10	0.30
<i>Lithax</i>	-0.84	0.24	-0.20
<i>Lymnaea</i>	0.74	-0.39	0.90
<i>Lype</i>	0.09	-0.28	0.25
<i>Macronychus</i>	0.38	0.41	0.31
<i>Mesovelgia</i>	0.11	-0.38	-0.05
<i>Metalype</i>	-0.06	-0.07	0.43
<i>Micrasema</i>	-1.04	0.44	0.01
<i>Micronecta</i>	0.31	0.12	-0.17
<i>Molanna</i>	0.52	-0.60	0.74
<i>Molannodes</i>	0.66	-0.48	0.81
<i>Mystacides</i>	0.14	0.04	0.16
<i>Nemoura</i>	-0.87	0.07	-0.47
<i>Neureclipsis</i>	0.98	0.06	-0.35
<i>Niphargus</i>	-0.15	-0.11	0.46
<i>Normandia</i>	0.36	0.28	0.84
<i>Notidobia</i>	0.05	-0.48	0.40
<i>Ochthebius</i>	-0.29	-0.21	-0.15
<i>Odontocerum</i>	-1.31	0.16	0.26
<i>Oecetis</i>	0.08	0.44	0.20
<i>Oecismus</i>	-1.02	0.37	0.47
<i>Oligoneuriella</i>	-0.29	0.89	0.06
<i>Oligoplectrum</i>	-0.26	0.78	0.01
<i>Onychogomphus</i>	0.07	0.47	-0.01
<i>Ophiogomphus</i>	0.24	0.42	0.08
<i>Orconectes</i>	0.48	0.18	-0.05
<i>Orectochilus</i>	-0.14	0.26	0.06
<i>Orthetrum</i>	1.03	0.33	0.17
<i>Orthotrichia</i>	0.92	0.59	0.34
<i>Oulimnius</i>	0.07	-0.13	0.20
<i>Oxyethira</i>	-0.48	-0.02	0.46
<i>Oxygastra</i>	0.54	0.65	0.39
<i>Pacifastacus</i>	-0.47	0.14	-0.71
<i>Paraleptophlebia</i>	-0.39	0.16	-0.33
<i>Peltodytes</i>	0.60	0.27	0.33
<i>Perla</i>	-1.00	0.45	-0.04
<i>Perlodes</i>	-1.09	0.42	0.01
<i>Philopotamus</i>	-1.53	0.20	0.02
<i>Phryganea</i>	0.83	-0.55	0.16
<i>Physa</i>	0.79	0.01	0.02
<i>Physella</i>	0.73	-0.63	-0.54
<i>Pisidium</i>	0.19	-0.64	0.16
<i>Platycnemis</i>	0.57	0.08	0.06
<i>Plea</i>	0.45	-1.43	-0.92
<i>Plectrocnemia</i>	-0.88	-0.46	-0.42

<i>Polycentropus</i>	-0.06	0.05	-0.01
<i>Pomatinus</i>	-0.04	0.15	0.15
<i>Potamanthus</i>	0.56	0.79	0.04
<i>Potamophilus</i>	0.69	0.84	0.20
<i>Potamopyrgus</i>	0.27	-0.48	0.23
<i>Potomida</i>	0.93	0.61	0.34
<i>Procloeon</i>	0.43	0.34	-0.01
<i>Protonemura</i>	-1.33	0.18	-0.18
<i>Pseudanodonta</i>	0.32	-0.06	0.39
<i>Pseudocentroptilum</i>	-0.32	0.28	-0.25
<i>Psychomyia</i>	0.11	0.59	-0.19
<i>Radix</i>	0.34	-0.31	0.35
<i>Raptobaetopus</i>	0.73	0.97	0.09
<i>Rhithrogena</i>	-1.08	0.15	-0.14
<i>Rhyacophila</i>	-0.59	0.06	-0.11
<i>Riolus</i>	0.23	-0.47	0.75
<i>Sericostoma</i>	-0.78	-0.20	0.34
<i>Setodes</i>	0.26	0.69	0.03
<i>Silo</i>	-0.59	-0.22	0.17
<i>Siphonurus</i>	-0.73	-0.59	-0.67
<i>Siphonoperla</i>	-1.26	0.31	-0.14
<i>Somatochlora</i>	0.43	-0.27	-0.49
<i>Sphaerium</i>	0.42	-0.46	0.16
<i>Stagnicola</i>	0.27	-1.00	0.98
<i>Stenelmis</i>	0.19	0.36	0.49
<i>Taeniopteryx</i>	-0.56	0.57	0.11
<i>Theodoxus</i>	0.36	-0.03	0.80
<i>Thraulius</i>	-0.19	0.80	0.32
<i>Thremma</i>	-1.91	0.45	0.02
<i>Tinodes</i>	0.29	-0.39	0.09
<i>Torleya</i>	-0.73	0.39	0.16
<i>Triaenodes</i>	0.37	0.60	0.41
<i>Unio</i>	0.40	-0.06	0.25
<i>Valvata</i>	0.46	-0.17	0.50
<i>Viviparus</i>	-0.36	-1.89	3.03
<i>Wormaldia</i>	-1.06	0.27	0.17
<i>Xanthoperla</i>	0.76	1.07	0.04
<i>Ylodes</i>	0.24	0.60	0.39

C. Procédure de jointure des réseaux T-NET et RHT

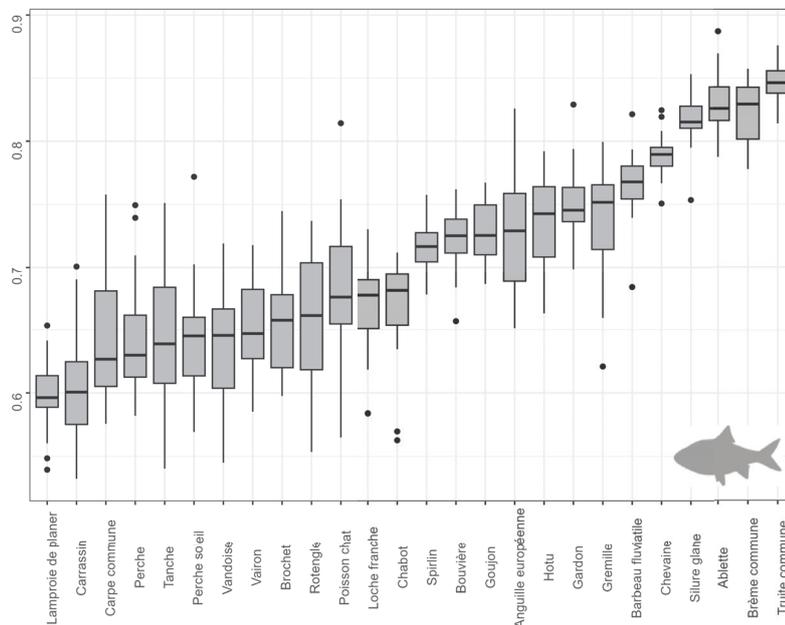
La probabilité d'intermittence des cours d'eau est disponible à l'échelle du réseau RHT. Il est conçu à partir d'un modèle numérique de terrain qui permet de construire un réseau hydrographique en déterminant les sections et direction où se produisent les écoulements à l'échelle de la France métropolitaine (Pella *et al.*, 2012). Il s'agit donc d'un réseau théorique, basé sur la jointure entre un premier réseau théorique construit à partir de la Base de Données sur la CARTographie THématique des AGences de l'eau et du ministère chargé de l'environnement (BD CARTHAGE) et un second réseau étendu pour orienter les écoulements (RHE). Il est constitué de 114 601 tronçons et 283 639 km linéaire de réseau à l'échelle nationale ; sur le bassin de la Loire cela représente 25 843 tronçons et 88 561 km linéaire de réseau hydrographique (Pella *et al.*, 2012).

Afin de pouvoir mobiliser la probabilité d'intermittence des cours d'eau sur le bassin de la Loire, un travail de correspondance entre les deux réseaux RHT et T-NET a dû être réalisé au préalable. La correspondance entre les réseaux hydrographiques T-NET et RHT a été réalisée en deux étapes : (i) une jointure spatiale entre les deux couches via des outils SIG, (ii) une vérification à la main des jointures sous certaines conditions avec ré-attribution à la main quand cela était possible.

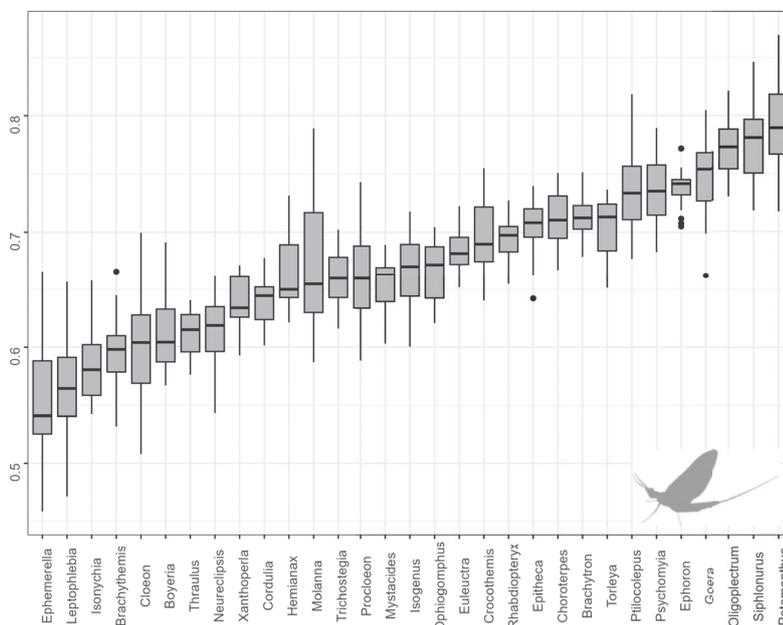
Trois points pris sur chaque tronçon T-NET ont été projetés sur le tronçon RHT le plus proche. Lorsque ces trois points ont été projetés sur le même tronçon, on considère que la correspondance est exacte à condition que la distance de projection soit inférieure à 100 m. Cela a permis de faire correspondre directement 25 541 tronçons entre eux. Lorsque la distance de projection est supérieure à 100 m et/ou que les points sont projetés sur différents tronçons RHT différentes conditions ont été vérifiées. Lorsque l'un des points étaient projetés sur un tronçon RHT à moins de 50 m et que les ordres de Strahler et les modules correspondaient entre les deux réseaux, c'est le tronçon RHT le plus proche qui a été conservé. Parmi les projections restantes, celles dont le ratio des modules de T-NET et du RHT sont comprises entre 0,5 et 2 sont également considérées comme valides. Les 105 autres tronçons projetés ne remplissant pas cette condition ont été vérifiés à la main et ont soit été ré-attribués au tronçon correspondant à la main soit la jointure initiale a été conservée. Un total de 18 566 tronçons T-NET n'ont pas trouvés de correspondance RHT.

