



**HAL**  
open science

# Etude des réponses des assemblages de poissons aux variations de l'environnement par modélisation hiérarchique bayésienne : Application aux juvéniles de cyprinidés du Haut-Rhône

Jeremy Piffady

► **To cite this version:**

Jeremy Piffady. Etude des réponses des assemblages de poissons aux variations de l'environnement par modélisation hiérarchique bayésienne : Application aux juvéniles de cyprinidés du Haut-Rhône. Ecologie, Environnement. AgroParisTech, 2010. Français. NNT : 2010AGPT0061 . pastel-00566444

**HAL Id: pastel-00566444**

**<https://pastel.hal.science/pastel-00566444>**

Submitted on 16 Feb 2011

**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.



## Doctorat ParisTech

# THÈSE

pour obtenir le grade de docteur délivré par

## L'Institut des Sciences et Industries du Vivant et de l'Environnement (AgroParisTech)

**Spécialité : Biostatistiques**

*présentée et soutenue publiquement par*

**Jérémy PIFFADY**

le 05 novembre 2010

## **Etude des réponses des assemblages de poissons aux variations de l'environnement par modélisation hiérarchique bayésienne: Application aux juvéniles de cyprinidés du Haut-Rhône**

Directeur de thèse : **Yves SOUCHON**

Directeur de thèse : **Eric PARENT**

### **Jury**

**M. Marco RODRIGUEZ**, Professeur, GREA, Université de Québec aux Trois Rivières

**M. Olivier GIMENEZ**, Docteur, CEFE UMR 5175,

**M. Etienne PREVOST**, Docteur, UMR INRA/UPPA Ecobiop

**M. Christian DERQUENNE**, Docteur, EDF R&D

**M. Eric PARENT**, Professeur, MORSE, AgroParisTech Engref

**M. Yves SOUCHON**, Professeur, LHQ, Cemagref

Rapporteur

Rapporteur

Examineur

Examineur

Directeur

Directeur



## Remerciements

Cette page sera sans doute la seule que certaines personnes prendront le temps de lire, c'est dire son importance... Pas de méthode miracle, pas de résultat extraordinaire ni de résumé de ce qui a été fait ou reste à faire... non, juste une pensée sincère pour toutes les personnes ayant compté dans la réalisation de ce travail!

Et si finalement, ce n'était pas toutes ces rencontres les plus importantes?

Je tiens à remercier, depuis le bureau d'à côté, Yves Souchon pour avoir cru depuis le départ en mes capacités d'adaptation et d'apprentissage. Merci de m'avoir confié ce sujet et de m'avoir soutenu tout au long de ce parcours. L'aventure au sein du pôle ne fait que commencer!

Je tiens à adresser mes plus sincères pensées et remerciements à Eric Parent, qui depuis le début s'est investi spontanément et dynamiquement dans l'encadrement de cette thèse. A Paris devant le tableau blanc ou en wadders sur les bords du Rhône pour t'imprégner du protocole expérimental, merci pour ton implication, ton soutien, ta confiance et ta patience! Pourvu que ça dure...

Merci également aux membres de mon jury de thèse pour avoir fait du moment important qu'est la soutenance, un échange intéressant, stimulant et très humain! Merci à Marco Rodriguez et Olivier Gimenez d'avoir accepté d'être rapporteurs, merci à Christian Derquenne pour sa bienveillante présidence et merci à Etienne Prévost pour sa franchise (tant en in qu'en off)

Merci aux membres du comité de pilotage d'avoir pris sur leur temps pour m'aiguiller sur les différents problèmes soulevés : merci à Jean-Pierre Raoult, Roger Arditi, Roger Jones, Etienne Rivot (dont le rôle dans la réalisation de cette thèse remonte à bien en amont, merci aussi pour les conseils de rédaction pour les rapports Engref) et à Marie-Pierre Etienne pour leur dynamisme et leur enthousiasme.

La vie au bureau ne serait pas la même sans les discussions quotidiennes, au détour du couloir ou de la salle café... merci à tous de supporter mes avis musicaux tranchés et mes appels discrets à boire le café... merci surtout pour la bonne humeur quotidienne et votre soutien dans les moments où il y en avait besoin (légers stress passagers, problèmes de dos, etc.) : Merci à Bertrand (j'attends toujours le resto sénégalais, je dis ça, je dis rien...), Nolwenn, André, Laurent, Martial, Laurent, Federica, Thierry, Kris, Jean-René, Anaïs, Damien et Souad.

Merci à tous ceux qui se sont investis dans la prise de données depuis de nombreuses années et qui m'ont permis de travailler dessus, mais surtout qui m'ont emmené me tremper les pieds (et parfois le reste) pour me rappeler que l'ordinateur, c'est bien mais que le Rhône, c'est mieux : Merci à Hervé Capra et à Pascal Roger.

Et parce qu'il y a aussi (surtout) une vie à côté du travail...

Merci à mes Parents de m'avoir toujours laissé le choix de ce que je voulais faire, tout en m'encourageant et en me donnant les moyens d'y arriver. Sans vous, je ne serais pas là !

Merci à Cathy qui a été ma première supportrice et mon plus grand soutien ! Merci d'être là pour moi et merci pour tes rires ! T'es trop forte Mich-mich !

Merci à ma sœur Audrey et à Fred, pour la joie, le soutien et le merveilleux cadeau qu'est ma ptite nièce, Lyne.

Merci à mes meilleurs amis, Julien et Diane ! Vous avez toujours été là quand j'en avais besoin, et ce bien avant la thèse ! Si on m'avait dit qu'un fou rire en cours de mécanique quantique nous amènerait là, j'aurais signé tout de suite ! Je souhaite à tout le monde de connaître des amis comme vous...

Merci à mes comparses musicaux grâce à qui la vie est une fête : Tof, Benji et Djaïls ! J'aime beaucoup ce que vous faites ! (je me mets un peu à part, un bassiste, ça sert à rien !)





*A mes Parents*



## Résumé

La compréhension des déterminismes des *communautés de poissons* et de leurs variations est un enjeu majeur, tant en écologie qu'en bioindication. Dans ce travail doctoral, nous avons mis en place des outils de *statistiques avancées* afin d'identifier les aspects des régimes thermique et hydrique, constituants essentiels de l'environnement naturel des poissons, responsables des variations des assemblages de *juvéniles* de cyprinidés du haut Rhône. En utilisant des chroniques de températures de l'eau, de débits et d'échantillonnages par pêche électrique sur la période 1980-2005, nous avons déterminé trois groupes d'espèces présentant des réponses synchrones aux variations d'environnement, et développé deux types de *modèles hiérarchique*, s'appuyant sur ces groupes, selon une perspective bayésienne :

- un modèle de *régression poissonnienne*, pour lequel nous avons utilisé une procédure de choix de modèle fondée sur le critère du *Facteur de Bayes*. Ce premier modèle nous a permis de sélectionner les variables d'intérêt et de quantifier leurs effets sur les variations interannuelles d'abondances des différentes espèces.
- un modèle à *facteur latent partagé*, résumant l'influence commune d'une variable latente, l'*hypersignal*, sur les variations de l'environnement et des proportions relatives des groupes d'espèces. Nous avons ainsi souligné les importances relatives des variables environnementales dans la structuration des communautés.

Alors que leurs effets sont généralement discutés séparément, nous avons mis en évidence l'influence conjointe des régimes thermique et hydrique sur la structuration et les variations interannuelles des assemblages de juvéniles de poissons en grandes rivières.

## Mots-clés

Communautés de poissons, régime thermique, régime hydrique, variations interannuelles, statistiques bayésiennes, modélisation hiérarchique.

## Abstract

Understanding how fish assemblages vary is a major goal of ecology and bioindication. In this PhD thesis, advanced statistical tools were used to identify which descriptors of temperature and flow regimes, that are among the most essential components of fish natural environment, do generate interannual variations of juvenile assemblages of cyprinid fish in the upper River Rhone. By using datasets of water temperature, flow and electric fishing samples over the 1980-2005 period, three groups of species have been determined based on their synchronic response to environmental variations and have been used to develop two types of Bayesian hierarchical models :

- a Poisson regression, for which a selection of variables was performed, using the Bayes Factor criterion. This first model allowed for the quantification of the effects of the selected variables on the interannual variations of the different species abundance.
- a model with shared latent variables, assuming the existence of a common source of variation for the environmental and the biological data, the Hypersignal. This model underlined the relative importance of the environmental variables in determining the assemblages structure.

While their effects are generally treated separately, the joint effect of the temperature and flow regimes on the structure and the interannual variations of juvenile fish assemblages was put forward here.

## Keywords

Fish assemblages, thermal regime, flow regime, interannual variations, Bayesian statistics, hierarchical modelling.

# Sommaire

<b>Sommaire</b>	<b>xv</b>
<b>Table des illustrations</b>	<b>xix</b>
<b>Liste des tableaux</b>	<b>xxiii</b>
<b>L'écologie</b>	<b>1</b>
<b>1 Introduction</b>	<b>3</b>
1.1 Le déterminisme des assemblages piscicoles, une question centrale . . . . .	3
1.1.1 En écologie . . . . .	3
1.1.2 Pour les besoins opérationnels . . . . .	7
1.2 Comment étudier la réponse des communautés aux variations du milieu ? . . . . .	12
1.2.1 La prise en compte de l'environnement . . . . .	12
1.2.2 La description d'une communauté . . . . .	15
1.2.3 En modélisation . . . . .	19
1.3 Un site pour une première analyse : le cas du Bugey . . .	24
1.3.1 Objectif du suivi des populations piscicoles de la CNPE de Bugey . . . . .	24
1.3.2 Description du dispositif d'échantillonnage . . . . .	24
1.3.3 En quoi est-ce un cas exemplaire ? . . . . .	26
1.3.4 Les sources de variabilités . . . . .	27
1.4 Conclusion . . . . .	29
<b>2 Les défis méthodologiques de la question de thèse</b>	<b>31</b>
2.1 Les défis écologiques . . . . .	32

2.1.1	Les approches classiques de la bioindication sont insuffisantes . . . . .	32
2.1.2	Les assemblages piscicoles en grands cours d'eau : un système peu étudié . . . . .	33
2.1.3	La prise en compte de l'environnement naturel . . . . .	36
2.1.4	L'étude des 0+ est-elle suffisante pour estimer l'état du système ? . . . . .	43
2.2	Les défis de la modélisation . . . . .	43
2.2.1	Un protocole expérimental préexistant et non dédié à la question posée . . . . .	43
2.2.2	La plasticité des espèces . . . . .	49
2.2.3	Peut-on quantifier par des lois ces phénomènes dont la connaissance est surtout qualitative ? . . . . .	50
2.3	Conclusions . . . . .	50

**La modélisation** **52**

**3 Un modèle linéaire généralisé logPoisson** **55**

3.1	Description des données . . . . .	55
3.1.1	Description des variables d'environnement explicatives . . . . .	55
3.1.2	Les variables à expliquer : données d'abondances de juvéniles de cyprinidés. . . . .	59
3.2	le modèle loglinéaire Poisson . . . . .	65
3.2.1	Une répartition poissonnienne des individus . . . . .	65
3.2.2	La loi de Poisson pour compter des poissons . . . . .	65
3.2.3	La régression loglinéaire Poisson . . . . .	66
3.2.4	Les clusters d'espèces comme structuration hiérarchique des modèles . . . . .	67
3.2.5	La question du choix de modèles ? . . . . .	69
3.2.6	Vérifier l'hypothèse de distribution Poissonnienne grâce aux résidus de Pearson . . . . .	74
3.2.7	Une méthode d'estimation de la capacité prédictive du modèle . . . . .	75
3.3	Application au cas du Rhône, sur le site d'échantillonnage de Bugey . . . . .	77
3.3.1	Un premier modèle pour stationnariser l'effet centrale . . . . .	77

3.3.2	Un second modèle intégrant la variabilité station- nelle et saisonnière : un essai pour limiter la sur- dispersion . . . . .	84
3.3.3	Vers une meilleure modélisation de la surdispersion	92
3.4	Conclusion . . . . .	103
<b>4</b>	<b>Un modèle à facteur latent partagé</b>	<b>105</b>
4.1	Introduction . . . . .	105
4.1.1	Travailler de façon conditionnelle à l'effectif total pour stationnariser le protocole expérimental . . .	105
4.1.2	Créer des variables explicatives ad-hoc . . . . .	106
4.1.3	Quelques approches bibliographiques de l'analyse multivariée . . . . .	106
4.2	Structure du modèle . . . . .	108
4.2.1	Facteur latent partagé . . . . .	108
4.2.2	Observations . . . . .	110
4.2.3	Prédictions . . . . .	114
4.2.4	Conclusion . . . . .	115
4.3	Application au cas du Rhône à Bugey . . . . .	115
4.3.1	Observables . . . . .	115
4.3.2	Structuration des données . . . . .	116
4.3.3	Modèle . . . . .	116
4.3.4	Résultats . . . . .	119
4.3.5	Interprétations écologiques . . . . .	128
4.3.6	Test de la capacité prédictive du modèle . . . . .	131
4.4	Conclusion . . . . .	131
	<b>Au final</b>	<b>136</b>
<b>5</b>	<b>Conclusion générale et perspectives</b>	<b>139</b>
5.1	Bilan des travaux de thèse . . . . .	139
5.1.1	Rappel des objectifs et des principales contraintes .	139
5.1.2	Apports statistiques . . . . .	140
5.1.3	Apports en écologie . . . . .	141
5.2	Perspectives dans un contexte complexe et incertain . . .	142
5.2.1	Approfondissement des méthodes . . . . .	142
5.2.2	Transposer les méthodes . . . . .	143
5.3	Vers un outil d'aide à la décision bayésien . . . . .	146
5.3.1	Une situation fortement entachée d'incertitudes . .	146

5.3.2	Faciliter la discussion autour des processus : les structures graphiques . . . . .	147
5.3.3	Notre démarche s'inscrit dans la mise en place d'un réseau de probabilité bayésien . . . . .	148
	<b>Références bibliographiques</b>	<b>153</b>
	<b>Annexes</b>	<b>171</b>
A	Publications scientifiques	171
B	Code WinBUGS : modèle logPoisson à "effet centrale" stationnarisé	213
C	Code WinBUGS : modèle logPoisson à variabilité stationnelle	217
D	Code WinBUGS : modèle logPoisson surdispersé	221
E	Code WinBUGS : modèle à facteur latent partagé	225

# Table des illustrations

1.1	Échelles spatiales et temps de persistance (a) et de réponse (b) des assemblages de poissons aux perturbations	6
1.2	Relations entre sociosystème et écosystème aquatique . . .	10
1.3	Processus et échelles successives associées déterminant le fonctionnement de l'écosystème . . . . .	20
1.4	Situation géographique du site d'étude de Bugey . . . . .	25
2.1	Chronique thermique annuelle et cycles de vie des compartiments de l'écosystème aquatique . . . . .	38
2.2	Influence du régime hydrique sur l'intégrité des assemblages piscicoles . . . . .	39
2.3	Exemple de variations interannuelles de débits du Rhône, au niveau de la CNPE de Bugey, entre 3 années . . . . .	41
2.4	Exemple de variations interannuelles de températures du Rhône, au niveau de la CNPE de Bugey, entre 3 années .	42
2.5	Graphe des corrélations des paramètres de densité et de capturabilité . . . . .	45
3.1	Variables explicatives résumant le régime thermique . . .	57
3.2	Variables explicatives résumant le régime hydrique . . . .	57
3.3	Chroniques des valeurs centrées et réduites des 9 variables environnementales. . . . .	58
3.4	Chroniques des Captures Par Unité d'Effort par espèces, sommées sur l'ensemble du site . . . . .	60
3.5	Dendrogramme des espèces, basé sur la mesure du Tau de Kendall . . . . .	63
3.6	Chronologies des CPUE par espèces (pêches d'été), ordonnées par groupe déterminés par clustering . . . . .	64

3.7	Graphe Acyclique Orienté du modèle à effet central stationnarisé. . . . .	78
3.8	Boxplot des densités de distribution des paramètres associés aux variables retenues dans les modèles sélectionnés sur la base du Facteur de Bayes. . . . .	80
3.9	Comparaison des valeurs observées par espèce et des intervalles prédits par le modèles pour les années 2006 à 2009 sous forme de boxplots. . . . .	82
3.10	Distribution de probabilité d'une variable aléatoire $a$ , suivant une loi de Poisson, pour différents paramètres $\lambda$	85
3.11	Structure du modèle à effet stationnel, présentée sous forme de Graphe Acyclique Orienté. . . . .	86
3.12	Distribution de probabilité a posteriori du paramètre lié à l'effet saisonnier automnal, présentée sous forme de boxplot.	89
3.13	Distributions de probabilité des valeurs des paramètres d'effets stationnels pour chaque espèce, présentées sous forme de boxplots. . . . .	90
3.14	Boxplots des distributions de probabilités des paramètres associés aux variables retenues dans les modèles sélectionnés sur la base du Facteur de Bayes. . . . .	91
3.15	Structure du modèle à effet stationnel et saisonnier, avec terme d'erreur, présentée sous forme de DAG. . . . .	93
3.16	Distributions de probabilité des paramètres liés aux effets stationnels sur chacune des 8 espèces, présentées sous forme de boxplot. . . . .	95
3.17	Distribution de probabilité a posteriori du paramètre lié à l'effet saisonnier automnal, présentée sous forme de boxplot.	96
3.18	Boxplots des distributions de probabilités des paramètres associés aux variables retenues dans les modèles sélectionnés sur la base du Facteur de Bayes. . . . .	97
3.19	Illustration de la capacité prédictive du modèle à mélange de loi. . . . .	101
3.20	Illustration de la capacité prédictive du modèle à mélange de loi. . . . .	102
4.1	L'hypersignal $Z$ , au coeur de la structure de covariation des données observées continues $X$ et catégorielles $Y$ . . .	109
4.2	Représentation théorique des méthodes d'analyse multivariée. . . . .	111
4.3	DAG du modèle à facteur latent. . . . .	117

---

4.4	Distributions de probabilités annuelles a posteriori de l'hypersignal. . . . .	120
4.5	Proportions relatives annuelles des trois groupes d'espèces pour les pêches d'été et les stations amont. . . . .	122
4.6	Proportions relatives annuelles des trois groupes d'espèces pour les pêches d'été et les stations aval. . . . .	122
4.7	Proportions relatives annuelles des trois groupes d'espèces pour les pêches d'automne et les stations amont. . . . .	123
4.8	Proportions relatives annuelles des trois groupes d'espèces pour les pêches d'automne et les stations aval. . . . .	123
4.9	Distributions a posteriori des coefficients du GLM multinomial à fonction de lien "logit" . . . . .	125
4.10	Évolution interannuelle des variables descriptives de l'environnement, et valeurs moyennes sur les périodes 1980-1992 et 1993-2005 . . . . .	126
4.11	Distributions de probabilités a posteriori des composantes principales . . . . .	127
4.12	Distributions de probabilités annuelles a posteriori de l'hypersignal $Z$ , pour les années 1980 à 2005 (à gauche de la ligne verte) et distributions prédictives a posteriori du facteur latent $Z$ , pour les années 2006 à 2009 (à droite de la ligne verte). . . . .	132
4.13	Distributions prédictives a posteriori (boxplots) et observations (lignes) des proportions relatives des trois groupes d'espèces lors des pêches d'été, pour les stations situées en amont et en aval de la centrale nucléaire. . . . .	133
4.14	Distribution prédictives a posteriori (boxplot) et observations (lignes) des proportions relatives des trois groupes d'espèces lors des pêches d'automne, pour les stations situées en amont et en aval de la centrale nucléaire. . . . .	134



# Liste des tableaux

2.1	Nombre d'articles référencés dans le Web of Science pour quelques espèces de poissons (tableau actualisé au 1er juin 2010) . . . . .	35
3.1	Tableau des corrélations entre variables explicatives . . . .	59
3.2	Résultats du test du Chi-2 de la distribution poissonnienne des données . . . . .	81
3.3	Résultats du test de l'hypothèse poissonnienne du modèle à effet stationnel : quantiles de la loi du chi-2 à 3070 degrés de libertés et moyennes et intervalles de confiances simulés et observés de la statistique de test pour le modèle à effet stationnel. . . . .	92
3.4	Résultats du test de l'hypothèse poissonnienne du modèle à effet stationnel : quantiles de la loi du chi-2 à 3070 degrés de libertés et moyennes et intervalles de confiances simulés et observés de la statistique de test, pour le modèle de mélange à effet stationnel. . . . .	94



# L'écologie



# Chapitre 1

## Introduction

### 1.1 Le déterminisme des assemblages piscicoles, une question centrale

#### 1.1.1 En écologie

Produire et synthétiser de la connaissance sur le monde naturel et être capable à terme de prévoir comment les structures ou les fonctions changent, tant au niveau de l'espèce que des populations, des communautés ou de l'écosystème dans son ensemble, sont deux objectifs fondamentaux des recherches en écologie (Power et al., 1988). Au niveau des communautés, il est nécessaire de comprendre comment les assemblages d'espèces se façonnent, en réponse aux variations de l'environnement. Un *assemblage* est un groupe d'espèces différentes occupant un espace commun à un instant donné. A l'échelle planétaire, on dénombre environ 8500 espèces de poissons d'eau douce, peuplant pour la majorité les cours d'eau et les plaines inondables (Lowe-McConnell, 1987). Mais la répartition de ces espèces n'est pas uniforme et toutes les espèces ne partagent pas les mêmes *traits biologiques* (un trait biologique est une caractéristique permettant de décrire le comportement biologique d'une espèce : par exemple, la stratégie de reproduction, ou le régime alimentaire). Plusieurs hypothèses écologiques ont été proposées pour expliquer la répartition des espèces et la constitution des assemblages locaux.

#### **Echelles spatiales et filtres successifs**

La *théorie des filtres écologiques* soutient que les assemblages piscicoles sont le résultat de l'adéquation entre les traits biologiques des

espèces et les contraintes locales. Elle suppose aussi qu'il existe une hiérarchie dans l'application des contraintes qui peuvent donc être vues comme des filtres environnementaux agissant à différentes échelles et s'appliquant successivement (Poff et al., 1997). Cette hiérarchie s'instaure depuis les grands ensembles zoo-géographiques jusqu'au microhabitat, et les conditions locales aux faibles échelles sont contraintes par celles aux échelles supérieures (Bayley and Li, 1992). Ces échelles d'application (figure 1.1) peuvent être définies sous un angle soit spatial, soit temporel, sachant que les deux descriptions sont très fortement liées. En effet, les conditions physiques de l'environnement aux échelles spatiales supérieures sont beaucoup plus stables dans le temps que celles aux échelles locales. Ainsi, en définissant un maillage de l'espace catégorisant les grandes zones zoo-géographiques ( $10^5$ - $10^7$  mètres), la période de persistance des conditions environnementales associée sera de l'ordre de  $10^7$ - $10^9$  ans. Ces échelles décroissent linéairement jusqu'à l'échelle du microhabitat ( $10^{-1}$  mètres) dont le temps de persistance est de  $10^{-1}$ - $10^0$  ans, et qui correspond à l'environnement local dans lequel évolue l'individu poisson. Dans le même temps, cette théorie suppose que les espèces répondent (via adaptation, disparition ou colonisation) aux variations de l'environnement de façon plus ou moins rapide selon l'échelle. A l'échelle de l'ensemble zoo-géographique, qui est l'échelle à laquelle la zone de répartition des espèces se définit, le temps de réponse aux variations sera de l'ordre du siècle, à l'échelle du tronçon de cours d'eau qui correspond à l'habitat général de l'individu poisson, le temps de réponse aux variations sera de l'ordre du mois, au gré des possibles migrations, tandis qu'à l'échelle microhabitat, qui est l'échelle d'activité de l'individu poisson, le temps de réponse sera de l'ordre de l'heure, avec une adaptation des comportements au moins au cycle nyctéméral. On voit dans cette théorie qu'il y a un glissement d'échelle progressif des trois compartiments considérés (espace géographique, persistance des conditions environnementales et temps de réponse biologique) et qu'il est donc nécessaire, selon la question étudiée, de bien appréhender cette hiérarchie pour se placer à la bonne échelle. L'intérêt, pour l'étude des systèmes écologiques, de la clarification des échelles par cette description hiérarchique a été montrée très tôt, notamment par Allen and Starr (1982). Ainsi, l'ensemble des espèces potentiellement présentes sur un tronçon, est déterminé aux échelles successives de la zone zoo-géographique (continentale), de l'hydroécocorégion (Wasson (2004), avec un rôle important de la latitude, de l'altitude et de la géologie) et du bassin versant. Un *bassin versant* se

définit comme une zone géographique au sein de laquelle toute eau de ruissellement passe par le même exutoire. Ainsi, un point d'un réseau hydrographique désigné comme étant l'exutoire suffit à définir un bassin versant. En général, pour parler du bassin versant d'un cours d'eau, on prendra soit la confluence avec un cours d'eau plus élevé, si c'est une rivière, soit son estuaire si c'est un fleuve. Puis au sein du bassin versant, la répartition des espèces est étroitement liée à la structure même du réseau hydrographique.

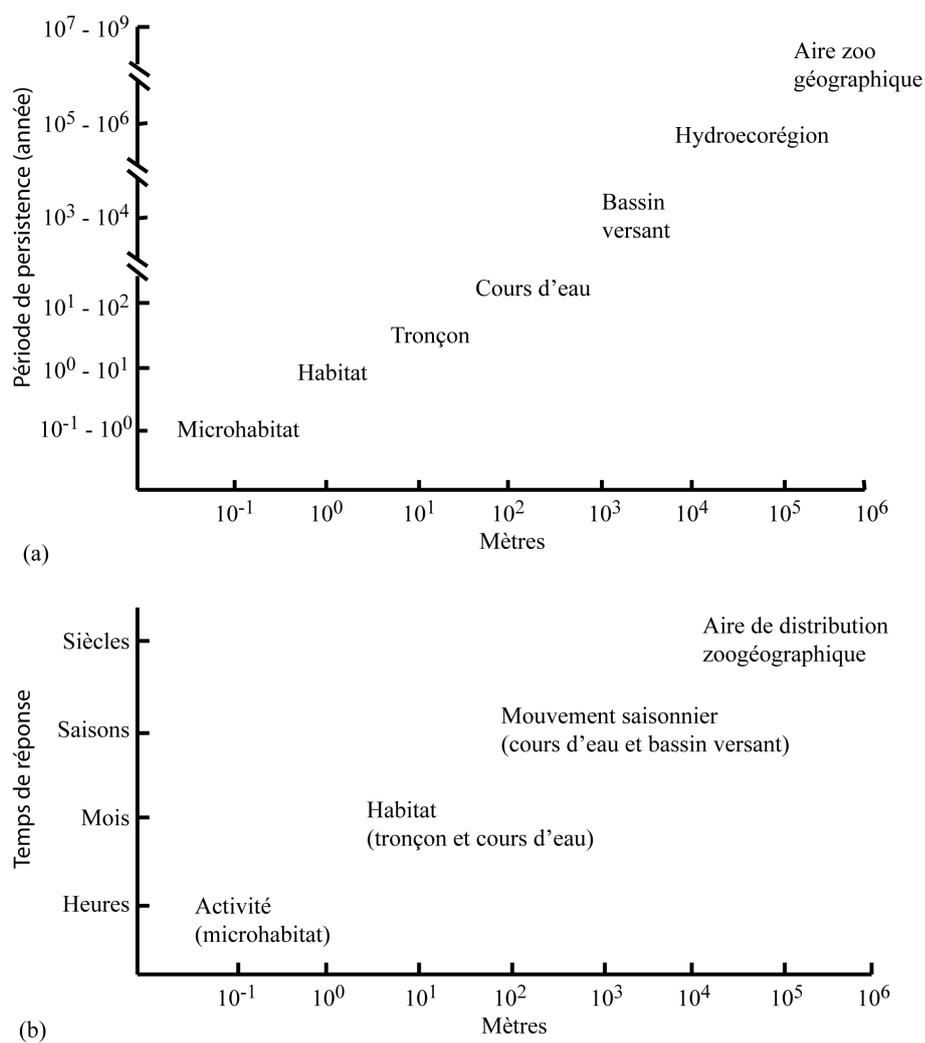
Conséquemment, la répartition des espèces à l'échelle du petit cours d'eau ou du tronçon est fortement régie par les conditions de pente, de température et de substrat. Les conditions d'habitat (échelle inférieure) et leur intégration va quant à elle jouer un rôle dans la définition des proportions relatives des différentes espèces présentes. On retrouve, dans cette description des facteurs régissant la répartition des habitats présents et in fine de la répartition fine des espèces et de leurs différents stades de développement au sein du bassin versant, les travaux de Huet (1949), qui, en identifiant l'évolution longitudinale progressive des peuplements piscicoles a déterminé des *zones* nommées en fonction de l'espèce emblématique présente (successivement zone à truite, à ombre, à barbeau et enfin à brème). Ces zones reflètent la gradation longitudinale des cours d'eau et les preferences des espèces pour certaines conditions locales. Ainsi les gradients de pente et de température de l'eau déterminent fortement la probabilité de présence d'une espèce à un point donné du cours d'eau.

### **Localement, des variations plus fines des assemblages**

Les espèces potentiellement présentes au niveau d'un tronçon de cours d'eau sont donc définies en réponse à une succession de contraintes qui s'appliquent à des échelles spatio-temporelles plus grandes. Mais entre deux tronçons similaires, situés de façon proche sur les grands gradients spatio-temporels, ces espèces peuvent connaître des variations d'effectifs plus ou moins importantes en réponses à des variations plus locales de leur environnement. Ces variations locales d'environnement peuvent prendre des formes très diverses, liées aux variations naturelles des conditions de température ou de débit, ou des processus physico-chimiques, mais également en lien avec les activités humaines qui exercent des pressions sur les cours d'eau.

On trouve au coeur de ces discussions le concept de *perturbation*, qui s'oppose aux hypothèses classiques de systèmes en équilibre (McIn-

Figure 1.1 – Échelles spatiales et temps de persistance (a) et de réponse (b) des assemblages de poissons aux perturbations (tiré de Bayley and Li, 1992)



tosh, 1987) dans un environnement quasi constant. Les modèles écologiques classiques (un des exemples les plus connu est le modèle de Lotka-Volterra) prévoient que les communautés ne se déterminent qu’au travers des relations biotiques entre les espèces, et n’intègrent pas les variations de l’environnement. L’importance des perturbations du milieu comme facteur de richesse spécifique a été introduite plus tard, notamment dans le cadre de la théorie des “perturbations intermédiaires” (Connell, 1978). Cette théorie s’appuie sur une hiérarchie compétitive des espèces en présence. En l’absence de perturbation, la plus compétitive s’impose ; mais lorsque le milieu est perturbé, les variations locales permettent aux autres espèces de se maintenir. Par la suite, Huston (1979) introduit la notion d’équilibre dynamique dans son modèle. La structure de la communauté est alors le résultat d’un compromis entre les taux de croissances spécifiques, des phénomènes d’exclusions compétitives, et de la fréquence des réductions des populations liées aux perturbations. Si ces épisodes sont très fréquents, l’exclusion compétitive n’arrive pas à terme et les espèces moins compétitives peuvent se maintenir.

De nombreux auteurs ont montré que les espèces sont localement adaptées aux variations naturelles “prévisibles” de l’environnement (Resh et al., 1988; Power et al., 1988; Konrad et al., 2008) et cherchent alors à décrire les temps de retour de certains épisodes, ainsi que la *résilience* des structures, définie comme la capacité de la communauté à se remettre d’une perturbation. Les événements rares, par essence imprévisibles, sont quant à eux capables de bouleverser cette structure.

### 1.1.2 Pour les besoins opérationnels

Les activités humaines peuvent exercer des pressions sur les conditions d’environnement aux différentes échelles spatiales, depuis le bassin versant jusqu’à l’échelon local, et en conséquence avoir une influence, tant sur les espèces présentes que sur leurs effectifs et leur proportion relative. Les effets découlent d’actions directes (ex. pollution ponctuelle ; altération de l’habitat par intervention dans le cours d’eau) ou indirectes au sein du bassin versant ou des ripisylves (ex. pollutions diffuses, matières en suspension, ...).

## La directive cadre et l’ingénierie

*Contexte politique et objectifs*

La Directive Cadre européenne sur l'Eau (DCE), adoptée le 23 octobre 2000, établit un cadre pour une politique communautaire dans le domaine de l'eau, qualifiée de ressource d'intérêt général. Elle vise à protéger de façon globale les ressources en eaux douces, saumâtres ou salées superficielles ou souterraines, de transition<sup>1</sup> et côtières. C'est une directive ambitieuse qui vise à prévenir et réduire la pollution des eaux, promouvoir son utilisation durable, protéger l'environnement, améliorer l'état des écosystèmes aquatiques et atténuer les effets des inondations et des sécheresses. Pour cela, elle impose comme objectif, à l'horizon 2015, l'atteinte du "bon état écologique" des milieux aquatiques et du bassin versant, seul moyen de garantir une gestion durable et soutenable de cette ressource vitale pour l'humanité et pour toutes les espèces vivantes connues.

Cette directive implique deux étapes, chronologiquement :

- diagnostic de l'état des masses d'eau : elle passe par l'identification et la caractérisation des eaux européennes par grand bassin régional et district hydrographique,
- restauration : adoption de plans de gestion et de programmes de mesures appropriées à chaque masse d'eau visant à rétablir le bon état. Ces plans de gestion doivent être en mesure de :
  - prévenir la détérioration, améliorer et restaurer l'état des masses d'eau de surface, atteindre un bon état chimique et écologique de celles-ci, ainsi que réduire la pollution due aux rejets et émissions de substances dangereuses ;
  - protéger, améliorer et restaurer les eaux souterraines, prévenir leur pollution, leur détérioration et assurer un équilibre entre leurs captages et leur renouvellement ;
  - préserver les zones protégées.

#### *Les indicateurs de qualité*

La DCE définit l'état comme étant le moins bon des deux états, chimique et écologique.

- physico-chimie : un certain nombre de molécules ou de composés sont identifiés pour leur potentiel polluant du milieu (voire la liste en annexe 8 de l'Arrêté du 25 janvier 2010 relatif aux méthodes et critères d'évaluation de l'état écologique, de l'état chimique et

---

1. Une eau de transition est définie par la DCE comme masse d'eaux de surface à proximité des embouchures de rivières, qui sont partiellement salines en raison de leur proximité d'eaux côtières, mais qui sont fondamentalement influencées par des courants d'eau douce.

du potentiel écologique des eaux de surface pris en application des articles R. 212-10, R. 212-11 et R. 212-18 du code de l'environnement). Pour chacune, un seuil de tolérance doit être fixé. L'état chimique est déterminé par la présence / absence de ces composés à des valeurs supérieures à ces seuils.

- écologie : l'état écologique est évalué par comparaison à un *état de référence*, qui doit être défini en intégrant les potentialités du bassin. Cet état doit être évalué pour quatre compartiments fonctionnels de l'écosystème des milieux aquatiques, appelés *éléments de qualité* :
  - les diatomées,
  - les macroinvertébrés,
  - les macrophytes,
  - les poissons.

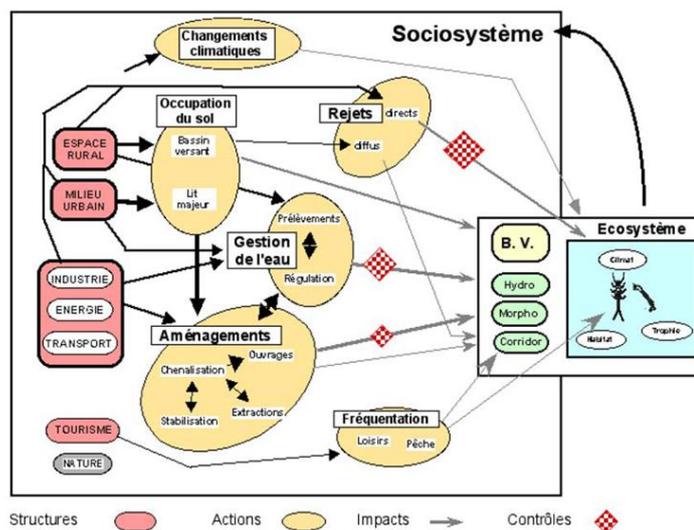
#### *Le cadre d'étude pression / impact*

De la lecture de la DCE, on retient également qu'elle vise à garantir durablement les usages humains du milieu. Cette considération replace de facto la DCE dans un cadre économique, et le modèle conceptuel englobant les relations pressions/impacts est le FPEIR, ce qui signifie Forces motrices - Pressions - Etat - Impact - Réponses). Il est construit sur un modèle socio économique initial de l'Organisation de Coopération et de Développement Economique (OCDE), le modèle PER (pressions - état - réponses). Il fonde les analyses de l'Agence Européenne pour l'Environnement. Ce schéma est basé sur les relations de causes à effets entre les différents compartiments des systèmes sociaux, économiques et environnementaux. On a donc deux grands systèmes juxtaposés et interdépendants : le sociosystème et l'écosystème (figure 1.2).

Les forces motrices, ou structures, correspondent aux compartiments du sociosystème qui interagissent avec l'eau : milieu urbain ou agricole, industries usages (énergie, transport, loisirs...). Ces structures génèrent des pressions : occupation du sol des bassins versants et des vallées, rejets polluants directs ou diffus, prélèvements et ouvrages (barrages, chenalisation...). Les pressions influent sur les composantes abiotiques du milieu, à la fois physico-chimiques (pollution organique, nutriments, toxiques...) et hydromorphologiques (structure, hétérogénéité, connectivité, variabilité de l'habitat physique), ce qui détermine in fine l'état écologique constaté au niveau des biocénoses.

Dans un second temps, un impact est défini comme la conséquence de la dégradation de l'état des masses d'eau pour l'homme, mesurée par des

Figure 1.2 – Relations entre sociosystème et écosystème aquatique (d'après Wasson et al, 1993 et 2005)



Les structures correspondent aux compartiments du « socio-système » qui interagissent avec l'eau : milieu urbain ou agricole, industries, usages (énergie, transports, loisirs...). Ces structures génèrent des pressions : occupation du sol des bassins et des vallées, rejets polluants directs ou diffus, prélèvements et ouvrages (barrages, chenalisation...). Les pressions induisent des impacts sur les composantes abiotiques du milieu, à la fois physico-chimiques (pollution organique, nutriments, toxiques...) et hydro-morphologiques (structure, hétérogénéité, connectivité, variabilité de l'habitat physique). Ces impacts déterminent in fine l'état écologique constaté au niveau des biocénoses.

Wasson et al.,  
1993 et 2005

critères économiques d'utilité et de valeur écologique. Enfin, la réponse est conçue comme l'action à mener pour redonner à la masse d'eau son utilité ou sa valeur écologique, de façon durable.

### Changement Global

Le changement global est une problématique majeure dans le monde scientifique actuel et comprendre ses effets sur les populations végétales ou animales, les structures des communautés ou encore sur le fonctionnement des écosystèmes est un des défis majeurs de l'écologie moderne (Daufresne and Boet, 2007). Bien que depuis le début des années 90, la littérature sur le sujet ait très rapidement augmentée, les scénarios et conséquences sont encore discutés, mais l'émergence, au début des années 2000, d'études par méta-analyse (Root et al., 2003; Parmesan and Yohe, 2003; Daufresne and Boet, 2007) semble avoir conduit à un certain consensus : le réchauffement climatique est un facteur de changements biologiques. Cet effet perturbateur est potentiellement important dans le cadre de la DCE, car il s'avère assez difficile d'évaluer les effets re-

latifs du changement climatique et des perturbations non climatiques sur l'écosystème aquatique (Daufresne and Boet, 2007). Il s'agit donc de tenir compte de ces effets, des scénarios acceptables et d'adapter les démarches DCE.

### **Le rôle du Cemagref et de ses partenaires**

Afin de répondre aux questions posées par la *Directive Cadre sur l'Eau*, différents acteurs sont mobilisés. En premier lieu, le Ministère en charge de l'écologie est chargé d'organiser politiquement l'application des décisions prises lors des négociations à la Commission Européenne par des voies législatives et réglementaires. Directement sous la tutelle de ce Ministère, l'Office Nationale de l'Eau et des Milieux Aquatiques (ex-Conseil Supérieur de la Pêche) est chargé d'animer et coordonner les recherches scientifiques et techniques. Pour cela, il s'appuie sur différents partenaires au sein des établissements publics à caractère scientifique et technique, tels que le Cemagref, qui développent des connaissances, à la fois fondamentales et appliquées, sur les milieux aquatiques et exercent des fonctions d'expertise. Enfin, en tant qu'organismes financeurs, les Agences de l'Eau (une agence pour chaque grand bassin fluvial) sont fortement impliquées. Ces dernières sont à la fois des organismes financeurs et opérationnels. Par exemple, elles participent à l'émergence des Schémas Directeurs d'Aménagement et de Gestion des Eaux (lignes directrices de l'aménagement au niveau des bassins versants) et à l'application des Schémas d'Aménagement et de Gestion des Eaux (gestion locales) en interface avec les Directions Régionales de l'Environnement, de l'Aménagement et du Logement et les Comités de bassin.

Pour répondre à ces questions, les partenaires scientifiques s'appuient à la fois sur des jeux de données préexistants, issus des anciens réseaux de surveillance du Conseil Supérieur de la Pêche : Réseau Hydrobiologique et Piscicole. Les nouveaux protocoles et réseaux d'échantillonnage sont aujourd'hui en place (Réseau de référence (400 stations nationales) et réseau de contrôle et de surveillance (1500 stations)) afin de permettre de mieux répondre aux questions actuelles, mais aussi futures.

Ainsi, si des jeux de données existent déjà, ceux-ci doivent être mis à jour progressivement. Mais à elles seules ces données de réseau ne portent pas ou ne porteront pas toutes les situations pressions/impacts écologiques nécessaires pour améliorer les connaissances, en partie en raison de leur jeunesse (manque de profondeur temporelle pour les nouveaux réseaux), de leur couverture spatiale, de la nature variée des situations

naturelles et des pressions combinées. C'est pourquoi soit des monitoring d'opérations de restauration (ex. arasement d'ouvrage, relèvement de débits réservés, abatement de pollution) et d'autres réseaux tels que ceux des industriels comme EDF revêtent également une grande importance.

## 1.2 Comment étudier la réponse des communautés aux variations du milieu ?

### 1.2.1 La prise en compte de l'environnement

#### Le rôle de la température

L'intégrité écologique des cours d'eau dépend de la dynamique naturelle du régime thermique (Olden and Naiman, 2010). La température de l'eau influence directement les activités métaboliques, la physiologie et les traits d'histoire de vie des espèces aquatiques et participe aussi de la détermination des rythmes des grands processus naturels tels que les cycles du carbone ou de l'azote (éléments essentiels de la productivité du milieu). Indirectement, elle régit les interactions biotiques (Caissie, 2006; Poole and Berman, 2001; Magnuson et al., 1979)

Les espèces de poissons et les invertébrés aquatiques sont ectothermes (leur température corporelle dépend de la température du milieu) et par conséquent, ils se déplacent afin de trouver les habitats thermiques correspondant à leurs besoins thermiques propres et leur permettant de réaliser certaines fonctions telles que la survie, la croissance ou la reproduction (Coutant, 1987; Vinson and Hawkins, 1998). Par conséquent, certains aspects du régime thermique ont une importance écologique très forte pour les organismes aquatiques au cours de leur cycle de vie (Olden and Naiman, 2010).

La température de l'eau interagit avec les températures létales (inférieures et supérieures) des espèces et participe donc directement de l'établissement des zones de répartitions des espèces à l'échelle continentale. De plus, elle régit directement la croissance des organismes aquatiques. En effet, ceux-ci sélectionnent les habitats dont la température optimise leur prise d'énergie relativement au coût physiologique (Olden and Naiman, 2010).

Le développement et les étapes du cycle de vie des espèces de poissons ou d'invertébrés localement adaptés sont très fortement coordonnés par les caractéristiques du régime thermique dominant (Huryn and Wallace, 2000; Vannote and Sweeney, 1980; Vinson and Hawkins, 1998).

Par exemple, certains signaux thermiques vont déclencher des comportements particuliers du cycle de vie tels que la migration, la ponte ou encore l'éclosion.

### **Le rôle du débit**

Le débit a été très tôt défini comme une variable majeure contraignant le fonctionnement des écosystèmes aquatiques, limitant les distributions et abondances des différentes espèces de poissons (Resh et al., 1988) et régulant l'intégrité des systèmes aquatiques courants (Poff, 1997). En effet, le débit est très fortement corrélé à plusieurs caractéristiques physicochimiques d'importance des rivières, comme la température, la géomorphologie du chenal et des berges, et la diversité des habitats (Poff, 1997).

L'importance potentielle du régime hydrique a toutefois été principalement occultée, dans la gestion de l'écologie des cours d'eau, par l'adoption quasi systématique du principe de débit minimum. Cette vision très réductrice du fonctionnement de l'écosystème considère que seules les périodes d'étiage peuvent être préjudiciables, et que par conséquent, la gestion de la quantité d'eau laissée dans les cours d'eau à un instant  $t$  n'est contrainte que par le fait de laisser une quantité minimum. Or ce principe occulte totalement les très fortes variations annuelles naturelles de débit et le fait que cette variabilité joue un rôle majeur dans le maintien d'un fonctionnement optimal de l'écosystème aquatique (Poff, 1997).

Dans les cours d'eau, la structure physique de l'environnement, et donc de l'habitat est largement définie par des processus physiques, et spécialement par le mouvement de l'eau et des sédiments, dans le chenal d'une part, mais aussi entre le chenal et le lit majeur, et Poff (1997) souligne que pour comprendre la biodiversité, la productivité et la durabilité d'un écosystème rivière, il est nécessaire d'apprécier le rôle organisateur central joué par les processus garants d'un habitat varié et dynamique. L'habitat physique d'une rivière inclut la taille et la diversité du sédiment, les morphologies du chenal et du lit majeur ainsi que d'autres caractéristiques géomorphologiques qui sont déplacées et déposées par le courant (quantité de sédiment disponible au transport, débris ligneux et autres matériaux charriables). Donc les conditions d'habitat varient entre rivières selon les caractéristiques de débit et le type et la quantité de matériaux disponibles. Au long d'un cours d'eau, différents éléments d'habitat sont créés et entretenus par des conditions variées de débit.

Ainsi les bancs, ou les séquences radier-mouille sont définies par un débit principal, ou débit de plein bord. Ce débit peut déplacer une grande quantité de matériaux et se produit suffisamment souvent pour modifier le chenal de façon continue (Wolman and Miller, 1960). La structure du lit majeur, quant à elle, est souvent liée à des périodes de hauts débits se produisant plus rarement (Miller, 1990 in Poff (1997)).