C'est cette jointure qui est à l'origine du réseau H, simplification du réseau T-NET, et qui est utilisé comme base géographique d'étude.

D. Boxplot des performances des modèles de la distribution des classes d'abondance pour les espèces de poissons et les genres de macroinvertébrés



(a)

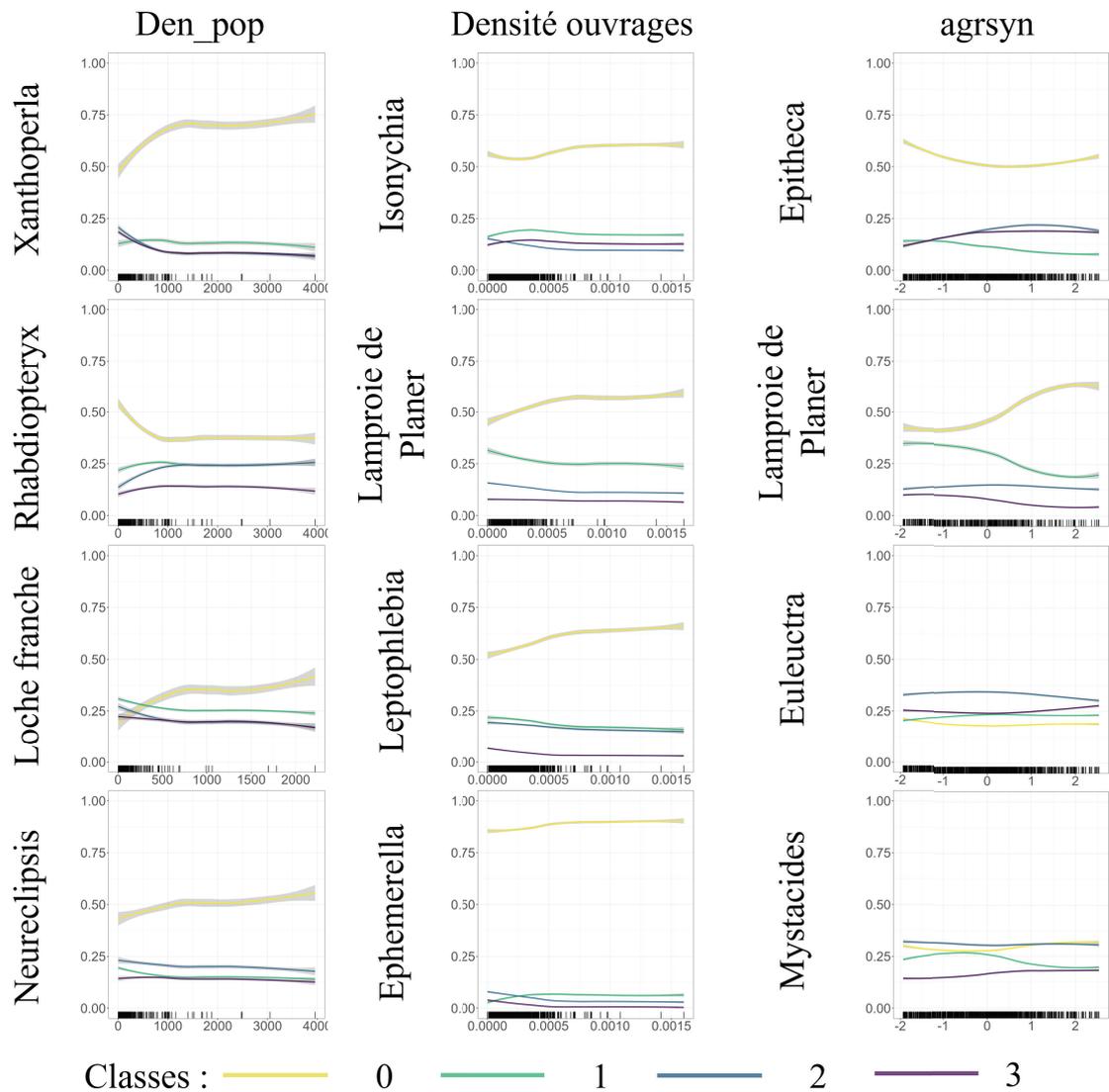


(b)

Figure

B.1. Boxplot des performances selon l'AUC multi-classes des modèles OF avec transformation SCUT des données d'entrées pour les modèles poissons (en haut) et pour les macroinvertébrés (en bas). Les valeurs d'AUC multi-classe ont été calculées pour les 20 répliquats de modèle. Les taxons sont ordonnés en fonction de la valeur médiane de leur AUC.

E. Exemples supplémentaires de *Partial Dependence Plot*



Figure

C.1. Exemples supplémentaires de *Partial Dependence Plot* pour les prédicteurs densité de population (den_pop), densité d'ouvrages (Densité ouvrages) et gradient agricole (agsyn).

F. Rang des prédicteurs avant intégration des indicateurs de ressources macroinvertébrés dans les modèles basés sur le jeu de données communs

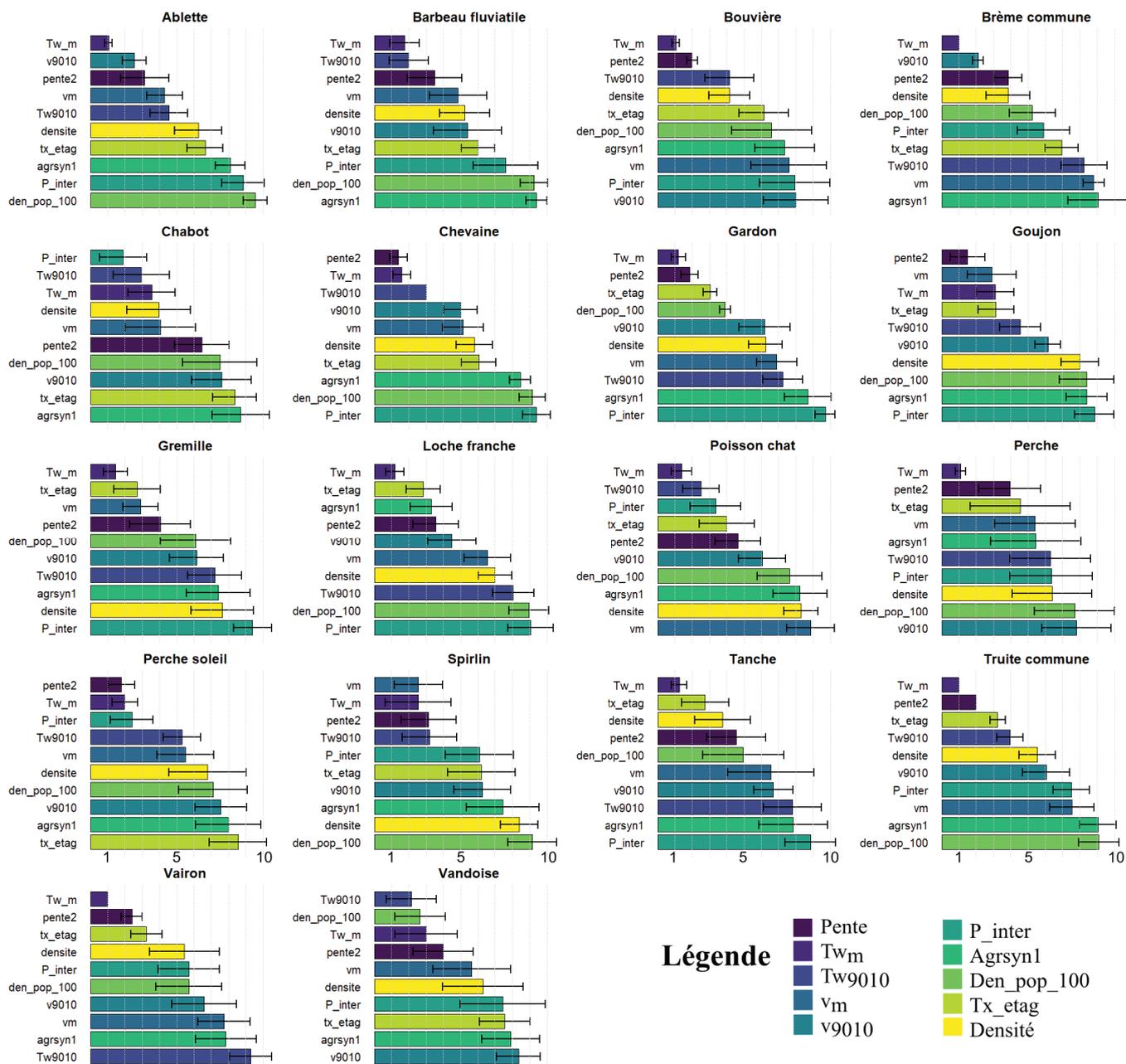


FIGURE 84

Rang des prédicteurs dans les modèles naïfs poissons, avant intégration des indicateurs de ressources macroinvertébrés. La correspondance avec les codes prédicteurs peut être consultée au tableau 9

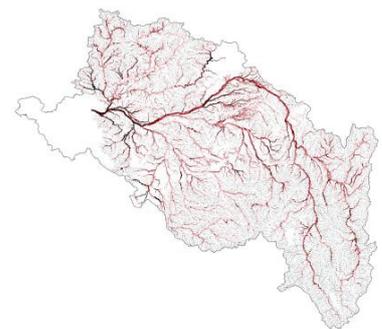
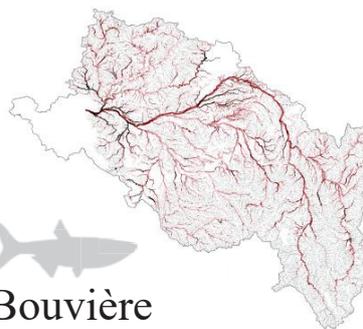
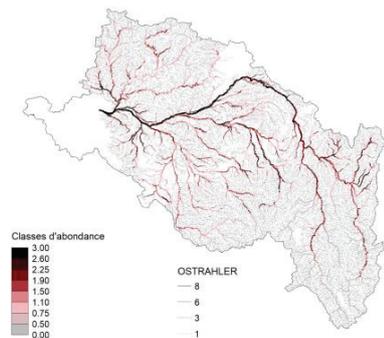
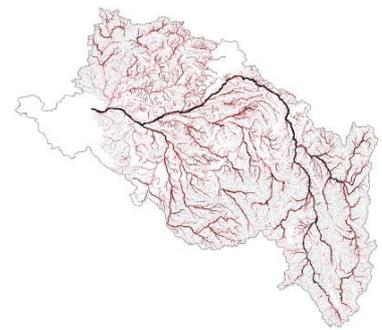
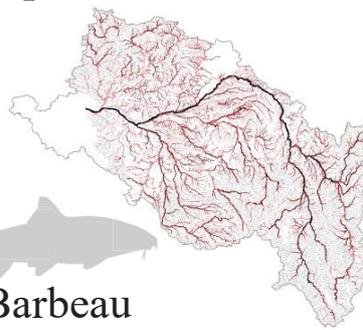
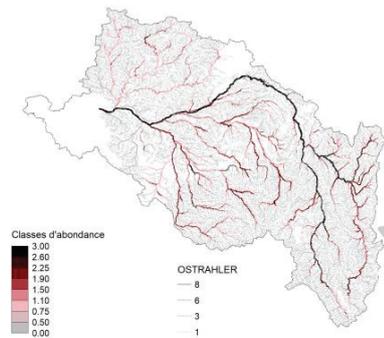
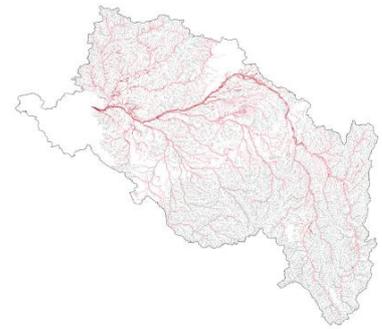
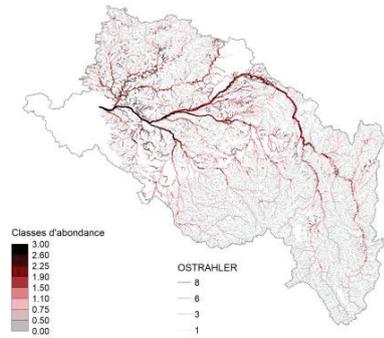
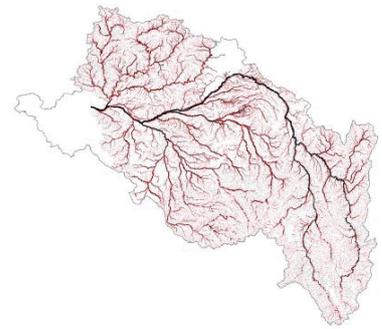
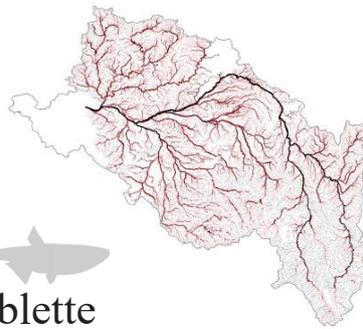
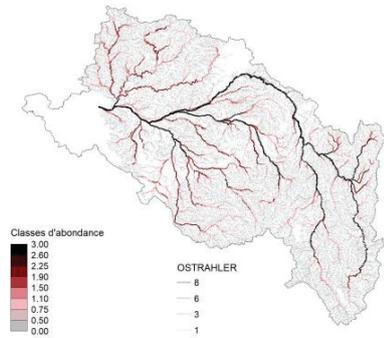
G. Cartographies des projections futures pour les poissons (n = 26 espèces) et les macroinvertébrés (n = 31 genres)

Cette annexe présente les cartographies des projections de chaque SDM de classes d'abondance pour chaque taxons aux trois horizons temporels, 2010, 2050 et 2080. Les classes d'abondances présentées sont la moyenne des classes projetées pour les 20 répliquats de modèles. Les projections sont d'abord présentées pour les 26 espèces de poissons modélisées, puis pour les 31 genres de macroinvertébrés, par ordre alphabétique au sein de chaque.

H2010

H2050

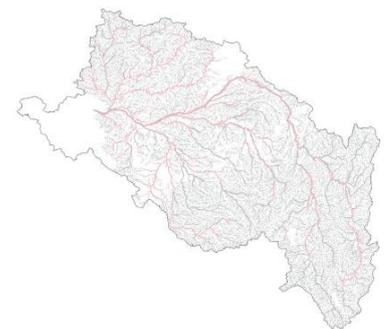
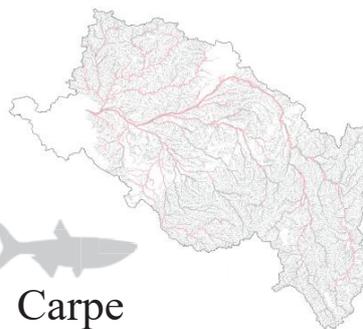
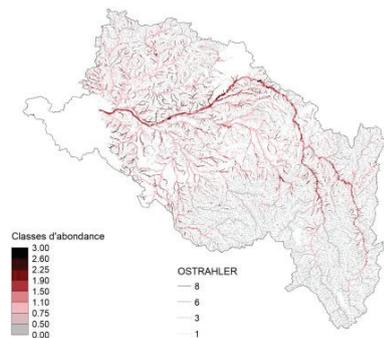
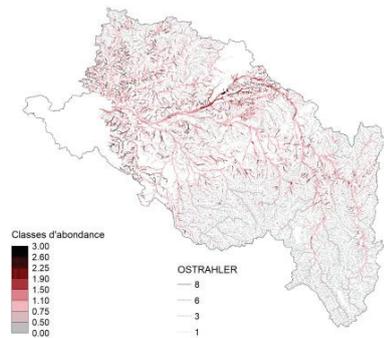
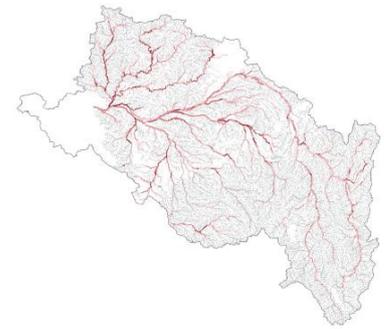
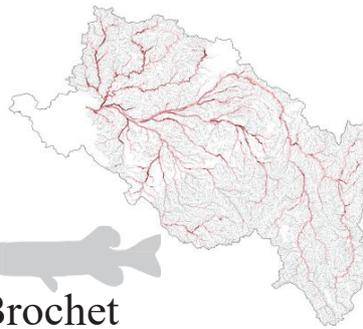
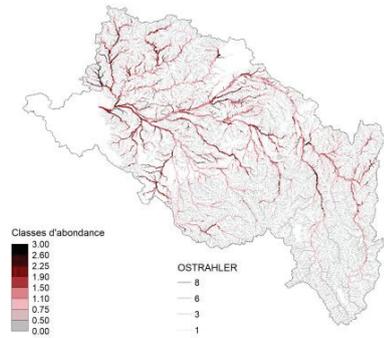
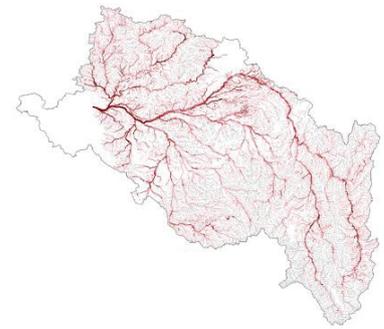
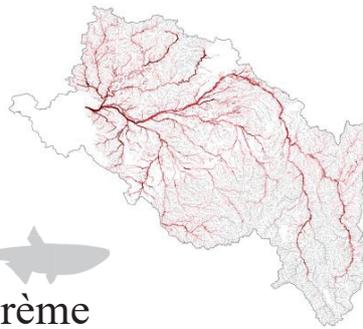
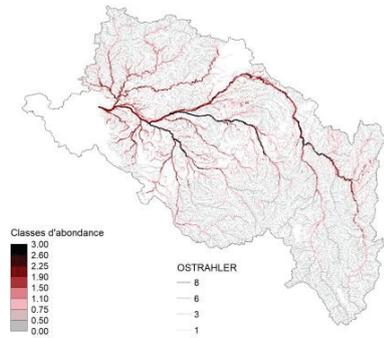
H2080