Sur une période, de l'ordre de quelques années à la décennie, un cours d'eau peut procurer des types d'habitat éphémères, saisonniers ou persistants, allant des eaux libres à l'étiage le plus sévère. Le caractère prévisible (bien qu'aléatoire) de la disponibilité d'habitats divers dans le chenal ou le lit majeur a favorisé la sélection naturelle d'espèces exploitant la mosaïque créée et maintenue par ces conditions hydrologiques variées. Pour un certain nombre d'espèces, la réalisation du cycle de vie requiert différents habitats aquatiques dont la disponibilité est régie par le régime hydrique (Sparks, 1995; Reeves et al., 1995; Greenberg et al., 1996). Le "Flood Pulse Concept" (concept du pic de crue) fait d'ailleurs des inondations régulières et prévisibles des zones inondables la principale force motrice du maintien de la biodiversité et de la productivité des écosystèmes aquatiques (Junk et al., 1989). En effet, les échanges latéraux entre les plaines inondables et le chenal principal d'énergie, de biomasse et de nutriments sont alors les plus importants. Cette adaptation au dynamisme de l'environnement permet aussi à certaines espèces de persister malgré des conditions environnementales sévères, comme les étiages ou les crues qui détruisent et recréent régulièrement des éléments d'habitat.

D'un point de vue évolutif, le patron de dynamique spatio-temporelle de l'habitat influence le succès d'une espèce dans des conditions d'environnement données. C'est la théorie de l'"habitat template" (Southwood, 1977) : dans le cas des cours d'eau, le débit serait la variable majeure expliquant la mosaïque d'habitats locaux et causant les variations subtiles ou profondes d'histoire d'individus d'une même espèce peuplant différentes parties de leur zone de répartition.

Enfin, le débit influence aussi le fonctionnement biochimique de l'écosystème, par des processus d'échanges (Poff and Allan, 1995) : latéraux entre le lit majeur et le lit mineur, entre l'eau et les berges ou les bancs de graviers, entre les zones d'infiltration et d'exfiltration. Tous ces espaces d'interface dont les communications sont régies par le débit et ses variations influent sur les capacités d'autoépuration des écosystèmes aquatiques.

## Il existe peu d'approches multi-pressions fines

Les effets des régimes thermique et hydrique sont en général étudiés séparément.

Concernant les températures, les effets des variations du régime thermique sont assez bien documentées, notamment en s'appuyant sur la détermination des cycles de vie par des aspects thermiques (Mills and Mann, 1985; Poole and Berman, 2001; Nunn et al., 2003; Wolter, 2007). Mais ces approches ne tiennent nullement compte de l'influence des variations du régime hydrique.

Cela peut s'expliquer par la complexité inhérente à la définition des caractéristiques influentes des aspects hydriques. Des auteurs tels que Poff et al. (1997) ont par exemple cherché à caractériser ce que serait un régime hydrique naturel. D'autres, tels Richter et al. (1996) ont proposé d'évaluer le niveau d'altération anthropique du régime hydrique en décrivant 32 paramètres hydrologiques, mais sans parvenir à réellement déterminer les aspects prépondérants qui feraient biologiquement du sens. De plus, Poff and Zimmerman (2010) pointent l'absence d'études statistiquement robustes de ces effets. Les principales études sur l'influence des variations de débit portent plutôt sur des variations drastiques de régime, comme par exemple l'établissement d'un débit réservé minimal sur une section de cours d'eau (Lamouroux et al., 2006). Des travaux sur le Rhône ont montré la possible relation entre un épisode de crue majeur et la modification de structure de l'assemblage de poissons (Cattanéo et al., 2001).

Par conséquent, il existe un très grand besoin de connaissance sur la structuration locale des assemblages de poissons par les variations inter-annuelles conjointes des régimes thermiques et hydriques.

### 1.2.2 La description d'une communauté

#### L'assemblage, une échelle adaptée

Les études ne visant qu'un faible nombre de bioindicateurs à la fois ont montré leurs limites quand il s'agit de discriminer des situations peu contrastées (Wasson et al., 2010). Tous les éléments de l'écosystème sont potentiellement porteurs d'information sur le fonctionnement du milieu, et donc bioindicateurs. Il faut considérer le système dans son ensemble. En étudiant des variations au sein d'un groupe d'espèces, on souhaite mettre en évidence les mécanismes responsables plus subtiles que ceux régissant la présence ou l'absence d'espèces cibles. Un *assemblage* se dé-

finit comme l'ensemble des espèces occupant un même espace donné. C'est un concept assez proche de celui de *communauté*, mais ce dernier est associé à des concepts écologiques théoriques s'attachant à décrire les interactions entre les espèces et leur impact sur la structure, notamment trophique. Le terme d'assemblage est donc plus neutre, lorsqu'il s'agit uniquement de décrire cet ensemble d'espèces. De nombreuses approches ont été proposées pour décrire les assemblages, donnant naissance à différents descripteurs, chacun s'attachant à un aspect particulier de l'assemblage. Toutes ces approches créent une richesse potentielle d'information qui semble particulièrement intéressante pour étudier les réponses du compartiment biotique aux variations de milieu. On peut s'attacher à l'étude :

- des effectifs des différentes espèces : c'est la description la plus évidente de l'assemblage. Il s'agit alors de compter le nombre d'individus de chacune des espèces composant l'assemblage,
- de la composition relative de l'assemblage : la composition est alors exprimée en pourcentages pour chacune des espèces. Cela permet de comparer plus facilement les structures de différents assemblages, mais perd le lien direct avec la biomasse totale de l'assemblage.

De plus, on peut regrouper les espèces au sein de *guildes fonctionnelles*, ou bien les subdiviser en classes d'âges afin d'étudier la dynamique ou la structure en âge de chaque espèce. Ces deux approches peuvent être également mêlées et être à leur tour étudiées en termes d'effectifs ou de compositions relatives.

### **Le concept de guildes**

On appelle *gilde* un ensemble d'espèces utilisant un même type de ressource environnementale de façon similaire. Ce terme regroupe des espèces, sans s'occuper du positionnement taxonomique, dont les *niches écologiques* se recoupent significativement (Root, 1967). C'est un concept classique en écologie des communautés et qui a été utilisé pour étudier, entre autres, les communautés de plantes (Sebastian, 2007) ou d'oiseaux (Bishop and Myers, 2005). Il a également déjà été introduit pour l'étude des poissons ((Aarts and Nienhuis, 2003, par exemple)). En effet, les espèces de poisson n'ont pas toutes les mêmes comportements, elles n'occupent pas toutes les mêmes habitats, avec les mêmes stratégies. Ainsi, des espèces partageant un trait d'histoire de vie commun peuvent être associées au sein d'une même gilde. La gilde se définira par rapport

au type de ressource utilisé : on parle de “guildes trophiques” lorsque les espèces sont classées selon leur régime alimentaire, et les espèces peuvent alors être invertivores (le régime alimentaire est principalement constitué d’invertébrés), piscivores (le régime alimentaire est principalement constitué d’autres poissons), omnivores (régime alimentaire varié, de type opportuniste), etc... Si l’on s’intéresse à l’habitat de reproduction, on pourra classer les poissons parmi les lithophiles (reproduction sur ou dans un substrat rocheux), les phytophiles (reproduction sur substrat végétal), les phytolithophiles (reproduction sur support mixte), etc...

Root (1967) soulignait déjà quelques avantages de ce concept :

- il permet de se concentrer sur toutes les espèces sympatriques sans se soucier de leurs relations taxonomiques,
- les guildes sont particulièrement intéressantes dans les études comparatives des communautés. En effet, il est en général impossible d’étudier toutes les espèces d’un écosystème, le concept de guildes permet de se concentrer sur des groupes d’espèces associées au regard d’une relation fonctionnelle particulière.

De plus, le besoin de réduire l’étude des communautés à des ensembles plus facilement manipulables a favorisé la généralisation des guildes. Cela constitue un avantage supplémentaire : les guildes peuvent être utilisées comme les “briques de base” des communautés (Hawkins and MacMahon, 1989). En soi, le concept de guildes donne une assise stable et établie pour décomposer un écosystème et en facilite l’étude en dégageant les fonctions écologiques de base : trophie, reproduction, ... Dans les approches plus appliquées, les guildes sont souvent utilisées comme des outils d’évaluation et sont considérées comme des indicateurs de l’état des ressources, ou de l’état des communautés (Oberdorff et al., 2002; Pont et al., 2006), et permettent d’aller un peu plus vers une bioévaluation plus fonctionnelle (Statzner and Bêche, 2010).

### **L’information apportée par les classes d’âge**

Les effectifs des différentes espèces de poissons ou des différentes guildes peuvent également être subdivisés en fonction de l’âge des individus observés. Les individus non matures sont regroupés dans la catégorie *juvénile*, qui est elle-même généralement subdivisée en deux catégories : les *0+* sont les individus dans leur première année de vie et les *1+* sont les individus dans leur deuxième année de vie. La plupart des espèces étant sexuées lors de la troisième année de vie, la dernière catégorie, appelée *>1+* regroupe tous les individus adultes, quel que soit leur âge.

Cela permet d'étudier la dynamique de ces espèces ou de ces guildes, qui constitue un aspect fondamental de la structuration des assemblages.

Les évaluations, tant quantitatives que physiologiques, des juvéniles s'avèrent être particulièrement intéressantes dans l'évaluation de l'état écologique (Copp et al., 1991; Schiemer et al., 1991) et le nombre de juvéniles des différentes espèces est un bon indicateur du succès reproducteur de celles-ci, qui est lui-même un facteur prépondérant dans le maintien des espèces (Schiemer et al., 2001). Les poissons sont très dépendants de la quantité d'habitat disponible (Souchon et al., 1995) et c'est sur cette assertion qu'est bâti le raisonnement des modèles de microhabitat (Souchon et al., 1988, 1989; Souchon and Valentin, 1991; Souchon, 1994; Pouilly et al., 1995; Ginot, 1995; Lamouroux et al., 1999). Or, les 0+ sont un indicateur de la quantité d'habitat disponible plus sensible que les adultes (Copp et al., 1991).

Mills and Mann (1985) ont introduit la notion de *force des classes d'âge* (Year-class strength) dans l'étude des espèces de poisson de rivière, comme étant l'importance relative des différentes cohortes (ensemble d'individus de même âge) dans l'effectif total. Ils ont aussi montré que la force de classe d'âge d'une cohorte, suivie dans le temps, se transmet d'année en année. Cela montre que chez les poissons de rivière, la force de classe d'âge d'une cohorte est déterminée très tôt, lors de la première année de vie. Copp et al. (1991) ont en plus montré que les stades oeuf, larve et juvénile étaient les phases les plus critiques dans le cycle de vie des poissons de rivière. Ce sont sur ces stades de vie que l'on observe les taux de mortalité les plus importants. Il apparaît donc que les 0+ sont probablement plus sensibles aux variations d'environnement que les classes d'âge ultérieures et que le succès reproducteur est un des facteurs influençant fortement les variations de populations de poissons (Houde, 1997). Ainsi, de nombreux auteurs ont souligné le besoin de mieux comprendre les mécanismes de variation du succès reproducteur (Gulland, 1982; Houde, 1997), et particulièrement l'importance des variations de l'environnement.

Enfin, dans le cas des grands cours d'eau (rang de Strahler  $> 7$ ) tels que le Rhône, l'échantillonnage des 0+ est plus aisée que celle des adultes. En effet, les 0+ peuplent les parties de cours d'eau moins profondes que celles colonisées par les adultes, qu'il s'agisse des berges, ou bien des hauts-fonds et bancs de graviers situés dans le chenal ou les îles. Les adultes, eux, occupent la majeure partie du temps les parties du chenal où la colonne d'eau est la plus importante, ce qui les rend particulière-

ment difficiles à échantillonner par la technique de pêche électrique. On a donc plus commodément accès à la description de l'état de la partie juvénile du système.

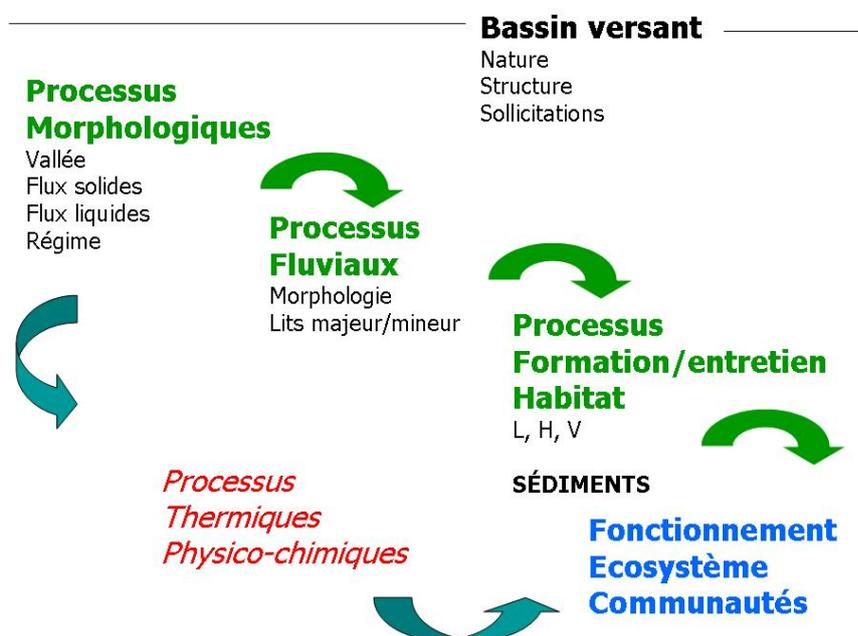
### 1.2.3 En modélisation

Dans la figure 1.3, on note un emboîtement d'échelles géographiques au niveau desquelles on considère les forces motrices et les pressions. Les forces motrices sont quantifiées à l'échelle large du bassin versant d'un cours d'eau. Les pressions sont discernables au niveau du tronçon de cours d'eau. Il existe donc un premier filtre spatial, et une situation de pressions locales ne peut pas être considérée indépendamment de la situation structurelle du bassin versant. De la même façon, il est impossible de quantifier toutes les pressions pour tous les tronçons de cours d'eau. Il s'agit donc, à partir de typologies large échelle, de développer des méthodes de modélisation permettant de prévoir les pressions probables qui s'appliquent aux échelles plus faibles. Ces modèles existent déjà dans le cadre de la prévision de l'état physico-chimique des masses d'eau, avec le modèle PEGASE (Planification Et Gestion de l'Assainissement des Eaux, Smitz et al. (1997)), par exemple.

La relation pression - état écologique est quant à elle définie à une seule échelle spatiale locale, celle du tronçon ou de la masse d'eau. En effet, les méthodes développées pour la détermination d'un état écologique le sont à ce niveau local, suffisamment petit pour permettre de détecter des variations fines du fonctionnement de l'écosystème. Il s'agit alors de développer des méthodes permettant de quantifier l'état écologique et ses variations en réponse aux pressions.

L'une des principales difficultés provient donc toujours de la mise en place du modèle à une échelle adaptée, et surtout, à terme, d'être capables de gérer au sein d'un même modèle, des phénomènes apparaissant à des échelles différentes. D'après les connaissances dont on dispose sur des phénomènes à large échelle qui se répercutent sur les échelles plus locales, il faudra réussir à traduire ces transferts d'influence entre échelle, en tenant compte des variations locales, des sources d'incertitude émergeant aux différentes échelles (Wu et al., 2006). Dans ce cadre où la gestion des échelles spatio-temporelles est primordiale, l'approche par modélisation hiérarchique est identifiée comme particulièrement adaptée pour rendre compte de la complexité des systèmes naturels (Royle and Dorazio, 2008; Cressie et al., 2009).

Figure 1.3 – Processus et échelles successives associées déterminant le fonctionnement de l'écosystème



**Réponse  $Y$  et impact  $X$  et fonction de réponse  $f$  ne sont pas définis**

Pour améliorer nos connaissances sur le fonctionnement du compartiment poisson de l'écosystème, nous adoptons une démarche de modélisation statistique. Un modèle statistique est une abstraction fonctionnelle visant à décrire les relations, présumées ou existantes, entre deux types de grandeurs observées, que l'on nomme conventionnellement variables à expliquer et variables prédictives.

$$Y = f(X, \varepsilon)$$

Les valeurs observées de la variable à expliquer sont considérées comme des réalisations d'une variable aléatoire  $Y$ , dont on cherche à préciser le comportement. Celui-ci est décrit à l'aide d'une partie expliquée par les variations d'une ou plusieurs variables prédictives  $X$ , ainsi que d'une partie aléatoire, ou erreur, notée  $\varepsilon$ , qui décrit la part de variabilité résiduelle qui n'est pas expliquée une fois les variables prédictives  $X$  prises en compte. Le lien entre toutes ces variables est fait à l'aide d'une fonction  $f$  qui reflète les hypothèses faites par le modélisateur sur les mécanismes réels sous-jacents qui font varier  $Y$  en même temps que  $X$ . La partie du modèle décrivant la variation de  $Y$  en fonction de  $X$  est appelée modèle structurel.

La fonction  $f$  peut être du type combinaison linéaire (on a alors des modèles de type régression). Elle peut aussi être une fonction de seuil, permettant de décrire des variations brutales dans la réponse de la variable prédite : en-dessous d'une valeur seuil, la fonction  $f$  a un fonctionnement donné, linéaire ou non ; au-dessus de cette valeur seuil, la fonction change. On pourrait aussi envisager d'introduire une fonction de spline, afin de traduire très fidèlement les variations de la variable prédite.

**Les sources de variation  $\varepsilon$** 

Dans une démarche de modélisation statistique, on confronte une vision du système, que l'on exprime par une structure de modèle, à des données observées, qui sont une réalisation du système. On désire à partir de ces observations apprendre sur le système. Toutefois, la vision du système n'est qu'une abstraction, et la confrontation aux données ne peut être parfaite. De cette rencontre naît l'incertitude, au coeur de l'analyse

statistique. Cette notion d'incertitude peut en outre se décomposer sous plusieurs formes :

– *l'incertitude par essence* :

Cette incertitude résulte de l'erreur de processus, de formalisme. En effet, il est impossible, dans un univers complexe, de prendre en compte tous les processus et tous les facteurs de détermination d'un phénomène. On ne propose qu'une représentation, que l'on souhaite la plus fidèle, mais qui de toute façon reste fautive. Cette erreur ne peut pas être réduite car elle est intrinsèque à la structure du modèle.

Elle provient de deux phénomènes : la stochasticité environnementale, qui provient du fait que l'on ne contrôle pas tous les facteurs environnementaux, et la stochasticité démographique, qui provient du fait que l'on ne peut observer la population complète, mais que l'on ne l'appréhende qu'au travers d'échantillons. La variabilité d'un échantillon à un autre traduit les différences de comportements individuels, au sein d'une même population.

– *l'incertitude par ignorance* :

L'incertitude par ignorance traduit les limites de ce que l'on peut apprendre d'un jeu de données, les limites d'un apprentissage par l'observation. Contrairement à l'incertitude par essence, on peut réduire l'incertitude par ignorance en augmentant la taille des échantillons observés, en améliorant la qualité des observations ou en modifiant la façon dont sont réalisées les observations afin de mieux les adapter aux hypothèses.

On distingue deux sources d'incertitude par ignorance : l'erreur statistique et les erreurs de mesure.

L'erreur statistique découle de la démarche d'inférence, consistant à déduire une propriété sur une population totale à partir de l'observation d'un échantillon, donc d'une sous-population. Comme les observations d'un échantillon à l'autre sont différentes, l'inférence de la propriété à l'échelle de la population est par essence incertaine. Cette erreur peut être corrigée en augmentant la taille de échantillons. Asymptotiquement, plus la taille de l'échantillon augmente, plus l'erreur statistique diminue.

L'erreur de mesure découle de la difficulté à mesurer précisément les variables d'état. Ainsi, il existe un écart entre la valeur réelle, inaccessible, et la valeur observée. Ce problème peut être particulièrement sensible dans les analyses statistiques basées sur des séries d'observations issues d'études dans le milieu naturel (Solow, 1998). En effet, les conditions d'observation sont souvent mal contrôlées et les méthodes d'échantillonnage peuvent nécessiter des protocoles d'observation indirectes qui

multiplient les sources d'incertitude. On distingue les erreurs de mesures systématiques, provenant de la qualité intrinsèque du matériel de mesure utilisé, et les erreurs d'observation, provenant de la façon dont on modélise le processus de recueil de données lors de l'observation.

### **Trier le naturel de l'artificiel**

L'utilisation du milieu naturel par l'homme impose d'en déterminer les effets sur l'environnement. Le travail d'analyse statistique va donc consister en la répartition de la variabilité totale observée entre les différentes sources d'incertitude. Dans le but de déterminer l'importance des pressions humaines, il faut donc déterminer les mécanismes naturels responsables de variations, et déterminer la part de variabilité résultant de ces variations naturelles d'environnement. Une fois cette part de variabilité soustraite (car expliquée), on s'attache à décomposer la variabilité résiduelle entre effets anthropiques et incertitude. L'étude des variations d'assemblage piscicole va donc résulter d'une décomposition successive, c'est-à-dire l'explication progressive, des sources d'incertitudes.

### **Les effets de mémoire**

Comme décrit au paragraphe 1.1, les assemblages (piscicoles dans ce cas mais le raisonnement vaut aussi pour les autres compartiments de l'écosystème) sont également déterminés par une succession de filtres agissant à différentes échelles spatiales. Plus l'échelle spatiale diminue et plus les perturbations évoluent rapidement, et plus les assemblages répondent rapidement. A l'échelle du tronçon, la période de persistance d'une espèce est de l'ordre de 10 à 100 ans. Il est donc nécessaire de ne pas réduire la dégradation d'un écosystème aquatique à l'observation de disparitions d'espèces. Ces disparitions, ne sont d'ailleurs bien souvent jamais observées. L'évaluation d'un état écologique doit donc tenir compte de ces temps de réponse assez longs et de possibles effets de mémoires.

## **1.3 Un site pour une première analyse : le cas du Bugey**

### **1.3.1 Objectif du suivi des populations piscicoles de la CNPE de Bugey**

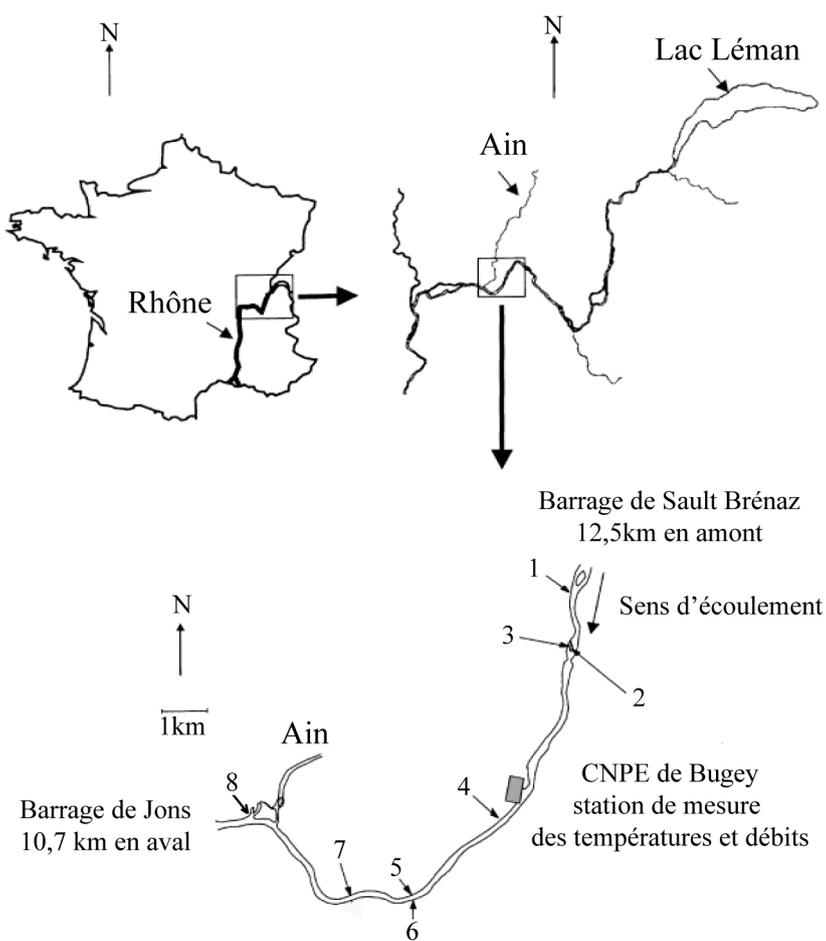
L'étude de ce site est née d'une convention passée entre Electricité de France (EDF) et le Cemagref. Le Cemagref s'est vu attribuer le suivi annuel de l'évolution des populations piscicoles suite à l'implantation du Centre Nucléaire de Production d'Énergie (CNPE) de Bugey, au sein d'un suivi environnemental poussé dont la nécessité s'est imposée à la fin des années 1970 pour accompagner le programme nucléaire. Cette centrale nucléaire utilise l'eau du Rhône pour refroidir ses réacteurs. Elle prélève de l'eau froide en amont, et la rejette sous forme gazeuse en grande partie, mais rejette aussi une partie de cette eau sous forme d'eau réchauffée d'une dizaine de degrés en moyenne par rapport à la température de l'eau amont. Pour suivre les éventuels effets de l'installation sur les poissons, il s'agit donc de décrire les structures de peuplement en amont et en aval du site d'implantation, ainsi que leurs évolutions en réponse à l'augmentation de température due au rejet d'eau réchauffée.

### **1.3.2 Description du dispositif d'échantillonnage**

Le CNPE de Bugey est implantée sur la partie française du Haut-Rhône (45.81N ; 5.21E), sur la commune de Saint-Vulbas dans l'Ain (01), à une trentaine de kilomètres en amont de Lyon. Elle a été mise en service en 1978 et les suivis des populations par le Cemagref ont débuté en 1979 et se poursuivent encore aujourd'hui. Chaque année, 4 campagnes d'échantillonnage sont menées (hiver, printemps, été, automne), selon un protocole établi initialement et stable dans le temps : 8 stations d'échantillonnage sont réparties sur les rives droite (6) et gauche (2) depuis 6 kilomètres en amont du CNPE jusque 11 kilomètres en aval, soit juste en aval de la confluence avec l'Ain. Trois stations sont situées en amont, et cinq en aval du CNPE.

Lors de chaque pêche saisonnière, chaque station est échantillonnée par pêche électrique en bateau pendant 30 minutes et uniquement sur les bords, où les plus faibles profondeurs rendent la pêche plus efficace. Chaque station représente ainsi une surface d'environ 500 mètres de long sur 2 à 4 mètres de large. Chaque occasion d'échantillonnage nécessite quatre opérateurs, un conducteur et trois épuiseteurs. La pêche électrique

Figure 1.4 – Situation géographique du site d'étude de Bugey



consiste en la génération d'un champ électrique de tension continue comprise entre 300 et 400 volts et d'intensité comprise entre 2 et 3 ampères, entre une cathode et une anode fixées en deux points du bateau. Lorsqu'un poisson pénètre dans le champ électrique à proximité de la cathode, il subit un effet répulseur, c'est pourquoi celle-ci est fixée à l'arrière du bateau et que la pêche se pratique préférentiellement de façon perpendiculaire à la rive. Au contraire, un poisson pénétrant le champ électrique à proximité de l'anode, constituée de deux anneaux métalliques concentriques et fixée à l'avant du bateau, il subit une tétanie momentanée et un effet attracteur du type "nage forcée" vers le centre des anneaux métalliques où il peut être prélevé à l'aide d'une épuisette. Le courant est généré à l'aide d'un groupe électrogène de chez Dream Electronics (type : Héron DC ; 300-400V ; 2-3A ; Pessac, France) embarqué.

A la fin de chaque échantillonnage stationnel, chaque individu prélevé est examiné pour en déterminer l'espèce et est mesuré. Une fois cette opération terminée, les individus sont remis à l'eau.

De plus, le dispositif expérimental est complété par la présence d'une sonde thermique et d'une station de mesure des débits. Ces deux systèmes d'enregistrement sont implantés au niveau du CNPE et fournissent des informations additionnelles sur l'environnement : la température moyenne quotidienne et le débit moyen quotidien sur la période 1980 à aujourd'hui.

### 1.3.3 En quoi est-ce un cas exemplaire ?

Le jeu de données de suivi du CNPE de Bugey est intéressant à plusieurs titres :

- c'est un jeu de données permettant d'étudier les effets des variations environnementales. En effet, Humphries et al. (2008) ont souligné l'importance de disposer de jeux de données couvrant une longue période, supérieure à 20 ans, pour pouvoir discriminer l'importance des variations d'environnement dans les phénomènes biologiques. Le jeu de données Bugey, couvrant une période de 30 ans (au moment de l'étude et qui se poursuit toujours) fait partie de ces rares jeux de données,
- la présence des stations de mesure de température et de débit au sein même du site d'évaluation biologique permet une confrontation directe des deux types d'information. Ce n'est pas toujours le cas, et certaines études reposent sur des reconstructions des données de températures via des modèles de régression entre la tem-

- pérature de l'air et la température de l'eau. Ces reconstructions introduisent alors une source d'erreur supplémentaire,
- du fait des plus grands problèmes d'échantillonnage et de la moindre présence d'espèces de premier intérêt halieutique, les jeux de données sur les grands cours d'eau sont relativement peu nombreux par rapport à ceux relatifs aux cours d'eau de plus petite taille, étudiant notamment les biotopes salmonicoles (truite, saumon).

### 1.3.4 Les sources de variabilités

#### Contrôlées

La modélisation permet de structurer l'information contenue dans le jeu de données en introduisant les deux sources de variabilité contrôlées dans le dispositif expérimental. On peut en effet étudier les effets suivants :

- effet saison : le système étudié étant dynamique, le nombre d'individus observés à différents moments de l'année pourrait varier de façon significative. Cette éventuelle influence de la saison d'échantillonnage sur les abondances de juvéniles de l'année peut être étudiée en introduisant une variable saisonnière dans les modèles.
- effet station : les 8 stations d'échantillonnage ne sont pas strictement équivalentes en termes de conditions d'habitat pour les différentes espèces de poisson. Il est possible d'étudier les variations d'abondances inter-stationnelles en introduisant une variable stationnelle dans les modèles. Les stations peuvent également être regroupées en fonction de certaines caractéristiques particulières pour tester des effets réalistes tels que la position amont/aval du CNPE. Dans ce cas, une différence de comportement des stations aval par rapport aux stations amont pourrait être interprétée comme un effet des activités de la centrale nucléaire.
- effet espèce : tous les individus sont déterminés jusqu'à l'espèce, cette effet peut donc être introduit. Toutefois, certaines erreurs peuvent survenir, du fait de fortes ressemblances entre certaines espèces, sur des individus peu âgés.

Évidemment, ces effets peuvent être introduits séparément ou simultanément. Par exemple, on peut aussi considérer un effet d'interaction "espèce x station amont/aval". Dans ce cas, toutes choses étant égales par ailleurs, le nombre moyen d'individus observés pour une espèce donnée sera différent entre les stations amont et les stations aval.

### Les autres sources d'incertitudes

Il subsiste néanmoins de nombreuses sources de variabilité qui ne sont pas contrôlables car elles ne sont pas testables au sein d'un plan expérimental.

Même à station et à saison fixées, le résultat des échantillonnages par pêche électrique est le fruit de nombreuses incertitudes. En effet, en dehors du fait que comme tout protocole d'échantillonnage il n'est pas exhaustif, tous les poissons observés dans l'eau ne sont pas prélevés lors de l'épuisetage. En effet en fonction de la force du courant, les poissons sont parfois entraînés et le champ électrique peut ne pas être suffisant pour les retenir assez longtemps dans la région centrale de l'anode. De plus, en fonction de la taille des individus et de l'angle de pénétration de leur corps dans le champ électrique, l'effet de tétanie n'est pas le même, et certains individus ne sont que faiblement touchés, bien qu'ils soient observés. De la même façon, manoeuvrer le bateau en bordures de cours d'eau souvent encombrées est assez difficile et il arrive que le bateau se retrouve parallèle à la berge sur une courte période, générant des effets répulsifs de la cathode sur les individus présents.

Les données de température et de débit ne donnent qu'une indication du régime quotidien. Si cela semble peu problématique pour les températures (les variations de température de l'eau au cours d'une journée sont relativement faibles), les non homogénéités peuvent être plus importantes pour les débits. En effet, le Haut-Rhône est exploité pour l'hydroélectricité depuis la fin du 19ème siècle et a vu sur cette période l'implantation de 9 barrages hydroélectriques entre Genève et Lyon. La gestion industrielle de ces ouvrages génère de grandes variations de débit au sein d'une même journée. Alors que le débit journalier moyen observé est d'environ  $500 \text{ m}^3 \cdot \text{s}^{-1}$ , ces variations atteignent parfois plusieurs centaines  $\text{m}^3 \cdot \text{s}^{-1}$ , en l'espace de quelques heures. Or, les poissons semblent être assez sensibles à ces variations rapides de débit, et, en cas de très faibles débits, ils pourraient ne repeupler que lentement les zones qui étaient découvertes. Les conditions de débit dans les heures qui précèdent les échantillonnages ont sans doute une importance particulière sur les variations d'abondance observées, sans que l'on puisse appréhender cet effet.

Malgré tout, on peut en première approche considérer que ces sources de variabilités sont relativement stables au cours de différentes campagnes d'échantillonnage, et donc que leurs effets ne sont pas particulièrement structurants et qu'ils font partie des sources d'erreurs résiduelles des modèles statistiques.

## 1.4 Conclusion

Que la question soit abordée d'un point de vue fondamental ou plus appliqué, l'un des points cruciaux dans la compréhension des réponses des communautés aux variations de l'environnement est de réussir à séparer la part de variabilité liée aux variations naturelles de celles liées à des pressions anthropiques. Cette thèse a pour objectif d'apporter de nouvelles méthodes de modélisation permettant une prise en compte de la complexité du système, liée à l'existence de ces multiples forces motrices dans un cadre quantifié.



## Chapitre 2

# Les défis méthodologiques de la question de thèse

Pour étudier le fonctionnement naturel de systèmes complexes tels que les cours d'eau, il faut résoudre un certain nombre de problèmes pratiques ou théorique, ou au moins, de clarifier les hypothèses fortes qui vont servir à simplifier le modèle d'étude. Se servir de la modélisation statistique pour mieux comprendre, ou préciser certaines hypothèses sur, le fonctionnement écologique de ces milieux nécessite de s'interroger tant sur la biologie que sur le protocole expérimental et les outils mathématiques.

Dans cette partie, nous listons les principaux défis à relever dans le cadre de cette thèse. Nous verrons dans un premier temps les principales interrogations écologiques : il nous semble nécessaire de proposer de nouvelles approches complémentaires aux principaux outils de la bioévaluation ; que le système étudié souffre d'un déficit de connaissances, par rapport à des systèmes plus restreints et mieux documentés ; et que la description de l'environnement naturel n'est pas évidente. Dans un second temps, nous verrons que les données dont on dispose ne permettent pas de développer tous les modèles auxquels nous aurions pu, a priori, penser.

## 2.1 Les défis écologiques

### 2.1.1 Les approches classiques de la bioindication sont insuffisantes

Les bioindicateurs sont des organismes ou ensembles d'organismes qui - par référence à des variables biochimiques, cytologiques, physiologiques, éthologiques ou écologiques - permettent, de façon pratique et sûre, de caractériser l'état d'un écosystème et de mettre en évidence aussi précocement que possible leurs modifications naturelles ou provoquées Blandin (1986). On peut donc caractériser des bioindicateurs à différentes échelles de complexité.

- les biomarqueurs sont identifiés au niveau cellulaire et tissulaire. Des polluants toxiques présents dans le milieu peuvent pénétrer dans les organismes et modifier le fonctionnement cellulaire, en fonction de l'intensité et de la durée d'exposition. Des modifications de composition, de concentration ou d'activité de certaines substances -marqueurs biochimiques - peuvent alors être détectées par des dosages spécifiques et les biomarqueurs sont des indicateurs d'exposition à des substances toxiques.
- les bioessais sont pratiqués sur des organismes tests. Mis en oeuvre dans un contexte expérimental selon des procédures standardisées, les bioessais mesurent la réponse physiologique ou comportementale induite par un polluant toxique sur un échantillon d'organismes sélectionnés d'après leur sensibilité au toxique. Les grandeurs les plus souvent utilisés sont les taux de mortalité et de reproduction, l'activité respiratoire, la mobilité chez des invertébrés et des poissons, etc...

Ces deux premières catégories d'indicateurs biologiques relèvent de l'éco-toxicologie qui peut être définie comme "discipline étudiant qualitativement et quantitativement les effets néfastes des polluants chimiques sur les écosystèmes, pouvant aller jusqu'à l'examen des impacts sur l'homme.

- les indicateurs biocénotiques sont basés sur la structure des peuplements, et se situent donc au niveau des communautés. Ils s'inscrivent dans un cadre conceptuel différent de celui de l'écotoxicologie dans la mesure où le système naturel est l'objet central de la connaissance pour l'hydrobiologie et l'écologie des milieux aquatiques.

Plusieurs approches ont été proposées pour décrire la structure biologique des communautés. En formulant l'hypothèse que les milieux de meilleure

qualité abritent une faune plus variée, la première approche a vu la mise en place d'indices de richesse spécifique ou de diversité, tels que l'indice de Shannon, afin de caractériser l'état du milieu par la richesse en espèces. Cependant, il a été montré qu'une légère détérioration du milieu pouvait aussi favoriser la cohabitation d'un plus grand nombre d'espèces. C'est la théorie des perturbations intermédiaires (Connell, 1978) et cela rend l'utilisation de ces descripteurs difficile. Ensuite, l'utilisation de taxons ou d'espèces cibles, identifiées comme répondant à la modification d'un aspect donné de l'écosystème a été proposée. Cependant ces méthodes fondées sur la présence/absence d'un petit nombre d'espèces cibles ne permettent souvent de caractériser que des états très contrastés (notamment pour le compartiment invertébrés) et peinent à apporter des réponses lorsque les variations sont plus fines. De plus, dans le cas du compartiment poisson, la forte plasticité (capacité à accepter des conditions environnementales diverses) des espèces cibles, notamment chez les individus adultes de cyprinidés, fait que les situations discriminées ainsi sont en général assez triviales.

Dans le cas du site de Bugey, comme pour un grand nombre de masses d'eau, l'état écologique doit être discriminé plus précisément et ce travail nécessite le développement de modèles d'étude plus complexes.

### **2.1.2 Les assemblages piscicoles en grands cours d'eau : un système peu étudié**

#### **Approches bibliographiques**

L'étude des assemblages nous met face à un niveau de complexité important, et les choix de représentation révèlent des parti-pris importants.

Face à cette complexité, les approches sont majoritairement très descriptives. Historiquement, les études de la structure des assemblages se concentraient essentiellement sur la description de la structure attendue, connaissant la position du site au sein de grands gradients régionaux. Il s'agissait donc d'une vision très globale et stationnaire des déterminismes de réponse des assemblages. Les grands descripteurs utilisés étaient la distance à la source et la température (Huet, 1949; Vannote et al., 1980), et se révèlent insuffisamment fins car n'intégrant pas de variables décrivant l'environnement local pour expliquer les variations fines d'assemblage.

Depuis le début des années 90, de nouvelles approches sont apparues afin de mieux prendre en compte les déterminismes locaux de la varia-

tion de structure des assemblages. Actuellement, deux grandes approches dominant le champ bibliographique de l'étude des communautés, cherchant toutes deux à introduire une meilleure prise en considération de la complexité du système :

– description par une méthode multimétrique :

Une métrique est une mesure calculée qui représente un ou plusieurs aspects de la structure, de la fonction ou toute autre caractéristique de l'assemblage biologique et qui varie avec l'augmentation des impacts anthropiques. C'est donc une caractéristique d'un assemblage qui évolue en réponse à un gradient de pression et dont on espère que le calcul permette de préciser la position du site étudié sur ce dit gradient. Un indice multimétrique n'est qu'une pondération des résultats obtenus pour différentes métriques afin de constituer une note qui permet de déterminer l'état plus ou moins impacté du site. On voit tout de suite les limites de cet indice multimétrique, qui, intégrant les effets des différentes pressions, ne permet plus de discerner quelle pression est responsable de quelle modification dans la structure biologique. Trebitz et al. (2003) ont montré que de tels indices manquent de sensibilité aux variations fine du milieu. L'analyse de métriques simples et sélectionnées sur la base de grandes caractéristiques biologiques bien choisies pourrait permettre de mieux discerner les différents effets.

– description statistique multivariée :

Les approches multivariées reposent sur des modèles statistiques de corrélation visant à comparer les assemblages observés sur les sites de test à ceux que l'on observe sur des sites de référence non, ou faiblement, impactés. Plusieurs méthodes statistiques sont utilisées dans ce contexte, de l'Analyse par Composantes Principales (Lima-Junior et al., 2006) aux réseaux de neurones (Gozlan et al., 1999). En utilisant ces méthodes, certaines études favorisent aussi la recherche de rupture dans les structures et leur explication par des événements concomitants (Cattanéo et al., 1999b,a).

### **Les assemblages sont peu étudiés selon notre approche**

L'utilisation de ces méthodes descriptives ne permet pas de rendre compte de l'évolution dynamique des assemblages, dimension pourtant très importante de l'intégrité du milieu. Dresser des corrélations par suivis de séries temporelles ou par mesure de déviation entre sites est insuffisant pour répondre aux questions posées dans une perspective de prédiction ou de prévision. La plupart des études ne disposent que de

Tableau 2.1 – Nombre d’articles référencés dans le Web of Science pour quelques espèces de poissons (tableau actualisé au 1er juin 2010)

mot clé	nom vernaculaire	nombre de références
“Salmo trutta”	truite fario	3868
“Rutilus rutilus”	gardon	1493
“Leuciscus cephalus”	chevesne	430
“Barbus barbus”	barbeau	219
“Gobio gobio”	goujon	171
“Alburnus alburnus”	ablette	164
“Chondrostoma nasus”	hotu	143
“Leuciscus leuciscus”	vandoise	94
“Alburnoides bipunctatus”	spirlin	30

jeux de données courts dans le temps, ne permettant pas de mettre en évidence les évolutions dynamiques de la structure et du fonctionnement de l’assemblage car ils peinent à appréhender de façon satisfaisante les variations internannuelles de structure et leurs déterminismes environnementaux, qui sont pourtant sans doute importantes.

### **Un déficit de connaissances de la biologie des espèces en présence**

Les espèces de poisson d’eau douce n’ont pas connu, historiquement, le même niveau d’intérêt de la part des communautés scientifiques. Certaines espèces, plus emblématiques car présentant un intérêt patrimonial lié à l’exploitation par pêche, soit professionnelle, soit de loisir, ont été bien plus étudiées. Ainsi, on dispose de bases d’informations très importantes sur la dynamique de population ou l’évolution des traits d’histoire de vie d’espèces telles que la truite, mais les études traitant des mêmes questions pour les espèces de cyprinidés, pourtant grandement majoritaires en grands cours d’eau, sont bien moins nombreuses. une simple recherche bibliographique menée sur le Web of Science avec par mot clé le nom latin de chacune des espèces est éloquent (Tableau 2.1). L’espèce la plus étudiée est le gardon, mais le nombre de publications le concernant n’atteint pas même la moitié de celui des articles sur la truite. De plus, les articles s’intéressant aux cyprinidés sont souvent des articles traitant de plusieurs espèces, au sein d’un assemblage, et non d’une espèce en particulier.

Des études sur les liens entre habitat et population existent, mais elles ont souvent pour sujet la truite ou le saumon et leur habitat dans les cours d'eau de plus petite taille (Gouraud et al., 2001, 2004, 2008; Baglinière and Champigneulle, 1982; Baglinière et al., 2004, 2005). Ces études ne sont pas directement généralisables aux grands cours d'eau. En effet, les deux types d'habitat sont très différents et, ne se situant pas aux mêmes échelles spatiales, n'ont pas les mêmes vitesses de réponse aux perturbations.

### **2.1.3 La prise en compte de l'environnement naturel**

#### **Quantification de l'environnement**

Les régimes annuels hydrique et thermique doivent être décrits de façon synthétique. En effet, la description complète et exhaustive d'un régime annuel revient à intégrer une valeur de température et de débit pour chaque jour, et donc autant de variables, ce qui est manifestement inutilisable. Il est beaucoup plus aisé de limiter la description des régimes à certaines de ses caractéristiques. En général, cinq notions sont particulièrement adaptées (Poff and Ward, 1989; Walker et al., 1995; Olden and Naiman, 2010; Resh et al., 1988; Poff et al., 1997) : magnitude (valeur absolue observée), de fréquence (taux de retour d'une condition donnée), de durée (nombre de jours durant lesquels une condition donnée est observée), de temporalité (à quel moment la condition est observée) et de soudaineté (décrit la rapidité avec laquelle la température ou le débit passe d'une magnitude à une autre). Décrire ces différentes composantes d'un régime permet de le décomposer en variables quantitatives concises, dont l'utilisation résume les effets généraux du régime.

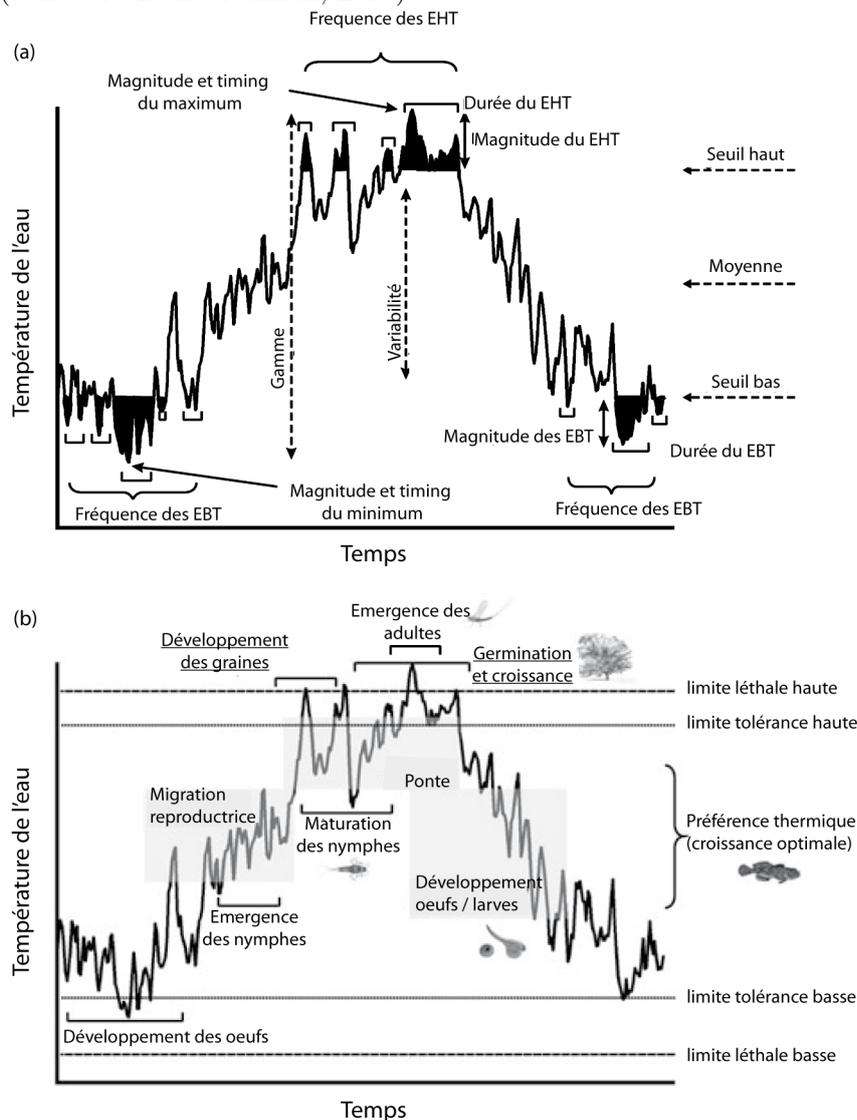
Afin de déterminer les aspects importants à prendre en compte pour les différents régimes, on s'appuie sur la connaissance de la biologie des espèces et de leurs cycles de vie. En effet, les parties les plus intéressantes du cycle de vie sont les phases de grande sensibilité de celles-ci à leur environnement, comme par exemple les phases embryonnaire ou larvaire, ou encor le passage à la nourriture exogène (Balon, 1984). La temporalité de ces phases détermine les périodes d'intérêt des régimes environnementaux. Ainsi, les périodes de ponte, les phases larvaires et les premiers stades de croissance sont particulièrement intéressants à repérer (figure 2.1), car elles correspondent à des périodes déterminant fortement le nombre d'individus annuellement recrutés pour une espèce, soit par le nombre d'oeufs émis, soit par la survie des oeufs et des larves

lors des premiers stades de vie.