H2010

H2050

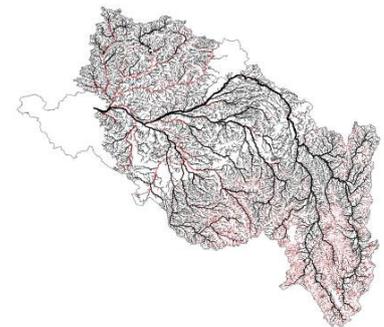
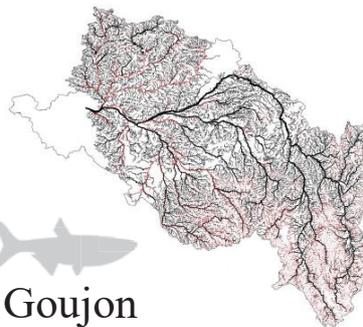
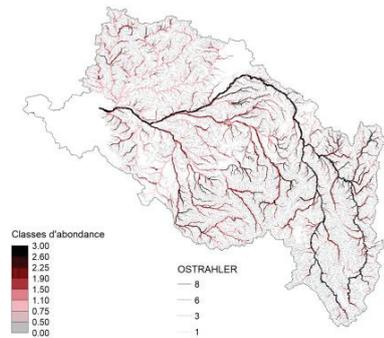
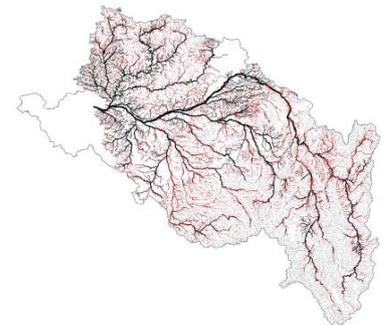
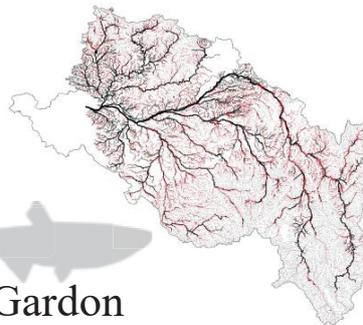
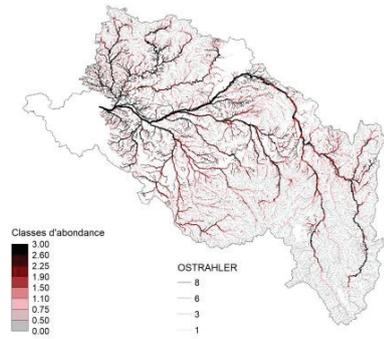
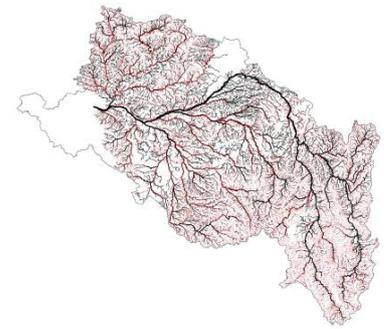
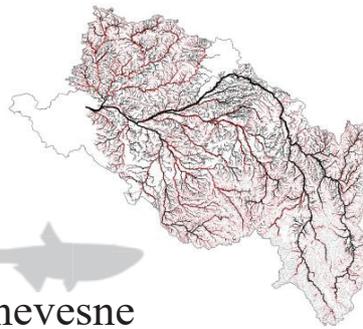
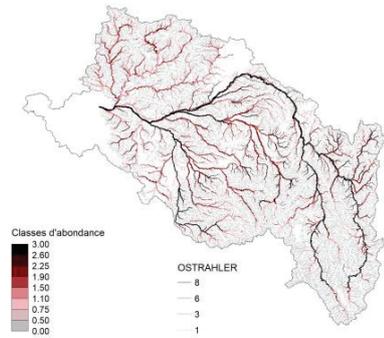
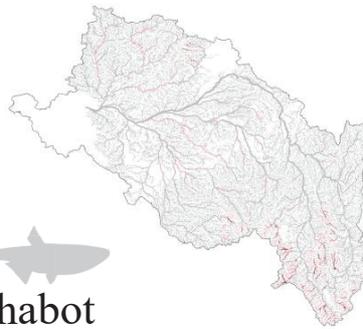
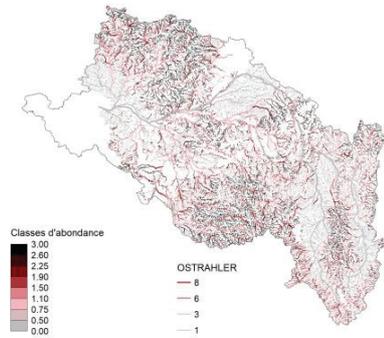
H2080



H2010

H2050

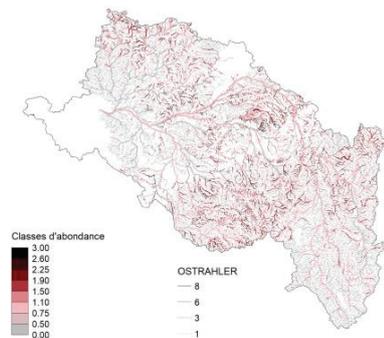
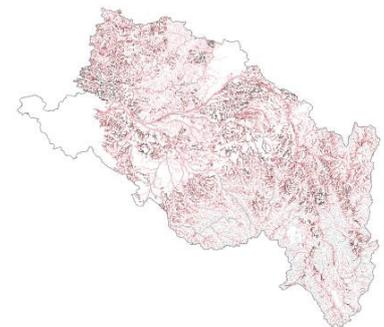
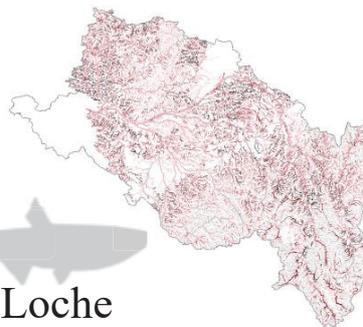
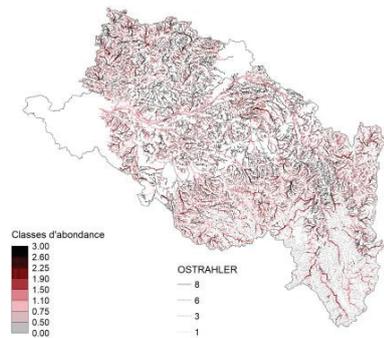
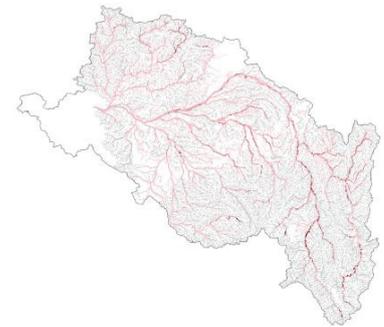
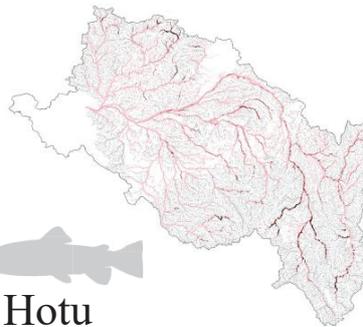
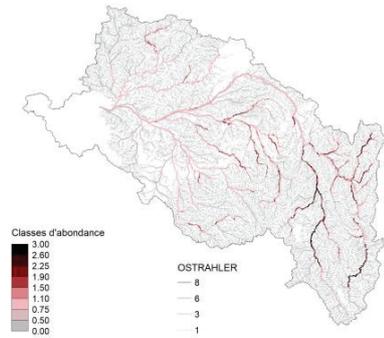
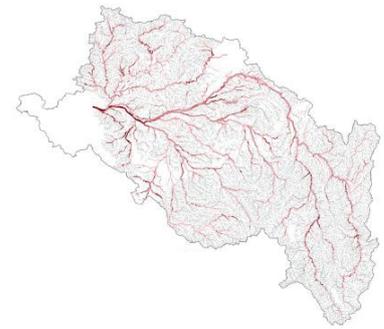
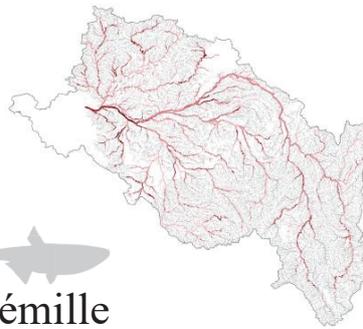
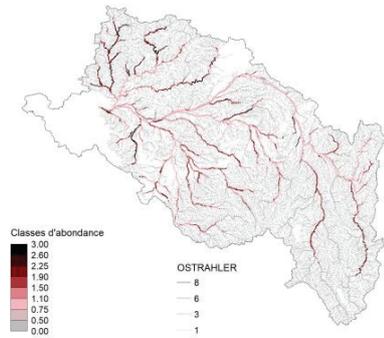
H2080



H2010

H2050

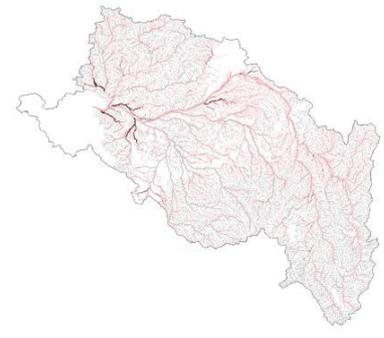
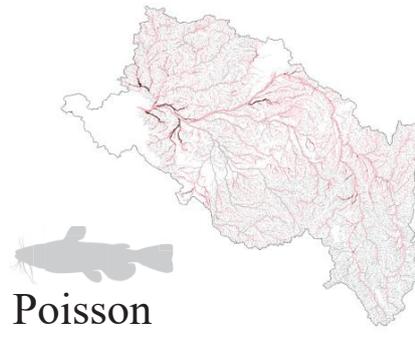
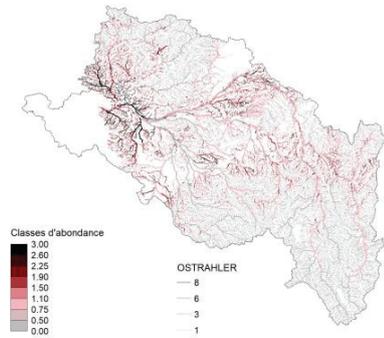
H2080