En général, on admet que la température de l'eau joue un rôle majeur dans la temporalité de la ponte des poissons, et chez les espèces de cyprinidés peuplant les cours d'eau français, la majorité des espèces commencent à se reproduire lorsque la température de l'eau atteint 12°C (Mills and Mann, 1985). La temporalité de l'épisode de ponte peut potentiellement avoir une forte influence sur le nombre de juvéniles (via des phénomènes de trade-off entre croissance et production d'oeufs chez les adultes, ou des plus faibles taux de survie des juvéniles dans le cas encore d'une moins longue période de croissance car des alevins trop petits survivent moins bien à la période hivernale), et il est donc intéressant d'introduire une variable décrivant la temporalité d'occurrence de cette température seuil de 12°C. De la même façon, le développement des macroinvertébrés aquatiques, qui constituent une part importante de l'alimentation des juvéniles de poisson est un épisode important. La maturation des nymphes s'accroît lorsque la température de l'eau atteint environ 18°C (voir figure 2.1).

De la même façon, Karr (1991) insiste sur l'importance des variations du régime hydrique sur la structure et le fonctionnement, et donc l'intégrité, de la biocénose aquatique (figure 2.2). Le régime hydrique joue un rôle majeur dans la détermination de l'habitat disponible, que ce soit à une échelle temporelle faible pour l'habitat local (repos, alimentation, ...) ou à l'échelle saisonnière pour par exemple la disponibilité d'habitat nécessaire à la reproduction lors des périodes favorables. De plus, la disponibilité de l'habitat sera différente selon le stade de développement atteint par les individus. Les préférences sont différents pour une larve, un juvénile ou un adulte. Les aspects énergétiques sont importants aussi, et le transport de substrat solide par les cours d'eau est au coeur du fonctionnement hydrogéomorphologique des rivières. Le transfert des substrats solides garantit une diversité d'habitats et leur renouvellement temporel. De la même façon, un régime hydrique trop faible, donc à très faible énergie, peut s'accompagner de dépôt dommageable de sédiments fins sur les zones naturellement propices à la reproduction, ou à la production d'invertébrés. Ce phénomène de colmatage perturbe fortement ces fonctions du milieu. Enfin, le régime hydrique joue un rôle fondamental dans la qualité chimique de l'eau en diluant plus ou moins, et en transportant plus ou moins vite vers l'aval les composés chimiques présents dans l'eau. De plus, les sédiments fins sont capables, dans certaines conditions, d'absorber ces composés, influençant directement leur

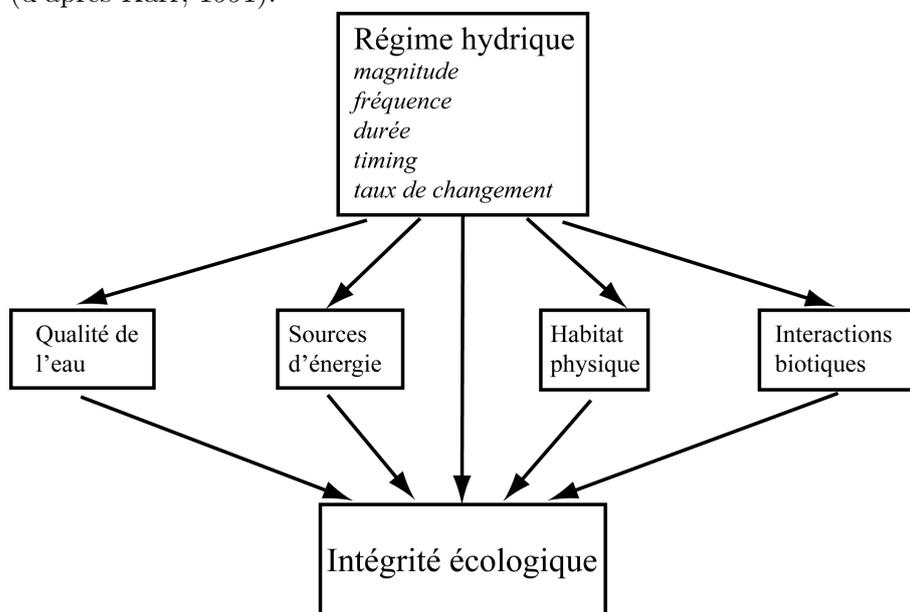
Figure 2.1 – Chronique thermique annuelle et cycles de vie des compartiments de l'écosystème aquatique  
(dans Olden and Naiman, 2010)



(a) Le régime thermique d'un cours d'eau définit la magnitude, la fréquence, la durée, le timing et la variabilité de la température de l'eau à différentes échelles spatiales et temporelles. Quelques descripteurs quantitatifs (d'autres pourraient être définis) de différentes composantes du régime thermique sont montrées : événements de haute température (EHT) et événements de basse température (EBT), fondé sur des seuils de températures hauts et bas prédéterminés. Les seuils peuvent provenir de connaissances écologiques (par exemple températures léthales haute et basse, seuil déclencheur des pontes ou de croissance) ou de distributions statistiques des températures annuelles (par exemple les 10èmes et 90èmes percentiles).

(b) Le régime thermique influence l'écosystème aquatique par le biais de multiples mécanismes qui opèrent à différentes échelles spatiales et temporelles. Quelques exemples de l'influence de la température sur les aspects bioénergétiques et les histoires de vie des poissons (fond gris), des insectes (police normale) et des plantes (souligné). Dans l'ensemble, les biotas natifs des écosystèmes aquatiques ont évolué en réponse au régime thermique naturel.

Figure 2.2 – Influence du régime hydrique sur l'intégrité des assemblages piscicoles (d'après Karr, 1991).



Le régime hydrique est d'une importance capitale dans le maintien de l'intégrité écologique des systèmes aquatiques. Les cinq compartiments du régime - magnitude, fréquence, durée, timing et taux de changement - influencent, tant directement qu'indirectement au travers de leurs effets sur des régulateurs primaires de l'intégrité. La modification du régime hydrique a par conséquent des effets en cascade sur l'intégrité écologique des cours d'eau.

biodisponibilité.

### Importance de la variabilité naturelle

La grande majorité des articles cherchant à lier les assemblages piscicoles à leur environnement se placent dans un contexte d'étude des pressions anthropiques, et très peu d'études sont menées pour évaluer l'importance de la variabilité naturelle, y compris dans les contextes très anthropisés. Pourtant, il semble indispensable d'évaluer, pour une situation de pression donnée, la variabilité résiduelle existant dans la composition des assemblages, afin de ne pas en tirer de conclusions hâtives. Il

est impératif de mesurer les poids respectifs de la structuration naturelle et des pressions.

L'importance des régimes naturels, tant hydrique Poff (1997) que thermique Olden and Naiman (2010), dans le fonctionnement des écosystèmes aquatiques et surtout dans la structuration des assemblages de poissons ont été mis en avant à la section précédente. Les processus écologiques de sélection naturelle et d'adaptation aux conditions locales font que dans un système non perturbé, les assemblages observés sont le résultat de ces conditions locales. Afin d'expliquer l'évolution des assemblages en réponse aux perturbations, on s'intéresse généralement à la déviation de l'assemblage par rapport à cet état initial, qu'il nous faut donc décrire et comprendre. Ainsi, on peut décrire en parallèle l'évolution de l'environnement due aux perturbations, et décrire la façon dont les assemblages répondent à cette évolution. En introduisant des variables décrivant l'environnement hydrique et thermique, on peut chercher à extraire la part de variabilité due aux variations interannuelles naturelles de l'environnement de la variabilité totale, et on peut ensuite se consacrer à l'explication de la variabilité résiduelle, qui sera peut-être expliquée partiellement par les perturbations du milieu.

On constate sur les figures 2.3 et 2.4 que les composantes des régimes hydrique et thermique peuvent être très variables d'une année à l'autre. Par conséquent, il est probable que les comportements induits par ces composantes puissent être décalés dans le temps, favorisés, pénalisés, voire même supprimés, induisant des modifications plus ou moins profondes de la structure de l'assemblage, et ce, malgré des conditions de pression anthropique stationnaires.

### **Des pressions non prises en compte**

En réduisant les régimes naturels à certaines de leurs composantes, on entre de plein pied dans le parti-pris de modélisation. Le choix des variables est très important, et malgré toute l'expertise que l'on peut réunir, il existe toujours un risque de négliger des aspects importants des variations du milieu.

De plus, en considérant que les pressions anthropiques sur le milieu sont localement stables dans le temps, on fait peut-être une hypothèse forte. Il se peut que les pressions à large échelle se répercutant jusqu'au niveau local, telles que l'occupation ou l'utilisation des sols sur le bassin versant qui influencent la physico-chimie de l'eau, la production d'algues et d'invertébrés, soient variables et aient une influence sur la structu-

Figure 2.3 – Exemple de variations interannuelles de débits du Rhône, au niveau de la CNPE de Bugey, entre 3 années (données EdF)

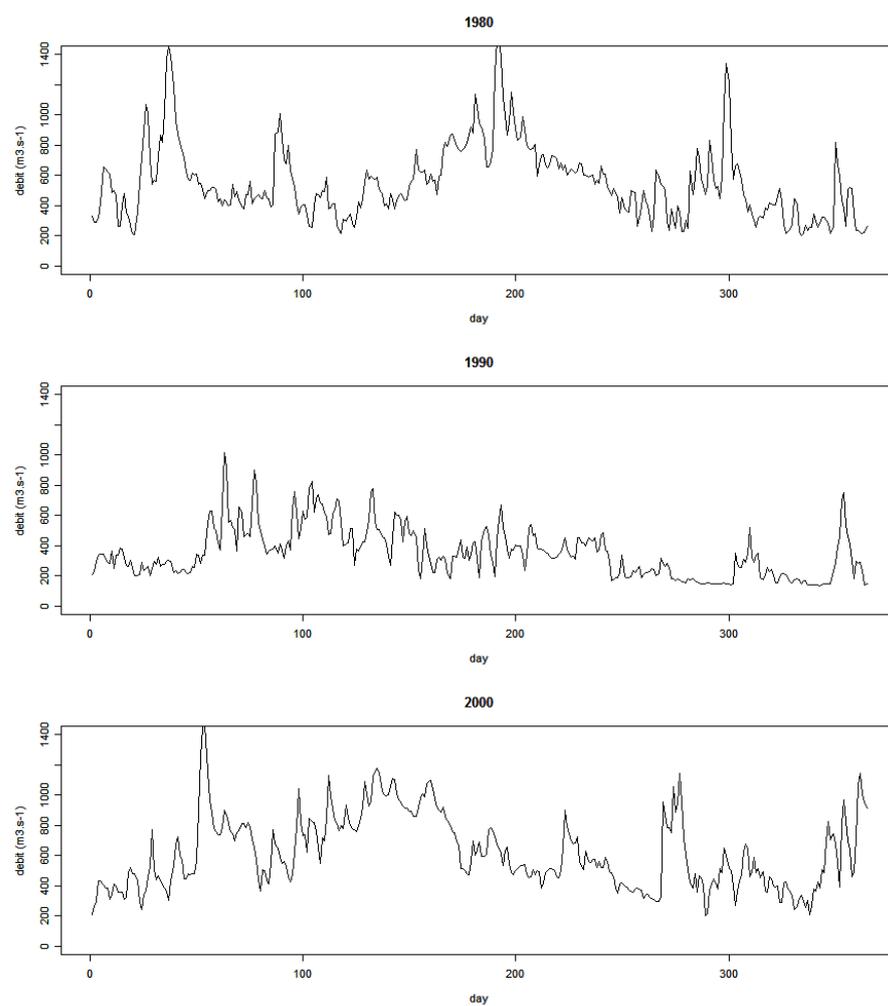
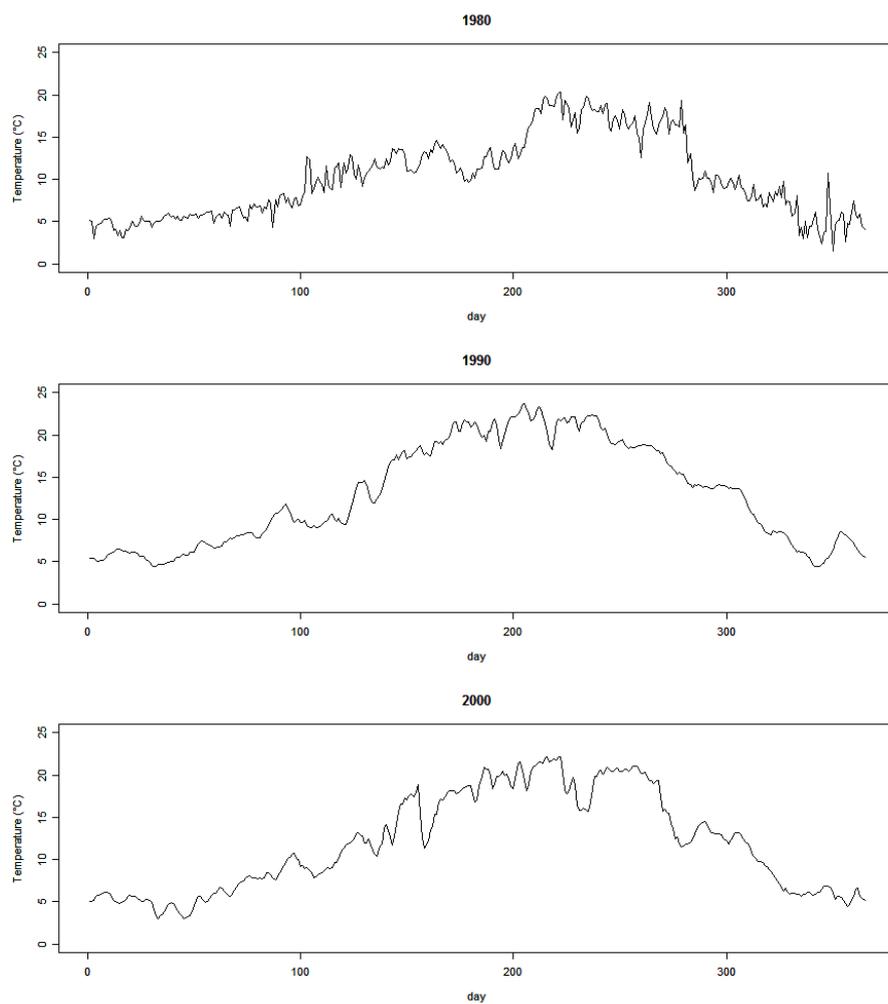


Figure 2.4 – Exemple de variations interannuelles de températures du Rhône, au niveau de la CNPE de Bugey, entre 3 années (données EDF)



ration locale des assemblages. Ces effets ne peuvent cependant pas être pris en compte, faute de données pertinentes disponibles.

#### **2.1.4 L'étude des 0+ est-elle suffisante pour estimer l'état du système ?**

L'importance et l'utilité d'étudier la composition des stades juvéniles a déjà été montrée. Cependant, il convient de considérer si l'étude de ces stades est suffisante pour caractériser le compartiment poisson pour les grands cours d'eau. On sait que les assemblages de poissons adultes sont porteurs d'une information qui intègre toute l'histoire de vie de plusieurs cohortes et les effets successifs des déterminants abiotiques et biotiques, tant en intra qu'en interspécifique. Toutefois, les limitations liées à l'échantillonnage ne permettent pas cette étude, du moins de façon directe. Il faut donc s'intéresser aux possibilités d'inférer la structure des assemblages adultes à partir de celle des assemblages juvéniles.

Une hypothèse forte que nous avons formulée est que le nombre d'adultes présents au moment du frai ne joue qu'un rôle négligeable pour déterminer les effectifs de juvéniles observés. Cette hypothèse, assez forte, d'une *résilience* (capacité d'un système à se rétablir à la suite d'une perturbation) importante du système a pour corollaire qu'il est impossible d'inférer le nombre d'adultes présents une année  $n$  à partir des effectifs de juvéniles observés cette même année. Cependant, les travaux de Mills and Mann (1985) ont mis en évidence que la *force de classe d'âge* (effectif relatif d'une classe d'âge par rapport à l'effectif total d'une espèce) se propage dans le temps et qu'une cohorte importante l'année  $n$ , reste successivement forte les années suivantes. Cela plaide pour un taux de mortalité stable pour les différentes classes d'âge non juvéniles. On pourrait donc envisager d'inférer la structure d'un assemblage adulte de l'année  $n$ , à partir des assemblages juvéniles des années  $n-5$  à  $n-1$ , par exemple.

## **2.2 Les défis de la modélisation**

### **2.2.1 Un protocole expérimental préexistant et non dédié à la question posée**

Le protocole de suivi des populations de poissons du site de Bugey a été mis en place, dès 1977, sans déterminer de plan d'expérience statistique propre. Étant donné que cette étude avait pour but essentiel

de déterminer les éventuels impacts de l'implantation de la CNPE de Bugey sur les populations piscicoles, l'accent a été mis sur une répartition de stations d'échantillonnage entre l'amont (stations prises comme référence, car non impactées) et l'aval (stations de test) de la centrale.

Une sous répartition entre rive droite et rive gauche du Rhône avait également pour but de tester des hypothèses concernant la propagation du panache d'eaux réchauffées par la centrale (mitigation plus ou moins rapide des deux types d'eau ou panache concentré sur la rive d'émission du panache). L'implantation d'une unique station de mesure des conditions environnementales de débit et de température en amont immédiat de la centrale nous oblige à formuler des hypothèses simplificatrices ou des parti-pris de modélisation.

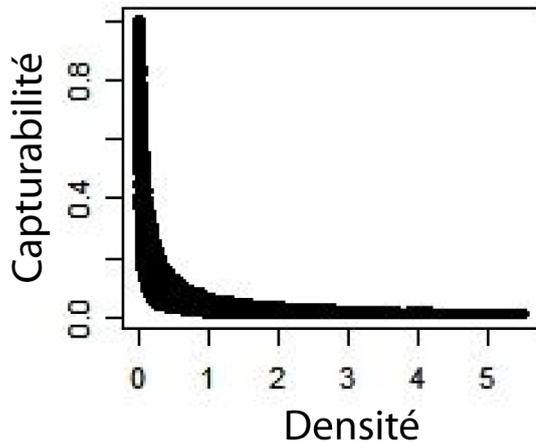
De plus, certaines limitations sont induites par le protocole expérimental lui-même.

### **On ne peut pas faire la différence entre la capturabilité et la taille des populations**

Le protocole d'échantillonnage ne prévoit qu'un seul passage à chaque occasion de pêche pour chaque station. C'est une limitation importante des modèles statistiques possibles. En effet, pour modéliser un échantillonnage aléatoire d'un nombre d'individus  $n$  parmi  $N$  possibles, on utilise en général un modèle du type binomial, régi par un couple  $(N, p)$  de paramètres, où  $p$  est la probabilité de capture de chaque individu. L'intérêt des méthodes d'échantillonnage constituées de plusieurs épisodes de capture, tels que les protocoles de capture-marquage-recapture ou de captures successives, est justement de permettre l'estimation conjointe des deux paramètres du modèle binomial car on a alors un système à deux équations pour deux inconnues.

En présence d'une seule occasion de pêche à chaque échantillonnage, on n'a qu'une seule équation pour deux inconnues et un problème d'identifiabilité : il est impossible de dégager l'abondance totale des espèces, de leur probabilité de capture. On identifie facilement ce problème en vérifiant a priori la capacité d'un modèle simple à inférer la valeur des paramètres sur un jeu de données simulé. Ce travail consiste en la génération d'un jeu de données à l'aide du modèle pour lequel des valeurs de paramètres ont été fixées. Ensuite, à partir de ce jeu de données, on infère les distributions de probabilité de ces paramètres sur lesquels on place alors un prior, selon les préceptes bayésiens. Le graphe de corrélation entre les paramètres est alors caractéristique (Fig. 2.5). En effet,

Figure 2.5 – Graphe des corrélations des paramètres de densité et de capturabilité



un faible nombre d'individus échantillonnés peut s'expliquer soit par un faible nombre d'individus sur le site, soit par une capturabilité faible, et en l'absence d'information supplémentaire, il est impossible de privilégier l'une de ces deux hypothèses.

Une solution serait alors de préciser les priors, et plus précisément celui que l'on place sur le paramètre de capturabilité, afin d'orienter le modèle vers la plus réaliste des deux hypothèses et l'on peut adopter pour cela deux stratégies :

- l'élicitation bibliographique ou à dire d'expert : cependant, les recherches bibliographiques ont révélé un déficit de connaissance important concernant la capturabilité des espèces en grands cours d'eau, avec un protocole similaire à celui de Bugey.
- développer un protocole expérimental dédié de captures successives ou de capture marquage recapture : en raison de grandes difficultés opératoires, notamment liées à l'impossibilité de naviguer dans un système restreint et fermé par des filets, cette option a du être abandonnée.

Par conséquent, il est impossible de travailler sur l'abondance absolue des espèces, mais uniquement sur des indices d'abondance, au facteur capturabilité (par conséquent supposé constant dans le temps) près.

### Les modèles théoriques de dynamique de population multi-spécifiques ne peuvent se contenter d'un indice d'abondance

L'impossibilité de préciser les paramètres d'abondance et de capturabilité empêche le développement de modèles d'écologie théorique trop complexes et l'inférence de leurs paramètres.

Les modèles classiques de dynamique de population reposent sur des équations théoriques. Ce sont souvent des équations différentielles qui décrivent l'évolution au cours du temps de l'effectif d'une population, régie par des paramètres décrivant les caractéristiques populationnelles, telles que la reproduction et la mortalité. On cherche alors à inférer les valeurs de ces paramètres afin de préciser le fonctionnement de la population. Des modèles plus évolués font intervenir un jeu d'équations couplées et permettent de décrire les interactions entre plusieurs populations ou plusieurs espèces. Le plus célèbre de ces modèles est le modèle proie-prédateur de Lotka-Volterra, décrit en 1925. Plus tard, la description du modèle quadratique a permis la prise en compte des interactions biotiques plus fines entre les espèces, telles que la compétition (Beltrami, 1997). Il s'écrit ainsi :

$$\begin{cases} \frac{dN_1}{dt} = N_1.(a_1.N_1 + b_1.N_2 + c_1) \\ \frac{dN_2}{dt} = N_2.(a_2.N_1 + b_2.N_2 + c_2) \end{cases}$$

où  $N_1$  et  $N_2$  sont les effectifs des espèces 1 et 2 ;  $a_1$ ,  $a_2$ ,  $b_1$ ,  $b_2$ ,  $c_1$  et  $c_2$  sont les paramètres démographiques et d'interactions entre les populations. Décrit pour deux espèces, il peut être généralisable à un système multiple d'espèces.

Le modèle quadratique présente l'avantage de pouvoir décrire plusieurs comportements selon les valeurs respectives des paramètres. Il peut tendre vers une cohabitation des deux espèces (compétition faible) ou vers l'extinction de la plus faible (compétition exclusive). Ainsi, son utilisation permet de déterminer la nature des relations entre les différentes espèces, ou les différents groupes d'espèces. Mais il fait intervenir des effets du second ordre, qui sont incompatibles avec le fait de n'avoir accès qu'aux indices d'abondances relatives.

### Un système ouvert

Le système étudié est un système ouvert vers l'amont et vers l'aval. Ainsi, il peut exister des transferts d'individus entre la zone étudiée et les compartiments amonts et aval, qui constituent des effets de fuite et

de recolonisation dans le temps. Ces effets pourraient avoir un rôle important sur la dynamique des populations suivies, mais ne peuvent pas être pris en compte. C'est une contrainte qui existe dans tous les travaux menés sur les populations aquatiques, y compris pour les systèmes plus restreints où des pêches exhaustives sont menées. Dans le cadre de notre étude, ce sont surtout les effets de fuite qui peuvent perturber les résultats. En effet, certains travaux font état d'une dévalaison (déplacement important vers l'aval) précoce des juvéniles de cyprinidés, sans que la question de savoir si ce comportement est génétiquement induit, ou si les conditions environnementales en sont principalement responsables (ou une combinaison des deux hypothèses) ne soit clairement tranchée (Reichard et al., 2002; Zitek et al., 2004,?). En effet, les capacités de nage des jeunes juvéniles étant limitées, une augmentation importante des débits pourrait avoir pour conséquence l'entraînement vers l'aval des larves et des juvéniles. On considérera que si l'hypothèse génétique est vérifiée, alors, ce comportement est indépendant des variations d'environnement et s'effectue en amont de notre étude ce qui ne perturbe pas nos résultats. Si l'hypothèse environnementale est juste, alors, elle rentre pleinement dans le cadre de notre étude, et nous devrions pouvoir en déceler certains indices.

### **Le contraste dans les données**

En l'absence d'un plan expérimental dédié, il faut que les données contiennent des situations suffisamment contrastées pour pouvoir discerner des effets sur des changements peu brutaux. Habituellement, ce sont des phénomènes très forts et marqués qui sont étudiés. Pour garantir ce contraste, deux solutions peuvent être envisagées :

- étudier un site, sur une longue période : dans le cadre des études environnementales par le biais de séries temporelles, on considère en général qu'une vingtaine d'années d'observations sont nécessaires pour mettre en évidence des effets (Humphries et al., 2008).
- étudier plusieurs sites : il faut alors trouver des sites comparables en termes de position dans les grands gradients environnementaux qui déterminent le peuplement global (hydroécocorégion, distance à la source, pente, latitude). En multipliant les stations, on augmente la variabilité des conditions environnementales annuelles observées. Cependant, on introduit également un niveau supplémentaire de complexité dans le modèle et des sources d'incertitude qu'il faut cadrer au mieux.

### **Retour sur le choix des régimes**

Du point de vue de la modélisation, il faut s'assurer que les régimes hydrique et thermique sont des entrées valables et constituent des variables d'impacts essentielles. Il s'agit en effet dans un premier temps de séparer les signaux forts responsables des variations populationnelles afin de progressivement pouvoir introduire des variables traduisant des comportements plus fins.

Les effets directs de la température sont assez bien détaillés dans la littérature scientifique. En effet, de nombreuses études ont montré son importance dans la régulation de l'investissement des individus adultes dans la production et la maturation des gonades au cours de l'automne, dans le déclenchement des comportements reproducteurs (Mills and Mann, 1985), dans le développement embryonnaire et larvaire avec des nombres de degré-jour relativement bien défini et nécessaire pour le passage d'un stade à un autre (Schiemer et al., 2003, 2004). Les possibles effets indirects, par contre, le sont beaucoup moins, et l'importance du régime thermique sur la productivité algale, planctonique et macroinvertébrée, source importante de nourriture pour une grande majorité des poissons (en fonction de leur stade de développement) est encore discutée (Bayley and Li, 1992).

Quant au rôle du débit, de nombreuses questions subsistent quant à son importance et son interprétation : son rôle dans la dévalaison des juvéniles de certaines espèces est encore discuté (Reichard et al., 2002; Zitek et al., 2004,?), mais il est considéré comme une variable de premier ordre dans la détermination de l'habitat disponible pour les différents stades de développement des espèces de poisson. Cette seconde considération est à la base des modèles de microhabitat, qui permettent de transformer un débit donné sur un cours d'eau donné en valeur d'habitat. Toutefois, si cette transformation est très bien documentée sur des cours d'eau de petite taille et des espèces emblématiques telles que la truite, la généralisation à d'autres espèces, mais surtout à des cours d'eau de grande taille, est une question encore ouverte, mais mieux documentée depuis quelques années (Lamouroux et al., 1999).

### **Un effet perturbateur non maîtrisé : la Centrale de Production d'Énergie Nucléaire du Bugey**

La présence de la CNPE du Bugey induit un certain nombre de contraintes dans l'étude du jeu de données. En effet, le fonctionnement

de la Centrale occasionne des modifications des régimes environnementaux. La Centrale prélève de l'eau en amont pour le refroidissement des réacteurs nucléaires. Il y a alors une partie de cette eau qui s'échappe du système sous forme de vapeur d'eau, l'autre partie est rejetée en aval de la station sous forme d'eau réchauffée. Ce sont ainsi environ  $90 \text{ m}^3 \cdot \text{s}^{-1}$  (le volume prélevé est établi en temps réel par le biais d'un modèle théorique de calcul du réchauffement induit, sur la base d'un mélange complet des eaux, et d'un seuil d'élévation autorisé déterminé dans un Arrêté de Rejet) d'eau réchauffées qui sont rejetés. Or, le débit de prise d'eau et la température de l'eau de rejet ne sont pas mesurés et ne peuvent être pris en compte directement dans les données. De plus, la structure de mixtion des eaux froides du Rhône et des eaux chaudes du rejet n'est que partiellement connue. Des études ont montré que le mélange se fait lentement (de l'ordre d'un degré par kilomètre) et que le panache des eaux chaudes se dilue peu latéralement. Ainsi, les stations présentes en rive droite (rive d'implantation de la centrale et du rejet) subissent une élévation de température non mesurée.

Pour tenir compte de cet effet, on peut soit recalculer les températures en aval, en fonction d'hypothèses sur le panache chaud, soit introduire dans les modèles un effet fixe "de position de la station", soit dans le panache chaud, soit en amont. C'est cette seconde solution que nous utiliserons dans la suite des travaux afin de ne pas introduire une nouvelle source d'erreur peu maîtrisée via la reconstruction de données.

### 2.2.2 La plasticité des espèces

Les espèces étudiées ne peuvent pas être considérées une à une comme des bioindicateurs. En effet, ces espèces sont adaptées à la variabilité environnementale et présentent donc des spectres de tolérance assez grands. Cette plasticité des espèces fait qu'en étudiant chaque espèce séparément, on ne pourrait espérer mettre en évidence que des variations très importantes et brutales de régimes environnementaux et il faut s'interroger sur l'opportunité, en étudiant ce cortège d'espèces, d'apporter des éléments de réponse suffisamment fins. Toutefois, on considère le peuplement dans son ensemble afin de déceler, dans les variations relatives des espèces au sein de cet assemblage, les modifications dues à des variations plus fines de l'environnement.

### 2.2.3 Peut-on quantifier par des lois ces phénomènes dont la connaissance est surtout qualitative ?

Le travail de modélisation consiste en une démarche de représentation simplifiée, mais fidèle, du fonctionnement du système étudié. Pour ce faire, il s'agit de bien identifier le phénomène biologique que l'on souhaite étudier, et les hypothèses que l'on peut formuler.

En général, il s'agit de décomposer le système en différents compartiments, observés ou latents, que l'on suppose impliqués dans les phénomènes étudiés. Il s'agit là d'une première difficulté : identifier ces compartiments et les transformer en variables chiffrées. La seconde difficulté réside alors dans la traduction des connaissances sur les phénomènes qui lient les différents compartiments, que ce soient des fonctions biologiques ou bien des représentations de protocoles d'échantillonnage, en équations. Pour ces dernières, la sélection d'une distribution de probabilité pour les différents compartiments concernés est un défi récurrent (Gray, 2005).

Il s'agit donc d'avoir une connaissance suffisante du système biologique, et une maîtrise des outils mathématiques permettant de modéliser son fonctionnement. Les échantillonnages qui servent à contrôler les états simultanés des différents compartiments introduits, ce qui permet de quantifier les équations que l'on place entre eux. Naturellement, plus le nombre de compartiments et de relations les liant augmente, plus le nombre d'échantillons expérimentaux doit être important. Il est donc également nécessaire de trouver la bonne échelle d'étude afin de ne pas introduire de mécanismes trop complexes ou trop fins dans le système, et qui ne puissent pas être suffisamment contrôlés.

Pour conclure, on peut citer Coquillard and Hill (1997) : "le processus de modélisation est donc complexe. C'est une entreprise de longue haleine, parsemée de pièges et de difficultés. L'établissement d'objectifs clairs dès la décision d'élaborer un modèle (qu'attend-on du modèle?) évitera bien des modifications ultérieures hasardeuses. Les choix du niveau d'abstraction et du degré de détail sont cruciaux."

## 2.3 Conclusions

Entre la volonté de développer des modèles complets et très précis, et la disponibilité des données, il est nécessaire de faire des hypothèses simplificatrices pour avancer malgré les difficultés de ma question de recherche.

Ainsi au chapitre 3, nous supposerons :

1. Que le procédé de mesure est propre : la capturabilité est la même pour toutes les espèces.
2. Que les seules variables explicatives du phénomène s'obtiendront à partir de résumés des données température et débits classiquement utilisées en écologie.

Au chapitre 4, nous tenterons de nous appuyer sur des hypothèses plus réalistes :

1. En ne considérant que les ratios annuels de captures, on élimine une partie de la variabilité en travaillant conditionnellement au nombre de captures, qui varie beaucoup d'une année à l'autre.
2. Au niveau des variables d'impacts, au lieu d'imposer notre choix de combinaison linéaire explicative, nous laissons les données parler et construisons elles-mêmes des facteurs explicatifs ad hoc les mieux reliés aux événements de captures.



# La modélisation



## Chapitre 3

# Un modèle linéaire généralisé logPoisson

### 3.1 Description des données

#### 3.1.1 Description des variables d'environnement explicatives

##### Résumer le régime thermique

Nous avons sélectionné cinq variables décrivant la structure du régime thermique annuel sur la période de l'année jugée comme déterminante pour l'installation d'une cohorte annuelle, c'est à dire entre la période de ponte et la période des pêches d'été. Suivant les travaux menés sur le Rhône par Fruget et al. (2006), nous avons retenu trois seuils de température correspondant à des périodes cibles, et nous avons pris comme trois premières variables le nombre de jours séparant le 1er janvier de l'année et le jour de première atteinte de cette température seuil :

- C12 : pour le nombre de jours avant la première atteinte de 12°C. Cette température est atteinte en moyenne au début du printemps et elle correspond plus ou moins à un seuil déclencheur de l'activité de reproduction pour un grand nombre d'espèces (parmi les espèces sélectionnées, seule la vandoise a une reproduction plus précoce, lorsque la température de l'eau atteint à peine 8°C.
- C18 : pour le nombre de jours avant la première atteinte de 18°C. Cette température est atteinte en moyenne à la fin du printemps.
- Cmx : pour le nombre de jours avant l'atteinte de la température annuelle maximum. On considère cette température comme

marquant globalement la fin des grandes contraintes thermiques estivales.

A partir de ces trois valeurs seuils, nous pouvons déterminer deux fenêtres temporelles, correspondant schématiquement au printemps et à l'été (par notation, dans les noms des variables, l'indice 1 sera attribué au printemps, l'indice 2 à l'été). Il est intéressant d'exprimer le caractère de ces deux périodes, en termes de chaleur relative par rapport à une année moyenne, afin de décrire à la fois la durée et l'intensité des contraintes thermiques sur les périodes définies. Pour cela, nous avons déterminé :

- S1 la somme des écarts quotidiens de la température annuelle à la moyenne interannuelle sur la période définie annuellement par C12 et C18, selon la formule, pour la  $i^{\text{ème}}$  année :

$$S1_i = \sum_{j=C12}^{C18} \left( T_{ij} - \frac{\sum_{i=1}^I T_{ij}}{I} \right) \quad (3.1)$$

- S2 la somme des écarts quotidiens de la température annuelle à la moyenne interannuelle sur la période définie annuellement par C18 et Cmx, selon la formule, pour la  $i^{\text{ème}}$  année :

$$S2_i = \sum_{j=C18}^{Cmx} \left( T_{ij} - \frac{\sum_{i=1}^I T_{ij}}{I} \right) \quad (3.2)$$

La figure 3.1 présente ces variables explicatives liées à la température.

### Résumer le régime hydrique

Sur chacune des fenêtres temporelles décrites, on calcule les débits moyens, Qm1 et Qm2, et les débits maximums, Qmx1 et Qmx2. Ces variables simples permettent d'exprimer le contraste entre le débit moyen, qui détermine l'habitat moyen disponible pour les poissons, et la contrainte maximum liée à la plus forte valeur de débit.

La figure 3.2 présente ces variables explicatives résumant le régime hydrique

Figure 3.1 – Variables explicatives résumant le régime thermique (exemple sur l'année 1995). La courbe bleue est la chronique de température de l'année 1995, la courbe rouge est la moyenne interannuelle. Les sommes des écarts quotidiens de température à la moyenne interannuelle sont représentées par les surfaces entre les courbes, en bleu lorsque la température annuelle est inférieure à la moyenne, en rouge lorsqu'elle est supérieure.

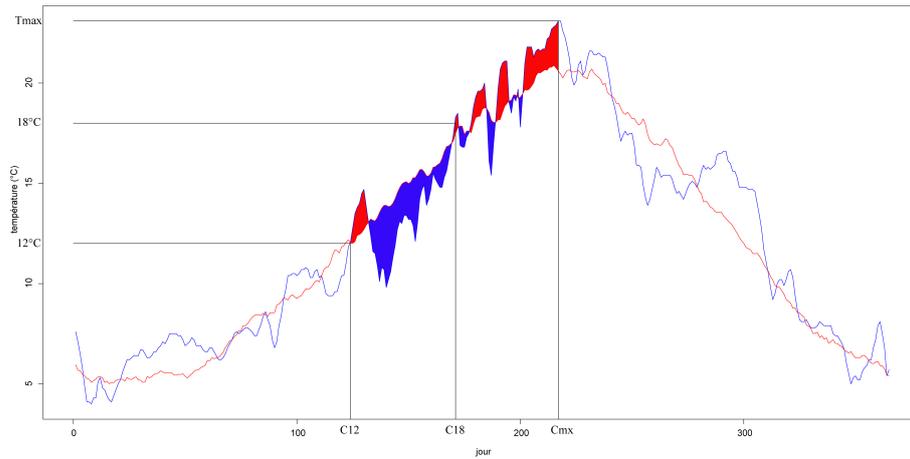


Figure 3.2 – Variables explicatives résumant le régime hydrique (exemple sur l'année 1995). Les barres noires signalent les jours de passage des seuils de température, délimitant les deux périodes temporelles.

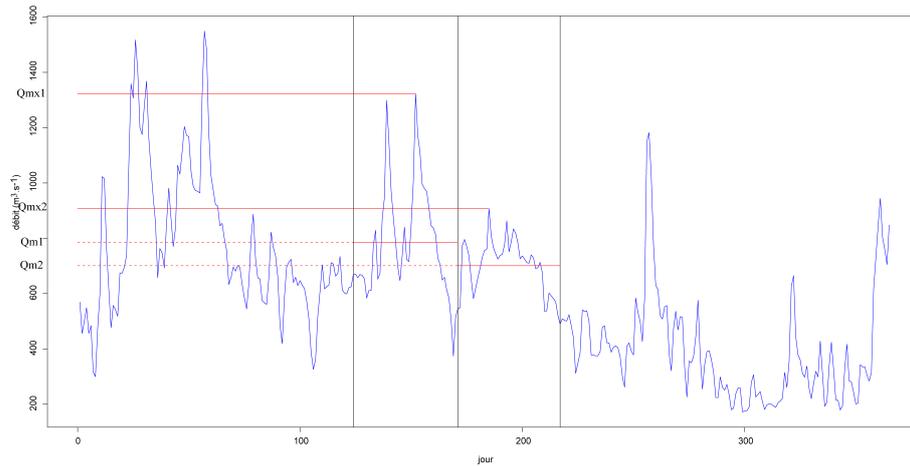
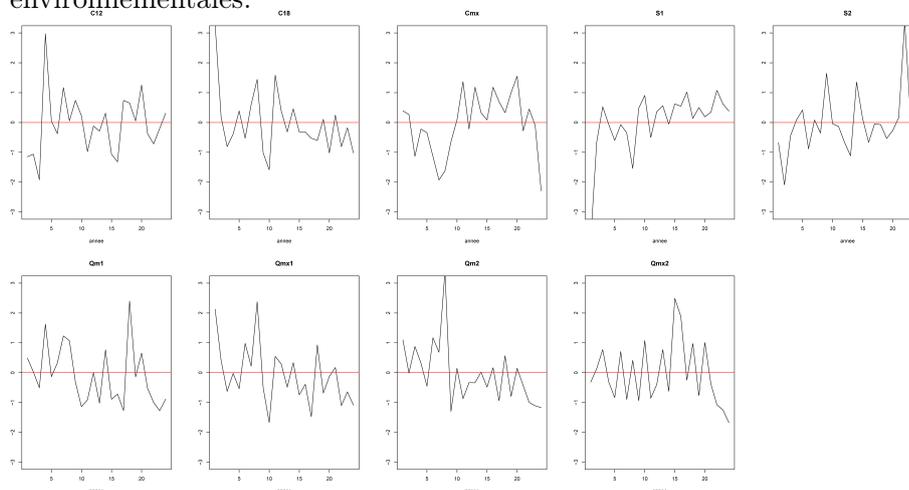


Figure 3.3 – Chroniques des valeurs centrées et réduites des 9 variables environnementales.



### Variations des variables descriptives des régimes thermique et hydrique

Les variations des variables descriptives de l'environnement (notées C12, C18, Cmx, S1, S2, Qm1, Qmx1, Qm2, Qmx2) sont présentées à la figure 3.3. On observe des variations interannuelles très fortes autour de leurs moyennes. Ce sont ces variations interannuelles que l'on cherche à corrélérer aux variations interannuelles d'abondances de juvéniles des différentes espèces de poissons.

### Structure de corrélation entre les variables explicatives

L'écriture d'un modèle linéaire faisant le lien entre une variable à expliquer et plusieurs variables explicatives nécessite au préalable de s'interroger sur l'indépendance entre ces variables explicatives. En effet, les estimateurs des paramètres associés à ces variables explicatives ne seraient alors pas non plus indépendants, pouvant gêner l'inférence, ou pire, fausser les effets relatifs ou leurs tendances. Le tableau 3.1 présente les coefficients de corrélation de Spearman associés à chaque couple de variables explicatives que nous avons défini. On constate que certaines corrélations sont assez élevées. Oberdorff et al. (2002) préconise dans ces cas de supprimer une variable si son coefficient de corrélation avec une autre variable est supérieur, en valeur absolue, à 0.8. Deux couples de va-

Tableau 3.1 – Tableau des corrélations entre variables explicatives

	C12	C18	Cmx	S1	S2	Qm1	Qmx1	Qm2	Qmx2
C12	1.000	-0.251	-0.125	0.130	0.191	0.499	-0.074	-0.023	-0.256
C18		1.000	0.046	-0.861	-0.248	0.239	0.743	0.371	-0.210
Cmx			1.000	0.103	-0.121	-0.194	-0.143	-0.321	0.283
S1				1.000	0.297	-0.419	-0.801	-0.544	0.104
S2					1.000	-0.124	-0.332	-0.367	-0.418
Qm1						1.000	0.645	0.585	0.059
Qmx1							1.000	0.728	0.048
Qm2								1.000	0.384
Qmx2									1.000

riables sont dans ce cas là. Il s'agit de la variable C18 et de la variable S1, ainsi que de la variable Qmx1 et de la variable S1. On pourrait décider de supprimer S1 et conserver ainsi les deux autres variables. Cependant, on s'aperçoit que la variable S1 est assez faiblement corrélée à chacune des autres variables (la corrélation est maximale avec Qm2), tandis que la variable C18 est elle fortement corrélée à la variable Qmx1, qui est elle-même également fortement corrélée avec la variable Qm1. Il paraît donc plus intéressant de supprimer les deux variables, C18 et Qmx1, afin de limiter l'ensemble des corrélations fortes entre les variables explicatives utilisées.

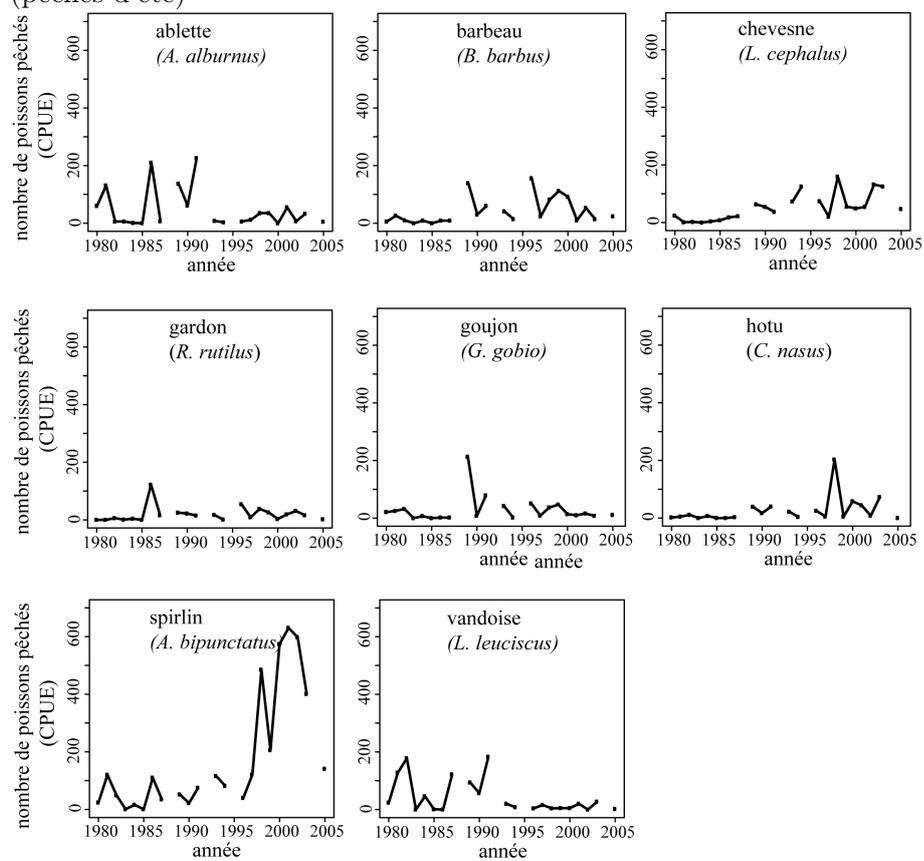
Seules les variables C12, Cmx, S1, S2, Qm1, Qm2 et Qmx2 seront donc ici considérées. Elles formeront les colonnes (avec le vecteur constant) d'un tableau de variables explicatives noté  $X$  dans la suite de ce chapitre.

### 3.1.2 Les variables à expliquer : données d'abondances de juvéniles de cyprinidés.

#### Variations interannuelles des effectifs

Les campagnes d'échantillonnage sur le site du Bugey mettent en évidence des variations interannuelles importantes d'effectifs pêchés des différentes espèces (3.4), de l'ordre d'un facteur 10 à 100.

Figure 3.4 – Chroniques des Captures Par Unité d'Effort par espèces, sommées sur l'ensemble du site (pêches d'été)



**Description du synchronisme des variations d'effectifs entre les espèces : la corrélation de rang, ou le Tau non-paramétrique de Kendall**

Afin de décrire l'influence des variations interannuelles d'environnement sur les variations interannuelles d'effectifs des populations juvéniles de poisson, on peut formuler l'hypothèse selon laquelle des espèces variant de façon synchrone sont sensibles aux variations des mêmes facteurs environnementaux. On pourra ainsi s'attacher à décrire quels facteurs environnementaux sont responsables des variations d'effectif de quelles espèces ou groupe d'espèces. On cherche donc dans un premier temps à décrire la structure du peuplement piscicole de Bugey, en termes de covariations interannuelles d'abondances de juvéniles.

Afin de regrouper les espèces en fonction de leur réponses synchrones, on peut calculer pour chaque couple d'espèce une valeur de la corrélation de leurs effectifs. Ici, nous souhaitons rapprocher les espèces dont les variations interannuelles vont dans le même sens, et nous utilisons donc le Tau de Kendall. On peut utiliser cette mesure non-paramétrique de corrélation car l'échelle de temps sur laquelle les échantillonnages sont considérés, définit elle-même une variable discrète et ordonnée. Pour chaque espèce, on classe alors les années en ordonnant les effectifs de façon croissante. Pour deux espèces données, on regarde si chaque couple de rangs possibles (couples d'années dans notre cas) sont ordonnés dans le même sens. On comptabilise le nombre  $n_c$  de paires concordantes (les couples d'années sont ordonnés dans le même sens), et le nombre  $n_d$  de paires discordantes (les couples d'années sont ordonnés en sens contraire). Le Tau de Kendall est donné par la formule suivante :

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)} \quad (3.3)$$

où  $n$  est le nombre total de couples d'années possibles. La valeur de la corrélation est normalisée par le dénominateur, et  $\tau$  varie entre -1 (les variations sont parfaitement opposées) et 1 (les variations sont parfaitement synchrones).

Cette mesure de corrélation présente l'avantage de confronter les effectifs observés à tous les pas de temps, deux à deux. Ainsi, on a une description à la fois des covariations interannuelles, mais aussi des tendances à plus long terme.

### Une procédure de clustering pour créer des groupes d'espèce "a priori"

La matrice des taux de corrélations obtenus pour chaque paire d'espèce est difficilement exploitable directement. Mais le tau de corrélation de Kendall peut servir de base au calcul d'une distance, utilisable pour générer un dendrogramme des espèces (Fig. 3.5), dans le cadre d'une procédure de clustering. Pour cela, on utilise la méthode de Ward, méthode de classification hiérarchique qui a pour principe la minimisation de la variance intra-groupe (Ward, 1963). La méthode commence avec un cluster par objet et cherche à chaque étape à grouper deux objets au sein d'un même nouveau cluster, sélectionnant les deux objets ayant la plus faible somme des carrés des écarts à la moyenne, selon la formule (pour un cluster X, ayant  $N_X$  valeurs) :

$$SSE(X) = \sum_{i=1}^{N_X} \left| x_i - \frac{1}{N_X} \sum_{j=1}^{N_X} x_j \right|^2 \quad (3.4)$$

et la distance entre deux clusters X et Y est alors, mathématiquement, en notant XY le cluster formé de la jonction des clusters X et Y :

$$D(X, Y) = SSE(XY) - [SSE(X) + SSE(Y)] \quad (3.5)$$

Le dendrogramme représente ces distances entre les espèces, déterminées sur la base de leur réponse synchrone. On peut déterminer ici trois groupes, ou clusters distinct, que l'on interprète en se référant aux évolutions temporelles enregistrées par les chroniques de pêche (Fig. 3.6) :

- le premier cluster regroupe l'ablette et la vandoise, espèces qui après avoir connu des variations fortes d'abondances entre les années 1980 et 1992, alternant des abondances très élevées et des abondances faibles, et pour lesquelles les variations d'effectifs capturés diminuent très fortement après 1993 avec des niveaux de captures très faibles. On se référera à ce groupe comme "groupe des espèces en déclin",
- le deuxième cluster regroupe le barbeau, le hotu et le goujon ; espèces qui semblent être en légère augmentation sur l'ensemble de la période. On se référera à ce groupe comme "groupe des espèces en légère augmentation",
- le troisième cluster regroupe le gardon, le chevesne et le spirilin ; espèces qui, pour le chevesne et le spirilin, connaissent une forte

Figure 3.5 – Dendrogramme des espèces, basé sur la mesure du Tau de Kendall

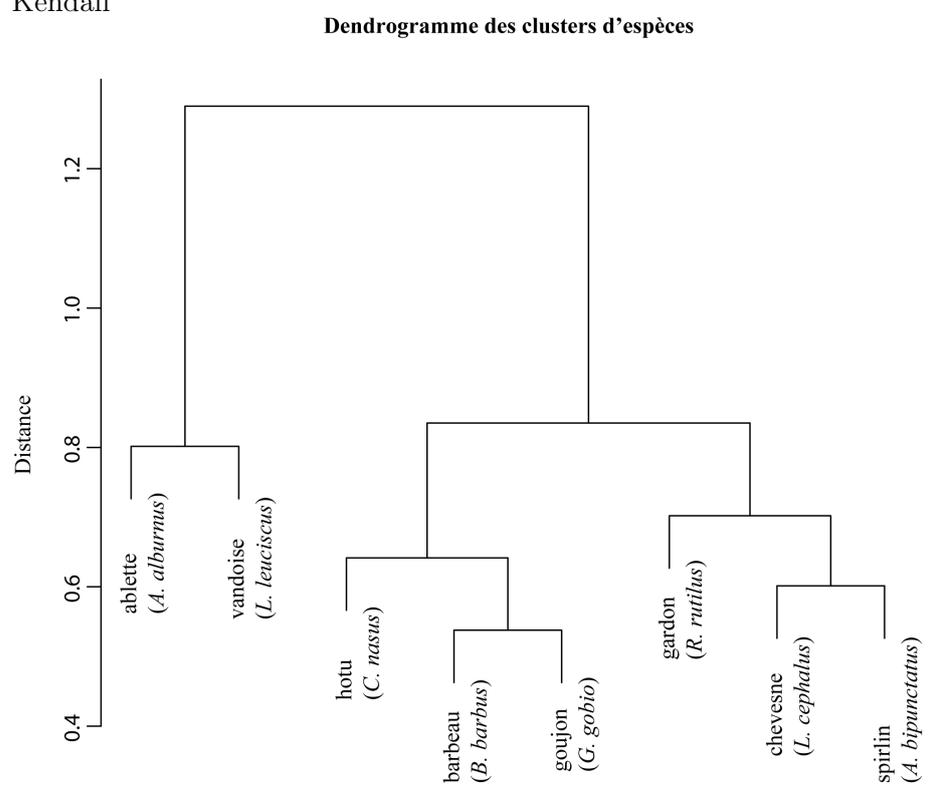
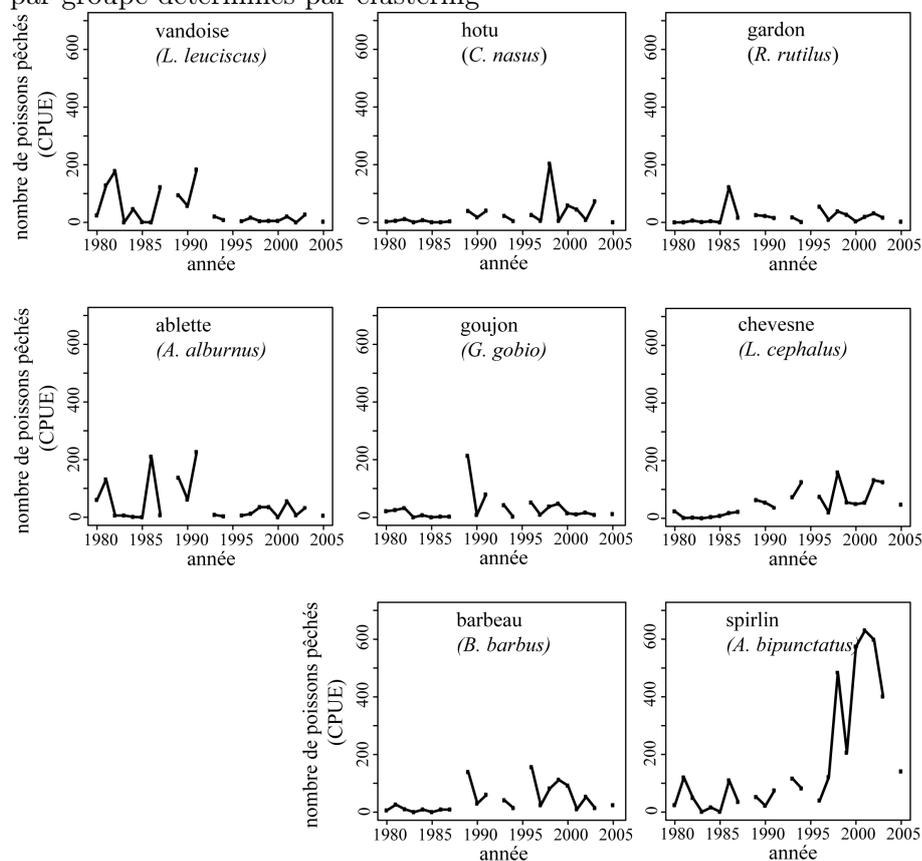


Figure 3.6 – Chronologies des CPUE par espèces (pêches d'été), ordonnées par groupe déterminés par clustering



augmentation sur la période, surtout après la seconde moitié des années 90. On se référera à ce groupe comme “groupe des espèces en forte augmentation”.

Il est intéressant de noter que le cluster 1 est plus éloigné des clusters 2 et 3, que les clusters 2 et 3 ne le sont entre eux. Cela signifie que les espèces des clusters 2 et 3 ont des réponses plus synchrones entre elles, qu’avec celles du cluster 1. La principale différenciation opérée par clustering semble opposer les espèces en déclin aux espèces en augmentation.

## 3.2 le modèle loglinéaire Poisson

### 3.2.1 Une répartition poissonnienne des individus

Dans ce chapitre, nous supposons :

- que l’ensemble des individus de chaque espèce sont répartis de façon aléatoire sur l’ensemble de la zone d’étude,
- que la capturabilité d’un individu est la même quelle que soit son espèce,
- et que cette capturabilité ne varie pas dans le temps.

On se place ainsi dans les conditions d’utilisation d’une loi de Poisson pour modéliser la variabilité associée au protocole d’échantillonnage. On considérera ainsi que le nombre total de poissons présents sur la zone d’échantillonnage se divise en une portion non capturable à laquelle on n’a pas accès avec le protocole expérimental, et d’une portion capturable, que l’on échantillonne. En supposant que la capturabilité est fixe dans le temps, on admet que la portion capturable est un indicateur de l’abondance totale, au facteur “capturabilité” près.

Il existe donc une abondance locale attendue d’individus capturables. Toutefois, il existe des disparités locales d’abondances et l’abondance locale observée n’est pas nécessairement égale à l’abondance locale attendue. On considère que dans ce cas, l’abondance locale observée est une réalisation de l’abondance locale attendue. Les abondances locales observées suivent une loi de Poisson, de paramètre  $\lambda$ , représentant l’abondance locale attendue.

### 3.2.2 La loi de Poisson pour compter des poissons

La loi de Poisson est une loi de probabilité discrète qui décrit l’occurrence d’un nombre d’évènements  $k$  se produisant dans un laps de temps fixé, sachant que ces évènements se produisent avec une fréquence moyenne connue et indépendamment du temps écoulé depuis l’évènement précédent. La loi de Poisson est également pertinente pour décrire le nombre d’évènements dans d’autres types d’intervalles, spatiaux plutôt que temporels, comme des segments, surfaces ou volumes.

Sa densité de probabilité s’écrit :

$$p(k) = p(Y = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!} \quad (3.6)$$

Le paramètre  $\lambda$  de la distribution de Poisson représente le nombre d’entités présentes (dans notre cas, le nombre de poissons) sur la surface

considérée (on peut considérer par exemple le  $\text{km}^2$ ). Ainsi, le paramètre  $\lambda$  représente la densité de poissons attendue.

### 3.2.3 La régression loglinéaire Poisson

Les données de comptage ne sont jamais distribuées selon une loi normale, et une transformation, comme celle de Box-Cox, afin de les normaliser ne peut pas être envisagée avec satisfaction, notamment dans le cas où de nombreux comptages nuls sont observés. De fait, ces transformations, destinées à se ramener au cas d'une régression ordinaire, ne sont pas intéressantes, car bien souvent, une régression loglinéaire offre un modèle simple et approprié.

Ce type de modèle appartient à la famille des modèles linéaires généralisés (McCullagh and Nelder, 1989). C'est un modèle à distribution discrète très souvent utilisé pour modéliser des données de comptage, notamment dans les études médicales, afin de déterminer l'influence d'un traitement, ou d'un facteur extérieur sur le nombre de cas cliniques (Haining et al., 2007). En écologie, c'est un modèle omniprésent dès lors qu'il s'agit d'étudier des abondances d'espèces, avec des protocoles d'échantillonnage par comptage, par exemple pour des oiseaux (Feria and Peterson, 2002; Thogmartin et al., 2004).

La régression loglinéaire est basée sur l'hypothèse de distributions de Poisson indépendantes pour les comptages. Ainsi, on ne considère plus que les résidus  $\varepsilon$  de la régression sont identiquement et indépendamment distribués selon une loi normale. De plus, dans ce type de modèle, ce n'est pas l'abondance locale attendue qui est linéairement dépendante de covariables  $X$ , mais son logarithme. Ainsi, si l'on considère le vecteur  $Y$ , représentant les  $n$  observations d'abondances, et si l'on suppose que l'abondance locale attendue  $\lambda$  varie en fonction de  $X$ , la régression loglinéaire de  $Y$  sur  $X$  s'écrira :

$$\begin{cases} \log(\lambda) &= X \cdot \beta \\ Y &\sim \text{Poisson}(\lambda) \end{cases} \quad (3.7)$$

Le choix de la loi de Poisson comme hypothèse "naturelle" dans ces cas pourrait être débattue, notamment parce qu'il impose une même valeur à la moyenne et à la variance des observations (c'est un trait caractéristique dû à la parcimonie de la représentation fondée sur une loi de Poisson). Quoi qu'il en soit, c'est le modèle standard pour représenter des comptages dans de nombreuses configurations pour lesquelles les

comptages proviennent de l'agrégation de processus ponctuels de Poisson (Royle and Dorazio, 2008).

Le choix d'un modèle loglinéaire est aussi conceptuellement plus satisfaisant, dans la mesure où il conduit à l'écriture d'un modèle à effets multiplicatifs, et non plus additifs, des covariables sur  $\lambda_i$ , l'effectif moyen attendu l'année  $i$ .

### 3.2.4 Les clusters d'espèces comme structuration hiérarchique des modèles

#### Apports du hiérarchique

La communauté des écologues est de plus en plus consciente de l'importance de bien identifier les multiples sources d'incertitude lorsqu'il s'agit de modéliser un phénomène écologique et d'en mener l'inférence (Cressie et al., 2009). Comme souligné par Clark (2005), les modèles bayésiens hiérarchiques traitent la complexité en permettant de subdiviser le problèmes en strates, et met à notre disposition un cadre mathématique fiable, autorisant le traitement de pratiquement tous les problèmes à grandes dimensions. Cette décomposition permet de décomposer et structurer les sources de variations sur les différentes strates. La vraisemblance est toujours au coeur de la structure, en tant que première strate,  $p(y | \theta) \equiv p(\text{observées} | \text{processus}, \text{paramètres})$ . La vraisemblance est le "modèle de données", conditionné par un modèle du processus et des paramètres. Mais on peut écrire des modèles pour eux aussi,  $p(\text{processus} | \text{paramètres})$  et  $p(\text{paramètres})$ . Le modèle complet est donc la distribution jointe des variables, que ce soit des paramètres ou des variables latentes.

$$\begin{aligned} p(\text{paramètres} | \text{modèle}, \text{données}) &\propto \\ p(\text{données} | \text{processus}, \text{paramètres}) & \\ \times p(\text{processus} | \text{paramètres}) & \\ \times p(\text{paramètres}) & \end{aligned}$$

Les variables latentes vont en général être incluses dans la strate du processus. Comme elles décrivent une partie du fonctionnement du système que l'on ne peut observer, elles dépendent de paramètres, et interviennent à leur tour comme une sorte de paramètre dans le modèle de vraisemblance des données.

En ajoutant un modèle décrivant la variation des paramètres (pour tenir compte de phénomènes de similarités liées à des proximités spatiales, par exemple), via une strate supplémentaire définissant des *hyperparamètres*, on introduit de la variabilité dans le processus. On écrit alors

$p(\text{paramètres} \mid \text{hyperparamètres})$ . Dans notre cas, la structure hiérarchique mise en place, tenant compte des évolutions interannuelles communes de certaines espèces, a permis de préciser l'influence des signaux environnementaux sur la structure de l'assemblage juvénile.

Dans le cas où l'on combine plusieurs sources de données observées, ces structures permettent de plus d'introduire le concept d'indépendance conditionnelle. Ainsi, si deux sources de données  $Y_1$  et  $Y_2$  sont indépendantes conditionnellement aux paramètres (c'est-à-dire que les observations de l'une n'ont pas d'influence sur les observations de l'autre, à paramètres fixés), alors, la vraisemblance du couple de variables sera simplement le produit des vraisemblances de chacune des variables.

Pour conclure, la structure hiérarchique autorise des sources de variations à différents niveaux. De plus, elle permet l'écriture d'un modèle cohérent et intégré dont on identifie et quantifie les sources d'incertitudes à chaque niveau (Webb et al., 2010). En travaillant sur des parties de modèles à plus petite dimension, on a en permanence une structure conditionnelle, et, plutôt que de poser la question "Comment fonctionne le processus dans son ensemble?", on regarde "Comment fonctionne cette partie, conditionnellement aux autres éléments qui l'affectent directement?" Les relations complexes, qu'elles soient spatiales, temporelles et à l'échelle individuelle ou du groupe, apparaissent lorsque l'on marginalise (c'est à dire que l'on intègre, donc on propage l'incertitude) sur l'espace des éléments, à l'aide généralement d'outils informatiques (Clark, 2005).

Bien sûr, des structures hiérarchiques peuvent être écrites en dehors d'un cadre bayésien (on pense aux modèles à effets mixtes, par exemple). Cependant, la flexibilité de l'approche bayésienne rend la formulation et la mise en oeuvre de tels modèles bien plus simple et directe, en comparaison de ces autres méthodes (Webb et al., 2010).

### **Application à notre cas d'étude**

Dans notre étude, nous allons supposer que les espèces présentant des réponses synchrones sont partiellement substituables. On peut expliquer cela par l'hypothèse selon laquelle deux espèces évoluant de façon similaire le font sous la contrainte d'un signal environnemental commun. Par conséquent, les trois groupes d'espèces définis à la section 3.1.2 peuvent servir de niveau hiérarchique sur lequel s'exercent les effets des variations de l'environnement.

### 3.2.5 La question du choix de modèles ?

#### Nombre de modèles possibles

Dans le cadre d'un modèle linéaire, le nombre de modèles possibles augmente de façon exponentielle avec le nombre de variables introduites. Ainsi, le fait d'étudier l'effet de  $k=7$  variables explicatives permet d'écrire  $2^k = 128$  modèles possibles, en considérant que l'on peut allumer ou éteindre l'effet de chacune des variables de façon indépendante les unes des autres. Toutefois, cette structure permet d'écrire assez facilement l'ensemble des modèles possibles, sous la forme d'une matrice de 0 et de 1 et donc de simplifier les mises en oeuvre informatiques.

#### Méthode bayésienne de choix de variables explicatives et Facteur de Bayes

Afin de comparer plusieurs modèles et de déterminer lequel d'entre eux reflète le mieux la structure des données, on utilise généralement le Facteur de Bayes (FB) comme critère. Le FB permet de confronter des modèles deux à deux. Considérons deux modèles  $\mathcal{M}_1$  et  $\mathcal{M}_2$ , ainsi que leur probabilité a priori d'être le meilleur modèle  $\Pi_1$  et  $\Pi_2$ . On peut donner également la vraisemblance des données observées, conditionnellement à chacun des modèles, respectivement  $\Pi(Y|\theta, M_1)$  et  $\Pi(Y|\theta, M_2)$ . Le facteur de Bayes s'écrit comme le rapport des vraisemblances marginales des données conditionnellement aux deux modèles en compétition, corrigé par le rapport des probabilités de chacun des modèles a priori. On a :

$$\frac{[M_1 | Y]}{[M_2 | Y]} = \frac{[Y | M_1] \cdot [M_1]}{[Y | M_2] \cdot [M_2]} \quad (3.8)$$

et

$$FB = \frac{[Y|M_1]}{[Y|M_2]} = \frac{[M_1|Y]}{[M_2|Y]} / \frac{[M_1]}{[M_2]}$$

Dans le cas où tous les modèles sont a priori équiprobables, le rapport des vraisemblances marginales, i.e. le FB, donne directement le rapport des probabilités a posteriori des modèles. C'est ce que nous considérerons ici.

Il existe également dans le paradigme fréquentiste des méthodes de sélection de variables, parfois plus rapides que le calcul des FB. Les plus connues et les plus couramment utilisées sont les procédures stepwise, qui sont des méthodes itératives qui, partant d'un modèle donné, proposent à chaque itération d'ajouter ou de supprimer une variable. Le critère AIC

(“Akaike Information Criterion”, défini comme égal à la vraisemblance du modèle pénalisée du double du nombre de paramètres du modèle) est alors calculé pour le modèle proposé. Si l’AIC du modèle proposé est supérieur à l’AIC du modèle actuel, alors le modèle proposé est retenu et une nouvelle proposition a lieu. La procédure s’arrête lorsqu’aucun modèle proche n’est meilleur. Toutefois, la procédure step wise peut se retrouver piégée dans un minimum local des valeurs d’AIC et donc ne pas sélectionner le modèle le plus performant. De plus, le calcul des FB permet de déterminer la probabilité a posteriori qu’a chaque modèle d’être le plus performant (finalement, à l’aide des FB, on ne fait que sélectionner le modèle dont la probabilité d’être le plus performant est la plus élevée). Cela constitue une information intéressante, notamment dans une optique d’utilisation prédictive du modèle, car ces probabilités a posteriori peuvent servir de base à une procédure de model averaging, qui consiste à calculer la somme des valeurs prédites par plusieurs modèles, pondérées par les probabilités de ces modèles. Le model averaging est plus performant que l’utilisation d’un seul modèle prédictif (Raftery, 1996; Raftery et al., 1997; Wintle et al., 2003).

### Application au cas de la régression logPoisson

Considérons un modèle  $\mathcal{M}_\gamma$ , auquel on associe la matrice  $X^\gamma$  des variables explicatives retenues pour ce modèle, et le vecteur  $\beta^\gamma$  des paramètres de régression associés. Dans le cadre d’un modèle logPoisson, on écrit la vraisemblance marginale des observations conditionnellement au modèle  $\mathcal{M}_\gamma$  comme l’intégrale double de la loi jointe des observations et des paramètres, définie sur les espaces des paramètres :

$$[Y|\mathcal{M}_\gamma] = \int_B [Y, \beta^\gamma | X^\gamma] d\beta \quad (3.9)$$

En considérant que les observations et les différents paramètres sont conditionnellement indépendants, on peut écrire :

$$[Y|\mathcal{M}_\gamma] = \int_B [Y|\beta^\gamma, X^\gamma] \cdot [\beta^\gamma | X^\gamma] d\beta \quad (3.10)$$

où l’on prendra une distribution poissonnienne des I données obser-

vées, conditionnellement aux paramètres :

$$[Y|\beta^\gamma, X^\gamma] = [Y|\beta^\gamma, X^\gamma] = \prod_{i=1}^I \frac{\exp(X^\gamma \beta^\gamma)_i}{y^i!} \exp(-\exp(X^\gamma \beta^\gamma)_i) \quad (3.11)$$

et une distribution normale des paramètres  $\beta^\gamma$ , centrée sur une valeur  $\beta_0^\gamma$  et de matrice de variance covariance  $V^\gamma$ .  $\beta_0^\gamma$  et  $V^\gamma$  sont ici les hyperparamètres de la loi a priori :

$$[\beta^\gamma|V^\gamma, X^\gamma] = (2\pi)^{\frac{p}{2}} \cdot \left(\frac{1}{\sigma^{2\gamma}}\right)^{\frac{p}{2}} \cdot |X^{\gamma T} X^\gamma|^{\frac{1}{2}} \cdot \exp\left(-\frac{1}{2}\right) \left((\beta^\gamma - \beta_0)^\top (V^\gamma)^{-1} (\beta^\gamma - \beta_0)\right) \quad (3.12)$$

enfin, on supposera un prior de type Zellner, c'est-à-dire que l'on écrira  $V^\gamma$  sous la forme  $(V^\gamma)^{-1} = \sigma^2 (X^{\gamma T} X^\gamma)$  avec une structure hiérarchique gamma de paramètres a et b pour le  $\sigma^2$  :

$$[\sigma^2|X^\gamma] = \left(\frac{1}{\sigma^2}\right)^{a-1} \cdot \frac{b^a}{\Gamma(a)} \cdot \exp\left(-\frac{b}{\sigma^2}\right) \quad (3.13)$$

Dans le cas de la régression ordinaire, ce choix de prior garantit :

- un certain équilibre a priori entre tous les modèles possibles (au sens de la distance de Kullback sur la capacité prédictive a priori de chacun des modèles en compétition),
- la possibilité de poursuivre la comparaison, même en cas de priors impropres ( $a \rightarrow 0, b \rightarrow 0$ ) puisque l'hyperparamètre  $\sigma$  est commun à tous les modèles,
- la mise en commun de tous les modèles dans un cadre prédictif de Bayesian Model Averaging a posteriori.

On obtient alors :

$$[Y|\mathcal{M}_\gamma] = \iint_{B\Sigma} \prod_{i=1}^I \frac{\exp(X^\gamma \beta^\gamma)_i}{y^i!} \exp(-\exp(X^\gamma \beta^\gamma)_i) \cdot (2\pi)^{\frac{p}{2}} \cdot \left(\frac{1}{\sigma^2}\right)^{\frac{p}{2}} \cdot |X^{\gamma T} X^\gamma|^{\frac{1}{2}} \cdot \exp\left(-\frac{1}{2}\right) \left(\frac{(\beta^\gamma - \beta_0)^\top X^{\gamma T} X^\gamma (\beta^\gamma - \beta_0)}{\sigma^2}\right) \cdot \left(\frac{1}{\sigma^2}\right)^{a-1} \cdot \frac{b^a}{\Gamma(a)} \cdot \exp\left(-\frac{b}{\sigma^2}\right) \cdot dB \cdot d\sigma^2 \quad (3.14)$$

que l'on peut écrire :

$$\begin{aligned}
[Y|\mathcal{M}_\gamma] &= \iint_{B\Sigma} \prod_{i=1}^I \frac{\exp(X^\gamma \beta^\gamma)_i}{y^i!} \exp(-\exp(X^\gamma \beta^\gamma)_i) \cdot \\
&(2\pi)^{\frac{p}{2}} \cdot |X^{\gamma T} X^\gamma|^{\frac{1}{2}} \cdot \frac{b^a}{\Gamma(a)} \cdot \left(\frac{1}{\sigma^2}\right)^{\left(\frac{p}{2}+a\right)-1} \cdot \\
&\exp\left(-\frac{b + \frac{(\beta^\gamma - \beta_0)^T X^{\gamma T} X^\gamma (\beta^\gamma - \beta_0)}{2}}{\sigma^2}\right) \cdot dB \cdot d\sigma^2
\end{aligned} \tag{3.15}$$

En utilisant le changement de variable :

$$a' = \frac{p}{2} + a \tag{3.16}$$

$$b' = b + \frac{(\beta^\gamma - \beta_0)^T X^{\gamma T} X^\gamma (\beta^\gamma - \beta_0)}{2} \tag{3.17}$$

on écrit :

$$\begin{aligned}
[Y|\mathcal{M}_\gamma] &= \iint_{B\Sigma} \prod_{i=1}^I \frac{\exp(X^\gamma \beta^\gamma)_i}{y^i!} \exp(-\exp(X^\gamma \beta^\gamma)_i) \cdot \\
&(2\pi)^{\frac{p}{2}} \cdot |X^{\gamma T} X^\gamma|^{\frac{1}{2}} \cdot \frac{b^a}{\Gamma(a)} \cdot \frac{\Gamma(a')}{b'^{a'}} \cdot \\
&\left(\frac{1}{\sigma^2}\right)^{a'-1} \cdot \frac{b'^{a'}}{\Gamma(a')} \cdot \exp\left(-\frac{b'}{\sigma^2}\right) \cdot dB \cdot d\sigma^2
\end{aligned} \tag{3.18}$$

$$\begin{aligned}
[Y|\mathcal{M}_\gamma] &= \int_B \prod_{i=1}^I \frac{\exp(X^\gamma \beta^\gamma)_i}{y^i!} \exp(-\exp(X^\gamma \beta^\gamma)_i) \cdot \\
&(2\pi)^{\frac{p}{2}} \cdot |X^{\gamma T} X^\gamma|^{\frac{1}{2}} \cdot \frac{b^a}{\Gamma(a)} \cdot \frac{\Gamma(a')}{b'^{a'}} \cdot d\beta \cdot \\
&\int_\Sigma \left(\frac{1}{\sigma^2}\right)^{a'-1} \cdot \frac{b'^{a'}}{\Gamma(a')} \cdot \exp\left(-\frac{b'}{\sigma^2}\right) \cdot d\sigma^2
\end{aligned} \tag{3.19}$$

On y reconnaît une densité de distribution de type gamma de paramètres  $a'$  et  $b'$  pour le paramètre  $\sigma^{2\gamma}$ , et par conséquent, la seconde

intégrale se somme à 1. On s'est ainsi ramené à une intégrale simple, sur l'espace des paramètres  $\beta^\gamma$  :

$$\begin{aligned}
 [Y|\mathcal{M}_\gamma] &= \int_B \prod_{i=1}^I \frac{\exp(X^\gamma \beta^\gamma)_i}{y^i!} \exp(-\exp(X^\gamma \beta^\gamma)_i) . \\
 &(2\pi)^{\frac{p}{2}} \cdot |X^{\gamma T} X^\gamma|^{\frac{1}{2}} \cdot \frac{b^a}{\Gamma(a)} \cdot \frac{\Gamma(a')}{b^{a'}} \cdot d\beta
 \end{aligned} \tag{3.20}$$

que l'on ne peut toutefois pas calculer analytiquement.

### Méthode de calcul approché de l'intégrale par Importance Sampling

L'Importance Sampling est une méthode numérique de calcul approché d'une intégrale. Pour une fonction  $f$ , définie sur un espace  $\theta$ , si l'on peut tirer directement  $n$  échantillons  $\theta^{(n)}$  selon la loi  $f$ , alors :

$$\int_{\Theta} f(\theta) \cdot d\theta = \lim_{n \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N f(\theta^{(n)}) \tag{3.21}$$

Si on ne peut pas échantillonner dans  $f$ , alors, on utilise une autre loi  $g$ , ou loi auxiliaire, connue, selon laquelle on sait tirer des échantillons  $\theta^{(n)}$ . En effet, on peut écrire :

$$\int_{\Theta} f(\theta) \cdot d\theta = \int_{\Theta} \frac{f(\theta)}{g(\theta)} \cdot g(\theta) \cdot d\theta \tag{3.22}$$

et donc, en appliquant l'équation 3.21, on obtient en tirant un échantillon  $\theta^{(n)}$  de taille  $N$ , selon la loi  $g$  :

$$\int_{\Theta} f(\theta) \cdot d\theta \approx \frac{1}{N} \sum_{n=1}^N \frac{f(\theta^{(n)})}{g(\theta^{(n)})} \tag{3.23}$$

La difficulté est de choisir une fonction  $g$  ad hoc pour le paramètre  $\theta$ . Plusieurs fonction peuvent être proposées. Nous utiliserons ici la loi proposée par Marin and Robert (2007) au chapitre 4, consistant en la loi normale paramétrée à l'aide des estimateurs fréquentistes du maximum de vraisemblance (MLE pour Maximum Likelihood Estimators). Il est en effet rapide, dans un cadre fréquentiste d'obtenir les MLE  $(\hat{\beta}, \hat{\sigma}^2)$  d'un

modèle logPoisson grâce à la routine Glm de R. En considérant cette loi normale comme loi d'importance, on obtient la formule suivante qu'il est possible d'implémenter sous R :

$$\int_{\Theta} f(\theta) . d\theta \approx \frac{1}{N} \sum_{n=1}^N \frac{\prod_{i=1}^I \frac{\exp(X^\gamma \beta^{\gamma(n)})_i}{y_i!} \exp(-\exp(X^\gamma \beta^{\gamma(n)})_i)}{(2\pi)^{-\frac{1}{2}} \cdot \left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}} \cdot \exp(-\frac{1}{2}) \left(\frac{(\beta^{\gamma(n)} - \hat{\beta})^T (\beta^{\gamma(n)} - \hat{\beta})}{\sigma^2}\right)} \cdot (2\pi)^{\frac{p}{2}} \cdot |X^{\gamma T} X^\gamma|^{\frac{1}{2}} \cdot \frac{b^a}{\Gamma(a)} \cdot \frac{\Gamma(a')}{\left(b + \frac{(\beta^{\gamma(n)} - \beta_0)^T X^{\gamma T} X^\gamma (\beta^{\gamma(n)} - \beta_0)}{2}\right)^{a'}} \quad (3.24)$$

### 3.2.6 Vérifier l'hypothèse de distribution Poissonienne grâce aux résidus de Pearson

Pour tester l'hypothèse d'une distribution poissonnienne des données, on utilise un test classique du chi-2 sur les résidus de la régression de Poisson. Considérons un vecteur  $Y$  de  $n$  observations. Sous l'hypothèse  $H_0$ , on considère  $Y$  distribué selon une loi de Poisson de paramètre  $\lambda$ . Alors la variable aléatoire  $e$ , définissant le résidu centré normé comme :

$$e = \frac{Y - \lambda}{\sqrt{\lambda}} \quad (3.25)$$

se comportera comme une loi normale centrée-réduite. Par conséquent, sous l'hypothèse  $H_0$ , la variable aléatoire  $SS$ , définie par :

$$SS = \sum_{i=1}^n e_i^2 \quad (3.26)$$

suit une loi du Chi-2 à  $n$  degrés de liberté.

Ce calcul peut très facilement être implémenté sous WinBUGS, ainsi que le calcul des quantiles approchés de la loi du Chi-2 associée, de la façon suivante. Soit un échantillon  $y^*$  de taille  $n$ , et tiré dans une loi de Poisson de paramètre  $\lambda$ . Alors la variable  $SS.prd$ , définie par :

$$SS.prd = \sum_{i=1}^n \left(\frac{y_i^* - \lambda}{\sqrt{\lambda}}\right)^2 \quad (3.27)$$

permet de calculer empiriquement les quantiles de la loi des sommes des carrés des résidus. On vérifie qu'elle suit asymptotiquement une loi du Chi-2 à  $n$  degrés de liberté.

### 3.2.7 Une méthode d'estimation de la capacité prédictive du modèle

Dans une optique de gestion des milieux, la capacité d'un modèle à prédire correctement l'évolution du système d'intérêt lorsqu'on le confronte à une évolution nouvelle (selon des scénarios de gestion, par exemple) est particulièrement intéressante. Cette utilisation fait partie intégrante de l'exploitation envisagée des outils de modélisation dans le cadre de la DCE. Cependant, avant d'exploiter ainsi les modèles, il s'agit de valider leur capacité à prédire les évolutions sur des jeux de données tests. L'étape de validation repose sur la confrontation des prédictions de quantités inconnues par le modèles à ces mêmes quantités observées par échantillonnage.

Pour obtenir des prédictions pour des quantités inconnues, il est nécessaire d'écrire la distribution conditionnelle a posteriori de cette quantité, conditionnellement au modèle choisi et à ses paramètres. Pour cela, il est recommandé d'avoir deux jeux de données indépendants. Un premier jeu, en général le plus conséquent, sur lequel l'inférence statistique, depuis le choix de modèle jusqu'à l'explicitation des distributions a posteriori des paramètres, a permis de conduire l'inférence des paramètres du modèle. Le second pour lequel on écrit la distribution conditionnelle a posteriori des quantités inconnues, conditionnellement au modèle inféré et aux nouvelles valeurs observées des covariables permet d'évaluer la performance prédictive.

Plusieurs solutions peuvent être envisagées pour obtenir ces deux jeux de données. La méthode "leave one out" préconise de réaliser l'inférence sur le jeu de données auquel on retranche un échantillon de taille variable. On calcule alors les valeurs prédites, et leurs intervalles de confiance, par le modèle pour cet échantillon et on compare les données observées et ces intervalles prédits. Cette méthode a l'avantage de pouvoir être menée de façon itérative, en faisant varier à chaque itération l'échantillon retiré du jeu de données. En multipliant ces tests de la capacité prédictive, on améliore la précision du test. Cette méthode présente l'avantage d'assurer que les prévisions sont faites dans le domaine d'application du modèle, puisque les données utilisées pour l'inférence et celles utilisées pour le test de la capacité prédictive sont étroitement mêlées.

Une autre méthode consiste à tester le modèle sur des nouvelles données, soit temporellement, soit spatialement séparées des données utilisées pour l'inférence. Dans le cadre de ce travail, ce sont les données piscicoles et environnementales pour les années 2006 à 2009 qui ont été

utilisées. En faisant l'hypothèse que les liens entre les conditions environnementales et la réponse des espèces de poissons n'ont pas radicalement changées entre l'année 2005 et les quatre années suivantes, on considère que l'on ne s'éloigne pas du domaine d'application du modèle, et que celui-ci peut être utilisé pour prévoir les peuplements piscicoles successifs.

Ainsi, nous proposons de réaliser les travaux de choix de modèle et d'inférence sur le jeu de données des années 1980 à 2005, et de prédire les abondances des différentes espèces pour les années 2006 à 2009. Pour cela, il est nécessaire de calculer les indicateurs des régimes thermique et hydrique pour ces 4 dernières années. Comme les variables utilisées pour l'inférence sont centrées et réduites, il faut également centrer et réduire les variables environnementales utilisées pour la prédiction. Cependant, afin de conserver pour ces nouvelles variables la même unité de variation autour de la même valeur moyenne que pour les variables utilisées pour l'inférence, chaque variable environnementale calculée pour les années 2006 à 2009 a été centrée et réduite à l'aide de la moyenne et de la variance calculées sur les années 1980 à 2005. Donc, pour déterminer les valeurs  $X_j'^{CR}$  centrées et réduites de la covariable  $X'$ , définie pour  $j$  variant de 27 à 30, on utilisera la formule :

$$X_j'^{CR} = \frac{\left( X_j' - \frac{1}{26} \sum_{i=1}^{26} X_i \right)}{\frac{1}{26} \sum_{i=1}^{26} \left( X_i - \frac{1}{26} \sum_{i=1}^{26} X_i \right)^2} \quad (3.28)$$

Alors, considérons les vecteurs  $X_*$  des covariables retenues dans le meilleur modèle, et  $\beta_*$  les paramètres de régression associés. Soit  $Y^{pred}$  le vecteur prédictif des nouvelles observations d'abondance, alors on a :

$$\begin{cases} \log(\lambda') & = X_*'^{CR} \cdot \beta_* \\ Y^{pred} & \sim Poisson(\lambda') \end{cases} \quad (3.29)$$

Techniquement, le tirage de l'échantillon  $\beta_*$  est réalisé lors de l'inférence, en utilisant les valeurs successives de la chaîne de Markov sous-jacente, une fois celle-ci stationnaire. Les calculs des distributions prédictives des  $\beta_*$  et des  $Y^{pred}$  sont donc simultanés.

Une fois les distributions des  $Y^{pred}$  obtenues, on leur confronte les données observées lors des échantillonnages. Un indicateur très simple de la qualité prédictive est alors le pourcentage d'observations appartenant à

un intervalle de crédibilité donné. On peut calculer les taux de couverture effectifs pour tous les intervalles de confiance possibles, en particulier, on utilise par convention l'intervalle de crédibilité à 95%.

### 3.3 Application au cas du Rhône, sur le site d'échantillonnage de Bugey

#### 3.3.1 Un premier modèle pour stationnariser l'effet centrale

##### Description du modèle

###### *Structure hiérarchique*

La structure du modèle est donnée sous forme de Graphe Acyclique Orienté (on utilise en général les initiales DAG, pour Directed Acyclique Graph) à la figure 3.7. Il s'agit d'un modèle dont les paramètres de régression sont définis pour chaque groupe d'espèces. Les effets environnementaux sont donc considérés comme identiques pour toutes les espèces d'un même groupe à la section 3.1.2. Cependant, si ces espèces appartenant au même groupe varient de façon similaire en réponse aux variations d'environnement, leurs abondances moyennes peuvent être très éloignées. On introduit donc une constante de normalisation  $Z_{0i}$ , définie indépendamment pour chaque espèce  $i$ , représentant le logarithme de l'abondance moyenne interannuelle spécifique.

Une variable latente  $Z_{i,j,k}$  représente l'abondance spécifique locale attendue pour chaque année  $j$ , chaque espèce  $i$  et en fonction du groupe  $k$  auquel l'espèce  $i$  appartient. Son logarithme est défini pour chaque espèce comme :

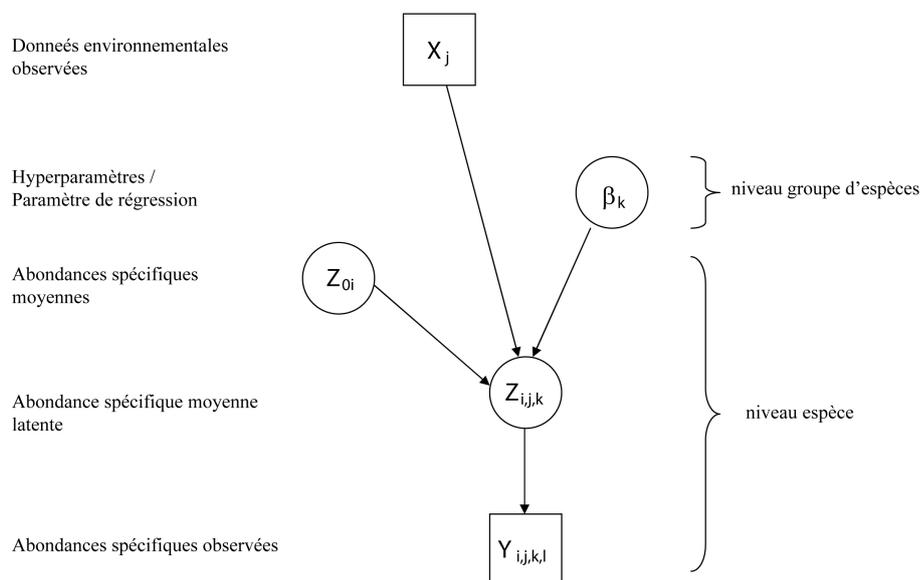
$$\log(Z_{i,j,k}) = Z_{0i} + \beta_{k(i)} \cdot X_j \quad (3.30)$$

Les abondances de juvéniles observées sont donc des réalisations d'un tirage dans une loi de Poisson de paramètre  $Z$ .

###### *Comment stationnariser la perturbation due à la CNPE ?*

Dans un premier temps, nous allons nous intéresser à l'influence des variables d'environnement sur les abondances de juvéniles, sans introduire d'effets perturbateurs supplémentaires. Nous ne prendrons donc pas en compte l'effet d'élévation de la température de l'eau par la CNPE. Cependant, nous ne pouvons ignorer la présence de cette dernière et nous

Figure 3.7 – Graphe Acyclique Orienté du modèle à effet central stationnarisé.



ne pouvons pas considérer que toutes les stations sont équivalentes. La solution la plus simple pour supprimer l'effet de la CNPE est de sommer sur l'ensemble des huit stations les effectifs de chaque espèce, à chaque occasion de pêche. Ainsi, on peut considérer que les variations interannuelles d'effectifs spécifiques sur l'ensemble du site ne sont pas influencées par la CNPE, ou plus précisément que l'effet station du à l'influence de la CNPE a été moyenné et donc éliminé.

### **Modèles sélectionnés et inférence sur les paramètres de la régression**

La méthode décrite aux sections 3.2.5 et 3.2.5 est appliquée à ce modèle afin de déterminer, pour chacun des groupes, le modèle le mieux ajusté aux données.

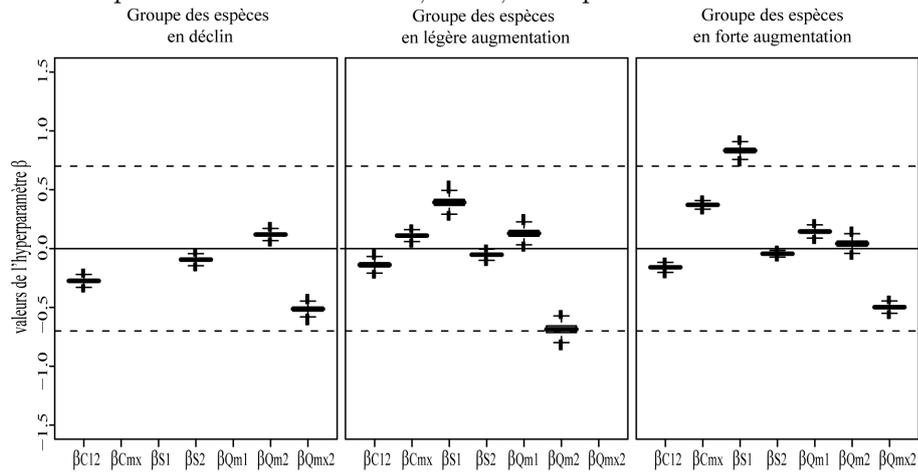
Pour chaque meilleur modèle retenu, l'inférence bayésienne sur les valeurs des paramètres de régression est menée sous WinBUGS 1.4 (Lunn et al., 2000). Conformément au paradigme bayésien, des priors sont attribués à chacun des paramètres inconnus du modèle. En l'absence d'informations complémentaires concernant les valeurs possibles des paramètres, des priors vagues ont été utilisés. Les paramètres de régression  $\beta_k$  et les constantes de normalisation  $Z_0$  suivent a priori des lois normales  $N(0, 10^{-3})$ . Le script du modèle sous WinBUGS est donné en annexe B.

La convergence du modèle est vérifiée à l'aide du test modifié de Gelman-Rubin (Brooks and Gelman, 1998). Elle est rapidement atteinte (le nombre d'itérations nécessaires est de l'ordre de quelques milliers) et par sécurité, une période de burnin de 50.000 itérations est employée avant de réaliser les 10.000 itérations retenues pour l'inférence.