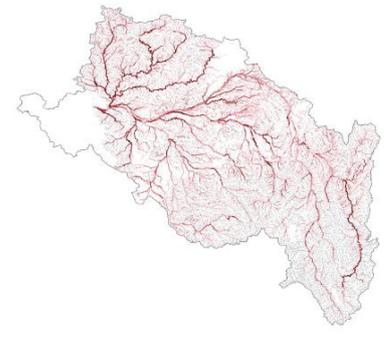
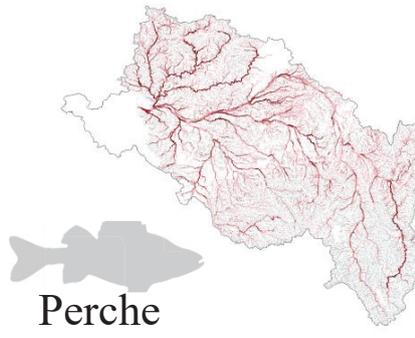
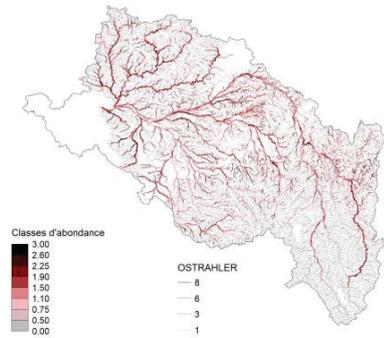
H2010

H2050

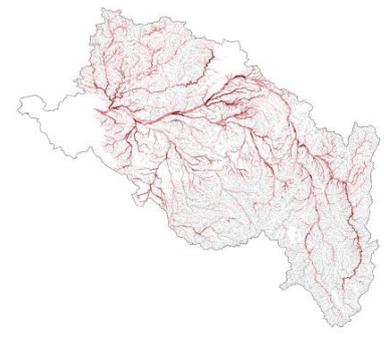
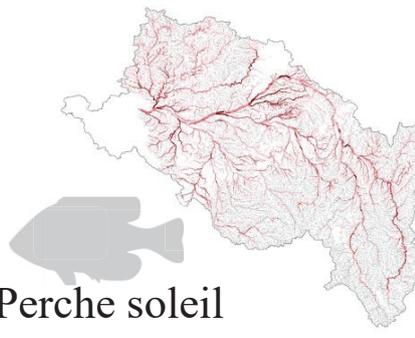
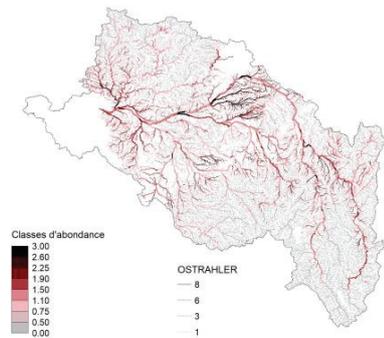
H2080



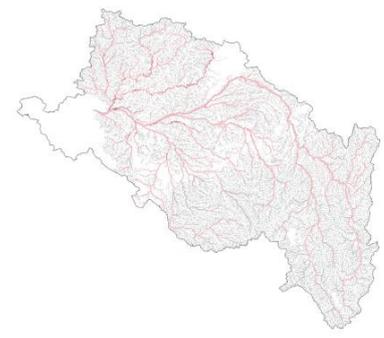
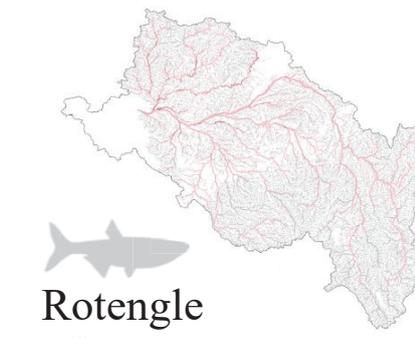
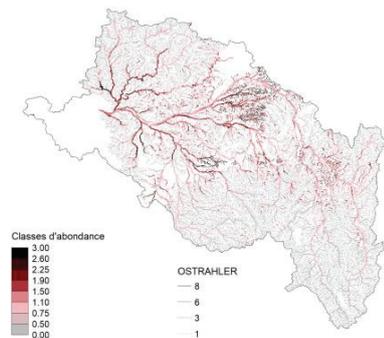

**Poisson
chat**




Perche




Perche soleil

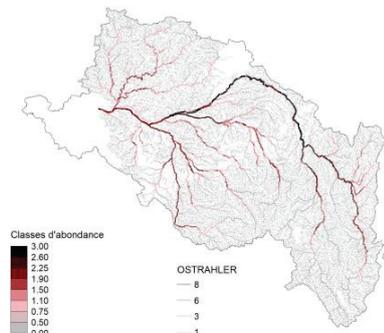



Rotengle

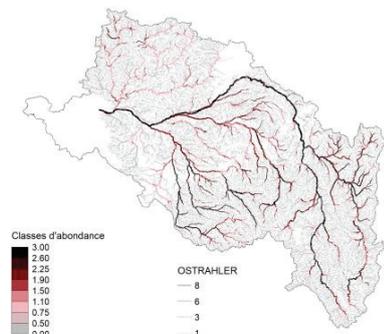
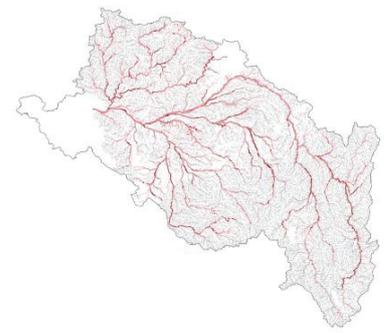
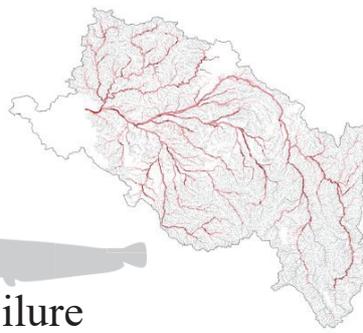
H2010

H2050

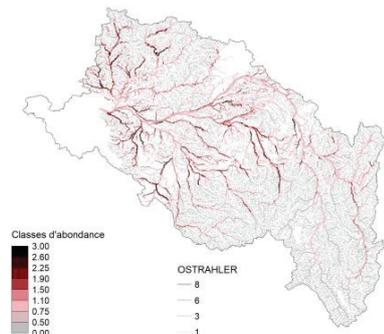
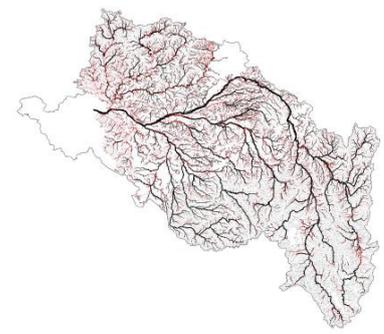
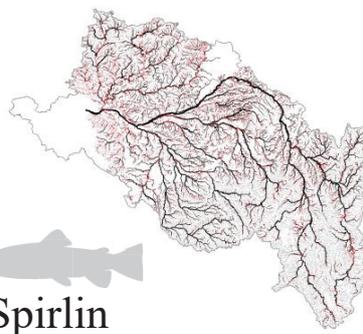
H2080



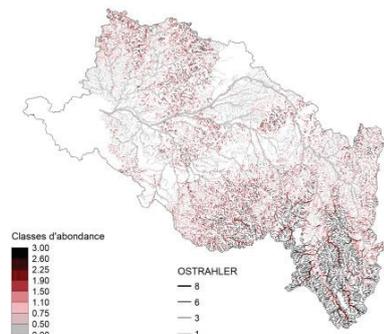
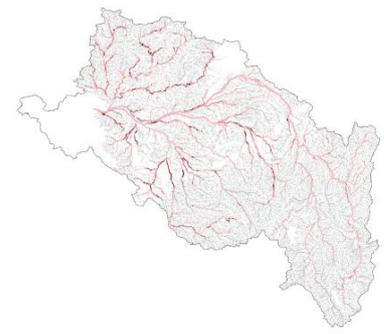
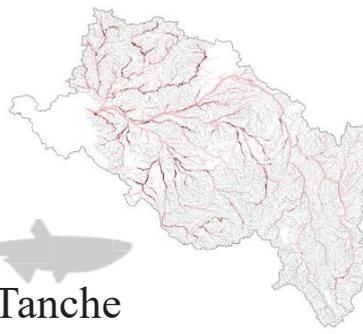
**Silure
glane**



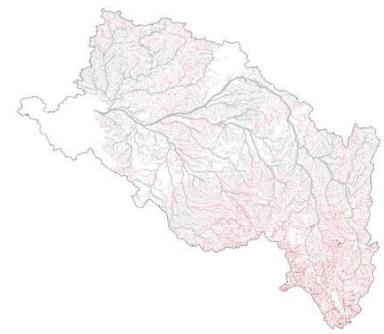
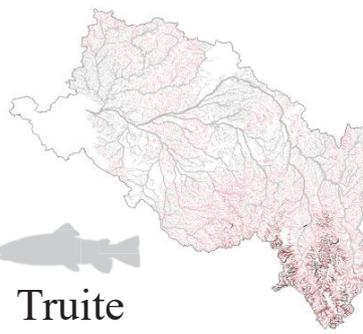
Spirilin