La figure 3.8 présente à la fois les variables retenues dans les modèles sélectionnés pour chacun des groupes spécifiques (la présence d'un boxplot signifie que la variable correspondante est retenue) et leurs densités de distribution sous la forme de boxplots. On s'aperçoit que les deux clusters qui étaient les plus proches ont des structures de variables sélectionnées assez proches, et que les effets conjointement sélectionnés vont en plus dans le même sens. La structure du modèle sélectionné pour le premier groupe est plus éloignée des deux autres.

Figure 3.8 – Boxplot des densités de distribution des paramètres associés aux variables retenues dans les modèles sélectionnés sur la base du Facteur de Bayes.

La ligne noire représente un effet nul pour les variables, les lignes pointillées représentent les valeurs  $-0,7$  et  $0,7$  des paramètres.



### Interprétation biologique

Pour une première discussion montrant le caractère très interprétable d'un point de vue écologique des coefficients liant les perturbations environnementales aux réponses des espèces selon les résultats de ce modèle, on se référera à la partie discussion de Piffady, et al. 2010 (in Press), en annexe A.

### Retour sur l'hypothèse de distribution poissonnienne

Les lignes de code correspondant à ce calcul sont présentées dans la partie "Calcul des sommes des résidus au carré" du programme WinBUGS présenté dans l'annexe B. Le tableau 3.2 présente les résultats de ce test. On y voit que l'intervalle de crédibilité approché par simulation d'échantillons tirés dans la loi de Poisson de paramètre  $\lambda$  est très proche de l'intervalle de confiance calculé pour une variable aléatoire suivant une loi du chi-2 à 360 degrés de liberté. A l'opposé, la valeur de la statistique de test calculée pour les données observées et son intervalle de crédibilité sont très éloignés de cet intervalle de confiance. L'hypothèse  $H_0$  d'une distribution poissonnienne des données est par conséquent rejetée et on

Tableau 3.2 – Résultats du test du Chi-2 de la distribution poissonnienne des données

	quantile(0.0275)	mean	quantile(0.975)
Chi-2 (360 ddl)	309.3278		414.4593
SS.prd	309.1	360.1	415.8
SS	23400	23950	24510

en déduit que les données sont surdispersées par rapport à la dispersion attendue d'une distribution de Poisson.

La surdispersion observée indique que la variabilité reflétée par le modèle est trop faible par rapport à la variabilité observée dans les données. Concrètement, cela signifie que le modèle peine à retranscrire les très fortes valeurs d'effectifs observées parfois, tandis que la plupart du temps, les effectifs observés sont faibles. Cela signifie également que la variabilité de l'environnement ne suffit pas à expliquer les grands écarts observés dans les abondances spécifiques. D'autres facteurs environnementaux, non pris en compte, ou d'origine différente, en particulier des facteurs biotiques, pourraient donc également expliquer les fortes variations.

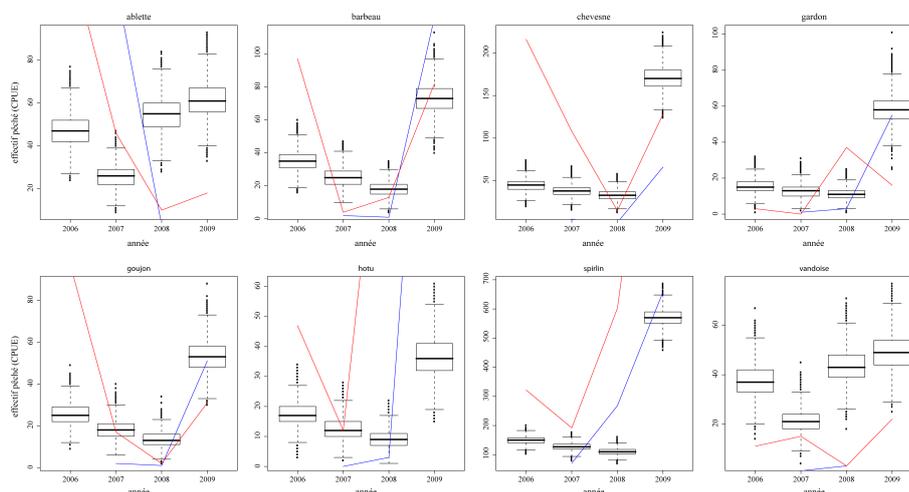
### Capacités prédictives du modèle

La figure 3.9 présente sous forme de boxplot les distributions de probabilité des valeurs prédites par le modèle pour chaque espèce et chaque année. Les courbes de couleur présentent les variations chronologiques d'effectifs observées pour ces espèces lors des échantillonnages ; en bleu pour l'échantillonnage estival, et en rouge pour l'échantillonnage automnal. Selon l'hypothèse formulée pour l'écriture du modèle, les échantillonnages estivaux et automnaux d'une même année sont considérés comme deux répétitions d'une même expérience. On compare donc ces deux valeurs au même intervalle de crédibilité.

Tout d'abord, on constate qu'au sein d'un groupe spécifique donné, et conformément à la structure hiérarchique décrite dans le modèle, les variations chronologiques d'effectifs prédites montrent les mêmes patrons d'évolution. De plus, alors même que cela n'est pas explicite dans le modèle, les espèces des clusters 2 et 3, qui sont les deux clusters regroupés en premier, exhibent des patrons de variations prédites très similaires, en très faible décroissance, voire même stabilisées, sur les années 2006 à 2008 et en très forte augmentation sur l'année 2009.

Figure 3.9 – Comparaison des valeurs observées par espèce et des intervalles prédits par le modèle pour les années 2006 à 2009 sous forme de boxplots.

(trait bleu : été, trait rouge : automne)



De façon générale, les variations interannuelles d'effectifs observés, pour chaque espèce, sont assez proches pour l'échantillonnage estival et l'échantillonnage automnal. On calcule la capacité prédictive comme étant le taux de valeurs observées, parmi les 64 possibles, appartenant à leur intervalle prédit. On comptabilise 9 de ces valeurs observées qui ont été correctement prédites, ce qui équivaut à une capacité prédictive calculée de 0,14. Toutefois, on s'aperçoit également qu'un certain nombre de valeurs observées, si elles n'appartiennent pas à leur intervalle prédit, s'en approchent très fortement et figurent parmi les valeurs d'outliers prédites par le modèle, en considérant que la surdispersion observée n'a pas été prise en compte. En corrigeant celle-ci, c'est-à-dire en élargissant les plages de valeurs prédites et en perdant en précision pour la prédiction, on devrait améliorer la capacité prédictive de ce modèle. C'est un point important, car on s'aperçoit ici qu'en ne prenant pas en compte la surdispersion existant dans les données, on risque de vouloir discriminer des situations d'assemblage qui en fait ne correspondent qu'à une variation aléatoire pour une réalisation commune de l'environnement.

### Discussion autour de ce modèle

Ce premier modèle d'étude de l'influence des variations interannuelles des régimes hydrique et thermique sur l'abondance de juvéniles des poissons du Rhône présente l'avantage majeur d'être très simple à mettre en oeuvre et à implémenter. Le fait de travailler sur des données annuelles sommées sur la totalité des sites d'échantillonnage permet de "stationnariser" un certain nombre d'effets environnementaux qui ne sont pas contrôlés par le plan d'échantillonnage, comme les effets potentiels du rejet d'eau réchauffée, ou bien les différences de qualité d'habitat entre les stations. Vis-à-vis d'un objectif de généralisation de ce modèle à l'étude de nouveaux sites comparables, ces arguments de simplicité et de stationnarité sont importants. Néanmoins, la mise en commun d'expériences menées dans des conditions différentes (que ce soit sur deux saisons différentes ou sur huit sites de pêche) contribue à introduire une hétérogénéité des données dont on se serait volontiers passé.

Il nous faut donc revenir sur l'hypothèse sous-jacente d'une distribution de Poisson pour les données de comptage car la surdispersion des données observées posent problème. En effet, même si la régression est couramment employée pour des jeux de données soit "enflés en zéro", soit violant l'hypothèse d'équidispersion (Cameron and Trivedi, 1998) (notamment en raison de sa simplicité d'utilisation et de sa présence dans les packages de base des logiciels statistiques), deux conséquences majeures de son utilisation dans ces cas sont à souligner. Premièrement, cela donne lieu à des prédictions locales fausses, même si la prédiction moyenne sur l'ensemble des sites sera cohérente avec les observations (Cameron and Trivedi, 1998; Barry and Welsh, 2002). Deuxièmement, cela engendre des conclusions plus optimistes quant à la significativité statistique des variables explicatives, via la réduction fallacieuse de la variance des paramètres associés à ces effets (Potts and Elith, 2006). Dans le cadre de procédures classiques de choix de modèle, comme les procédures stepwise, des variables incorrectes ont plus de chances d'être retenues (Fitzmaurice, 1997). Ces deux conséquences sont particulièrement importantes dans une optique de gestion environnementale, dans la mesure où les prédictions et/ou la structure d'un modèle peuvent influencer la prise de décision.

Dans notre cas, il semble que la procédure de sélection de variables fondée sur l'utilisation du facteur de Bayes ne souffre pas des lacunes du stepwise puisque nous ne cherchons pas le modèle le plus parcimonieux, mais juste le modèle le mieux adapté à la structure des données. Les

conséquences sur la prédiction sont plus préoccupantes et il s'agit par conséquent de corriger au mieux la surdispersion observée et de redonner de la liberté au modèle dans ses prédictions a posteriori. Plusieurs méthodes sont classiquement proposées pour corriger la surdispersion observée dans des données de comptage. Lorsque le protocole expérimental le permet, il peut être efficace de réduire la dimension des cellules sur lesquelles les échantillonnages sont faits. Cette méthode sera efficace essentiellement si la surdispersion est principalement due à l'existence d'un grand nombre de zéros dans le jeu de données (voir figure 3.10). On peut aussi adapter la structure du modèle de mélange, en utilisant soit un modèle adapté aux données enflées en zéro (mélange d'un modèle d'une distribution paramétrique type Poisson et d'une distribution dégénérée en zéro), soit en adoptant une loi de probabilité elle-même issue d'un mélange de lois pour décrire la structure d'erreur du GLM. Un exemple typique de ce mélange de loi est la loi Binomiale Négative, dérivée d'un mélange Poisson-gamma (le paramètre de moyenne de la loi de Poisson n'est pas fixé, mais est tiré dans une loi gamma spécifiée).

### **3.3.2 Un second modèle intégrant la variabilité stationnelle et saisonnière : un essai pour limiter la surdispersion**

#### **Description du modèle**

La structure de ce modèle est donnée sous la forme d'un DAG dans la figure 3.11.

La structure hiérarchique de base du modèle, concernant les effets environnementaux exprimés au niveau du groupe d'espèces reste la même qu'à la section 3.3.1. Toutefois, deux effets à tester sont ajoutés dans la partie loglinéaire du modèle pour contrôler l'influence de la saison et celle de la station..

#### *Introduction de l'information stationnelle de l'échantillonnage pour éliminer la surdispersion*

Dans ce nouveau modèle, on subdivise l'information issue des échantillonnages, ce qui a pour conséquence de réduire le nombre de captures réalisées à chaque occasion. En multipliant le nombre de points échantillonnés mais en limitant alors le nombre de captures, on fait en sorte de permettre l'augmentation du coefficient de variation du modèle de Poisson. Ce rapport de l'écart-type sur la moyenne augmente car il est

Figure 3.10 – Distribution de probabilité d'une variable aléatoire  $a$ , suivant une loi de Poisson, pour différents paramètres  $\lambda$

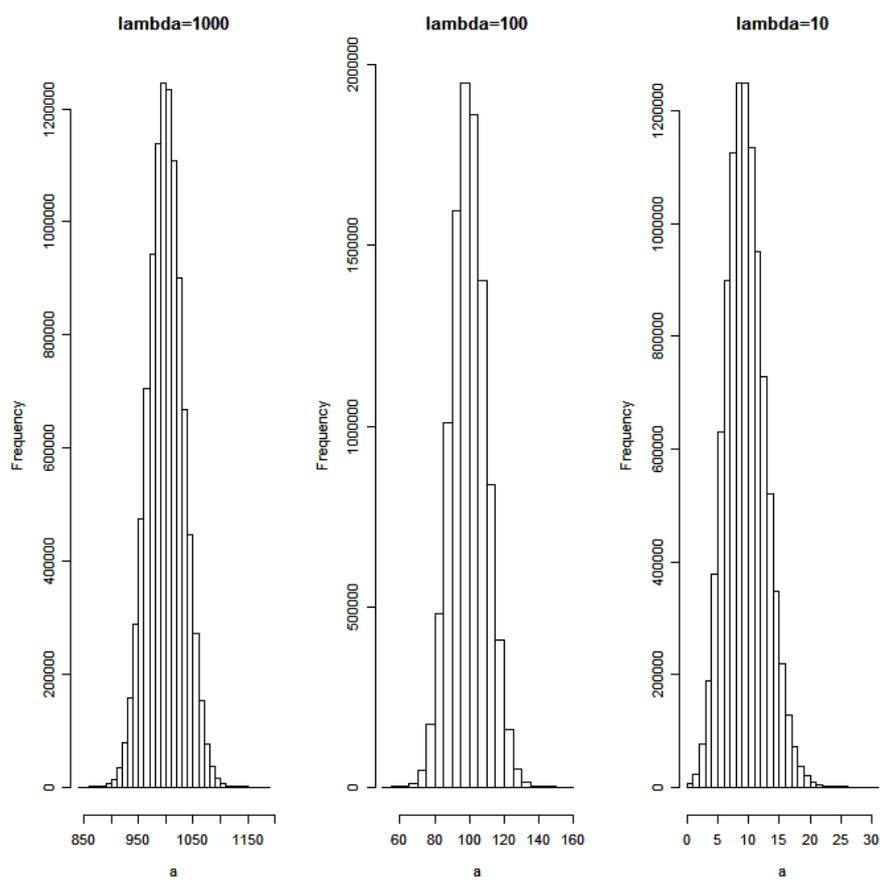
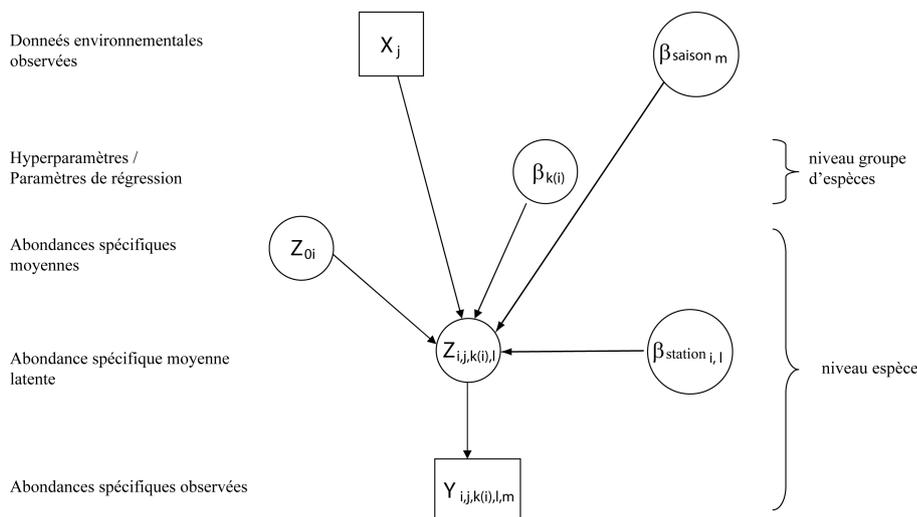


Figure 3.11 – Structure du modèle à effet stationnel, présentée sous forme de Graphe Acyclique Orienté.



l'inverse de la racine carrée du nombre d'individus attendus (que l'on diminue ici). Dans la section précédente, les échantillons réalisés sur chaque station étaient additionnés tous les ans, et l'on avait donc un écart très marqué entre les années de forte abondance et les années de faible abondance, que les variations naturelles de l'environnement ne suffisaient pas à expliquer. En considérant tous ces échantillonnages de façon indépendante, on étudiera des échantillons de plus petite taille et on devrait limiter cette variation résiduelle. On espère ainsi ne plus violer l'hypothèse d'égalité de la moyenne et de la variance de la loi de Poisson.

On considère alors un effet environnemental commun à toutes les stations d'échantillonnage, pour chacun des groupes d'espèces, auquel on ajoute un effet fixe local, tenant compte de la répartition géographique des stations sur le site d'étude. Pour ne pas contraindre cette structure géographique, on introduit un effet différent pour chaque station, qui permettra des variations individualisées pour chacune des stations. On pourra alors voir si une structure particulière se dégage parmi les stations, et l'interpréter. Par exemple, on peut espérer déceler un effet structurant de la CNPE du Bugey. Des études ont montré que les eaux chaudes du rejet ne se mélangent que très peu latéralement avec les eaux plus froides

du fleuve, décrivant une sorte de panache qui se cantonne à la moitié droite du Rhône en aval de la centrale (Ginot et al., 1996). De plus, Ginot et al. (1996) ont montré que les eaux du rejet se refroidissent lentement, selon un gradient d'environ 1°C par kilomètre. L'élévation de température du rejet étant en moyenne de dix degrés par rapport aux eaux non réchauffées, on considérera que la station n°8, en aval de la confluence avec l'Ain à une douzaine de kilomètres en aval de la centrale, n'est pas impactée. Par conséquent, on considérera que les stations 1, 2, 3, 6 et 8 sont des stations témoins et que les stations 4, 5 et 7 subissent potentiellement une perturbation. Si la centrale a réellement un effet perturbateur sur le peuplement, on devrait donc discerner une structure de l'assemblage des poissons juvéniles pour les stations non impactées, et une structure différente pour les stations de pêche situées en rive droite aval de la centrale.

Nous faisons également l'hypothèse que cet effet station est aléatoire, et que les effets sont différents pour chaque espèce, afin de permettre la prise en compte de variations indépendantes dans le fonctionnement du système.

#### *Introduction d'un effet saisonnier*

Nous avons considéré dans la section précédente que les résultats des échantillonnages étaient interchangeable d'une saison à l'autre, au sein d'une année. Cette hypothèse peut être testée, pour un coût très faible en nombres de paramètres, car il suffit d'introduire un effet saison qui n'aura que deux modalités possibles, été et automne. on considérera que les échantillonnages réalisés en été sont les témoins de l'expérience, et l'on testera la déviation des échantillonnages automnaux par rapport aux échantillonnage estivaux.

#### *Abondance locale attendue*

On exprime donc le logarithme de l'abondance locale attendue  $Z_{i,j,k(i),l}$  pour l'espèce  $i$  appartenant au groupe  $k$ , l'année  $j$ , la saison  $m$  et la station  $l$  comme :

$$\log(Z_{i,j,k(i),l}) = Z_{0i} + \beta_m^{saison} + \beta_l^{station} + .X_j \beta_{k(i)} \quad (3.31)$$

### **Modèles sélectionnés et inférence sur les paramètres de la régression**

Sur ce second modèle, on applique également la procédure de sélection des modèles les mieux ajustés décrite aux sections 3.2.5 et 3.2.5.

Une fois le meilleur modèle déterminé pour chaque groupe, on réalise l'inférence bayésienne sur les paramètres du modèle. Pour cela, des priors vagues sont introduits : des lois normales  $N(0, 10^{-3})$  sont attribuées aux  $\beta_k$ , ainsi qu'au paramètre  $\beta^{saison}$  et aux paramètres  $\beta^{station}$  de chaque espèce. Par convention, on attribue au premier paramètre des vecteurs  $\beta^{saison}$  et  $\beta^{station}$  la valeur 0. Les autres effets sont donc interprétés comme des déviations par rapport à cette situation de référence.

#### *Effet de la saison sur les abondances de juvéniles*

La figure 3.12 présente sous forme de boxplot la distribution de probabilité a posteriori du paramètre saisonnier représentant la variation d'effectif pêché lors des pêches d'automne par rapport aux pêches estivales. La valeur zéro n'étant pas incluse dans l'intervalle de crédibilité a posteriori, on ne peut rejeter l'hypothèse d'un effet saison positif. Les effectifs pêchés sont donc en moyenne une fois et demie plus importants lors des pêches d'automne que lors des pêches d'été ( $e^{0.4} \simeq 1.49$ ).

#### *Effets des stations d'échantillonnage sur les abondances de juvéniles*

La figure 3.13 montre, espèce par espèce, les variations liées à chacune des stations. La station 1 sert de référence, son effet est par convention fixé à 0 et les effets des autres stations sont exprimés par rapport à cette norme.

Aucun patron particulier ne se détache quant aux effets des stations sur les abondances de juvéniles et aucun indice ne semble traduire une influence du réchauffement des eaux. En effet, si l'on considère une espèce particulière, les effets stationnels des stations impactées (stations 4, 5 et 7) ne sont que très rarement tous de même signe. De plus, lorsque c'est le cas, comme pour le chevesne, espèce pour laquelle toutes les stations impactées ont un effet négatif, cet effet est de même signe et du même ordre de grandeur que les effets des autres stations non impactées.

#### *Effets des variables environnementales*

La figure 3.14 présente à la fois les variables retenues dans les modèles sélectionnés pour chacun des groupes spécifiques (la présence d'un boxplot signifie que la variable correspondante est retenue) et leurs densités de distribution sous la forme de boxplots. On s'aperçoit que les deux clusters qui étaient les plus proches ont des structures de variables sélectionnées assez proches, et que les effets conjointement sélectionnés vont en plus dans le même sens. La structure du modèle sélectionné pour le premier groupe est plus éloignée des deux autres.

Figure 3.12 – Distribution de probabilité a posteriori du paramètre lié à l'effet saisonnier automnal, présentée sous forme de boxplot.

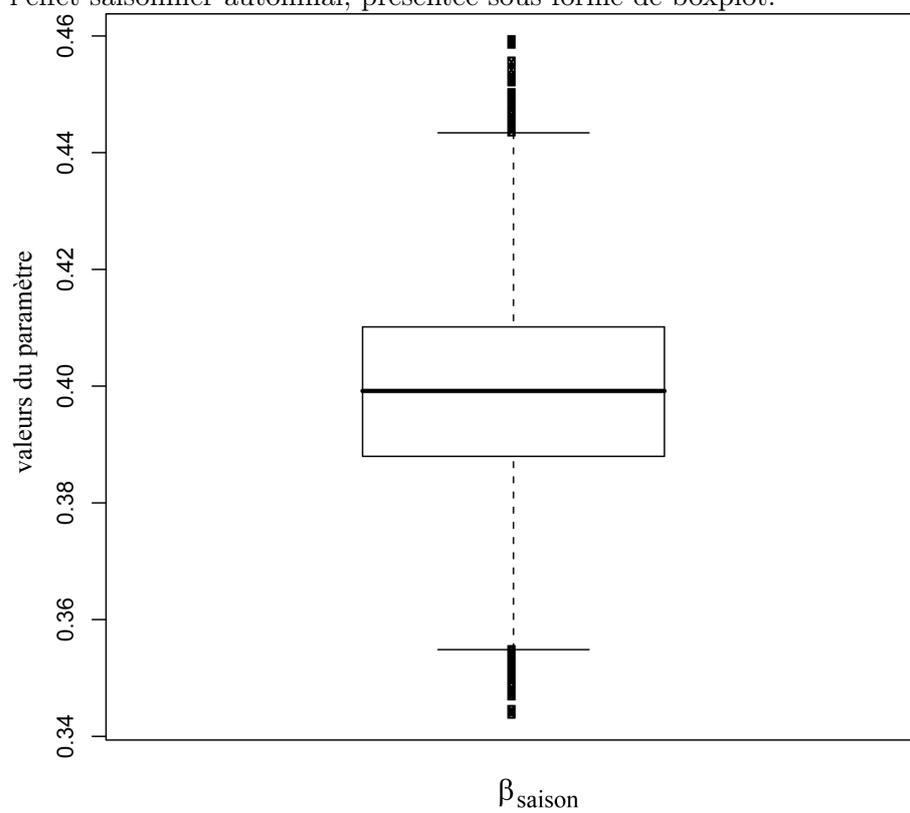


Figure 3.13 – Distributions de probabilité des valeurs des paramètres d'effets stationnels pour chaque espèce, présentées sous forme de boxplots.

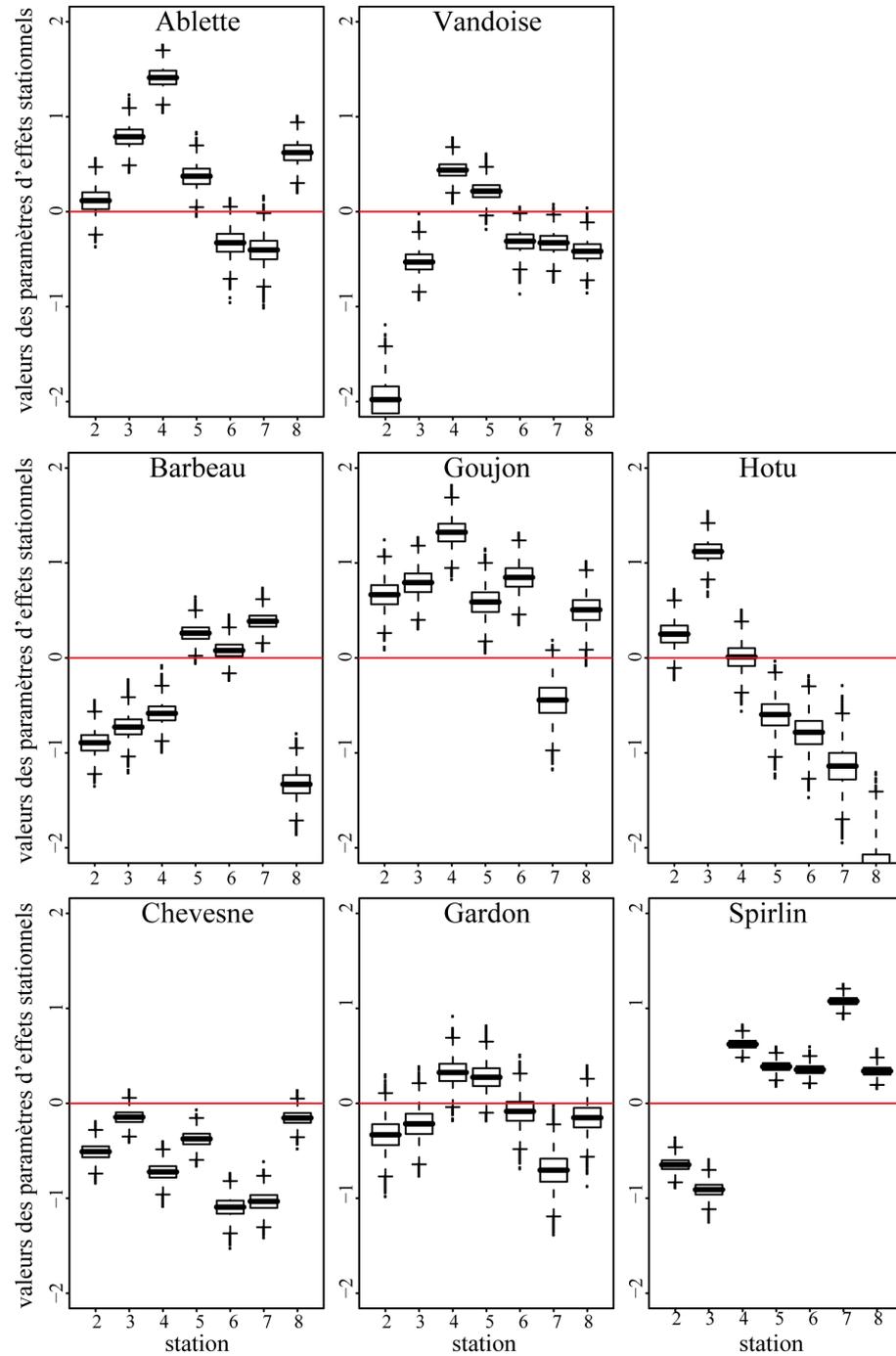
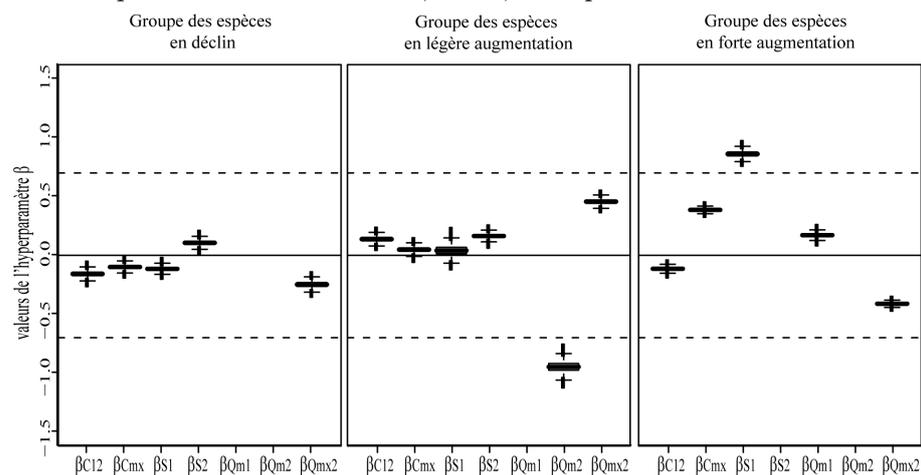


Figure 3.14 – Boxplots des distributions de probabilités des paramètres associés aux variables retenues dans les modèles sélectionnés sur la base du Facteur de Bayes.

La ligne noire représente un effet nul pour les variables, les lignes pointillées représentent les valeurs  $-0,7$  et  $0,7$  des paramètres.



Une fois la sélection des modèles réalisée, on mène l'inférence sur les paramètres de régression de ces modèles à l'aide du logiciel d'inférence bayésienne WinBUGS. Le code WinBUGS de ce modèle est donné en annexe C. Des priors vagues sont donnés aux paramètres de régression, sous la forme de lois normales centrées, de variance 1000.

La convergence de ces modèles est vérifiée dans WinBUGS à l'aide du test de convergence de Gelman-Rubin, modifié par Brooks and Gelman (1998).

### Test de l'hypothèse poissonnienne

La validité de l'hypothèse d'une distribution poissonnienne de l'erreur est testée selon la procédure présentée à la section 3.2.6. Les résultats de ce test sont présentés dans le tableau 3.3

Le fait d'avoir réduit la taille des cellules sur lesquelles on s'appuie pour mener l'inférence n'a pas suffi à corriger la très importante surdispersion que l'on constate dans les données. Il est nécessaire, si l'on veut améliorer la qualité des prédictions réalisées, d'utiliser un modèle de mélange pour corriger ce phénomène.

Tableau 3.3 – Résultats du test de l'hypothèse poissonnienne du modèle à effet stationnel : quantiles de la loi du chi-2 à 3070 degrés de liberté et moyennes et intervalles de confiances simulés et observés de la statistique de test pour le modèle à effet stationnel.

	quantile(0.0275)	mean	quantile(0.975)
Chi-2	2920.27		3227.52
SS.prd	2883	3071	3274
SS	74190	76140	78260

### 3.3.3 Vers une meilleure modélisation de la surdispersion

#### Introduction d'un terme d'erreur dans la modélisation du paramètre de la loi de Poisson

Il est classiquement proposé d'utiliser une loi binomiale négative pour modéliser des données surdispersées par rapport à la loi de Poisson. La loi binomiale négative peut être modélisée comme une loi de Poisson dont le paramètre  $\lambda$  n'est plus fixé, mais distribué à son tour selon une loi gamma. Selon cette description, on peut voir la loi binomiale négative comme une loi de Poisson surdispersée, car en introduisant une loi de distribution sur son paramètre, on ajoute du bruit autour de la distribution de Poisson et on augmente sa dispersion.

Toutefois, d'autres solutions sont possibles et Zuur et al. (2009) proposent d'introduire un terme d'erreur normal dans l'expression du paramètre  $\lambda$ . On note alors :

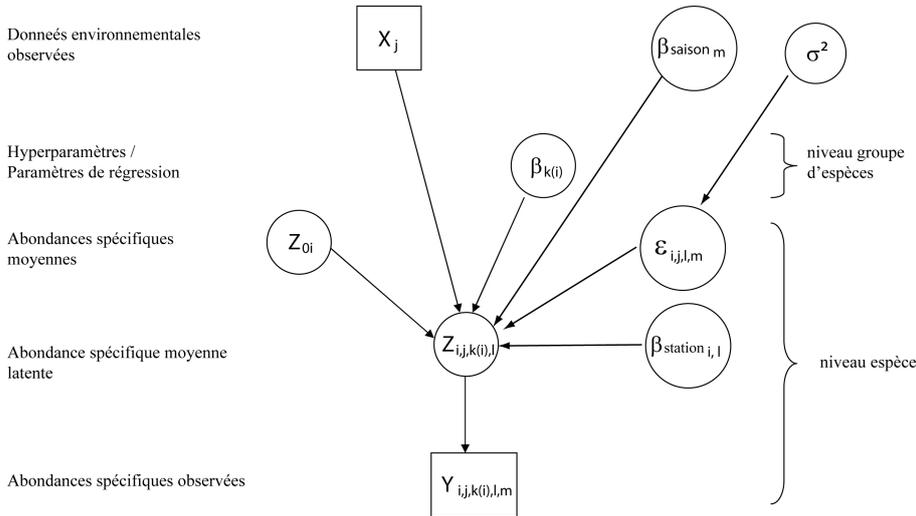
$$\begin{cases} \log(\lambda) &= X.\beta + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2) \end{cases} \quad (3.32)$$

où les  $\varepsilon$  sont distribués selon une loi normale d'espérance nulle et de variance  $\sigma^2$ .

Cette méthode est appliquée sur le modèle à effet stationnel et saisonnier présenté à la section 3.3.2 et l'on fait intervenir ce paramètre d'erreur  $\varepsilon$  dans l'équation loglinéaire définissant les variables latentes  $Z$ . On considérera que les  $\varepsilon$  suivent une même loi  $N(0, \sigma^2)$  pour l'ensemble des espèces, des saisons et des stations. Un prior gamma vague est introduit pour ce paramètre de variance.

L'ajout d'une structure d'erreur dans la partie loglinéaire du modèle doit absorber une partie de la variabilité qui, en son absence, est reportée

Figure 3.15 – Structure du modèle à effet stationnel et saisonnier, avec terme d’erreur, présentée sous forme de DAG.



sur les coefficients de régression liés aux effets stationnels, saisonniers ou environnementaux. En corrigeant les problèmes de surdispersion, on admet qu’il existe une part de variabilité non expliquée, correspondant désormais à ce terme en  $\varepsilon$ , et que l’on ne contraint plus. Ce faisant, on doit mieux discriminer les effets réellement significatifs.

Toutefois, pour des raisons de simplifications mathématiques (il faudrait introduire le terme  $\varepsilon$  dans l’équation 3.14 et intégrer aussi sur l’espace de ce paramètre), nous n’introduisons pas ce nouveau terme dans la procédure de sélection des variables et nous utiliserons les trois modèles déjà sélectionnés dans la section 3.3.2.

La structure de ce modèle est donnée sous forme de DAG à la figure 3.15. Le code WinBUGS de ce modèle est donné en annexe D.

Le logarithme de l’abondance locale attendue  $Z_{i,j,k(i),l}$  pour l’espèce  $i$  appartenant au groupe  $k$ , l’année  $j$ , la saison  $m$  et la station  $l$  s’exprime comme :

$$\log(Z_{i,j,k(i),l,m}) = Z_{0i} + \beta_m^{saison} + \beta_l^{station} + X_j \beta_{k(i)} + \varepsilon_{i,j,l,m} \quad (3.33)$$

Tableau 3.4 – Résultats du test de l'hypothèse poissonnienne du modèle à effet stationnel : quantiles de la loi du chi-2 à 3070 degrés de libertés et moyennes et intervalles de confiances simulés et observés de la statistique de test, pour le modèle de mélange à effet stationnel.

	quantile(0.0275)	mean	quantile(0.975)
Chi-2	2920.27		3227.52
SS.prd	2524	3075	4194
SS	2468	2977	4023

### Test de l'hypothèse poissonnienne

La validité de l'hypothèse d'une distribution poissonnienne de l'erreur est testé selon la procédure présentée à la section 3.2.6. Les résultats de ce test sont présentés dans le tableau 3.4

L'introduction d'une variable d'erreur au sein de la partie loglinéaire de la régression de Poisson a permis de corriger les problèmes de surdispersion des données par rapport à la loi de Poisson simple.

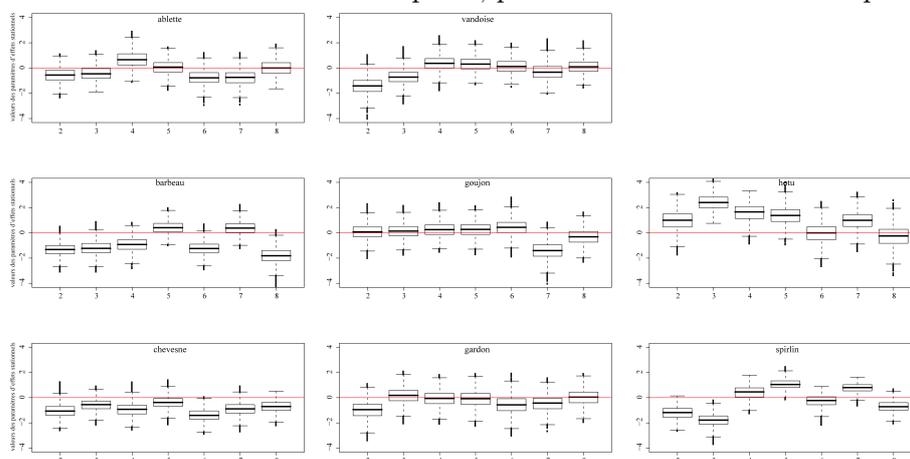
### Analyse des effets stationnels

Par rapport aux résultats obtenus avec la mise en oeuvre du modèle de la section 3.3.2, les effets sont assez largement gommés. Seules la station 2 pour le hotu et les stations 3 et 5 pour le spirin semblent avoir un effet significatif. Les autres stations pourraient donc être considérées comme interchangeables et donc comme des répétitions d'une même expérience. C'est un résultat intéressant car selon ce modèle, il semble donc que le rejet d'eaux réchauffées par la CNPE de Bugey n'a pas de conséquence sur les abondances de juvéniles. Il n'apparaît pas, au vu de ces résultats, et contrairement à ce que nous pouvions attendre, qu'aucune espèce n'ait été défavorisée par le réchauffement, ou bien n'ait privilégié les zones réchauffées pour assurer la croissance de ses juvéniles.

### Analyse de l'effet saisonnier

La figure 3.17 présente sous forme de boxplot la distribution de probabilité a posteriori du paramètre saisonnier représentant la variation d'effectif pêché lors des pêches d'automne par rapport aux pêches estivales. La valeur zéro n'étant pas incluse dans l'intervalle de crédibilité a posteriori du paramètre, on ne peut rejeter l'hypothèse d'un effet sai-

Figure 3.16 – Distributions de probabilité des paramètres liés aux effets stationnels sur chacune des 8 espèces, présentées sous forme de boxplot.



son positif. Pour ce modèle de mélange, le paramètre lié à l'effet saison garde une distribution de probabilités assez proche de celle issue de son estimation dans le cadre du modèle sans mélange. Les effectifs pêchés sont donc en moyenne environ une fois et demie plus importants lors des pêches d'automne que lors des pêches d'été ( $e^{0.36} \simeq 1.43$ ). Ce phénomène peut s'expliquer par une possible augmentation de la probabilité de capture des individus plus tard dans la saison. En effet, les individus plus gros en automne sont sans doute plus faciles à repérer lors des épisodes d'échantillonnage.

### Influence de l'environnement sur l'abondance des juvéniles

La figure 3.18 présente sous forme de boxplot les distributions de probabilités a posteriori des valeurs des paramètres de régression des effets environnementaux. Du fait du passage au logarithme, les valeurs de  $-0.7$  et de  $0.7$  des paramètres sont importantes, car elles signifient qu'une variation positive d'une unité de standard déviation de la variable environnementale correspondante entraîne respectivement une diminution de moitié ou un doublement de l'effectif observé. Contrairement aux résultats observés à la section 3.3.1, aucune variable environnementale ne semble avoir une telle influence, et les valeurs des paramètres reflètent plutôt une somme de petites variations expliquant les fortes disparités d'abondances entre les années.

Figure 3.17 – Distribution de probabilité a posteriori du paramètre lié à l'effet saisonnier automnal, présentée sous forme de boxplot.

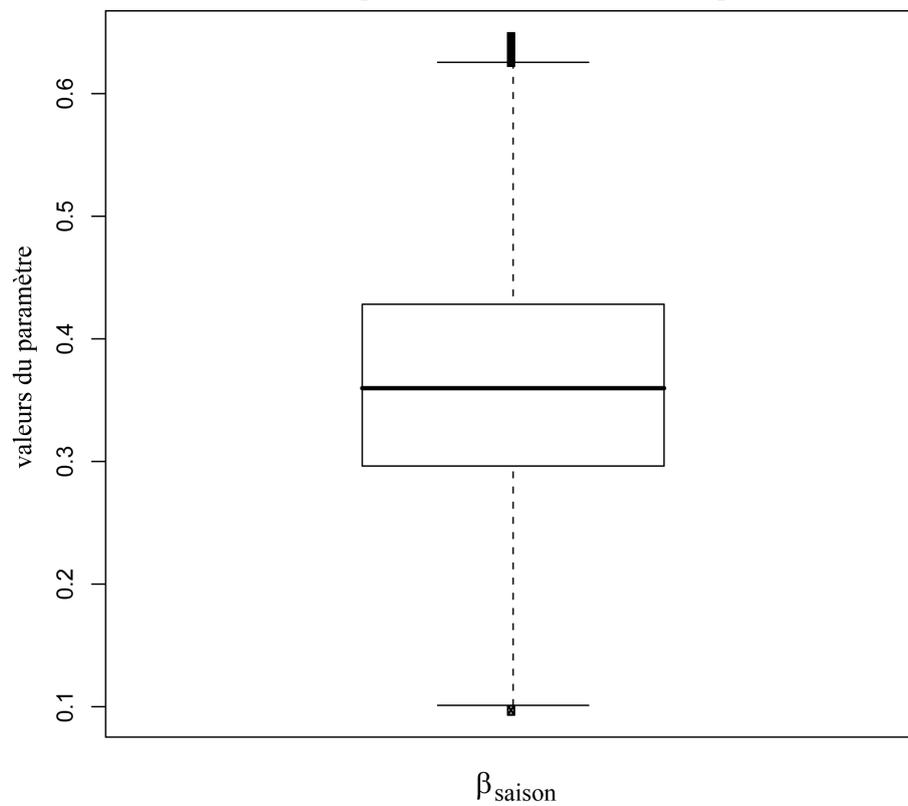
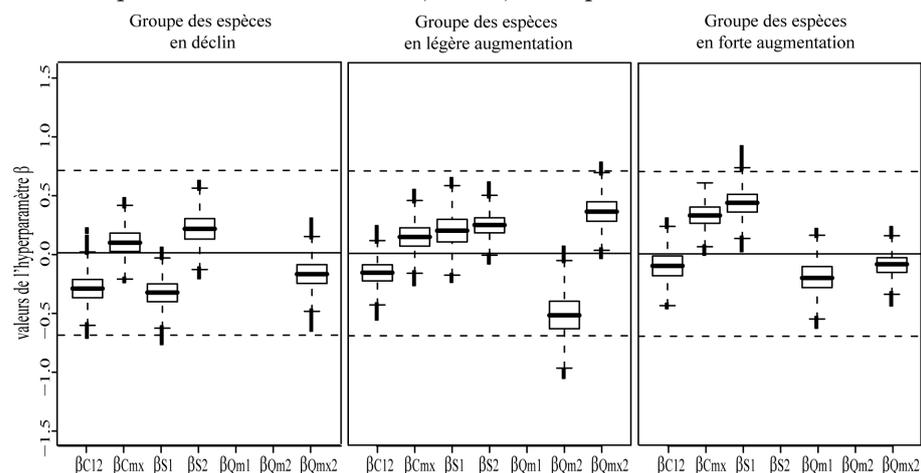


Figure 3.18 – Boxplots des distributions de probabilités des paramètres associés aux variables retenues dans les modèles sélectionnés sur la base du Facteur de Bayes.

La ligne noire représente un effet nul pour les variables, les lignes pointillées représentent les valeurs  $-0,7$  et  $0,7$  des paramètres.



#### *Effets communs à toutes les espèces*

Deux des variables environnementales introduites testées sont sélectionnées et ont un effet de même signe pour l'ensemble des groupes d'espèces : il s'agit des jours de franchissements de températures seuils C12 et Cmx. L'effet globalement négatif de C12 indique que le plus tôt les eaux atteignent  $12^{\circ}\text{C}$ , plus le nombre de juvéniles, toutes espèces confondues, est important. Les effets positifs d'une élévation précoce des températures au printemps avaient déjà été montrés pour la vandoise (Philippart, 1981; Mills and Mann, 1985). La reproduction de la vandoise a lieu de façon précoce au tout début du printemps (Teletchea et al., 2008), en général lorsque les eaux atteignent  $10^{\circ}\text{C}$  (Tissot and Souchon, 2008) et les larves de vandoise bénéficient de cette élévation précoce de température qui favorise leur croissance. Concernant l'ablette, il a été montré que sa stratégie de reproduction consistait en une ponte fractionnée (Mackay and Mann, 1969), notamment en se référant à la présence de différents stades de développement, jusqu'à quatre simultanément, d'ovocytes dans les ovaires avant la période de reproduction (Barus and Prokes, 1993). L'ablette pond plus tard que la vandoise, lorsque la température de l'eau a atteint environ  $15-16^{\circ}\text{C}$ , ce qui arrive sur le site d'étude en moyenne la

première semaine de juin. Une élévation plus précoce de la température pourrait favoriser une meilleure maturation des gonades, ainsi qu'une meilleure croissance des larves et améliorer la survie des fractions de ponte successives.

Teletchea et al. (2008) ont identifié le hotu comme espèce à reproduction précoce et fractionnée, le barbeau comme espèce à ponte tardive et éventuellement fractionnée, tandis que le goujon est considéré comme espèce à ponte tardive fractionnée (comportement également constaté par Mann, 1980). L'élévation précoce de température au début du printemps pourrait donc conduire à une meilleure croissance et donc survie des larves de hotu. Concernant les deux autres espèces, qui se reproduisent plus tard et de façon fractionnée, cette élévation précoce s'accompagne probablement de premières pontes plus précoces qui, dans de bonnes conditions, donnent des larves dont la croissance et la survie sont meilleures.

Le gardon pond généralement en une seule fois dans une période de 2 à 4 jours (Mackay and Mann, 1969), lorsque la température de l'eau atteint 15°C (Tissot and Souchon, 2008), tandis que le chevesne se reproduit à la fin du printemps (Teletchea et al., 2008) et est reconnu comme capable de pontes fractionnées (Economou et al., 1991; Vriese et al., 1994) jusqu'à trois fois (Libosvarsky, 1979). Le spirilin est probablement capable de ponte fractionnée dès lors que la température de l'eau a atteint 14°C (Tissot and Souchon, 2008). Globalement, pour ces trois espèces, une élévation précoce de la température de l'eau induit probablement une première ponte rapide, permettant aux larves et juvéniles une meilleure croissance et donc une meilleure survie.

L'effet globalement positif du passage tardif du seuil de maximum de température peut être expliqué par le fait que cela signifie un allongement de la période estivale, période durant laquelle l'écosystème aquatique est particulièrement productif. Cela induit un allongement de la période durant laquelle les juvéniles trouvent microalgues et microinvertébrés en quantité, et donc de meilleures croissance et survie des larves et des juvéniles.

*Un facteur environnemental prépondérant commun aux espèces en augmentation*

Une composante de l'environnement semble jouer un rôle majeur dans l'évolution différentielle des espèces. En effet, une augmentation de la variable S1, représentant la chaleur relative du printemps, a un effet négatif pour le groupe d'espèces en déclin, et un effet positif pour les deux

groupes en augmentation. Cet effet positif d'augmentation de la somme de degrés-jours sur la période printanière est en accord avec les précédents travaux qui tendent à montrer que des températures plus élevées favorisent une meilleure croissance des juvéniles de cyprinidés (Mills and Mann, 1985; Daufresne et al., 2003; Wolter, 2007). Il est probable qu'une augmentation du facteur S1 induise une diminution du temps d'incubation pour les six espèces en augmentation, provoquant un développement plus précoce des larves, ce qui se traduit plus tard par une meilleure survie des juvéniles pendant les quelques premiers mois de vie.

L'effet négatif de ce facteur sur les abondances de juvéniles de van-  
doise et d'ablette tend à montrer qu'une montée trop rapide en température de l'eau a un effet préjudiciable sur la survie ou le développement des embryons de ces espèces, qui sont considérées comme des espèces en limite sud de leur zone de répartition dans le Rhône et préférant les eaux plus fraîches Daufresne et al. (2003).

*Le débit influence différemment les espèces benthiques et pélagiques*

Si les variables décrivant le régime thermique structurent fortement le succès reproductif des cyprinidés, le régime hydrique semble également jouer un rôle. En effet, on observe des structures d'influences assez différentes pour les trois groupes d'espèces. Néanmoins, les espèces du groupe 1, en déclin, et les espèces du groupe 3, en forte augmentation, qui sont toutes pélagiques (de pleine eau), semblent répondre de façon assez proche aux variables de débit. Notamment, l'existence de phénomènes de crue marqués en été ont un effet très négatif sur ces espèces. Trois hypothèses peuvent être formulées pour expliquer cela. Premièrement, ces forts débits ponctuels peuvent favoriser un phénomène de dévalaison des larves et juvéniles. Carter and Reader (2000) suggéraient que la migration rapide vers l'aval des 0+ fait probablement partie des cycles migratoires et serait un moyen de disperser depuis les zones de pontes vers les zones de grossissement. Ils montraient que la dévalaison apparaissait surtout pour les individus âgés de quelques semaines seulement. Dans notre cas, sur la période estivale considérée, les 0+ doivent être considérés comme de tout jeunes juvéniles aux capacités de nage restreintes et donc peu à même de résister à des courants violents générés par ces pics de débits. Ensuite, Bolland et al. (2008) ont montré que dans les rivières de plaine, un grand nombre de juvéniles étaient retrouvés dans les zones humides latérales, temporairement créées lors des épisodes de crues fortes, lorsque le niveau d'eau dépasse les limites du lit majeur. Ces juvéniles accèdent alors à ces zones et sont incapables

de rejoindre le chenal principal lorsque le niveau d'eau baisse. Les principales espèces touchées sont alors les espèces eurytopiques, telles que l'ablette et le gardon, mais on trouve également en grands nombres des individus d'espèces rhéophiles comme la vandoise ou le chevesne. Enfin, concernant les espèces à ponte fractionnée, ces phénomènes de crues, s'ils interviennent tôt en été, pourraient perturber les pontes tardives, et causer une diminution du nombre de juvéniles engendrés.

Dans le même temps, les espèces du groupe 2, en faible augmentation, répondent de façon différente. Les forts débit estivaux ont un effet positif sur leur abondance. Cela peut s'expliquer par le fait que les espèces benthiques sont moins touchées par l'augmentation de débit lors des crues car leurs conditions locales de vie changent peu. En effet, en liaison avec la rugosité du substrat, les vitesses des courants au fond des cours d'eau sont plus faibles. Dans un second temps, les abondances des autres espèces diminuant à la suite de ces épisodes de crue, les possibles phénomènes de compétition pour la nourriture pourraient également diminuer et permettre aux espèces benthiques un meilleur développement. En revanche, un plus fort débit moyen durant l'été a un effet négatif sur ces espèces benthiques. Les conditions de courant plus fort sur des périodes plus élevées pourraient détériorer les conditions locales nécessaires à l'alimentation des juvéniles, soit par la nécessité d'une plus grande dépense énergétique pour se nourrir, soit par des conditions de hauteur d'eau plus importante empêchant la production en quantité suffisante d'algues primaires, source importante de nourriture.

### **Capacité prédictive du modèle**

Sur les 448 échantillonnages réalisés entre 2006 et 2009, le modèle prédit désormais correctement 352 valeurs, soit un ratio de 0,786 (figures 3.19 et 3.20). La capacité prédictive a très significativement augmenté grâce à la correction de la surdispersion.

### **Conclusion sur ce modèle**

Les résultats observés pour ce modèle sont en accord avec les éléments de discussion apportés par Cameron and Trivedi (1998). On a vu que l'utilisation d'une loi de Poisson surdispersée, via un mélange de lois normale-Poisson, permet de corriger les défauts des modèles présentés aux sections 3.3.1 et 3.3.2. Au niveau de l'inférence des valeurs des paramètres de régression, cette correction engendre effectivement des in-

Figure 3.19 – Illustration de la capacité prédictive du modèle à mélange de loi.

Les boxplots sont les distributions de probabilité prédites pour les abondances estivales des huit espèces (en ligne), sur la période 2006-2009 (en colonne), pour chaque station. Les abondances observées sont tracées en rouge.

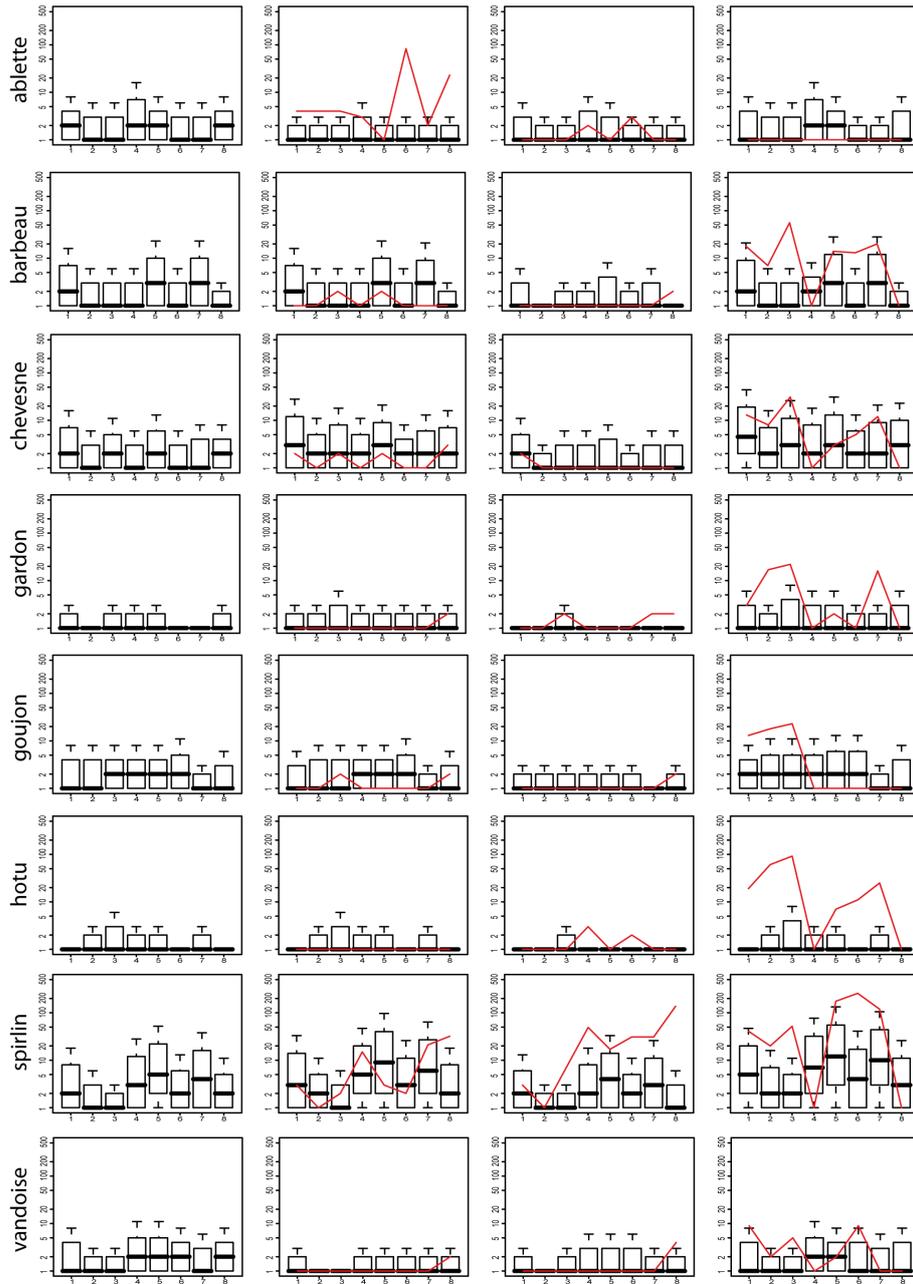
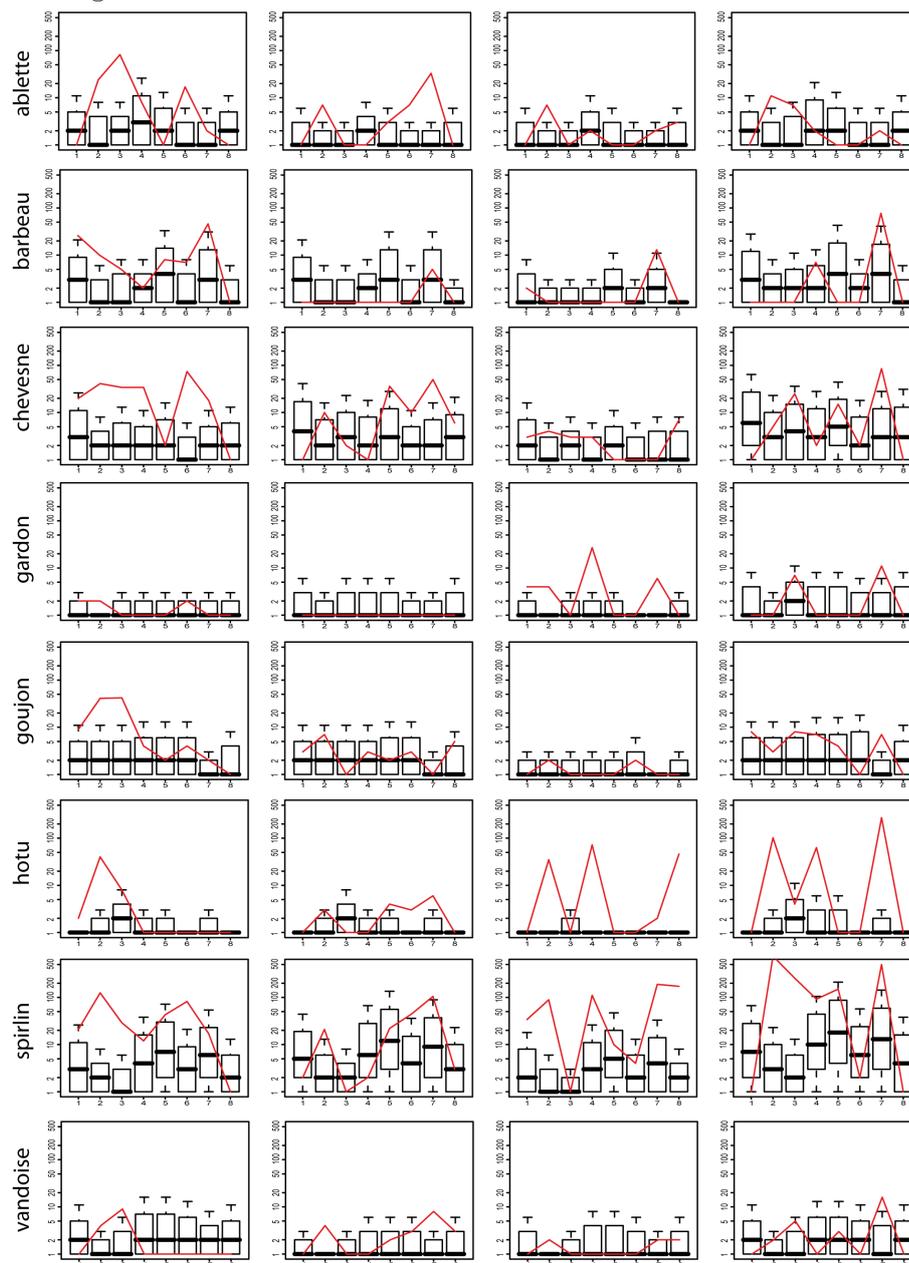


Figure 3.20 – Illustration de la capacité prédictive du modèle à mélange de loi.

Les boxplots sont les distributions de probabilité prédites pour les abondances automnales des huit espèces (en ligne), sur la période 2006-2009 (en colonne), pour chaque station. Les abondances observées sont tracées en rouge.



tervalles de crédibilité plus larges. Dans le cas d'un test de significativité des variables d'un modèle, on serait donc plus sévères dans le choix des variables à retenir. Concernant les prévisions de valeurs sur un jeu de données indépendant, nous avons vu les avantages de l'utilisation d'une telle loi.

### Pistes de modélisation

Dans ce modèle, nous avons supposé l'existence de groupes d'espèces, fondés sur leurs évolutions interannuelles d'effectifs de juvéniles, qui auraient des réponses similaires aux variations de l'environnement. La modélisation hiérarchique permet de prendre en compte facilement ces hypothèses a priori sur le fonctionnement du système. Toutefois, il serait possible de hiérarchiser davantage le modèle, notamment en structurant le terme d'erreur  $\varepsilon$ , qui représente la part de variabilité non expliquée par les variables environnementales dans le déterminisme des abondances de juvéniles.

Ainsi, le terme d'erreur pourrait être hiérarchisé par groupes d'espèces. Le terme d'erreur permettant de corriger la surdispersion observée indépendamment pour chaque espèce,  $\varepsilon_i$ ,  $k=1,2,\dots,8$  suivrait une loi normale centrée sur la variance moyenne  $\sigma_k^2$ , définie pour chaque groupe et avec une variance interspécifique  $\sigma^{2*}$  qu'on considérerait par exemple identique pour les trois groupes. Il s'agit alors de définir la distribution a priori des variances par groupe  $\sigma_k^2$ , à l'aide d'une loi gamma, dont les hyperparamètres  $(a_k, b_k)$  sont définis pour chaque groupe  $k$ . On obtient ainsi la structure hiérarchique suivante :

$$\begin{cases} \varepsilon_i \sim \text{Normale}(\sigma_{k(i)}^2, \sigma^{2*}) \\ \sigma_{k(i)}^2 \sim \text{gamma}(a_k, b_k) \end{cases} \quad (3.34)$$

## 3.4 Conclusion

Dans ce chapitre, nous avons développé une approche innovante dans l'étude des variations interannuelles d'assemblages de poissons. Nous avons montré qu'il est possible, dans un cadre statistique robuste, de quantifier les effets simultanés des variations interannuelles des régimes thermique et hydrique sur les abondances de différentes espèces de cyprinidés du Rhône. Au final, cette approche permet de mettre en relation les influences de ces deux régimes sur les abondances de juvéniles, allant

au-delà des études ne prenant en compte qu'un aspect des variations du milieu.

Cependant, nous avons du, pour cela, faire l'hypothèse d'une distribution poissonnienne des données, hypothèse qui s'est avérée fautive. Mais, moyennant l'ajout d'un terme d'erreur pour tenir compte de l'existence d'une part de variabilité non expliquée par les variations de débit et de température autour des effectifs attendu, nous avons pu corriger les phénomènes de surdispersion observés dans les données.

## Chapitre 4

# Un modèle à facteur latent partagé

### 4.1 Introduction

#### 4.1.1 Travailler de façon conditionnelle à l'effectif total pour stationnariser le protocole expérimental

Le chapitre précédent faisait l'hypothèse de comptages Poissonniens du nombre d'occurrences de chaque espèce. Or l'hypothèse Poissonnienne a été remise en cause car de nombreuses sources d'incertitudes, en particulier la non stationnarité du protocole de captures d'une année à l'autre, semblent pouvoir entraîner une surdispersion des résultats. Travailler conditionnellement à l'effectif total observé, c'est opérer une standardisation que nous pensons propre à réduire cette surdispersion en rendant le processus plus homogène, en gommant une certaine variabilité inter-annuelle : certaines années on capture beaucoup d'individus (pêche plus longue, débits plus propices, etc...), d'autres années la pêche est moins efficace. Dans ce chapitre on focalisera donc sur les ratios de chacun des trois groupes d'espèces  $(p_1, p_2, p_3)$  pour étudier leurs variations.

On notera qu'une telle approche est de plus très complémentaire à la modélisation des abondances, en termes d'information apportée. En effet, dans le cas d'une modélisation des comptages, les proportions relatives des groupes peuvent être calculées à partir des abondances respectives. Dans le cas présent, elles constituent la variable cible et on travaille conditionnellement aux nombres de captures de chaque campagne.

### 4.1.2 Créer des variables explicatives ad-hoc

La façon la plus directe de relier une réponse multinomiale à un jeu de  $p$  variables explicatives  $X_1, X_2, \dots, X_p$  consiste à mettre en place une relation avec  $2 \times p$  paramètres du type log linéaire :

$$\log\left(\frac{p_1}{p_3}\right) = \sum_{j=1}^p \beta_1^j X_j = X\beta_1$$
$$\log\left(\frac{p_2}{p_3}\right) = \sum_{j=1}^p \beta_2^j X_j = X\beta_2$$

Ces deux équations (avec la troisième équation  $p_1 + p_2 + p_3 = 1$ ) donnent facilement l'expression des probabilités  $(p_1, p_2, p_3)$  associés à chacune des trois communautés. Si l'on considère le problème de la sélection de modèle, en toute généralité, il y a cette fois  $2^p \times 2^p$  modèles alternatifs possibles en considérant que l'on peut les construire "en allumant ou en éteignant" l'effet de chaque variable explicative indépendamment dans chaque combinaison linéaire explicative (à comparer aux  $2^p$  modèles en compétitions avec les comptages Poissonniens au chapitre précédents). Pour notre application  $p = 9$ , il y aurait  $2^{18} = 262144$  modèles dont il faudrait réaliser l'inférence, ce qui rend les temps de calculs rédhibitoires sur un ordinateur personnel (plus de 182 jours si chaque inférence MCMC prenait une seule minute).

Par l'intermédiaire de facteurs explicatifs latents que nous introduisons pour faire un pont entre variables explicatives et réponses multinomiales, nous proposons ici de résoudre dans un même temps les problèmes d'inférence statistique et de sélection de variables explicatives.

### 4.1.3 Quelques approches bibliographiques de l'analyse multivariée

Considérant un jeu de  $p$  variables explicatives  $X_i, i = 1, \dots, p$ , d'une variable  $Y$  multivariée, discriminer l'influence particulière d'une variable  $X_i$  particulière est un problème ancien. Il a été montré que juxtaposer des modèles de régression simple sur chacune des composantes n'était pas satisfaisant. Les régressions linéaires multiples (MLR), où la variable  $Y$  est prise comme combinaison linéaire de l'ensemble des variables  $X_i, i = 1, \dots, p$  a été développée pour résoudre ce problème. Toutefois, cette méthode est prise en défaut lorsque les variables explicatives sont trop fortement corrélées entre elles (Woody and Brown, 2003). Dans le

cas de colinéarité des variables prédictives, les coefficients de régression de la MLR deviennent instables, et fluctuent fortement selon l'échantillon utilisé (Cramer et al., 1988) et leur interprétation devient difficile.

Des approches multivariées ont été développées afin de créer des variables explicatives non-corrélées à partir des variables explicatives initiales. Ces méthodes proposent donc de coupler une méthode d'analyse de données multivariées (par ex. l'Analyse en Composantes Principales (Saporta, 1990)) qui permet de décrire de façon synthétique la structure des données par leur projection dans un nouveau repère défini par des axes principaux, combinaisons linéaires des variables explicatives initiales, orthogonaux. Ces axes principaux sont alors utilisés dans une régression linéaire simple ou multiple (selon le nombre d'axes principaux conservés) (Tomassone et al., 1983). La régression sur composantes principales (PCR) utilise les axes principaux d'une ACP, la régression PLS (Partial Least Square) quant à elle utilise un algorithme qui construit des axes en maximisant la covariance entre les  $X$  et  $Y$ . L'avantage de ces méthodes est que les coefficients des régressions sur les axes principaux deviennent comparables (en signes et en amplitude) à ceux d'une simple corrélation (Tenenhaus et al., 1995; Wold et al., 2001). La régression PLS permet donc d'extraire directement les composantes qui sont significatives au regard des deux jeux de données  $X$  et  $Y$  (Wasson et al., 2010). Le cas général des régressions linéaires normales a été généralisé au cas des modèles linéaires généralisés afin de pouvoir faire le lien entre différents types de données continues ou catégorielles.

Ces méthodes développées dans un cadre fréquentiste ont été adaptées au paradigme bayésien. Les structures analogues de modèles de régression bayésiens sont couramment dénommés "modèles bayésiens factoriels de régression". Ils tirent ce nom de l'introduction d'un ou plusieurs facteurs latents dans la structure du modèle, qui sont les analogues des axes principaux fréquentistes. Ces modèles bayésiens ont été utilisés notamment dans le cadre des sciences sociales (Tsonaka and Moustaki, 2007), de la chémométrie (Woody and Brown, 2003), et de l'épidémiologie (Biggeri et al., 2005; Richardson, 1990; Clifford et al., 1989) mais n'ont pas encore trouvé écho au sein de la communauté écologue. Nous proposons ici le développement et l'application d'un modèle bayésien logistique multinomial à facteur latent partagé à l'étude des effets des variations de régimes thermique et hydrique sur la composition de l'assemblage de cyprinidés juvéniles du Rhône.

## 4.2 Structure du modèle

### 4.2.1 Facteur latent partagé

Nous allons définir un facteur latent  $Z$ , que nous allons considérer comme une *cause* commune aux variations environnementales et aux variations de proportions relatives des groupes d'espèces les uns par rapport aux autres. Conceptuellement, ce facteur latent peut être interprété comme un signal de grande envergure, pouvant s'appliquer à une large échelle spatiale et dont les variations dans le temps entraîneraient les variations de l'environnement climatique local et les variations d'abondance (ce pourrait être par exemple la position du *Gulf stream*, dont les conséquences sur le climat ont déjà été montrées, ainsi que son éventuelle implication sur les variations de succès reproducteur chez les poissons (Nunn et al., 2007), ou sur les variations hydrologiques (Kelly and Gore, 2008)). West (2003) suggère que l'utilisation d'un facteur latent est particulièrement adapté lorsqu'une multitude d'effets sous-jacents, non identifiés et/ou non contrôlables, gouvernent les patrons de variations des variables observées. Nous appelons ce facteur latent l'*hypersignal* dans la suite du document.

En absence d'informations supplémentaires sur l'évolution possible de l'hypersignal, nous proposons une variation purement indépendante normale. On dote donc les  $Z(i)$ ,  $i = 1, \dots, T$  de priors  $N(0, 1)$ .

D'autres structures pourraient être proposées, comme par exemple en supposant a priori une évolution lente et lisse de l'hypersignal au cours du temps. On impose alors que le facteur  $Z$  soit marginalement  $N(0, 1)$  et qu'il varie au cours du temps, par exemple sous forme autorégressive.

$$\begin{aligned} Z(t) &= \rho Z(t-1) + \sqrt{1 - \rho^2} \eta_t \\ \eta_t &\sim N(0, 1) \\ -1 &< \rho < 1 \end{aligned}$$

Dans ce cas, les paramètres  $\rho$  et  $\eta$  doivent être considérés dans le DAG, comme à la figure 4.1.

On pourrait aussi proposer une évolutions avec sauts aléatoires décrivant par exemple une rupture à une date  $\tau$  telle que :

$$\begin{aligned} Z(t) &= \varepsilon(t) \quad t \leq \tau \\ Z(t) &= \mu + \varepsilon(t) \quad t > \tau \end{aligned}$$

. Le paramètre  $\tau$  doit être muni d'un prior, et on pourra prendre par exemple un prior uniforme sur une plage temporelle plus ou moins large.

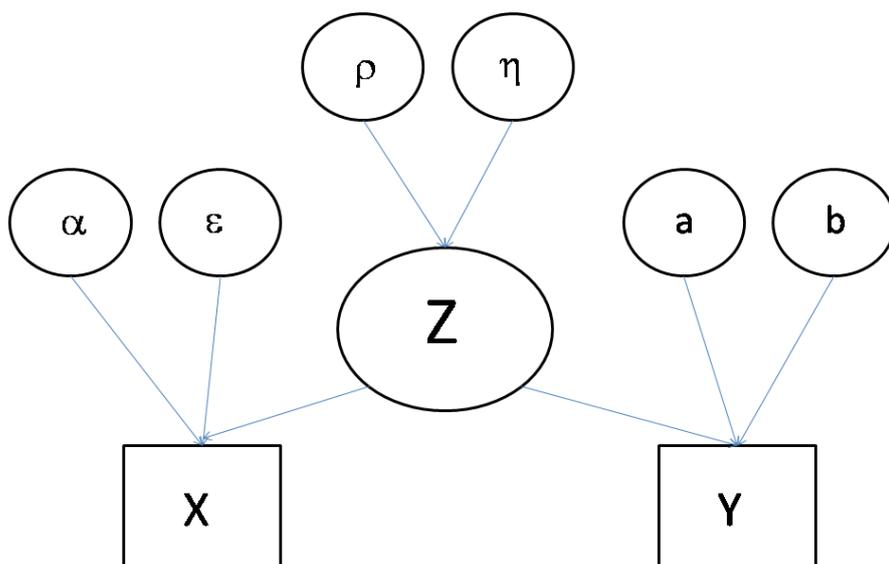


Figure 4.1 – L’hypersignal  $Z$ , au coeur de la structure de covariation des données observées continues  $X$  et catégorielles  $Y$

### 4.2.2 Observations

Les deux sources de données doivent être reliées à ce facteur commun  $Z$ .

#### Etat de l'environnement

Une structure possible pour les impacts est une forme de décomposition linéaire factorielle

$$\begin{aligned} X_1(t) &= \alpha_1 Z(t) + \sigma_1 \varepsilon_1(t) \\ &\dots \\ X_p(t) &= \alpha_p Z(t) + \sigma_p \varepsilon_p(t) \\ &\dots \\ X_P(t) &= \alpha_P Z(t) + \sigma_P \varepsilon_P(t) \end{aligned}$$

Prise indépendamment des observations sur les données de comptage, cette structure permet d'exprimer les corrélations existant entre les variables descriptives de l'environnement.  $Z(t)$  exprime la part de variabilité commune, donc le signal commun, existant au sein de ces covariables.  $Z$  pourrait être vu comme un analogue du premier axe principal d'une ACP classique (figure 4.2). Les  $\alpha_p$  sont les composantes de chaque variable explicative sur le facteur, les  $\varepsilon_p$  sont des *résidus*  $N(0, 1)$  non corrélés au facteur  $Z$ , leur variance  $\sigma_p^2$  peut être contrainte (si on a préalablement centré et réduit les données  $X$ ) en imposant  $\sigma_p^2 + \alpha_p^2 = 1, \forall p$ .

Dans le cas présent, où le facteur latent  $Z$  est "branché" à un modèle de régression sur  $Y$ , on n'est plus dans le cas d'une ACP, mais dans l'étude des corrélations entre deux tableaux de données, et donc de la régression PLS.

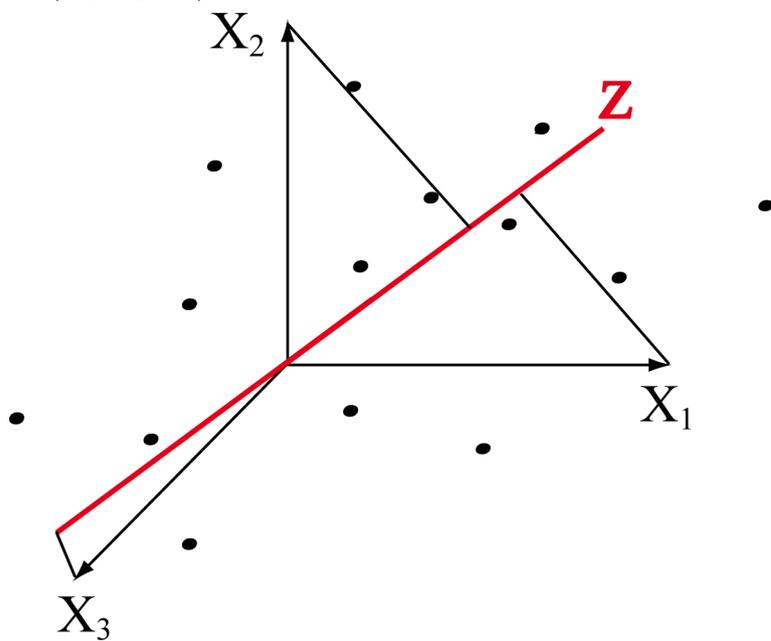
Dans le cas général, les  $\alpha_i$  et les  $\sigma_i^2$  sont à doter de priors. Faute d'information a priori, on prendra des normales centrées sur 0 pour  $\alpha$  et des gamma très vagues pour  $\sigma^{-2}$ .

Dans le cas du système contraint, on prendra des normales  $N(0, 1)$  pour chacun des  $\alpha_p$ .  $\sigma_p$  étant contraint par  $\alpha_p$ , il n'est plus nécessaire de leur attribuer de prior.

#### Communautés piscicoles

La partie droite du DAG de la figure 4.1 décrit la relation existant entre le facteur latent et les proportions relatives des différentes

Figure 4.2 – Représentation théorique des méthodes d'analyse multivariée. Ici,  $Z$  est le premier axe principal, permettant un changement de repère, de  $(X_1, X_2, X_3)$  vers  $Z$ .



espèces, ou des différents groupes d'espèces. Ces proportions, par définition, doivent être bornées sur l'intervalle  $[0, 1]$ , il n'est donc pas possible d'utiliser une régression linéaire normale. On se place par conséquent dans le cadre d'un modèle linéaire généralisé où les proportions sont liées linéairement au facteur latent, via une fonction de lien. Classiquement, on utilise une fonction de lien logistique.

*Modèle à deux groupes d'espèces*

On relie le nombre  $N_1$  d'individus appartenant au groupe 1 observés au nombre total de poissons pêchés  $N$ , par un modèle binomial avec une probabilité  $p_1$ . On lie le paramètre  $p_1$  au facteur  $Z$  :

$$N_1 \sim \text{multinomiale}(N, p_1, 1 - p_1)$$
$$p_1 = F(aZ + b)$$

où  $F$  est une fonction de répartition à choisir,  $a$  et  $b$  deux paramètres à munir d'un prior (des lois normales ont été utilisées pour un premier essai). Classiquement on prend souvent  $F$  logistique soit

$$p_1 = \frac{\exp(aZ + b)}{1 + \exp(aZ + b)}$$

que l'on écrit encore  $\text{logit}(p) = aZ + b$

*Modèle à trois groupes d'espèces*

On relie les nombres d'individus  $N_1, N_2, N_3$  d'individus observés dans chacun des groupes d'espèces  $i, i = 1 : 3$ , par un modèle multinomial de probabilités  $p_1, p_2, p_3$  :

$$(N_1, N_2, N_3) \sim \text{multinomiale}(N, p_1, p_2, p_3)$$

où l'on doit relier  $p_1, p_2, p_3$  aux variables latentes. Introduisons les paramètres  $(a_1, a_2, b_1, b_2)$  et posons :

$$\eta_1 = a_1Z + b_1$$
$$\eta_2 = a_2Z + b_2$$
$$\eta_3 = 0$$

On a le choix entre deux fonctions de lien

1. le modèle log-linéaire :

$$p_i = \frac{\exp(\eta_i)}{\sum_{j=1}^3 \exp(\eta_j)}$$

où le facteur commun  $Z$  régit communément les probabilités d'observer tel ou tel groupe d'espèces de la façon suivante :

$$\begin{aligned} \log\left(\frac{p_1}{p_3}\right) &= \eta_1 = \text{logit}\left(\frac{p_1}{p_1 + p_3}\right) \\ \log\left(\frac{p_2}{p_3}\right) &= \eta_2 = \text{logit}\left(\frac{p_2}{p_2 + p_3}\right) \end{aligned}$$

$\eta_1$  donne la probabilité conditionnelle d'observer un individu du premier groupe d'espèces sachant que le seconde groupe d'espèces n'est pas observé,  $\eta_2$  donne la probabilité conditionnelle d'observer un individu du deuxième groupe d'espèces sachant que le premier groupe d'espèces n'est pas observé. Cela donne deux probabilités conditionnelles relatives et donc plus difficilement interprétables.

2. le modèle logit plus directement interprétable. On travaille sur la probabilité  $p_1$  d'observer un individu du premier groupe d'espèces (expliquée par  $\eta_1$ ) et sur  $p_{2|-1} = \frac{p_2}{p_2+p_3}$ , la probabilité conditionnelle d'observer un individu du second groupe d'espèces sachant qu'il n'appartient pas au premier groupe (expliqué par  $\eta_2$ ) :

$$\begin{cases} \text{logit}(p_1) &= \eta_1 \\ \text{logit}(p_{2|-1}) &= \text{logit}\left(\frac{p_2}{p_2+p_3}\right) = \log\frac{p_2}{p_3} = \eta_2 \end{cases}$$

### Structure et indépendance conditionnelle

Une caractéristique de cette modélisation est de faire disparaître le lien direct entre les variables explicatives  $X_i$ , et les variables à expliquer  $Y$ . En effet, on a une structure d'indépendance de  $Y$  par rapport à  $X$ , conditionnellement au facteur latent  $Z$ . On a :

$$[Y | Z, X] = [Y | Z]$$

Cela induit que, structurellement, les variables  $X_i$  procurent de l'information sur  $Z$ , mais ne rentrent pas dans la régression (West, 2003).

### 4.2.3 Prédiction

Il est possible de tester la capacité prédictive d'un tel modèle, une fois l'inférence faite sur un jeu de données test. En effet, la structure linéaire exprimant les variables descriptives de l'environnement en fonction du facteur latent  $Z$  permet un calcul simple de la loi prédictive des proportions sachant les facteurs d'environnement. Dans le cas de variables centrées et réduites, on contraint  $\sigma_i, i = 1, \dots, p$  et on a :

$$X_1 = \alpha_1 Z + (1 - \alpha_1^2) \varepsilon_1$$

et :

$$\frac{X_1}{\alpha_1} = Z + \frac{1 - \alpha_1^2}{\alpha_1^2} \varepsilon_1$$

d'où :

$$\frac{X_1}{\alpha_1} \sim N \left( Z, \left( \frac{1 - \alpha_1^2}{\alpha_1^2} \right)^{-1} \right)$$

Par conséquent, on cherche à inférer la moyenne inconnue  $Z$  d'une variable normale  $\frac{X_1}{\alpha_1}$  dont le prior est normal. On sait alors que la loi a posteriori de  $Z$  est normale :

$$[Z | X_1, \alpha_1] \sim N(\mu^{(1)}, (v^{(1)})^{-1})$$

on applique les formules de la conjugaison normale :

$$\begin{cases} (v^{(1)})^{-1} &= 1 + \frac{\alpha_1^2}{1 - \alpha_1^2} \\ \mu^{(1)} \cdot v^{(1)} &= \frac{\alpha_1 X_1}{1 - \alpha_1^2} \end{cases}$$

La précision a posteriori est égale à la somme des précisions, et la moyenne a posteriori est le barycentre de l'observation et de la moyenne a priori, pondérées par leurs précisions respectives.

En itérant successivement sur les  $X_i, i = 1, \dots, p$ , et en posant

$$[Z | X_1, \dots, X_p, \alpha_1, \dots, \alpha_p] \sim N(\mu^{(p)}, v^{(p)})$$

on montre aisément que :

$$\begin{cases} v^{(p)} &= 1 + \sum_{i=1}^p \frac{\alpha_i^2}{1 - \alpha_i^2} \\ \mu^{(p)} &= \frac{\sum_{i=1}^p \frac{\alpha_i X_i}{1 - \alpha_i^2}}{1 + \sum_{i=1}^p \frac{\alpha_i^2}{1 - \alpha_i^2}} \end{cases}$$

En tirant un échantillon de grande taille des  $(\alpha_1, \dots, \alpha_p)$  dans leurs distributions de probabilités a posteriori, puis en appliquant cette formule qui donne  $[Z | X, \alpha]$ , on peut, pour les  $p$  données descriptives de l'environnement inconnues pour  $q$  nouvelles années  $X_i^{(j)}$ ,  $j = 1, \dots, q$ , exprimer les facteurs latents  $Z^{(j)}$  pour les  $q$  années nouvelles. Puis, on exprime :

$$\begin{cases} \text{logit}(p_1^{(j)}) & = a_1 Z^{(j)} + b_1 \\ \text{logit}(p_2^{(j)} |_{-1}) & = a_2 Z^{(j)} + b_2 \\ p_3 & = 1 - p_1 - p_2 \end{cases}$$

pour un échantillon de  $(a_1, b_1, a_2, b_2)$  de grande taille tiré dans les distributions de probabilités a posteriori de ces paramètres. Il faut que les échantillons  $(a_1, b_1, a_2, b_2)$  et  $(\alpha_1, \dots, \alpha_p)$  soient tirés conjointement.

On obtient donc une distribution prédictive a posteriori des proportions relatives des trois groupes d'espèces, auxquelles on peut confronter les proportions relatives observées.

#### 4.2.4 Conclusion

En introduisant cette structure de modèle à facteur latent partagé, on contraint le facteur latent par les deux types d'observation. En faisant remonter l'information contenue dans les deux types de données jusqu'au facteur  $Z$ , on exprime par  $Z$  les signaux communs que l'on trouve dans les deux sources de données. On recherche donc à extraire les parties communes des variables descriptives de l'environnement qui sont corrélées avec l'évolution des proportions des différents groupes d'espèces, au sein de la communauté. En termes de concept, on est proche des méthodes d'analyse multivariées classiques, et notamment de la régression PLS (pour Partial Least Square) logit multinomiale. Cette méthode consiste en l'écriture d'un (ou plusieurs) axe principal en utilisant la covariance entre les variables descriptives puis à utiliser un modèle linéaire généralisé multinomial avec une fonction de lien logit pour faire une régression des observations sur l'axe principal (Wasson et al., 2010).

### 4.3 Application au cas du Rhône à Bugey

#### 4.3.1 Observables

Le principe d'un tel modèle à facteur latent partagé étant d'exprimer les corrélations entre les variables, il n'est pas nécessaire dans ce cas de

faire une élimination préalable de variables parmi les couples de variables trop fortement corrélées. On appelle  $X_1, \dots, X_9$  les  $p = 9$  vecteurs des variables décrivant l'état environnemental du Rhône au niveau du site d'échantillonnage de Bugey présentés dans le chapitre 3. Ces variables sont observées pour les  $i = 1, \dots, T = 26$  années d'échantillonnage. On travaille sur les données centrées et réduites.

On appelle  $Y_{i,j,k} = (N_{1,i,j,k}, N_{2,i,j,k}, N_{3,i,j,k})$  le vecteur des nombres d'individus observés pour chacun des groupes d'espèces 1, 2 et 3, l'année  $i$  ( $i = 1, \dots, 26$ ), la saison  $j$  ( $j = 1, 2$ ) et la station  $k$  ( $k = 1, \dots, 8$ ). Les groupes d'espèces sont ceux définis au chapitre 3 à l'aide de la procédure de cluster.

### 4.3.2 Structuration des données

En tenant compte du protocole expérimental, on introduit deux effets supplémentaires dans la modélisation des proportions relatives des différents groupes d'espèce.

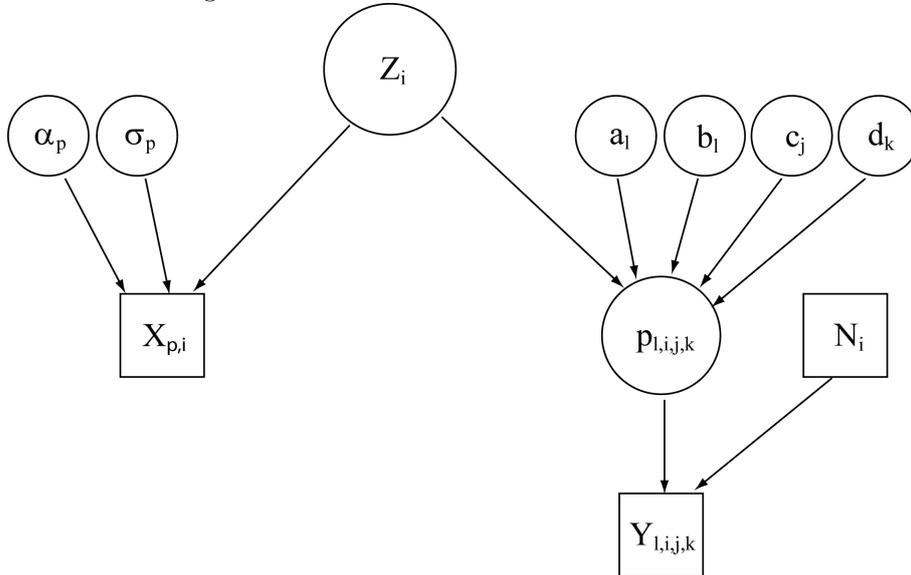
- Un effet saison, permettant de décrire une éventuelle déviation fixe des résultats des pêches d'automne par rapport aux pêches d'été.
- Un effet station permettant de décrire une éventuelle déviation de proportions observées pour certaines stations.

Cependant, on n'introduit pas ici un effet spécifique pour chaque station, mais une structure hiérarchique, tenant compte de la connaissance du système. On rappelle que les travaux de Ginot et al. (1996) ont montré que les eaux réchauffées du rejet de la centrale nucléaire ne se mélangaient que peu avec les eaux plus froides du Rhône. On observe donc une panache d'eaux plus chaudes confiné à la rive droite du Rhône (rive d'implantation de la CNPE) et dont la température diminue selon un gradient d'environ  $1^\circ\text{C}$  par kilomètre. On considérera donc que les stations d'échantillonnage 4, 5 et 7 subissent un effet commun, dû à ce réchauffement des eaux. Dans la suite du texte, la modalité  $k$ , de l'effet station, ne prendra donc que les modalités 1 ou 2.

### 4.3.3 Modèle

La structure du modèle est donnée sous la forme d'un DAG à la figure 4.3. Le code WinBUGS est donné en annexe E.

Figure 4.3 – DAG du modèle à facteur latent.



### Facteur latent

On considérera un facteur latent  $Z$ , dont les répétitions dans le temps  $Z_i$ ,  $i = 1, \dots, 26$ , sont a priori indépendantes et dotées de priors  $N(0, 1)$ .

$$\forall i, Z_i \sim N(0, 1)$$

### Observations environnementales

La partie gauche du DAG à la figure 4.3 représente la liaison entre le facteur latent et les variables descriptives de l'environnement. On prend une structure linéaire factorielle entre les variables descriptives centrées réduites et le facteur latent  $Z$ .

$$\begin{aligned}
 X_{1,i} &= \alpha_1 Z_i + \sigma_1 \varepsilon_{1,i} \\
 &\dots \\
 X_{p,i} &= \alpha_p Z_i + \sigma_i \varepsilon_{p,i} \\
 &\dots \\
 X_{9,i} &= \alpha_9 Z_i + \sigma_9 \varepsilon_{9,i}
 \end{aligned}$$

### Observation de la communauté

Nous allons considérer un modèle linéaire généralisé de type régression logit multinomial pour relier les probabilités d'observations de chacun des groupes d'espèces au facteur latent  $Z$ . Tenant compte de la structure hiérarchique décrite à la section 4.3.2, on pose :

$$\begin{aligned}\eta_{1,i,j,k} &= a_1 Z_i + b_1 + c_{1,j} + d_{1,k} \\ \eta_{3,i,j,k} &= a_3 Z_i + b_3 + c_{3,j} + d_{3,k} \\ \eta_2 &= 0\end{aligned}$$

où  $c_{l,j}$ ,  $j = 1, 2$  représente l'effet de la saison sur la proportion du groupe  $l$  dans la communauté piscicole et  $d_{l,k}$ ,  $k = 1, 2$  représente l'effet de la position amont/aval de la station de pêche sur la proportion du groupe  $l$  dans la communauté piscicole.

En utilisant la fonction de lien logistique, on exprime les quantités :

$$\begin{cases} \text{logit}(p_{3,i,j,k}) &= \eta_{3,i,j,k} \\ \text{logit}(p_{1,i,j,k} | -3,i,j,k) &= \eta_{1,i,j,k} \\ p_{2,i,j,k} &= 1 - p_{1,i,j,k} - p_{3,i,j,k} \end{cases}$$

### Quels priors ?

Les paramètres  $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8, \alpha_9)$ ,  $(\sigma_1, \sigma_2)$ ,  $(\varepsilon_1, \varepsilon_2)$ , et  $(a_1, a_3, b_1, b_3, c_{1,1}, c_{1,2}, c_{2,1}, c_{2,2}, d_{1,1}, d_{1,2}, d_{2,1}, d_{2,2})$  doivent être munis de priors. Après des tests de convergence du modèle sur jeux de données simulées, il apparaît que l'utilisation de priors trop vagues sur l'ensemble des paramètres fait systématiquement tomber l'échantillonneur MCMC de WinBUGS dans des *traps*.

Les variables descriptives de l'environnement ont été préalablement centrées et réduites et on dote les paramètres d'erreur  $\varepsilon$  de priors normaux  $N(0, 1)$ . On se place donc en situation d'utiliser le système contraint pour définir les paramètres  $(\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6, \sigma_7, \sigma_8, \sigma_9)$ .

$$\forall i = 1, \dots, 9 \quad \sigma_i^2 = 1 - \alpha_i^2$$

Pour des raisons de symétrie dans le modèle, entre d'une part les paramètres  $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8, \alpha_9)$  et d'autre part les paramètres  $(a_1, a_3)$ , il faut introduire une contrainte supplémentaire pour orienter les distributions a posteriori. Arbitrairement, on fixe  $a_3$  positif et on attribue à ce paramètre un prior gamma de paramètres  $(1, 1)$ . Par convention, on

contraint donc la proportion d'individus du groupe 3 dans le peuplement à augmenter lorsque la valeur du facteur latent augmente. Cela permet d'orienter l'interprétation des résultats.