Tanche



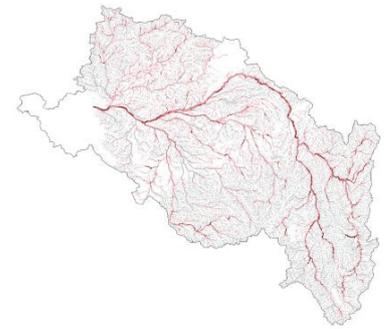
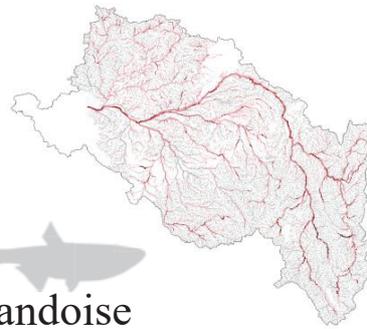
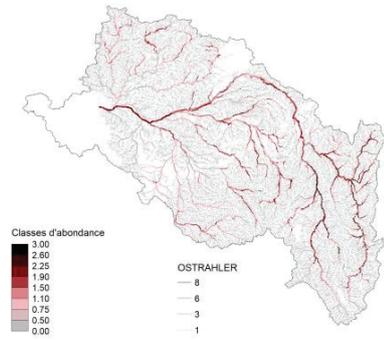
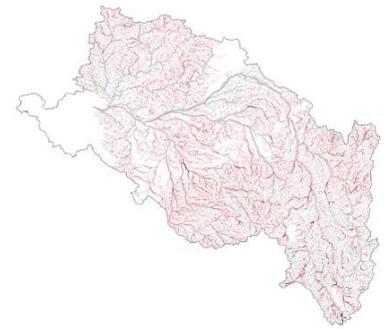
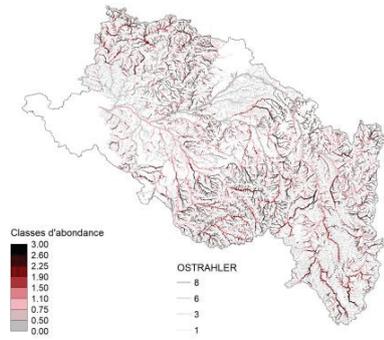
**Truite
commune**



H2010

H2050

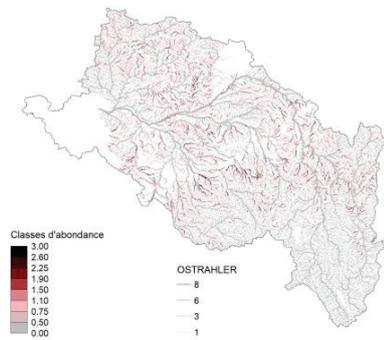
H2080



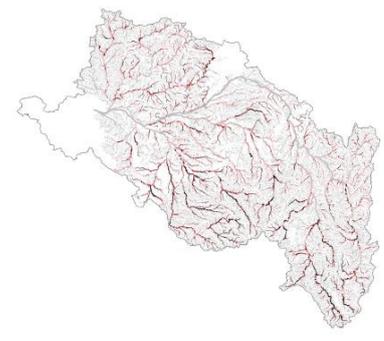
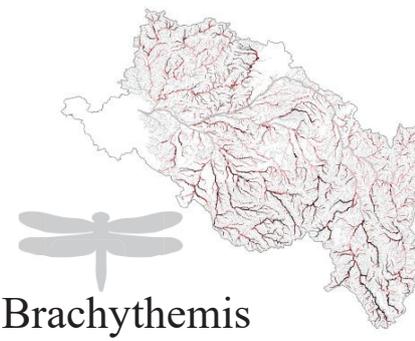
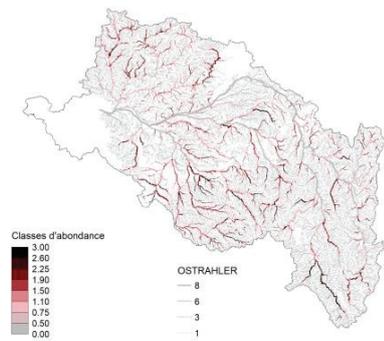
H2010

H2050

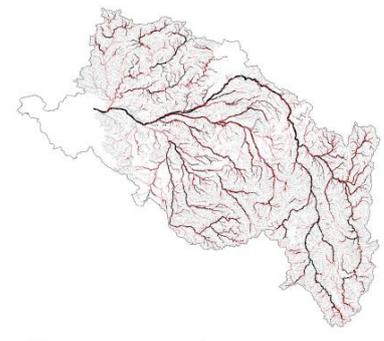
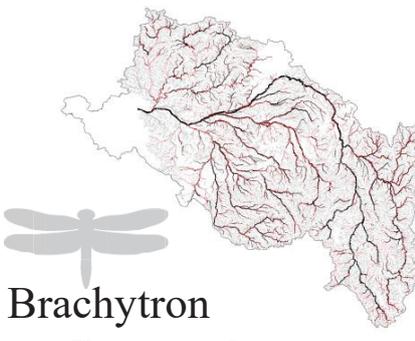
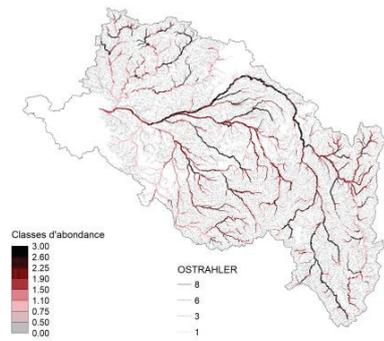
H2080



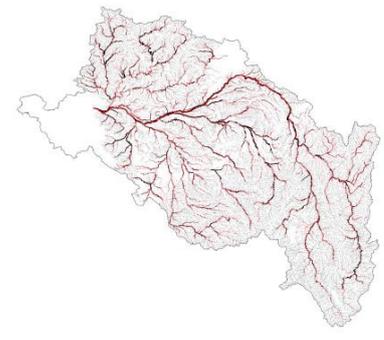
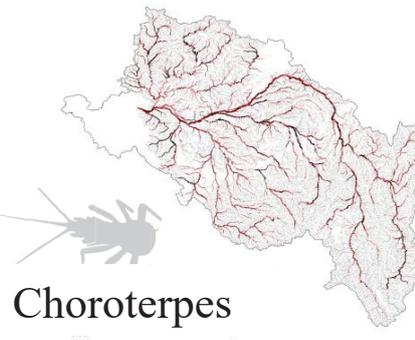
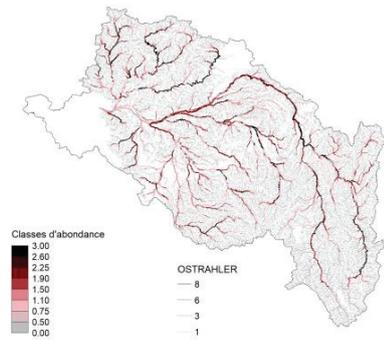
Boyeria



Brachythemis



Brachytron

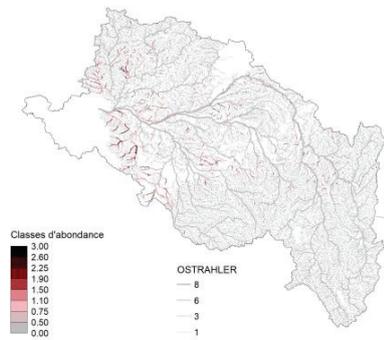


Choroterpes

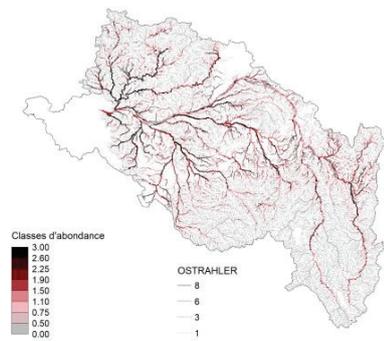
H2010

H2050

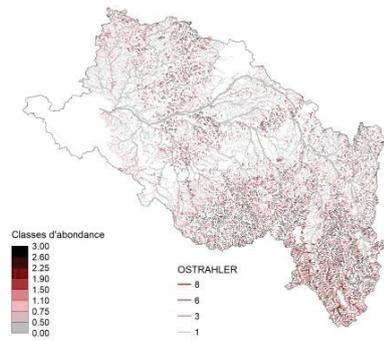
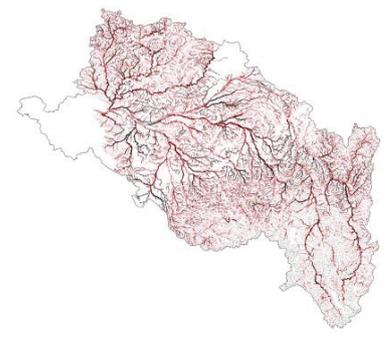
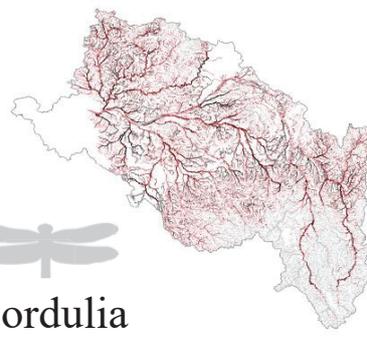
H2080



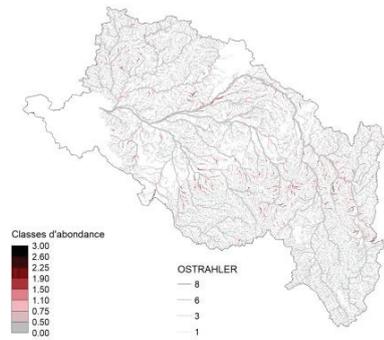
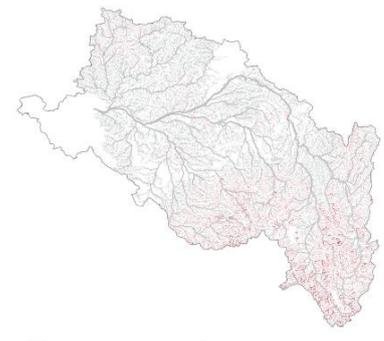
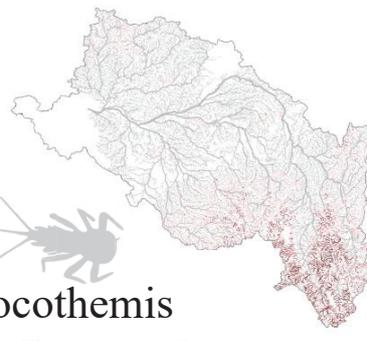
Cloeon



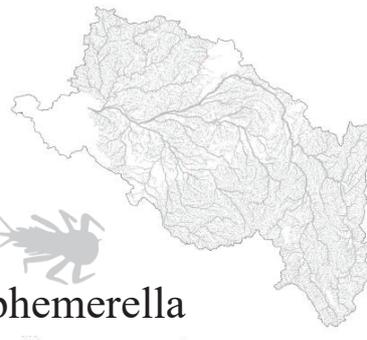
Cordulia



Crocothemis



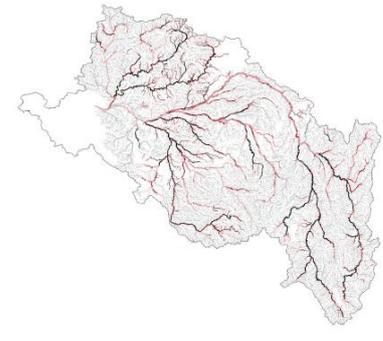
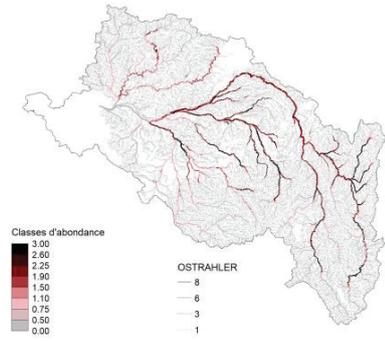
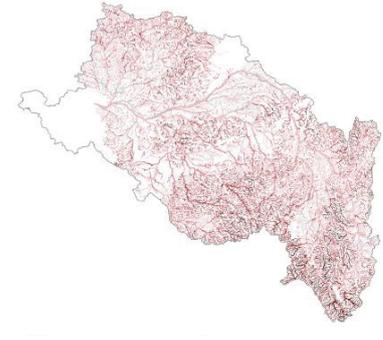
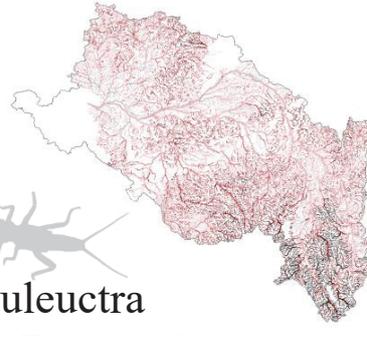
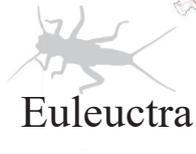
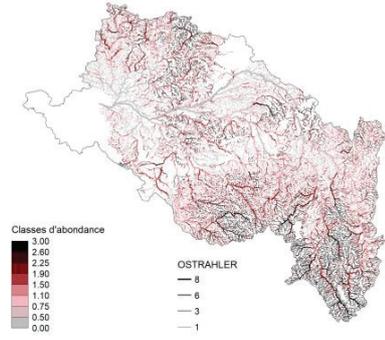
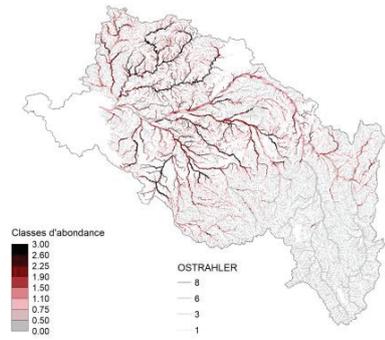
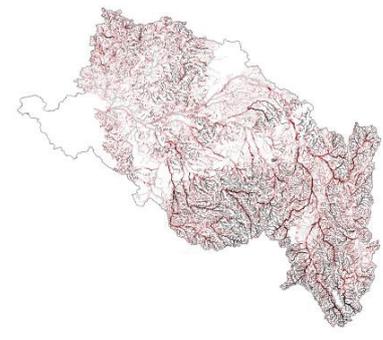
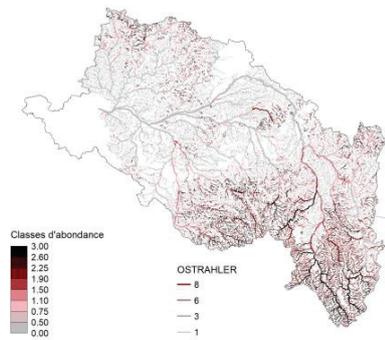
Ephemerella



H2010

H2050

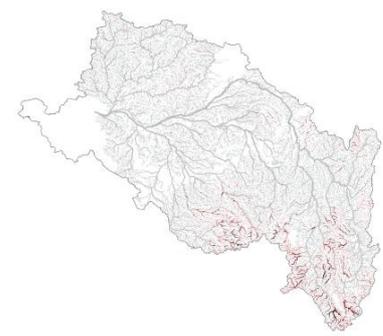
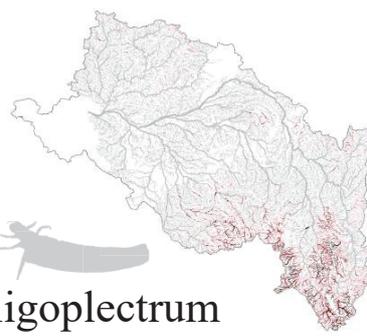
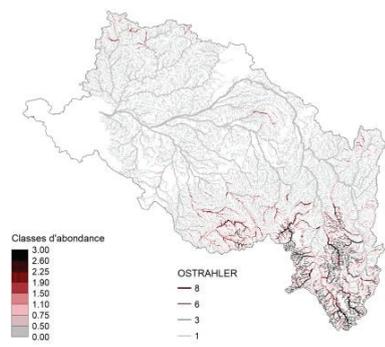
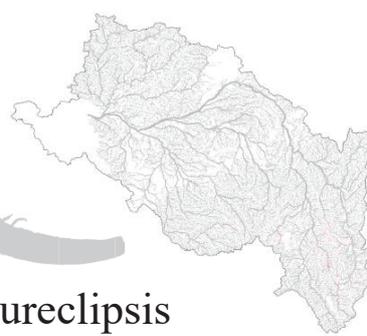
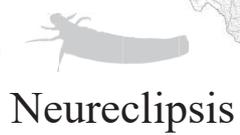
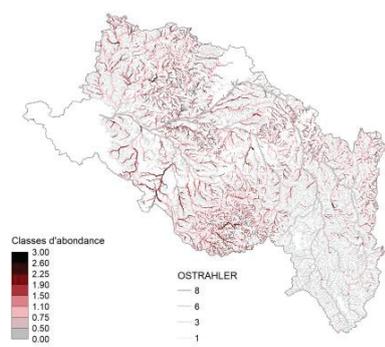
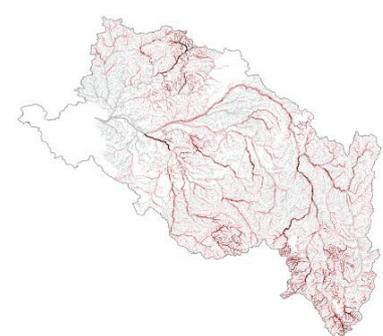
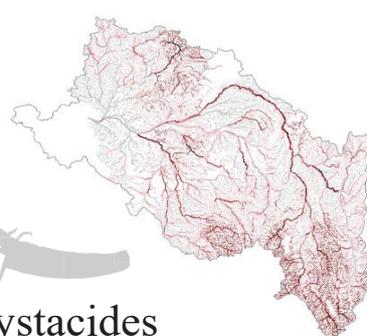
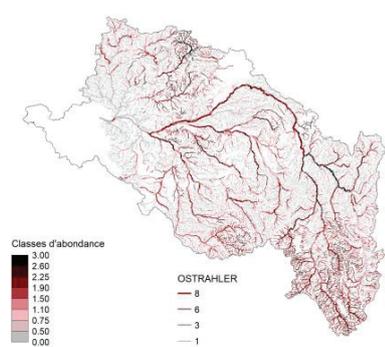
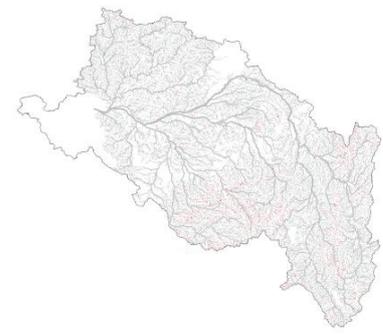
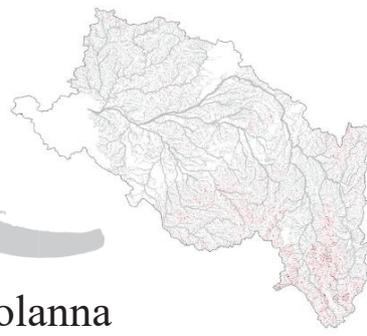
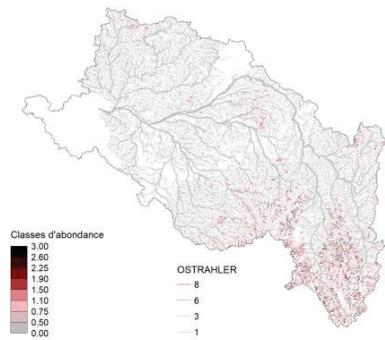
H2080