Dans le même temps, afin de permettre une évolution contrastée des proportions du groupe 1 (groupe dont les effectifs ont fortement décliné sur la période d'étude) par rapport à celles du groupe 3 (dont les effectifs ont augmenté sur la période d'étude), on ne contraint pas le paramètre  $a_1$  que l'on dote d'un prior normal  $N(0, 0.1)$ . Les paramètres  $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8, \alpha_9)$  sont dotés de priors uniformes sur l'intervalle  $[-1, 1]$ .

Les autres paramètres sont dotés de priors normaux  $N(0, 0.1)$ .

#### 4.3.4 Résultats

##### Évolution de l'hypersignal

La figure 4.4 montre sous forme de boxplot les distributions de probabilités a posteriori des valeurs annuelles successives du facteur latent  $Z$ .

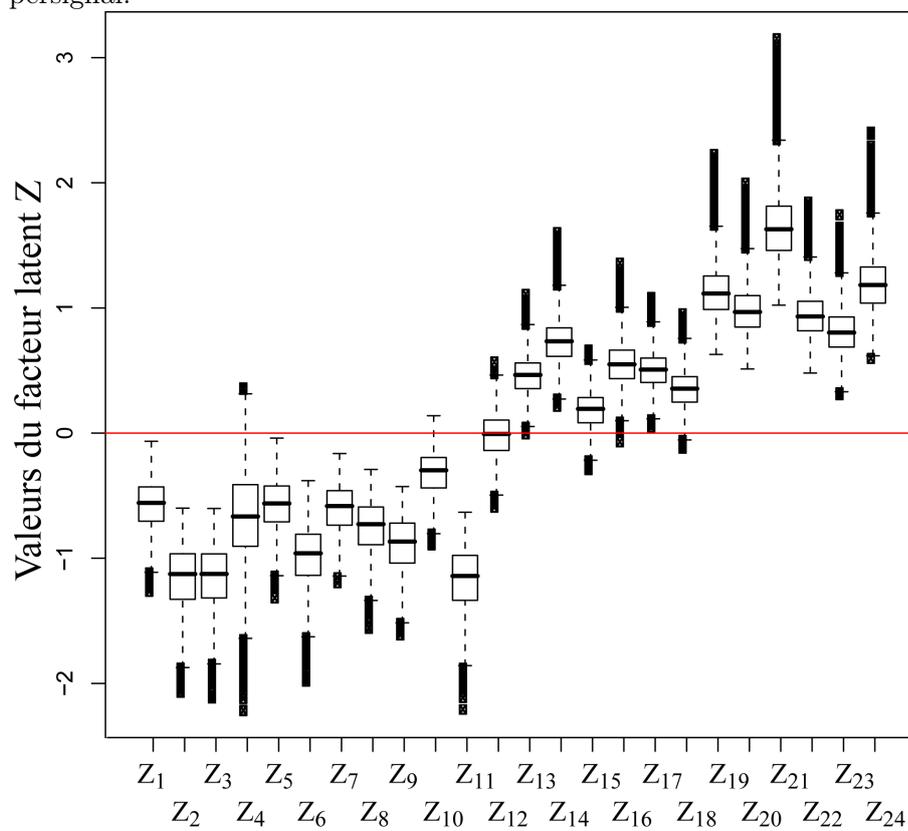
On remarque une très nette évolution dans le temps de l'hypersignal. Les onze premières années (1980-1991) de l'étude sont marquées par des valeurs négatives de ce facteur latent, et par une relative stabilité, tandis que les années 12 à 24 (1992 - 2005) voient leurs valeurs de  $Z$  être positives. Cette seconde période est toutefois plus instable que la première et on constate une réaugmentation du facteur latent entre les années 18 et 19. Avant ce second seuil, une courte période de transition de sept ans (1992-1999), où les valeurs du facteur latent sont variables en restant légèrement positives ; après, une stabilisation à un régime plus élevé.

Il s'agit donc de vérifier de quelle façon cette évolution positive fortement marquée de l'hypersignal s'accompagne de variations des variables descriptives de l'environnement d'une part et des proportions des différents groupes d'espèces dans l'assemblage d'autre part. Notamment, il faut vérifier si de telles variations brutales sont observées dans les deux chroniques de données.

##### Évolution des proportions des différents groupes d'espèces dans le peuplement

*Evolution de la structure de l'assemblage*

Figure 4.4 – Distributions de probabilités annuelles a posteriori de l'hy-persignal.



Les proportions relatives des différents groupes d'espèces constituent une autre variable latente du modèle. Les figures 4.5, 4.6, 4.7 et 4.8 montrent sous forme de boxplot les distributions a posteriori des proportions relatives annuelles des groupes pour les différentes modalités de saison et de positionnement des stations.

Afin de vérifier l'adéquation du modèle aux données, les proportions observées sont visibles sur ces figures, ainsi que la moyenne annuelle de ces observations. Les tendances observées de déclin des effectifs des espèces du groupe 1 et de forte augmentation des espèces du groupe 3 sont bien rendues par le modèle, bien que certains comportements interannuels ne soient pas parfaitement maîtrisés.

Les proportions a posteriori d'espèces du groupe 2 sont relativement stables dans le temps, ce qui semble bien rendre compte de la réalité, malgré quelques variations fortes interannuelles observées. Il apparaît donc que la modification de l'assemblage provient essentiellement d'un remplacement des espèces du groupe 1, par les espèces du groupe 3, et ce, quelle que soit la saison ou la situation amont/aval des stations échantillonnées.

On observe bien également une rupture dans la composition de l'assemblage entre les onzième (1991) et douzième (1992) années de l'étude. Avant cette rupture, le peuplement juvénile est clairement dominé par les espèces du groupe 1 dont les individus représentent plus de la moitié de l'effectif total. Les deux autres groupes composant à quasi égalité le reste du peuplement. Après l'année 1992, la proportion d'individus du groupe 1 dans le peuplement diminue rapidement, tandis que celle des individus des espèces du groupe 3 augmente. On observe également un deuxième seuil d'évolution entre les années 1999 et 2000, marqué par une réaugmentation de la proportion d'individus des espèces du groupe 3 dans le peuplement, accompagné cette fois d'une diminution conjointe des proportions des deux autres groupes. On aboutit donc sur les six dernières années à une domination complète du peuplement par le groupe 3.

#### *Paramètres de la régression linéaire multinomiale*

Le lien entre les données de comptage et le facteur latent est exprimé au sein du modèle linéaire généralisé multinomial. Les distributions a posteriori des coefficients de cette régression sont données sous forme de boxplot dans la figure 4.9. On remarque que la proportion d'individus du groupe d'espèces 1 est corrélée négativement ( $a_1 < 0$ ) avec le facteur latent  $Z$  : une augmentation de l'hypersignal s'accompagne d'une

Figure 4.5 – Proportions relatives annuelles des trois groupes d'espèces pour les pêches d'été et les stations amont.

Les boxplots représentent les distributions a posteriori des paramètres  $p$ , les points, les proportions observées et les lignes, les moyennes des proportions observées.

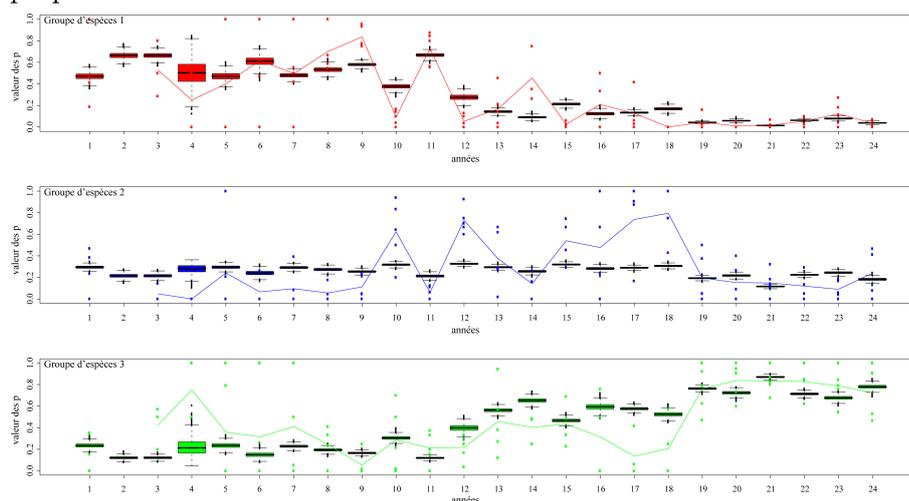


Figure 4.6 – Proportions relatives annuelles des trois groupes d'espèces pour les pêches d'été et les stations aval.

Les boxplots représentent les distributions a posteriori des paramètres  $p$ , les points, les proportions observées et les lignes, les moyennes des proportions observées.

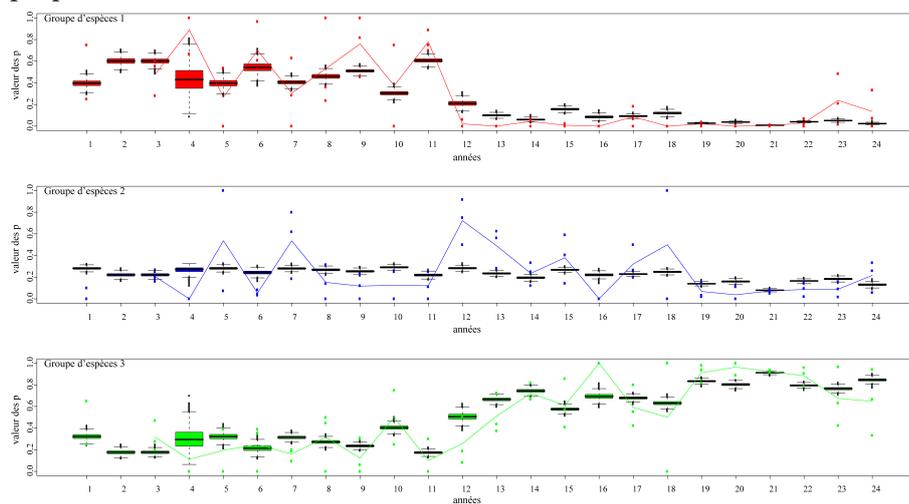


Figure 4.7 – Proportions relatives annuelles des trois groupes d'espèces pour les pêches d'automne et les stations amont.

Les boxplots représentent les distributions a posteriori des paramètres  $p$ , les points, les proportions observées et les lignes, les moyennes des proportions observées.

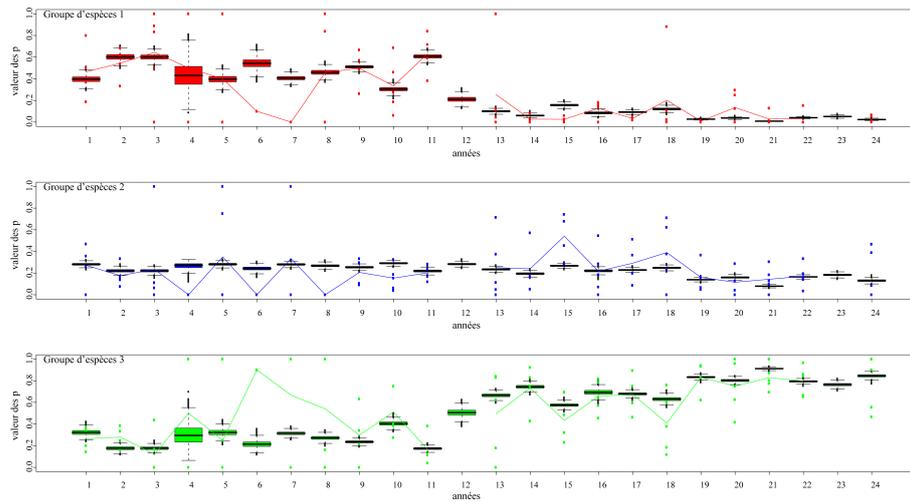
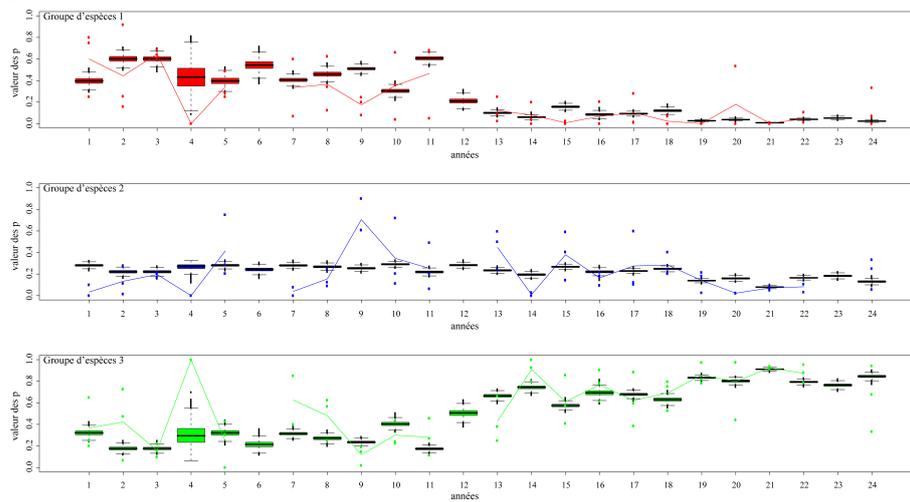


Figure 4.8 – Proportions relatives annuelles des trois groupes d'espèces pour les pêches d'automne et les stations aval.

Les boxplots représentent les distributions a posteriori des paramètres  $p$ , les points, les proportions observées et les lignes, les moyennes des proportions observées.



diminution de l'importance des espèces du groupe 2 dans le peuplement.

Toutefois, on remarque que pour un état "nul" de l'hypersignal, chacune de ces deux proportions a tendance à être légèrement inférieure à la moyenne ( $(b1, b3) \leq 0$ ), et par conséquent, dans cette situation, le groupe 2 devrait être en légère augmentation.

La saison de pêche a un effet différent sur les groupes d'espèces 1 et 2. Lors des pêches d'automne, la proportion d'individus du groupe 1 sera légèrement inférieure ( $c1 < 0$ ) et la proportion d'individus du groupe 3 sera légèrement supérieure ( $c3 > 0$ ) dans le peuplement, que lors des pêches d'été.

De même, pour la situation amont/aval de la station, les individus appartenant au groupe 1 sont en plus faible proportion dans les stations situées dans le panache de rejet chaud ( $d1 < 0$ ), en aval de la centrale nucléaire, tandis que les individus du groupe 3 y sont proportionnellement plus nombreux ( $d3 > 0$ ).

## Evolution des composantes de l'environnement

### *Structure temporelle des données descriptives de l'environnement*

En observant l'évolution temporelle des proportions respectives des trois groupes d'espèces dans le peuplement, il apparaît qu'un seuil existe entre l'année 11 et l'année 12. Ce seuil peut s'expliquer par les variables environnementales si on constate également un tel décrochement dans l'évolution interannuelle d'une ou plusieurs variables environnementales entre ces deux années. La figure 4.10 résume l'évolution interannuelle de ces variables descriptives.

Deux variables, C12 et Qmx2, ne connaissent qu'une très faible évolution tendancielle entre les deux périodes, malgré des variations inter-annuelles parfois assez marquées.

Les autres variables descriptives de l'environnement se séparent en deux groupes : quatre variables décroissent en moyenne entre les deux périodes (C18, Qm1, Qmx1 et Qm2), les trois autres augmentant (Cmx, S1 et S2).

### *Contributions relatives des variables descriptives de l'environnement à l'hypersignal*

La figure montre sous forme de boxplot les distributions a posteriori des valeurs des composantes principales  $\alpha_p$ , et donc les contributions relatives des différentes variables descriptives de l'environnement au facteur latent  $Z$ .

Figure 4.9 – Distributions a posteriori des coefficients du GLM multinomial à fonction de lien “logit”

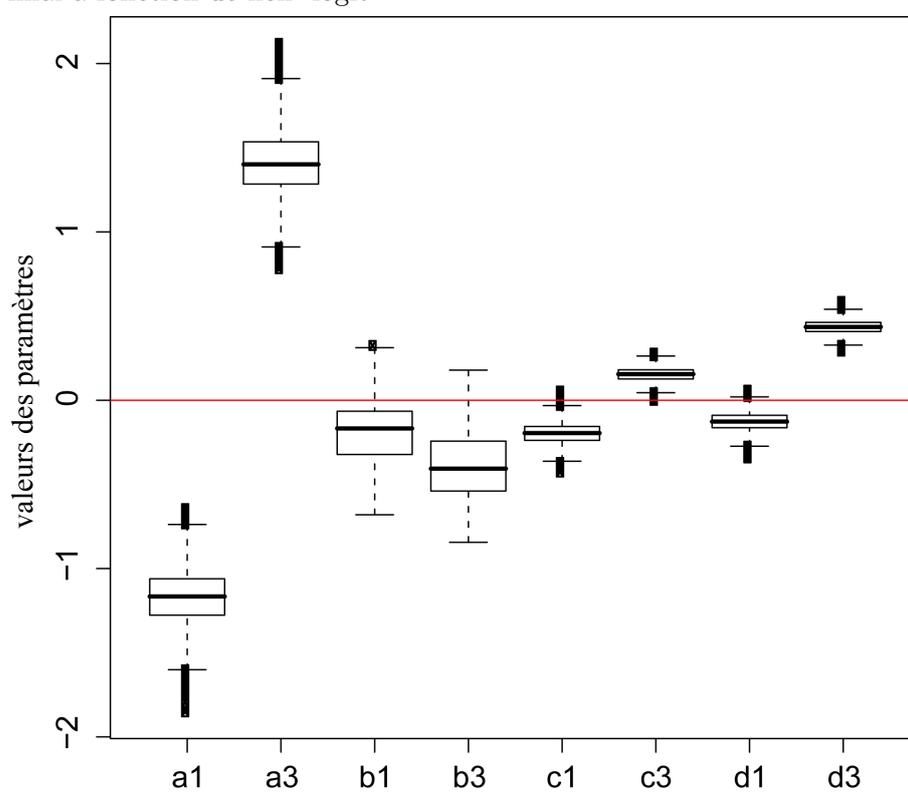
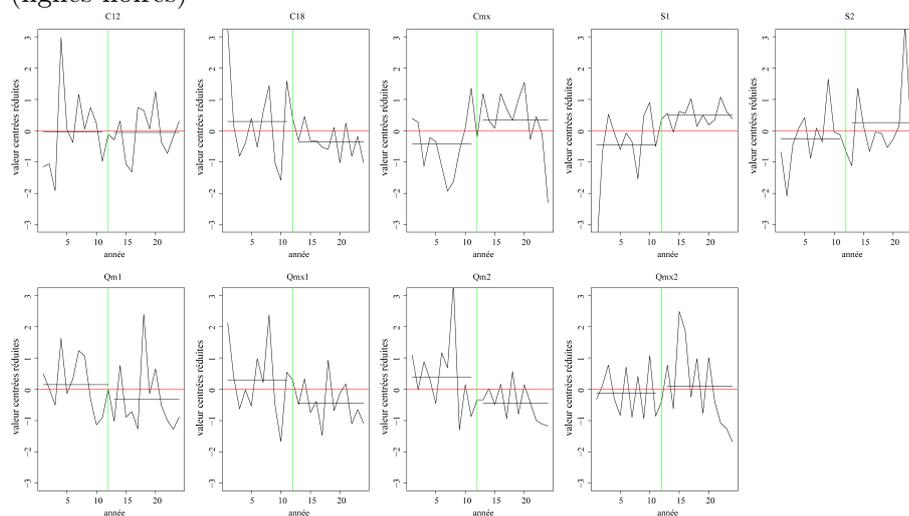


Figure 4.10 – Évolution interannuelle des variables descriptives de l'environnement, et valeurs moyennes sur les périodes 1980-1992 et 1993-2005 (lignes noires)

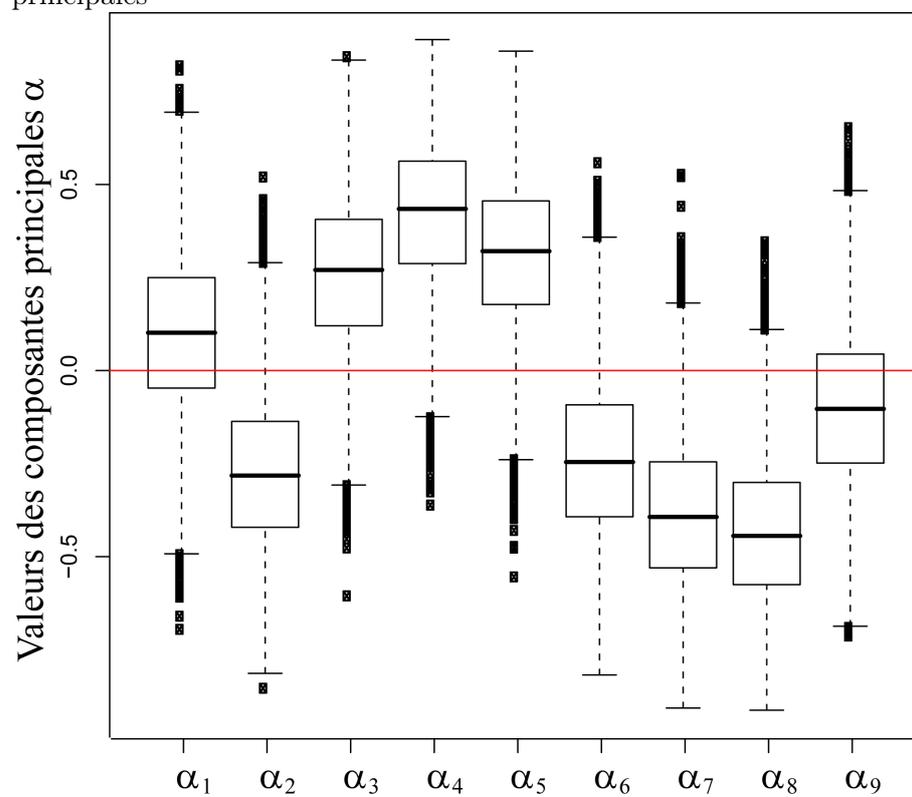


Pour une variable donnée, un coefficient de régression associé positif signifie une contribution positive de la variable au facteur latent. Une augmentation de cette variable s'accompagne d'une augmentation de la valeur du facteur latent. Une telle augmentation entraîne une augmentation de la proportion d'individus des espèces du groupe 3 dans le peuplement (puisque le coefficient de régression de  $Z$  sur cette proportion est positif) et conjointement une diminution de la proportion d'individus des espèces du groupe 1 dans le peuplement (puisque le coefficient de régression de  $Z$  sur cette proportion est négatif)

Les corrélations des variables descriptives de l'environnement avec les facteurs latents ne sont pas très fortes et les variables C12 et Qmx2 ne contribuent que faiblement aux variations du facteur latent, la médiane des distributions a posteriori de leurs coefficients respectifs ( $\alpha_1, \alpha_9$ ) étant très proches de zéro. On distingue trois variables qui semblent corrélées plutôt positivement avec l'hypersignal (Cmx, S1 et S2). Les variables décrivant le débit sont quant à elles corrélées plutôt négativement avec le facteur latent, de même que C18, qui est la seule variable du régime thermique négativement corrélée avec  $Z$ .

Concernant les variables du régime thermique, la variable contribuant le plus au facteur latent est la variable S1, suivie de la variable S2 et de

Figure 4.11 – Distributions de probabilités a posteriori des composantes principales



la variable Cmx. Ces trois variables contribuent positivement au facteur latent. Une augmentation d'une de ces variables s'accompagne donc d'une augmentation de la de la proportion d'individus des espèces du groupe 3 dans le peuplement ainsi qu'à une diminution de la proportion d'individus des espèces du groupe 2.

Les variables C18, Qm1, Qmx1 et Qm2 contribuent négativement au facteur latent. Leur augmentation a donc un effet négatif sur la proportion d'individus des espèces du groupe 3 et positif sur la proportion d'individus des espèces du groupe 1.

Le facteur latent caractérise donc l'existence de trois groupes de variables, qui, comme nous l'avons vu à la section précédente, se séparent par leur évolution tendancielle entre les deux périodes 1980-1991 et 1992-2005.

#### 4.3.5 Interprétations écologiques

##### *Vers une explication de l'hypersignal*

L'hypersignal doit être interprété comme un signal non observé qui, à large échelle, coordonnerait les variations locales observées. Un phénomène bien connu et documenté est l'influence de la position du Gulf Stream sur le climat en Europe. Lorsque le Gulf Stream se déplace plus vers le nord, l'Europe connaît en général des étés très chauds et secs. Des études ont ainsi montré l'influence de la position du Gulf Stream sur les communautés végétales (WILLIS et al., 1995), ou sur le succès reproducteur des poissons (Nunn et al., 2003, 2007). La structure observée dans les données pourrait partiellement correspondre à une manifestation de la position du Gulf Stream.

Toutefois, l'existence d'un saut net dans cette structure pourrait aussi cacher un phénomène plus brutal. Un phénomène local brutal pourrait avoir déstabilisé l'équilibre de l'assemblage en défavorisant les espèces du groupe 1 et le déplacement de l'équilibre aurait pu par la suite avoir été entretenu et/ou amplifié par les conditions climatiques. Par exemple, une crue d'amplitude exceptionnelle (estimée comme crue de période de retour de 110 ans (Cattanéo et al., 2001)) a été observée sur le Rhône en octobre 1993. Celle-ci est considérée comme suffisamment puissante pour modifier les conditions locales d'habitat. Le passage d'une communauté dominée par les espèces du groupe 1 à une communauté dominée par les espèces du groupe 3, plus généralistes, pourrait être partiellement une conséquence de cette modification d'habitat, renforcée par l'évolution des régimes thermiques et hydriques que l'on observe dès 1992.

*Effets du régime thermique*

La faible contribution de la variable C12 au facteur latent doit provenir du fait que, comme nous l'avons vu au chapitre précédent, cette variable a un effet similaire sur les le nombre de juvéniles des espèces des groupes 1 et 3. Par conséquent, les proportions relatives de ces deux groupes ne doit que peu varier avec les variations de C12.

Un régime thermique plus élevé, caractérisé par des périodes printanières plus courtes (passage du seuil 18°C plus tôt) et estivales plus longues, et des périodes printanières et estivales plus chaudes a un effet favorisant une forte proportions de juvéniles des espèces du troisième groupe. Les trois espèces qui composent ce groupe ont été identifiées comme des espèces généralistes, des parties sud de l'Europe (Daufresne et al., 2003). La tendance au réchauffement des eaux que l'on observe dans le régime thermique leur est donc bénéfique. Le franchissement précoce d'un seuil de température de 18°C pourrait s'accompagner d'une première ponte plus précoce de ces espèces. Notamment, dans le cas des espèces capables de pontes fractionnées, comme cela a été montré pour le chevesne (Economou et al., 1991; Vriese et al., 1994) et est probable pour le spirin (Tissot and Souchon, 2008), un premier épisode de ponte plus précoce accompagné d'une forte survie de ces premiers juvéniles pourrait avoir un effet très positif sur la production de larves et de juvéniles.

Par la suite, une période estivale plus longue, donc une période de production primaire (microalgues et invertébrés) par l'environnement plus longue, favoriserait également une meilleure croissance des juvéniles de ces espèces et une meilleure survie subséquente. Cette production primaire du milieu est sans doute meilleure lorsque l'année (printemps comme été) est plus chaude. L'effet positif de l'augmentation de la somme de degrés-jours des périodes printanière et estivale est en accord avec les précédents travaux qui ont montré que des températures plus élevées favorisent une meilleure croissance des juvéniles de cyprinidés (Mills and Mann, 1985; Daufresne et al., 2003; Wolter, 2007). Il est probable que cette augmentation induise une diminution du temps d'incubation, provoquant un développement plus précoce des larves, ce qui se traduit plus tard par une meilleure survie des juvéniles durant les premiers mois de vie.

Le comportement opposé pour les espèces du groupe 1 s'expliquerait par le fait que ces espèces (ablette et vandoise) sont en limite sud de leur zone de répartition et préféreraient donc les eaux plus fraîches (Daufresne et al., 2003). Le réchauffement climatique s'accompagnant de printemps

et d'étés plus chauds défavoriserait ces espèces. En l'absence d'information sur les effectifs d'adultes, il est toutefois impossible de savoir si cette sélection par la température ne se fait qu'au niveau des juvéniles avec des taux de survie très faibles, ou bien si c'est le nombre d'individus adultes qui, diminuant fortement en réponse à ces nouvelles contraintes, ne serait plus capable de générer des effectifs juvéniles importants.

#### *Effets du régime hydrique*

Un régime hydrique plus élevé, caractérisé par des débits moyens printaniers et estivaux plus importants et des pics de débit printaniers plus élevés a un effet dépréciateur sur la proportion d'espèces du groupe 3 dans la communauté. Notamment, des crues printanières et un débit moyen estival plus importants sont les effets les plus importants.

Plusieurs phénomènes pourraient expliquer cela. L'augmentation des débits pourrait entraîner un phénomène de dévalaison accru (Carter and Reader, 2000), surtout sur de jeunes spécimens dont la capacité de nage est réduite, et peu à même de résister à des conditions de courants plus forts. Pour les individus un peu plus âgés ayant de meilleures capacités, l'augmentation des débits pourrait entraîner une dépense énergétique accrue, pouvant se solder par la mort des individus les plus faibles, d'autant plus que ces situations de courant pourraient être préjudiciables à la production primaire du milieu et occasionner des périodes de diète pour les juvéniles. Les épisodes de crues importants pourraient également créer, par débordement dans le lit majeur, des zones de fuites pour les juvéniles cherchant des conditions de courant plus faibles. Cependant, lorsque le niveau d'eau baisse, un grand nombre de ces juvéniles pourraient se retrouver piégés dans ces zones exondées. Bolland et al. (2008) ont montré que le gardon et le chevesne faisaient partie des espèces les plus touchées par ce phénomène. Enfin, ces modifications du régime hydrique pourraient perturber les épisodes de ponte tardifs des espèces à ponte fractionnée, et causer une diminution forte des juvéniles de ces espèces.

#### *Effets relatifs de la température et du débit*

L'importance relative de ces deux facteurs et la prépondérance de l'un et de l'autre dans la structuration des assemblages piscicoles a été débattu et l'est toujours (Olden and Naiman, 2010; Poff and Allan, 1995; Poff et al., 1997). L'étude des relations entre ces deux régimes et les proportions relatives des espèces de cyprinidés, rassemblées en groupes, apporte un éclairage nouveau sur cette question. On note ainsi que les

régimes thermique et hydrique jouent un rôle antagoniste favorisant chacun un des groupes d'espèces. Surtout, on ne retrouve pas de signes de prépondérance de la température par rapport au débit. Au contraire, les deux variables dont les composantes principales sont les plus fortes sont des variables de débit ( $Q_{m1}$  et  $Q_{m2}$ ). Il apparaît donc que le débit joue un rôle important dans la structuration des assemblages de juvéniles en grands cours d'eau.

#### 4.3.6 Test de la capacité prédictive du modèle

L'inférence des paramètres du modèle a été réalisée sur les données des années 1980 à 2005. On utilise les données observées pour les années 2006 à 2009 pour tester la fiabilité du modèle en termes de prédictions.

En appliquant la méthode présentée à la section 4.2.3, on génère les distributions de probabilités prédictives du facteur latent pour les quatre nouvelles années. Elles sont présentées dans la figure 4.12. Les intervalles de crédibilité pour les prédictions du facteur latent sont très larges. En effet, le terme de précision des distributions du facteur latent, pour chaque année, est égale à la somme des précisions des régressions linéaires des  $X_i$ ,  $i = 1, \dots, p$ , sur  $Z$ . On sait que plus  $|\alpha_i|$  est grand, plus la précision de cette régression est grande. Or, comme montré à la figure 4.11, les différentes composantes principales sont petites en valeurs absolues, et donc la précision sur  $Z$  est faible.

A partir des distributions prédites des  $Z$ , on calcule les proportions prédites pour les trois groupes d'espèces. Les distributions prédites des proportions pour ces quatre années et pour les pêches réalisées en amont et en aval sont présentées aux figures 4.13 (pêches d'été) et 4.14 (pêches d'automne).

Les données observées appartiennent dans environ 2 cas sur 3 (28 points sur 42) aux intervalles de crédibilité à 95% prédits.

## 4.4 Conclusion

Les modèles d'analyse de données multivariées sont des méthodes classiques d'exploration de données, permettant de révéler les structures de corrélations au sein d'un jeu de données, ou entre deux jeux de données. L'extension des méthodes de régression linéaire aux modèles linéaires généralisés, permet également d'explorer les structures de corrélation entre des données de nature différente.

Figure 4.12 – Distributions de probabilités annuelles a posteriori de l'hyper-signal  $Z$ , pour les années 1980 à 2005 (à gauche de la ligne verte) et distributions prédictives a posteriori du facteur latent  $Z$ , pour les années 2006 à 2009 (à droite de la ligne verte).

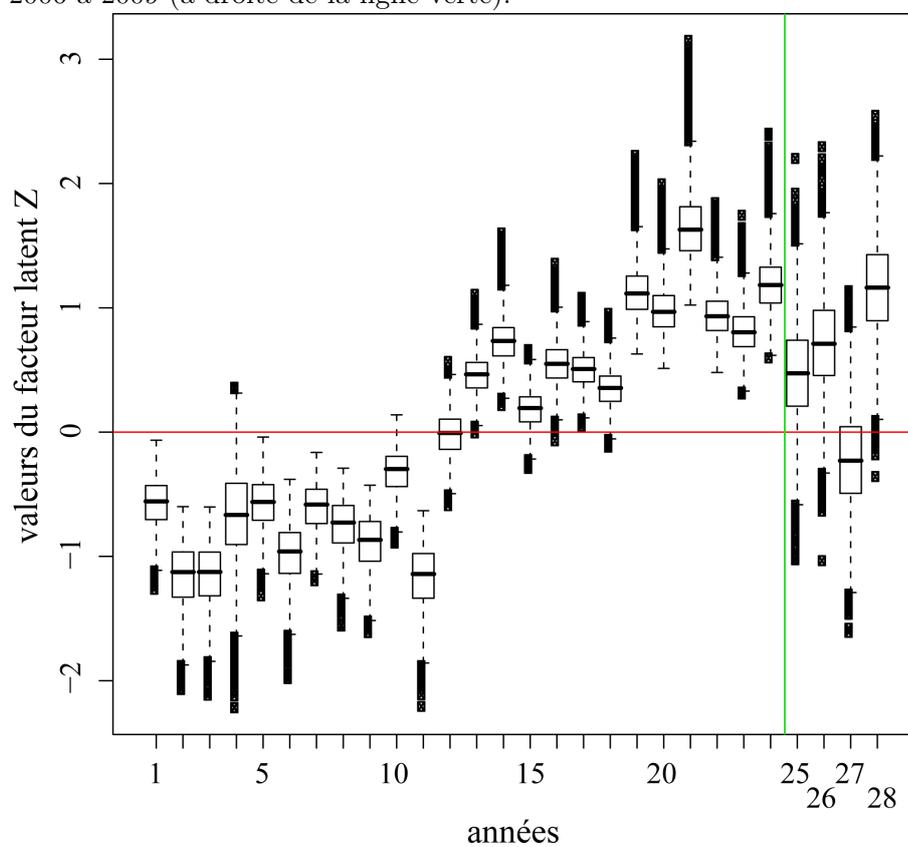


Figure 4.13 – Distributions prédictives a posteriori (boxplots) et observations (lignes) des proportions relatives des trois groupes d'espèces lors des pêches d'été, pour les stations situées en amont et en aval de la centrale nucléaire.

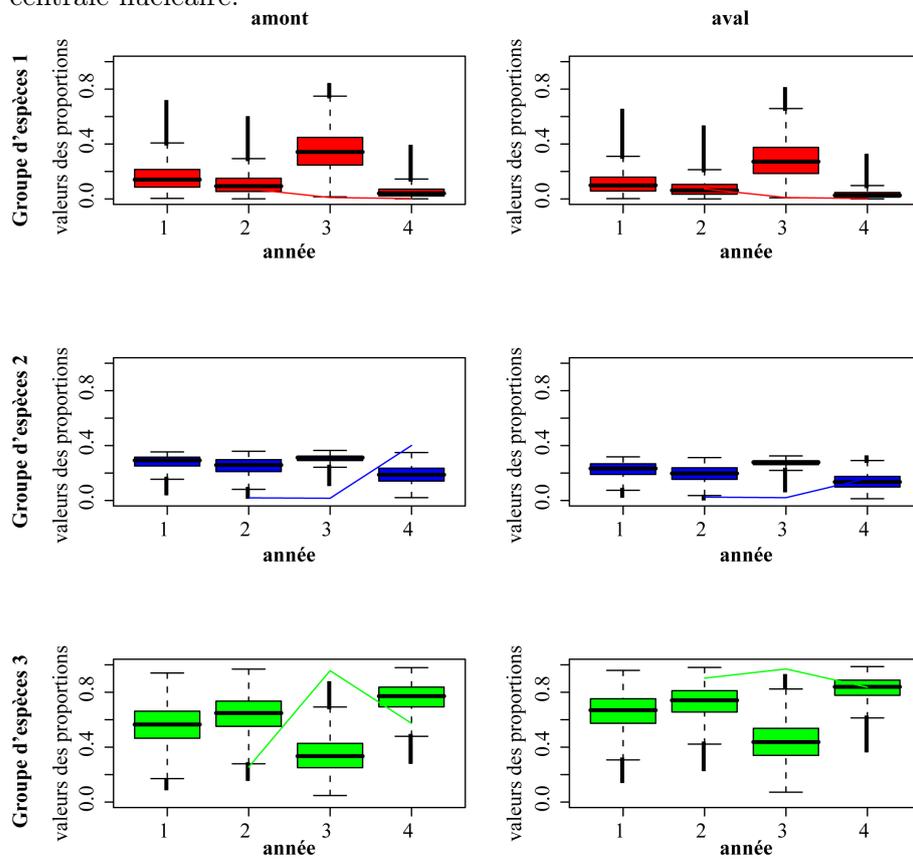
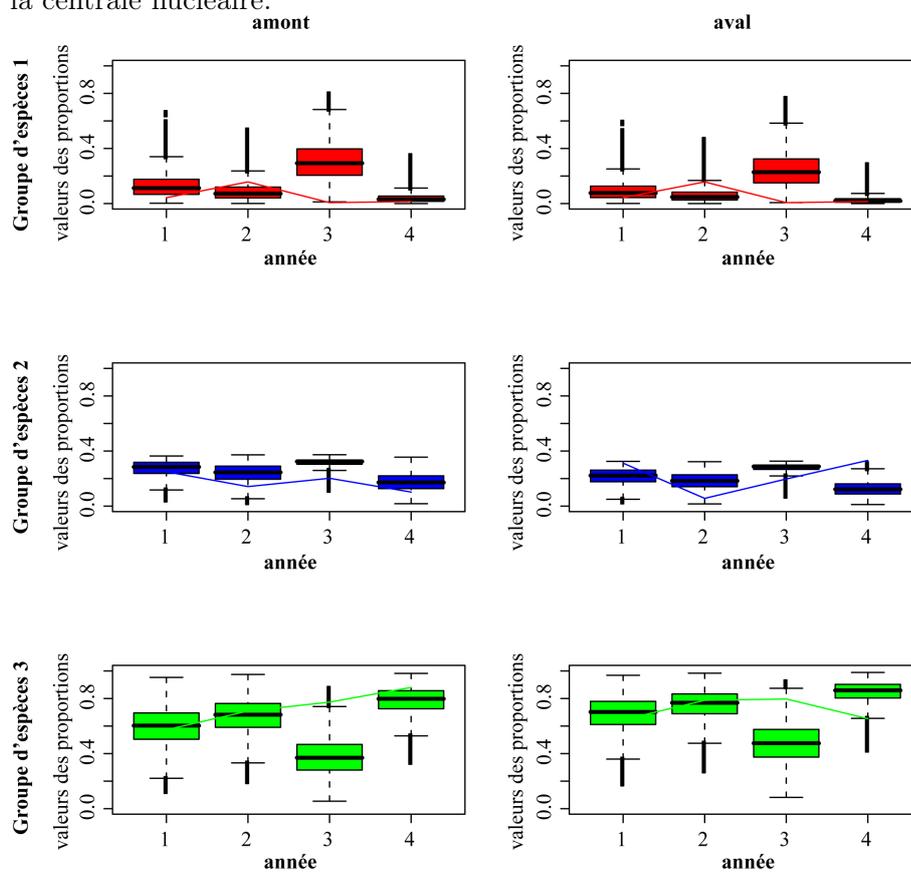


Figure 4.14 – Distribution prédictives a posteriori (boxplot) et observations (lignes) des proportions relatives des trois groupes d'espèces lors des pêches d'automne, pour les stations situées en amont et en aval de la centrale nucléaire.



Sur l'exemple de la communauté de cyprinidés juvéniles du Rhône, nous avons montré l'intérêt que peuvent présenter ces méthodes, utilisées dans un cadre bayésien. L'évolution des proportions relatives des trois groupes d'espèces qui la compose est particulièrement bien représentée par ce modèle, qui a en outre permis l'expression des relations avec des variables simples décrivant l'environnement annuel thermique et hydrique. De plus, la structure de la communauté est bien prédite pour des nouvelles valeurs d'environnement dans 2 cas sur 3, ce qui est particulièrement intéressant, même s'il apparaît que des mécanismes restent mal maîtrisés.

Trois principales améliorations du modèle peuvent être proposées :

- structurer le comportement du facteur latent

En l'absence de toute structure prédéfinie, le modèle a révélé une évolution de l'hypersignal par seuil. Il serait donc intéressant d'étudier l'apport d'une structure a priori de saut à valeur aléatoire en une année  $\tau$ , sur laquelle un prior serait fixé, afin de déterminer en quelle année le saut apparaît. Un modèle de type "Shifting Level", comme proposé dans Fortin et al. (2004) pourrait servir de squelette à l'hypersignal. D'autres structures a priori favorisent un comportement plus lisse (autorégressif, splines, etc...)

- faire jouer plusieurs facteurs latents

Le modèle présenté ici ne comporte qu'un seul facteur latent, régissant les évolutions conjointes de tous les groupes d'espèces. L'apport d'un second facteur latent pourrait être étudié. En effet, l'information portée par une ou plusieurs variables explicatives qui ne se corrèleraient pas aux autres et ne s'exprimerait donc pas dans le premier facteur latent, pourrait s'avérer essentielle à la compréhension du système. Deux questions se posent alors : comment écrire ce second facteur latent, conditionnellement au premier, afin de respecter une structure orthogonale des deux facteurs ? Comment modéliser les proportions des différents groupes, via la fonction logistique : chaque proportion doit-elle être combinaison linéaire des deux facteurs latents, ou bien chaque facteur doit-il exprimer principalement l'évolution d'un seul groupe ?

- structure libre de groupes d'espèces

Au lieu de constituer a priori les groupes d'espèces, on pourrait proposer une classification a posteriori (bayésienne) en s'appuyant sur le modèle reliant les variables  $X$  et  $Y$ .



**Au final**



## Chapitre 5

# Conclusion générale et perspectives

### 5.1 Bilan des travaux de thèse

#### 5.1.1 Rappel des objectifs et des principales contraintes

Au cours de cette thèse, nous souhaitons améliorer la connaissance des déterminismes naturels de la structure des assemblages de poissons. Cette question est d'un grand intérêt, tant d'un point de vue écologique fondamental, que d'un point de vue plus opérationnel, dans le cadre de la gestion et la restauration des milieux aquatiques. La compréhension des mécanismes responsables des variations interannuelles de structure nécessite l'utilisation de jeux de données suffisamment longs, afin de garantir l'existence de contrastes suffisants et de discerner des tendances. A ce titre, les jeux de données issus du suivi de l'implantation des Centrales Nucléaires de Production Électriques comme celui de Bugey, mis en place à la fin des années 1970, sont particulièrement intéressants car ils couvrent désormais une période d'une trentaine d'années.

Cependant le protocole de pêche électrique des bords, défini initialement dans ce cadre, tend à échantillonner préférentiellement les jeunes stades de poissons, juvéniles de l'année ou d'un an. Nous avons donc choisi de restreindre notre cadre d'étude aux déterminismes naturels des variations d'assemblage de juvéniles de l'année (ou 0+). De plus, l'échantillonnage en un seul passage dans un système ouvert ne permet pas de déterminer à la fois la densité absolue de poissons et leur probabilité de capture. Nous avons donc dû supposer que les conditions de pêches sont

identiques dans le temps : sous cette hypothèse, les indices d'abondances obtenus sont alors comparables et nous renseignent sur les variations d'effectifs des espèces. Enfin, comme nous n'avions que peu d'information sur les effectifs adultes, et connaissant les stratégies reproductrices des cyprinidés étudiés, nous avons formulé l'hypothèse selon laquelle le nombre de 0+ observés une année  $n$  ne dépend pas du nombre d'adultes présents cette même année. Ce faisant, nous supposons que les contraintes abiotiques du milieu constituent les facteurs principaux structurant les assemblages.

Toutefois, les grandes variables de l'environnement responsables de ces variations ne sont pas bien documentées. En nous appuyant sur les quelques travaux existant sur le sujet, nous avons utilisé des descripteurs assez simples des régimes thermique et hydrique. En utilisant une procédure de cluster, nous avons également défini trois groupes d'espèces sur la base de leurs réponses synchrones aux variations d'environnements thermique et hydrique. Nous avons mis en place des procédures statistiques permettant l'exploration de l'influence de ces variables synthétiques sur les assemblages juvéniles. La réponse des groupes d'espèces et les descripteurs synthétiques ont ensuite été mis en relation grâce une approche bayésienne hiérarchique, permettant de prendre en compte les phénomènes de similarités interannuelles des chroniques de capture.

### 5.1.2 Apports statistiques

Dans un premier temps, nous avons proposé une procédure de sélection de variables dans le cadre d'un modèle de régression linéaire poissonnienne, afin de déterminer les rôles relatifs de ces variables dans les variations interannuelles d'abondance des juvéniles. Les modèles de type Poisson constituent une approche assez classique pour des données de comptage, mais ce travail n'avait encore jamais été mené sur les espèces de poissons en grande rivière. La sélection des variables d'intérêt a été menée en utilisant les techniques du Facteur de Bayes pour comparer les modèles entre eux. Un premier modèle, portant sur les effectifs annuels observés pour chaque espèce, a été proposé pour stationnariser l'effet mal contrôlé de la CNPE de Bugey. Ce modèle a permis une première approche de l'influence du milieu naturel sur le succès reproducteur des espèces concernées, mais ne prenait pas en compte la surdispersion observée dans les données. Un second modèle a été construit, modélisant les occasions de pêche station par station, pour limiter la surdispersion incompatible avec l'hypothèse poissonnienne, mais cette correction a été

insuffisante. Un dernier modèle a enfin été développé, introduisant un effet aléatoire dans la description des effectifs attendus par espèce et par station. L'introduction d'une part de variabilité résiduelle, non expliquée par les variations d'environnement, dans les effectifs observés a permis de rendre compte de la surdispersion que l'on observait jusque là. Ce modèle n'a pas permis de mettre en évidence de patron significatif d'influence du rejet d'eau chaude par la CNPE sur les effectifs de juvéniles de cyprinidés du Rhône.