H2010

H2050

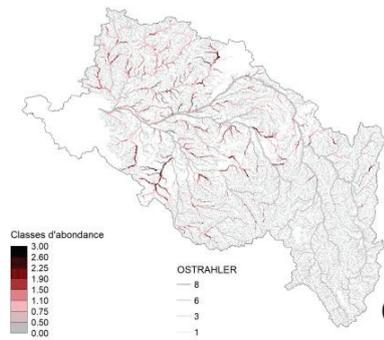
H2080



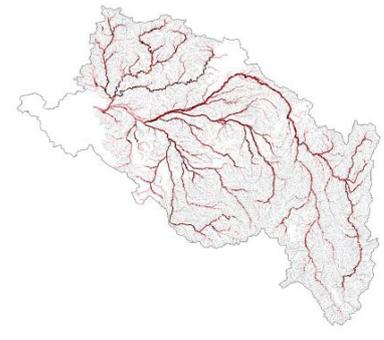
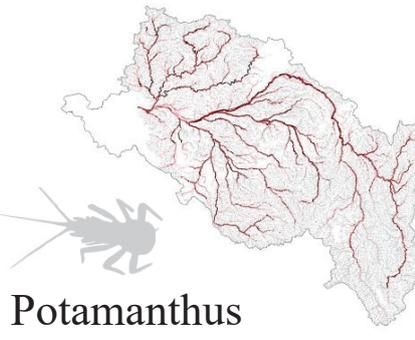
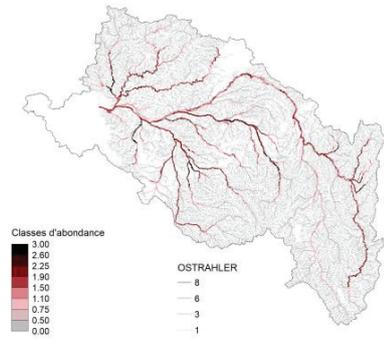
H2010

H2050

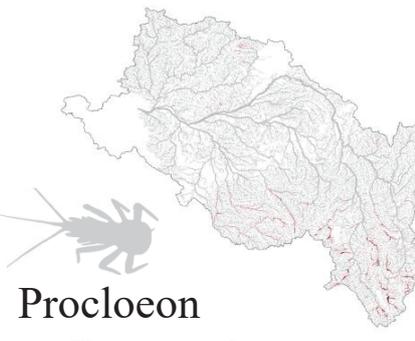
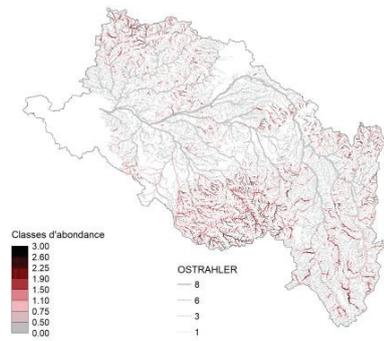
H2080



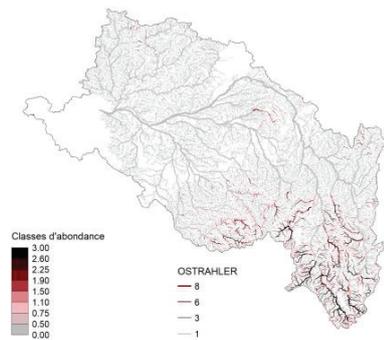

Ophiogomphus




Potamanthus




Procloeon

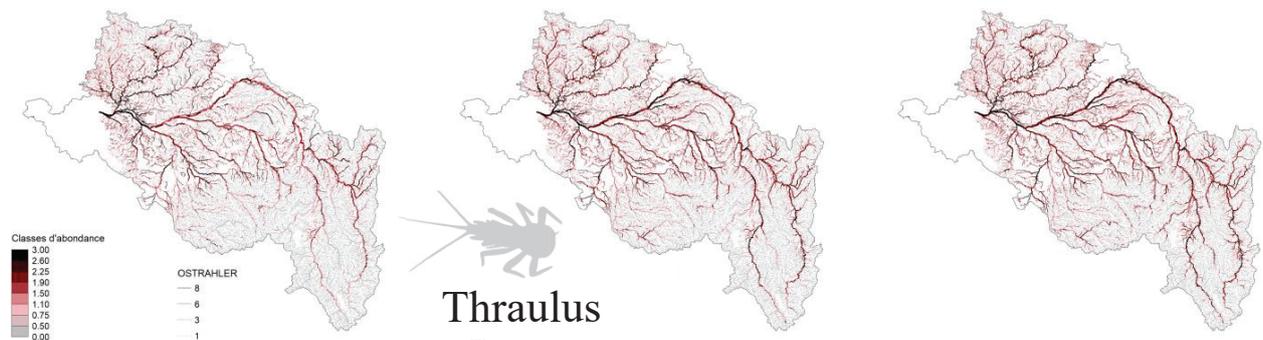
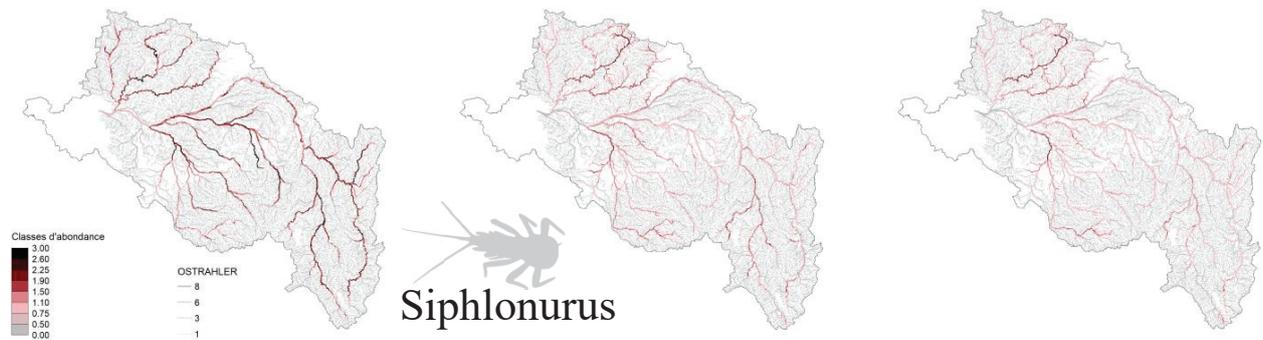
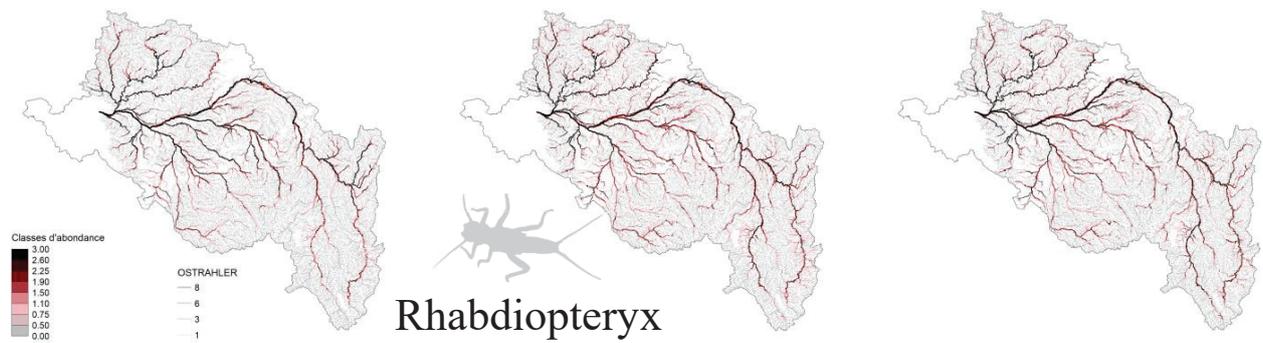
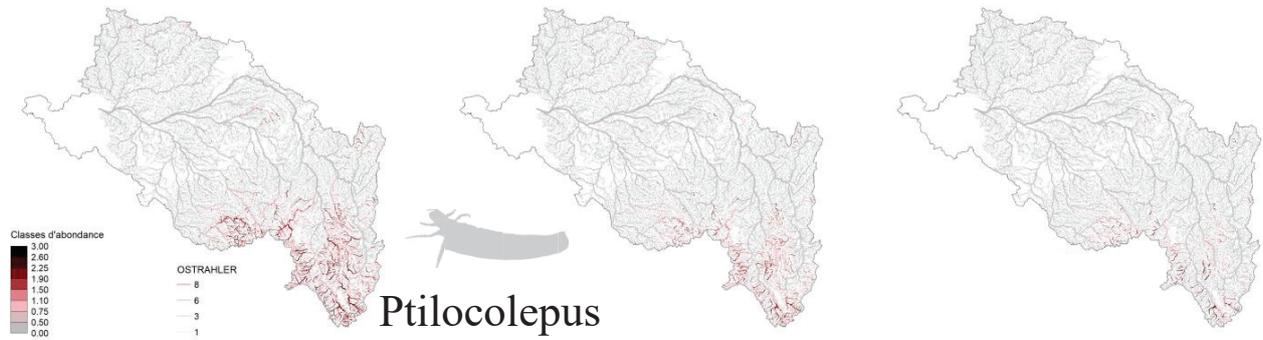



Psychomyia

H2010

H2050

H2080



H2010

H2050

H2080