Dans un deuxième temps, nous avons essayé de nous affranchir des problèmes de stationnarité dans le temps du protocole expérimental en travaillant conditionnellement aux effectifs capturés, c'est-à-dire en décrivant les variations de proportions relatives des différentes espèces. Pour cela, nous avons mis en place un modèle à facteur latent partagé. Un facteur latent est une variable qu'on ne peut pas observer directement à l'aide du protocole expérimental de prise de données, mais qui est utile à la représentation du processus étudié, tel qu'on le conçoit. Nous l'avons baptisé hypersignal pour mettre en évidence ce rôle au chapitre 4. Sa fonction est de réduire la dimension de l'espace des variables explicatives en faisant exprimer à ces variables la partie commune de leur signal. Ce résumé évite d'engager une procédure de sélection des variables explicatives comme au chapitre précédent. Les variables d'origine, les observées, sont donc corrélées car elles dérivent, conceptuellement, d'un même groupe de facteurs latents. L'hypersignal présente l'avantage d'autoriser l'expression des corrélations existant entre des données de natures différentes, comme dans une analyse discriminante ou dans une régression PLS : les variables continues décrivant les principales composantes des régimes thermique et hydrique, d'un côté ; les données catégorielles d'effectifs par groupe d'espèces, de l'autre. Les corrélations particulières de chacune des variables descriptives de l'environnement et de l'hypersignal décrivent l'importance relative de chacune d'elles dans les variations interannuelles des proportions des groupes d'espèces.

### 5.1.3 Apports en écologie

Ces outils de statistique avancée ont permis l'analyse de l'influence des régimes thermique et hydrique sur un compartiment biologique particulier de l'écosystème : les assemblages juvéniles de poissons. Grâce à eux a été mise en évidence une importance conjointe de ces deux régimes sur le succès reproducteur des espèces de cyprinidés du Rhône, et par conséquent, sur la structure même de cet assemblage. Si le premier mo-

dèle révèle une légère prépondérance de l'influence du régime thermique sur les abondances de juvéniles, par rapport au régime hydrique, le modèle à facteur latent partagé ne confirme pas ce résultat. Il semble donc que dans l'ensemble, les deux régimes jouent un rôle, mais probablement à des moments différents, et pour des raisons différentes.

Le régime thermique a une influence directe sur les processus biologiques, telles que la maturation des gonades des individus adultes, sur le timing de ponte (probablement en relation avec la rapidité de maturation des gonades), le développement embryonnaire et larvaire et la croissance des juvéniles. Tous ces éléments ont un impact direct sur l'abondance de juvéniles, soit via des pontes plus faibles, soit via un développement ou une croissance plus lente entraînant des mortalités plus importantes. En adaptant les descripteurs du régime thermique aux grandes étapes du cycle de vie, on peut mettre en évidence l'importance des variations interannuelles de régime.

Le régime hydrique quant à lui a un rôle plus indirect. Il régit les conditions d'habitat, tant en quantité qu'en qualité, durant les phases biologiques importantes du cycle de vie, de l'habitat de ponte jusqu'à la disponibilité et l'accessibilité de zones de grossissement ou d'abris accessibles. Il semble néanmoins que le débit puisse avoir un rôle plus direct d'entraînement des larves et des jeunes juvéniles. Cependant, si le rôle de la température a déjà été bien documenté, des études supplémentaires concernant l'adaptation des espèces au régime hydrique, et surtout la description plus fine de ce régime seront nécessaires. On pense notamment aux variables introduites par Poff and Ward (1989) du timing, de la durée et de l'amplitude des différents événements hydrologiques.

## **5.2 Perspectives dans un contexte complexe et incertain**

### **5.2.1 Approfondissement des méthodes**

Comme nous l'avons vu, certains choix ont dû être faits, compte tenu des différentes contraintes rencontrées. En particulier, l'absence d'influence des classes d'âge supérieures dans le déterminisme de l'assemblage juvénile semble être une hypothèse très forte. La dynamique des espèces, si possible prises en relation les unes des autres, serait sans doute porteuse de nouvelles informations. Le développement d'un modèle de dynamique de population associé à un modèle statistique d'influence

de l'environnement sur les paramètres démographiques permettrait de prendre en compte la totalité du cycle de vie et d'intégrer l'information portée par l'assemblage de poissons dans son ensemble. Cependant, cette approche ne pourra être mise en place que sur des systèmes aquatiques plus facilement échantillonnables grâce à des protocoles adaptés.

Concernant le modèle à facteur latent partagé, l'introduction d'une structure temporelle a priori sur l'hypersignal pourrait permettre de le relier à des hypothèses de modifications de l'environnement auxquelles nous n'aurions pas pensé pour l'heure, tandis que l'introduction d'un facteur latent multivarié permettrait sans doute de décrire des variations plus fines des proportions relatives des trois groupes d'espèces.

## 5.2.2 Transposer les méthodes

### La transposabilité géographique de la méthode

L'étude menée au cours de cette thèse n'utilise les données que d'un seul site. Il s'agit donc d'une étude géographiquement très localisée, dont les résultats ne peuvent être interprétés que comme tel. Dans le cadre de la DCE, il faut développer des modèles généralisables à des unités géographiques qui soient au moins régionales, voire nationales. Le contexte communautaire et les structures hiérarchiques nous encouragent même à envisager des modèles à l'échelle européenne, comme c'est déjà le cas pour les indices multimétriques pour la bioindication à partir des poissons (Pont et al., 2006). Avant de voir aussi grand, il faudra déjà vérifier que les résultats obtenus sont transposables à d'autres cours d'eau semblables. Dans le cadre des suivis d'implantations de CNPE, EDF dispose de données de suivis sur de nombreux sites, notamment sur la Loire et sur le Rhône aval. Ces jeux de données permettraient de vérifier que les effets environnementaux déterminant les assemblages juvéniles suivent des lois plus générales.

Les effets de l'environnement sur les assemblages de juvéniles devront également être testés pour d'autres types de communautés, probablement sur des cours d'eau de plus petite taille. Les jeux de données issus du Réseau Hydrobiologique et Piscicole (-2004) ou du Réseau de Contrôle de Surveillance (2005-) pourront sans doute apporter de nouveaux éclairages.

## **Il y a quatre compartiments biologiques**

Les poissons ne constituent pas le seul compartiment biologique visé par la DCE pour servir de support à la bioévaluation des milieux aquatiques. La DCE préconise des outils de bioévaluation faisant intervenir les poissons, les macroinvertébrés, les macrophytes et les diatomées. Chacun de ces compartiments correspond à un indice observable du fonctionnement de l'écosystème dans son ensemble. De plus, la DCE retient le principe du "one out, all out". Cela signifie que si l'état écologique d'une masse d'eau est jugé mauvais sur la base de l'un des quatre compartiments, alors la masse d'eau est considérée en mauvais état, quels que soient les classements obtenus par l'étude des autres compartiments. Pour classer une masse d'eau dans la catégorie "bon état écologique", il faut que les indicateurs biologiques des quatre compartiments donnent un tel jugement.

Une telle préconisation induit de fait une multiplication des relations pression-impact à définir.

A l'état pristine, chacun de ces compartiments est adapté aux conditions "naturelles" du milieu, induisant des cycles de vie particuliers et non synchronisés. Il faut donc, pour chacun de ces compartiments, quantifier la gamme de variations naturelles des assemblages, et identifier les grandes composantes de l'environnement naturel responsables de cette variabilité, afin de ne pas confondre variabilité naturelle et effet des pressions anthropiques. Dans le même temps, il faut identifier pour chaque compartiment, les nouvelles métriques qui varient le moins selon les facteurs naturels, mais le plus selon les gradients d'altération (Wasson, 2001). De même que pour l'étude des poissons, plusieurs pistes devront être explorées, les paramètres de diversité écologique et les traits biologiques des organismes pourront être testés, ainsi que des approches multivariées permettant de s'abstraire de la notion d'indice (Wasson, 2001).

## **Et de nombreuses pressions possibles**

Si la DCE recommande l'étude de plusieurs compartiments biotiques dans le cadre de la bioévaluation, c'est pour voir le fonctionnement des écosystèmes aquatiques dans leur ensemble, mais aussi parce qu'on s'attend à ce que chacun des compartiments permettent de mettre en évidence les effets de certains types de pressions particuliers. Ainsi, il est avéré que les invertébrés ainsi que les diatomées sont particulièrement

sensibles aux conditions chimiques du milieu, et sont donc bien adaptés pour témoigner des pressions responsables des pollutions, plus ou moins diffuses. Les poissons et les macrophytes semblent quant à eux, à dire d'experts, moins sensibles à ces types de pression, mais répondent par contre fortement aux modifications d'habitat, et donc par extension, aux pressions exercées sur l'hydrogéomorphologie des cours d'eau.

Ces types de pressions se déclinent bien évidemment en une multitude d'impacts comme la modification des berges ou du chenal (par exemple à des fins de navigation ou de protection), la régulation du débit et la fragmentation longitudinale (mise en place de barrages hydro-électriques, de seuils pour l'adduction d'eau, les prises d'eau pour l'irrigation, etc), les pollutions chimiques (métaux lourds, pesticides, fertilisants) et les pollutions organiques (domestiques ou liées aux activités agricoles du bassin) (Tejerina-Garro et al., 2005). A ces pressions "classiques" s'ajoutent un certain nombre de pressions émergentes dont les possibles retombées ne sont absolument pas connues, telles que le changement climatique, les espèces non-indigènes introduites et colmatage des substrats par les limons en raison des régulations de débits et de l'utilisation intensive des terres agricoles. La tâche est donc particulièrement complexe. Il s'agit de déterminer les grandes catégories de forces motrices, de déterminer et quantifier les pressions induites par ces forces motrices, et enfin quantifier les effets sur les compartiments biotiques, que ce soit à l'aide de métriques, d'indices ou directement sur la base de compositions spécifiques. Wasson (2001) préconise l'utilisation de modèles de réponses biologiques complexes : à l'échelle d'un tronçon, comment faire la part entre les différentes causes d'altération (physiques, chimiques, toxiques) dans la réponse d'un peuplement ? peut-on hiérarchiser ces impacts ? en déduire le niveau de réduction optimal de chacune des causes pour restaurer le "bon état" ?

### **La prise en compte d'autres indicateurs ?**

Nous avons ici travaillé au niveau de l'espèce car nous pensons qu'il est important de ne pas perdre le contact de cette information spécifique. De plus, une étude très localisée se prêtait bien à cela. Toutefois, les approches de modélisation des réponses biologiques utilisant des regroupements d'espèces fondés sur leurs traits biologiques communs semblent tout de même intéressantes afin de gagner en généralité. En effet, il semble que les assemblages de poissons, étudiés à ce niveau d'information, montrent une certaine stabilité géographique, à conditions environ-

nementales comparables. Les modèles développés permettent aisément de travailler sur des données tenant compte de cette information.

Malgré tout, il apparaît clairement qu'un des problèmes issu de la prise en compte unique des traits biologiques est que l'on perd le lien avec l'espèce en elle-même et que la réponse n'est pas sensible au remplacement d'une espèce par une autre ayant le même trait biologique. Là encore, une approche hiérarchique serait à même d'associer les deux approches. En hiérarchisant en tenant compte d'une part de cette stabilité aux grandes échelles géographiques et d'autre part en tenant compte de la ressemblance des espèces partageant les mêmes traits biologiques, on peut introduire des réponses similaires d'assemblages, tout en prédisant localement (au niveau du cours d'eau) des abondances (absolues ou relatives) d'espèces partiellement substituables. Il est donc possible de coupler les deux approches, permettant de replacer l'espèce au coeur du concept de communauté (Olden et al., 2006), et d'intégrer l'information provenant des différentes échelles spatiales.

## 5.3 Vers un outil d'aide à la décision bayésien

### 5.3.1 Une situation fortement entachée d'incertitudes

Les incertitudes opérationnelles liées à la DCE s'expriment à plusieurs niveaux. L'étape préalable à la réalisation des objectifs de la DCE concerne l'évaluation initiale de l'état des masses d'eau. Des études ont montré que les résultats des classifications sont influencés par plusieurs sources d'erreur, comme par exemple les variations de protocole d'échantillonnage ou de laboratoires d'analyse, les variations saisonnières naturelles ou encore les variations liées à l'échelle géographique (p.ex Carsensen, 2007), et la communauté scientifique impliquée insiste désormais sur le besoin de toujours associer cette évaluation à une probabilité (Hering et al., 2010), ainsi que sur le besoin de nouveaux jeux de données susceptibles de permettre l'étude des différentes sources d'incertitude, qu'elles soient temporelles (heure de la journée, saison, ou influence du climat) ou spatiales (identifier les stations d'échantillonnage représentatives) (Hering et al., 2010). Ces variations doivent d'autant mieux être maîtrisées que toutes les masses d'eau ne pourront être échantillonnées, et qu'il faudra avoir recours à des outils d'extrapolation de l'état écologique. Ces outils devront permettre de prédire un état écologique via l'observation d'un certain nombre d'éléments de l'environnement non biotique. La théorie des plans d'expérience dans un contexte bayésien (Cha-

loner and Verdinelli, 1995) peut ici aider à choisir de façon optimale vis à vis d'objectifs à définir, via l'introduction de fonctions d'utilité, les endroits et les espèces pour lesquels l'information collectée se révélera la plus pertinente.

Dans un second temps, l'étape de restauration de l'état écologique des cours d'eau pourra avoir lieu. Dans un système complexe et incertain, il s'agit de déterminer quels grands leviers sont à la disposition des gestionnaires du milieu pour maintenir ou restaurer le bon état écologique, vis-à-vis de chacun des compartiments biotiques. L'estimation des effets attendus des mesures de restauration sur les compartiments biotiques est au coeur de ce programme, car il s'agit de déterminer quelles mesures sont les plus à même d'atteindre les objectifs fixés. Si la liaison entre données écologiques et mesures de restauration se fait de façon assez directe lorsque il existe une relation quantifiée déjà maîtrisée (du style dose-réponse, dans le cas des pollution organique, par exemple), il est en revanche nettement plus difficile de déterminer cette réponse dans le cas d'impacts multiples et peu connus ou mal maîtrisés (Hering et al., 2010). Cette tâche ne s'annonce pas simple, car une action ayant un effet positif sur l'état d'un des compartiments pourrait impacter négativement l'un ou les autres compartiments. L'effet de chaque action de restauration possible devra donc être jugée dans l'ensemble, en prédisant ses possibles conséquences sur l'ensemble de l'écosystème. C'est pourquoi il faut anticiper sur le développement de modèles multiéchelles, et multicritères, hiérarchiques et prédisant les effets de chaque impact et de chaque action de restauration sur l'ensemble des compartiments. Il faut également adjoindre à ces modèles prenant en charge à la fois les différentes sources de données et les différentes sources d'incertitude un second modèle synthétisant les conséquences socio-économiques et écologiques des prises de décision (Keeney and Raiffa, 1993).

Dans ce cadre, la modélisation hiérarchique devrait donner la pleine mesure de son potentiel (Clark, 2005).

### **5.3.2 Faciliter la discussion autour des processus : les structures graphiques**

Dans ce contexte complexe, il est important de disposer d'outils permettant d'avoir une vue d'ensemble des relations pressions-impacts introduites dans les modèles. Les structures graphiques, ou réseaux, où ces relations sont présentées sous forme de boîtes pour les variables et de flèches indiquant les relations de causalités, sont particulièrement adap-

tées pour cela. Ils rendent explicites des hypothèses causales entre variables qui pourraient rester dans l'ombre dans d'autres approches. Cette représentation explicite permet aux biologistes, aux hydrologues, aux chimistes et aux écotoxicologues de tous discerner comment leurs recherches propres peuvent s'intégrer dans une synthèse quantitative (Borsuk et al., 2002). J'ajouterais que cela permet également à ces mêmes chercheurs de dialoguer plus facilement avec les modélisateurs statisticiens.

Or, la statistique bayésienne s'accommode parfaitement de ces structures graphiques, via les relations conditionnelles qui sous-tendent ce paradigme. D'ailleurs, bon nombre de logiciels permettent d'écrire un modèle directement sous la forme de DAG (par ex. WinBUGS, Bayesia-Lab, etc...). En ce début de 21ème siècle survient un profond changement dans la perception de ce que l'on peut apprendre des modèles et des données. Le développement des outils à disposition et la prise de conscience de la complexité des systèmes sont deux tendances complémentaires qui devraient permettre de passer des études cloisonnées de chacune des différentes composantes d'un système à une meilleure compréhension, plus intégrée, des processus. Les modèles multiniveaux s'appuyant sur des approches graphiques doivent en être les plateformes (Clark and Gelfand, 2006).

### **5.3.3 Notre démarche s'inscrit dans la mise en place d'un réseau de probabilité bayésien**

Dans l'ensemble de ce manuscrit, nous nous sommes placés dans un cadre de modélisation statistique, selon une perspective bayésienne. Nous avons vu l'on s'appuie en permanence sur des modèles graphiques et la notion de réseau. Dans cette recherche doctorale, les noeuds qui le composent représentent des variables définies sur un espace continu et nous avons dû mettre en place des techniques avancées de calcul bayésien pour conduire l'inférence (Boreux et al., 2009). Si l'on avait pu travailler sur des variables uniquement discrètes, les calculs auraient pu être menés de façon plus simplifiée, en se cantonnant à l'écriture de tables de probabilités, bien évidemment conditionnelles. Par souci de réalisme, nous avons refusé l'approximation consistant à discrétiser les grandeurs décrivant les inconnues des modèles traités dans cette recherche. Cependant, une fois l'inférence effectuée grâce aux performances du logiciel WinBUGS (et seulement alors), nous pourrions discrétiser les lois a posteriori obtenues, à des fins de démonstration et de communication avec les gestionnaires des milieux aquatiques.

Un réseau bayésien à variables discrètes écrit sous cette forme porte le nom de Réseaux de Probabilités Bayésiens. Historiquement, ils proviennent de la volonté d'introduire de l'incertitude dans les systèmes experts (Naïm et al., 2004). Ils ont trouvé leurs premières applications dans les domaines industriels, notamment dans le cadre de diagnostic de pannes, industrielles (par exemple chez EDF) ou logiciel (par exemple Microsoft). Mais dernièrement, leur potentiel dans les problématiques de gestion environnementale a été souligné (Stewart-Koster et al., 2010; Borsuk, 2004; Borsuk et al., 2006); et si nous préférons réaliser l'inférence bayésienne sur variables d'origine continues, nous ne contestons bien évidemment pas son emploi en mode prédictif.

Il existe donc un intérêt réel pour cette approche, très conviviale et interactive, comme outil d'aide à la décision dans le cadre de la gestion et de la restauration des cours d'eau, notamment pour réaliser des tests de scénarios plus ou moins complexes sur un système qui lui l'est particulièrement. Cependant, il faudra sûrement l'envisager comme une ultime étape, après l'analyse statistique des relations pressions-impacts. En effet, une fois ces relations mieux maîtrisées, il sera sans doute possible de discrétiser à bon escient les variables, en créant les modalités ad-hoc pour chacune des variables. L'avantage des méthodes par réseaux fondés sur des relations conditionnelles sera alors au coeur de l'écriture de ces modèles complexes, puisque chaque relation pourra être étudiée séparément avant d'être ajoutée comme nouvelle pièce au modèle de décision.



## Références bibliographiques



# Bibliographie

- Aarts, B. G. W. and P. H. Nienhuis (2003). Fish zonation and guilds as the basis for assessment of ecological integrity of large rivers. *Hydrobiologia* 500(1-3), 157–178.
- Allen, T. and T. Starr (1982). *Hierarchy : perspectives for ecological complexity*, Volume III. Chicago : University of Chicago Press.
- Baglinière, J. and A. Champigneulle (1982). Densité des populations de truite commune (*salmo trutta* l.) et de juvéniles de saumon atlantique (*salmo salar* l.) sur le cours principal du scorff (bretagne) : preferendums physiques et variations annuelles (1976-1980). *Acta Oecologica* 3(3), 241–256.
- Baglinière, J., L. Denais, E. Rivot, J. Porcher, E. Prévost, F. Marchand, and V. Vauclin (2004). Length and age structure modifications of the atlantic salmon (*salmo salar*) populations of brittany and lower normandy from 1972 to 2002. Technical report.
- Baglinière, J., F. Marchand, and V. Vauclin (2005). Interannual changes in recruitment of the atlantic salmon (*salmo salar*) population in the river oir (lower normandy, france) : relationships with spawners and in-stream habitat. *Ices Journal of Marine Science* 62(4), 695–707.
- Balon, E. (1984). Reflections on some decisive events in the early life of fishes. *Transactions of the American Fisheries Society* 113, 178–185.
- Barry, S. and A. Welsh (2002). Generalized additive modelling and zero inflated count data. *Ecological Modelling* 157(2-3), 179–188.
- Barus, V. and M. Prokes (1993). Fecundity of the bleak (*alburnus-alburnus*) in the vestonice reservoir. *Folia Zoologica* 42(3), 281–288. Life history.

- Bayley, P. and H. Li (1992). Riverine fishes. In P. Calow and G. Petts (Eds.), *The Rivers Handbook*, Volume 1, pp. 251–281. Oxford : Blackwell Scientific Publications.
- Beltrami, E. (1997). *Mathematics for dynamic modeling*. Elsevier.
- Biggeri, A., M. Bonannini, D. Catelan, F. Divino, E. Dreassi, and C. Lagazio (2005, DEC). Bayesian ecological regression with latent factors : Atmospheric pollutants emissions and mortality for lung cancer. *Environmental and Ecological Statistics* 12(4), 397–409.
- Bishop, J. A. and W. L. Myers (2005). Associations between avian functional guild response and regional landscape properties for conservation planning. *Ecological Indicators* 5(1), 33–48.
- Blandin, P. (1986). Bioindicateurs et diagnostic des systèmes écologiques. Technical Report 82160, Station Biologique de Foljuif.
- Bolland, J., I. Cowx, and M. Lucas (2008). Factors affecting the dispersal of coarse fish. Technical report, Environmental Agency, UK.
- Boreux, J., E. Parent, and J. Bernier (2009). *Pratique du calcul bayésien*.
- Borsuk, M. (2004). Predictive assessment of fish health and fish kills in the neuse river estuary using elicited expert judgment. *Human and Ecological Risk Assessment* 10(2), 389–408.
- Borsuk, M., P. Burkhardt-Holm, and P. Reichert (2002, 24-27 June 2002). A bayesian network for investigating the decline in fish catch in switzerland. In t. I. E. M. Society and Software (Eds.), *Integrated Assessment and Decision Support*, Lugano, Switzerland, pp. 108–113.
- Borsuk, M. E., P. Reichert, A. Peter, E. Schager, and P. Burkhardt-Holm (2006). Assessing the decline of brown trout (*salmo trutta*) in swiss rivers using a bayesian probability network. *Ecological Modelling* 192(1-2), 224–244.
- Brooks, S. P. and A. Gelman (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7(4), 434–455.
- Caissie, D. (2006). The thermal regime of rivers : a review. *Freshwater Biology* 51(8), 1389–1406.

- 
- Cameron, A. and P. Trivedi (1998). *Regression analysis of count data*. Cambridge University Press.
- Carstensen, J. (2007). Statistical principles for ecological status classification of water framework directive monitoring data. *Marine Pollution Bulletin* 55, 3–15.
- Carter, K. and J. Reader (2000). Patterns of drift and power station entrainment of 0+ fish in the river trent, england. *Fisheries Management and Ecology* 7, 447–464.
- Cattanéo, F., P. Breil, and G. Carrel (1999a). Stream flow and fish assemblages inter-annual variability in the lower rhône river, france. In *4ème congrès international de limnologie-océanographie*, Bordeaux, 7-10 septembre 1999.
- Cattanéo, F., P. Breil, and G. Carrel (1999b). Variabilité inter-annuelles des assemblages piscicoles en relation avec le régime hydrologique : le bas-rhône à montélimar. In *4<sup>th</sup> Congrès International Limnologie - Océanographie. Variabilités temporelles au sein des hydrosystèmes : résumés des communications*, Volume 27, 7-10 septembre 1999, Université de Bordeaux 1 (Talence, Gironde, France). Bulletin de la Société Linnéenne de Bordeaux.
- Cattanéo, F., G. Carrel, N. Lamouroux, and P. Breil (2001). Relationship between hydrology and cyprinid reproductive success in the lower rhône at montélimar, france. *Archiv für Hydrobiologie* 151(3), 427–450.
- Chaloner, K. and I. Verdinelli (1995). Bayesian experimental design : a review. *Statistical Science* 10, 273–304.
- Clark, J. (2005). Why environmental scientists are becoming bayesians. *Ecology Letters* 8, 2–14.
- Clark, J. and A. Gelfand (2006). A future for models and data in environmental science. *TRENDS in Ecology and Evolution* 21(7), 375–380.
- Clifford, P., S. Richardson, and D. Hemon (1989). Assessing the significance of the correlation between two spatial processes. *Biometrics* 45(1), 123–134.
- Connell, J. (1978). Diversity in tropical rain forests and coral reefs. *Science* 199(4335), 1302–1310.

- Copp, G., J. Olivier, M. Penaz, and A. Roux (1991). Juvenile fishes as functional describers of fluvial ecosystem dynamics : applications on the river rhone, france. *Regulated Rivers : Research and Management* 6(2), 135–145.
- Coquillard, C. and D. Hill (1997). *Modélisation et simulation d'écosystèmes : des modèles déterministes aux simulations à évènements discrets*. Masson.
- Coutant, C. (1987). Thermal preference : when does an asset become a liability? *Environmental Biology of Fishes* 18(3), 161–172.
- Cramer, R., J. Bunce, D. Patterson, and I. Frank (1988). Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional qsr studies. *Quantitative Structure-Activity Relationships* 7, 18–25.
- Cressie, N., C. A. Calder, J. S. Clark, J. M. V. Hoef, and C. K. Wikle (2009). Accounting for uncertainty in ecological analysis : the strengths and limitations of hierarchical statistical modeling. *Ecological Applications* 19(3), 553–570.
- Daufresne, M. and P. Boet (2007, DEC). Climate change impacts on structure and diversity of fish communities in rivers. *Global Change Biology* 13(12), 2467–2478.
- Daufresne, M., M. Roger, H. Capra, and N. Lamouroux (2003). Long-term changes within the invertebrate and fish communities of the upper rhône river : effects of climatic factors. *Global Change Biology* 10, 124–140.
- Economou, A., C. Daoulas, and T. Psarras (1991). Growth and morphological development of chub, *leuciscus cephalus* (l.), during the first year of life. *Journal of Fish Biology* 39(3), 393–408.
- Feria, T. P. and A. T. Peterson (2002). Prediction of bird community composition based on point-occurrence data and inferential algorithms : a valuable tool in biodiversity assessments. *Diversity And Distributions* 8(2), 49–56.
- Fitzmaurice, G. (1997). Model selection with overdispersed data. *Statistician* 46(1), 81–91.

- 
- Fortin, V., L. Perreault, and J. Salas (2004, AUG 20). Retrospective analysis and forecasting of streamflows using a shifting level model. *Journal of Hydrology* 296(1-4), 135–163.
- Fruget, J., P. Bady, J. Olivier, G. Carrel, Y. Souchon, B. Villeneuve, M. Daufresne, and H. Capra (2006). Etude thermique globale du rhône phase iii, lot 2 : Etude à l'échelle du rhône des compartiments biologiques. Technical report, Cemagref.
- Genot, V. (1995). Evha : Un logiciel d'évaluation de l'habitat du poisson sous windows. *Bulletin Français de la Pêche et de la Pisciculture* 337-338-339, 303–308.
- Genot, V., Y. Souchon, and P. Roger (1996). Impact de l'élévation artificielle de température induite par le fonctionnement du centre nucléaire de production électrique du buges (fleuve rhône) sur les communautés de poissons. *Hydroécologie Appliquée* 8(1-2), 1–33.
- Gouraud, V., J. Baglinière, P. Baran, C. Sabaton, P. Lim, and D. Ombredane (2001). Factors regulating brown trout populations in two french rivers : Application of a dynamic population model. *Regulated Rivers : Research & Management* 17(4-5), 557–569.
- Gouraud, V., H. Capra, C. Sabaton, L. Tissot, P. Lim, F. Vandewalle, G. Fahrnera, and Y. Souchon (2008). Long-term simulations of the dynamics of trout populations on river reaches bypassed by hydroelectric installations-analysis of the impact of different hydrological scenarios. *River Research and Applications* 24(9), 1185–1205.
- Gouraud, V., C. Sabaton, and H. Capra (2004). Role of habitat variability in trout populations dynamics : Application of a dynamic population model to three french rivers. *Hydroécologie Appliquée* 14(1), 221–244.
- Gozlan, R. E., S. Mastroiello, G. H. Copp, and S. Lek (1999). Predicting the structure and diversity of young-of-the-year fish assemblages in large rivers. *Freshwater Biology* 41(4), 809–820.
- Gray, B. R. (2005). Selecting a distributional assumption for modeling relative densities of benthic macroinvertebrates. *Ecological Modelling* 185(1), 1–12.

- Greenberg, L., P. Svendsen, and A. Harby (1996). Availability of microhabitats and their use by brown trout (*salmo trutta*) and grayling (*thymallus thymallus*) in the river vojman, sweden. *Regulated Rivers : Research and Management* 12, 287–303.
- Gulland, J. A. (1982). Why do fish numbers vary. *Journal of Theoretical Biology* 97(1), 69–75.
- Haining, R., J. Law, R. Maheswaran, T. Pearson, and P. Brindley (2007, AUG). Bayesian modelling of environmental risk : example using a small area ecological study of coronary heart disease mortality in relation to modelled outdoor nitrogen oxide levels. *Stochastic Environmental Research and Risk Assessment* 21(5), 501–509.
- Hawkins, C. P. and J. A. MacMahon (1989). Guilds : The multiple meanings of a concept. *Annual Review of Entomology* 34(1), 423–451.
- Hering, D., A. Borja, J. Carstensen, L. Carvalho, M. Elliott, C. K. Feld, A.-S. Heiskanen, R. K. Johnson, J. Moe, D. Pont, A. L. Solheim, and W. van de Bund (2010). The european water framework directive at the age of 10 : A critical review of the achievements with recommendations for the future. *Science of The Total Environment In Press, Corrected Proof*, –.
- Houde, E. D. (1997). Patterns and trends in larval-stage growth and mortality of teleost fish. *Journal of Fish Biology* 51, 52–83.
- Huet, M. (1949). Aperçu des relations entre la pente et les populations piscicoles des eaux courantes. *Zeitschrift für Hydrologie* 11(3-4), 332–351.
- Humphries, P., P. Brown, J. Douglas, A. Pickworth, R. Strongman, K. Hall, and L. Serafini (2008). Flow-related patterns in abundance and composition of the fish fauna of a degraded australian lowland river. *Freshwater Biology* 53(4), 789–813.
- Hurny, A. and J. Wallace (2000). Life history and production of stream insects. *Annual Review of Entomology* 45, 83–110.
- Huston, M. (1979). A general hypothesis of species diversity. *American Naturalist* 113(1), 81–101.

- Junk, W., P. Bayley, and R. Sparks (1989). The flood pulse concept in river-floodplain systems. In D. Dodge (Ed.), *Proceedings of the international large river symposium*, Honey Harbour, Ontario, Canada, pp. 110–127. Canadian Special Publication Fisheries Aquatic Sciences.
- Karr, J. (1991). Biological integrity : a long-neglected aspect of water resource management. *Ecological Applications* 1(1), 66–84.
- Keeney, R. and H. Raiffa (1993). *Decisions with Multiple Objectives : Preferences and Value Tradeoffs*.
- Kelly, M. and J. Gore (2008). Florida river flow patterns and the atlantic multidecadal oscillation. *River Research and Applications* 24, 598–616.
- Konrad, C., A. Brasher, and J. May (2008). Assessing streamflow characteristics as limiting factors on benthic invertebrate assemblages in streams across the western united states. *Freshwater Biology* 53(10), 1983–1998.
- Lamouroux, N., H. Capra, M. Pouilly, and Y. Souchon (1999). Fish habitat preferences in large streams of southern france. *Freshwater Biology* 42(4), 673–687.
- Lamouroux, N., J. Olivier, H. Capra, M. Zylberblat, A. Chandesris, and P. Roger (2006). Fish community changes after minimum flow increase : testing quantitative predictions in the rhône river at pierre-bénite, france. *Freshwater Biology* 51(9), 1730–1743.
- Lamouroux, N., J. M. Olivier, H. Persat, M. Pouilly, Y. Souchon, and B. Statzner (1999). Predicting community characteristics from habitat conditions : fluvial fish and hydraulics. *Freshwater Biology* 42(2), 275–299.
- Libosvsky, J. (1979). Gonad weight and egg numbers in chub, *leuciscus cephalus*, from the rokytna stream. *Folia Zoologica* 28(1), 35–42. Life history.
- Lima-Junior, S. E., I. B. Cardone, and R. Goitein (2006). Fish assemblage structure and aquatic pollution in a brazilian stream : some limitations of diversity indices and models for environmental impact studies. *Ecology Of Freshwater Fish* 15(3), 284–290.
- Lowe-McConnell, R. (1987). *Ecological studies in tropical fish communities*. Cambridge Tropical Biology Series. Cambridge University Press.

- Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter (2000). Winbugs - a bayesian modelling framework : Concepts, structure, and extensibility. *Statistics and Computing* 10(4), 325–337.
- Mackay, I. and K. Mann (1969). Fecundity of two cyprinid fishes in the river thames, reading, england. *Journal of the Fisheries Research Board of Canada* 26(11), 2795–2805.
- Magnuson, J., L. Crowder, and P. Medvick (1979). Temperature as an ecological resource. *American Zoologist* 19, 331–343.
- Mann, R. (1980). The growth and reproductive strategy of the gudgeon, *gobio gobio* (l.), in two hard-water rivers in southern england. *Journal of Fish Biology* 17, 163–176.
- Marin, J.-M. and C. Robert (2007). *Bayesian core : a practical approach to computational Bayesian statistics*. Springer-Verlag, New York.
- McCullagh, P. and J. Nelder (1989). *Generalized Linear Models. 2nd edition*. Chapman & Hall, London.
- McIntosh, R. (1987). Pluralism in ecology. *Annual Review of Ecology and Systematics* 18, 321–341.
- Mills, C. and R. Mann (1985). Environmentally-induced fluctuations in year-class strength and their implications for management. *Journal of Fish Biology* 27(Suppl. A), 209–226.
- Naïm, P., P.-H. Wuillemin, P. Leray, O. Pourret, and A. Becker (2004). *Réseaux bayésiens*. Eyrolles.
- Nunn, A., I. Cowx, P. Frear, and J. Harvey (2003). Is water temperature an adequate predictor of recruitment success in cyprinid fish populations in lowland rivers? *Freshwater Biology* 48, 579–588.
- Nunn, A. D., J. P. Harvey, J. R. Britton, P. A. Frear, and I. G. Cowx (2007). Fish, climate and the gulf stream : the influence of abiotic factors on the recruitment success of cyprinid fishes in lowland rivers. *Freshwater Biology* 52(8), 1576–1586.
- Oberdorff, T., D. Pont, B. Hugueny, and J. Porcher (2002). Development and validation of a fish-based index for the assessment of "river health" in france. *Freshwater Biology* 47, 1720–1734.

- 
- Olden, J. D., M. K. Joy, and R. G. Death (2006). Rediscovering the species in community-wide predictive modeling. *Ecological Applications* 16(4), 1449–1460.
- Olden, J. D. and R. J. Naiman (2010, JAN). Incorporating thermal regimes into environmental flows assessments : modifying dam operations to restore freshwater ecosystem integrity. *Freshwater Biology* 55(1), 86–107.
- Parmesan, C. and G. Yohe (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature* 421(6918), 37–42.
- Philippart, J. (1981). Ecologie d'une population de vandoise, leuciscus leuciscus (l.) dans la rivière ourthe (bassin de la meuse, belgique). *Annales de Limnologie* 17(1), 41–62.
- Poff, N. and J. Allan (1995). Functional organization of stream fish assemblages in relation to hydrological variability. *Ecology* 76(2), 606–627.
- Poff, N., J. Allan, M. Bain, J. Karr, K. Prestegard, B. Richter, R. Sparks, and J. Stromberg (1997). The natural flow regime. a paradigm for river conservation and restoration. *BioScience* 47(11), 769–784.
- Poff, N. and J. Ward (1989). Implications of streamflow variability and predictability for lotic community structure : a regional analysis of streamflow patterns. *Canadian Journal of Fisheries and Aquatic Sciences* 46, 1805–1818.
- Poff, N. and J. Zimmerman (2010). Ecological responses to altered flow regimes : a literature review to inform the science and management of environmental flows. *Freshwater Biology* 55, 194–205.
- 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(2), 391–409.
- Pont, D., B. Hugueny, B. Beier, D. Goffaux, A. Melcher, R. Noble, C. Rogers, N. Roset, and S. Schmutz (2006). Assessing river biotic condition at a continental scale : a european approach using functional metrics and fish assemblages. *Journal of Applied Ecology* 73, 70–80.

- Poole, G. and C. Berman (2001). An ecological perspective on in-stream temperature : Natural heat dynamics and mechanisms of human-caused thermal degradation. *Environmental Management* 27(6), 787–802.
- Potts, J. M. and J. Elith (2006). Comparing species abundance models. *Ecological Modelling* 199(2), 153 – 163. Predicting Species Distributions - Results from a Second Workshop on Advances in Predictive Species Distribution Models, held in Riederalp, Switzerland, 2004.
- Pouilly, M., S. Valentin, H. Capra, V. Ginot, and Y. Souchon (1995). Méthode des microhabitats : principes et protocoles d'application. *Bulletin Français de la Pêche et de la Pisciculture* 336, 41–54.
- Power, M., R. Stout, C. Cushing, P. Harper, F. Hauer, W. Matthews, P. Moyle, B. Statzner, and I. Wais De Badgen (1988). Biotic and abiotic controls in river and stream communities. *Journal of the North American Benthological Society* 7(4), 456–479.
- Raftery, A. (1996). Approximate bayes factors and accounting for model uncertainty in generalised linear models. *Biometrika* 83, 251–266.
- Raftery, A., D. Madigan, and J. Hoeting (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* 92, 179–191.
- Reeves, G., L. Benda, K. Burnett, P. Bisson, and J. Sedell (1995). A disturbance-based ecosystem approach to maintaining and restoring freshwater habitats of evolutionarily significant units of anadromous salmonids in the pacific northwest. *American Fisheries Society Symposium* 17, 334–349.
- Reichard, M., P. Jurajda, and M. Ondrakova (2002). The effect of light intensity on the drift of young-of-the-year cyprinid fishes. *Journal of Fish Biology* 61(4), 1063–1066.
- Resh, V., A. Brown, A. Covich, M. Gurtz, H. Li, G. Minshall, S. Reice, A. Sheldon, J. Wallace, and R. Wissmar (1988). The role of disturbance in stream ecology. *Journal of the North American Benthological Society* 7(4), 433–455.
- Richardson, S. (1990). A method for testing the significance of geographical correlations with application to industrial lung-cancer in france. *Statistics in Medicine* 9, 515–528.

- 
- Richter, B., J. Baumgartner, J. Powell, and D. Braun (1996). A method for assessing hydrologic alteration within ecosystems. *Conservation Biology* 10, 1163–1174.
- Root, R. (1967). The niche exploitation pattern of the blue-gray gnat-catcher. *Ecological Monographs* 37, 37–350.
- Root, T., J. Price, K. Hall, S. Schneider, C. Rosenzweig, and J. Pounds (2003). Fingerprints of global warming on wild animals and plants. *Nature* 421, 57–60.
- Royle, J. and R. Dorazio (2008). *Hierarchical Modeling and Inference in Ecology The Analysis of Data from Populations, Metapopulations and Communities*. Academic Press.
- Saporta, G. (1990). *Probabilités, Analyse Des Données et Statistique*. Editions TECHNIP.
- Schiemer, F., H. Keckeis, and L. Flore (2001). Ecotones and hydrology : Key conditions for fish in large rivers. *Ecohydrology and Hydrobiology* 1(1-2), 49–55.
- Schiemer, F., H. Keckeis, and E. Kamler (2003). The early life history stages of riverine fish : ecophysiological and environmental bottlenecks. *Comparative Biochemistry and Physiology A-Molecular and Integrative Physiology* 133(3), 439–449.
- Schiemer, F., H. Keckeis, H. Nemeschkal, E. Schludermann, G. Winkler, and I. Zweimüller (2004). Ontogenetic patterns in thermal adaptation of fish vs. long-term temperature trends in large rivers.
- Schiemer, F., T. Spindler, H. Wintersberger, A. Schneider, and A. Chovanec (1991). Fish fry associations : important indicators for the ecological status of large rivers. *Verhandlungen der Internationale Vereinigung für Theoretische und Angewandte Limnologie* 24, 2497–2500.
- Sebastia, M. T. (2007). Plant guilds drive biomass response to global warming and water availability in subalpine grassland. *Journal Of Applied Ecology* 44(1), 158–167.
- Smitz, J., E. Everbecq, J. Deliege, J. Descy, R. Wollast, and J. Vanderborght (1997). Pégase, une méthodologie et un outil de simulation prévisionnelle pour la gestion de la qualité des esaux de surface. *La tribune de l'eau* 50, 3–72.

- Solow, A. (1998). On the fitting of population dynamic model in the presence of observation error. *Ecology* 79, 1463–1466.
- Souchon, Y. (1994). Development of fish habitat modelization in french streams in recent years. *Bulletin Francais De La Peche Et De La Pisciculture* (332), 57–71.
- Souchon, Y., C. Leveque, and P. Gaudin (1995). Le poisson et son habitat : introduction du symposium. *Bulletin Français de la Pêche et de la Pisciculture* 337-338-339, 3–8.
- Souchon, Y., F. Trocherie, E. Fragnoud, and C. Lacombe (1989). Les modèles numériques des microhabitats des poissons : application et nouveaux développements. *Revue des Sciences de l'Eau* 2(4), 807–830.
- Souchon, Y., F. Trocherie, J. Malavoi, J. Wasson, and E. Fragnoud (1988). Improvements of biological responses of instream flow habitat models. In *4th Regulated Streams Symposium*, Loughborough (England), pp. 31. Petts G.
- Souchon, Y. and S. Valentin (1991). Instream flow incremental methodology (ifim) : study cases, technical aspects, research in 3 countries : Usa, uk, france. report of a scientific meeting held at the laboratoire d'hydroécologie quantitative, cemagref bea/lhq. lyon france (13-17 may 1991). Technical report, CEMAGREF BEA/LHQ.
- Southwood, T. (1977). Habitat, the templet for ecological strategies? *Journal of Animal Ecology* 46(16), 337–365.
- Sparks, R. (1995). Need for ecosystem management of large rivers and their floodplains. *BioScience* 45(3), 168–182.
- Statzner, B. and L. Bêche (2010). Can biological invertebrate traits resolve effects of multiple stressors on running water ecosystems? *Freshwater Biology* 55, 80–119.
- Stewart-Koster, B., S. Bunn, S. Mackay, N. Poff, and R. Naiman (2010). The use of bayesian networks to guide investments in flow and catchment restoration for impaired river ecosystems. *Freshwater Biology* 55, 243–260.

- 
- Tejerina-Garro, F., M. Maldonado, C. Ibanez, D. Pont, N. Roset, and T. Oberdorff (2005). Effects of natural and anthropogenic environmental changes on riverine fish assemblages : a framework for ecological assessment of rivers. *Brazilian Archives of Biology and Technology* 48(1), 91–108.
- Teletchea, F., A. Fostier, E. Kamler, J. Gardeur, P. Le Bail, B. Jalabert, and P. Fontaine (2008). Comparative analysis of reproductive traits in 65 freshwater fish species : application to the domestication of new fish species. *Reviews in Fish Biology and Fisheries*, 1–28.
- Tenenhaus, M., J. Gauchi, and C. Ménardo (1995). Régression pls et applications. *Revue de Statistique Appliquée XLIII*(1), 7–63.
- Thogmartin, W. E., J. R. Sauer, and M. G. Knutson (2004). A hierarchical spatial model of avian abundance with application to cerulean warblers. *Ecological Applications* 14(6), 1766–1779.
- Tissot, L. and Y. Souchon (2008). Etat des connaissances sur les préférences thermiques des principales espèces de poissons des eaux continentales françaises. Technical Report H-P76-2008-01399-FR, EDF Chatou.
- Tomassone, R., E. Lesquoy, and C. Millier (1983). *La régression : nouveaux regards sur une ancienne méthode statistique*. Masson, Paris.
- Trebitz, A., B. Hill, and F. McCormick (2003). Sensitivity of indices of biotic integrity to simulated fish assemblage changes. *Environmental Management* 32(4), 499–515.
- Tsonaka, R. and I. Moustaki (2007). Parameter constraints in generalized linear latent variable models. *Computational Statistics and Data Analysis* 51, 4164–4177.
- Vannote, R., G. Minshall, K. Cummins, J. Sedell, and C. Cushing (1980). The river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences* 37, 130–137.
- Vannote, R. and B. Sweeney (1980). Geographic analysis of thermal equilibria : a conceptual model for evaluating the effect of natural and modified thermal regimes on aquatic insect communities. *American Naturalist* 115, 667–695.

- Vinson, M. and C. Hawkins (1998). Biodiversity of stream insects : variation at local, basin, and regional scales. *Annual Review of Entomology* 43, 271–293.
- Vriese, F., S. Semmekrot, and A. Raat (1994). Assessment of spawning and nursery areas in the river meuse. *Water Science and Technology* 29(3), 297–299.
- Walker, K., F. Sheldon, and J. Puckridge (1995). A perspective on dryland river ecosystems. *Regulated Rivers : Research and Management* 11, 85–104.
- Ward, J. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 58(301), 236–244.
- Wasson, J. (2001). Les questions de recherche posées par la directive cadre européenne sur l'eau : problématique pour les eaux de surface continentales. In C. Editions (Ed.), *Actes du séminaire : "Etat écologique des milieux aquatiques continentaux", Lyon, 20 & 21 mars 2001* (Cemagref Editions ed.), pp. 11–26. Lyon.
- Wasson, J. (2004, November 2004). Hydro-ecoregions - a natural framework for water management. Technical report, Cemagref.
- Wasson, J.-G., B. Villeneuve, A. Iital, J. Murray-Bligh, M. Dobiasova, H. Bacikova, S. Timm, H. Pella, N. Mengin, and A. Chandesris (2010). Large-scale relationships between basin and riparian land cover and the ecological status of european rivers. *Freshwater Biology* 55, 1465–1482.
- Webb, J., M. Stewardson, and W. Koster (2010). Detecting ecological responses to flow variation using bayesian hierarchical models. *Freshwater Biology* 55, 108–126.
- West, M. (2003). Bayesian factor regression models in the "large p, small n" paradigm. In J. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, and A. Smith (Eds.), *Bayesian Statistics 7*, pp. 733–742. 7th Valencia International Meeting on Bayesian Statistics, Valencia, SPAIN, JUN 02-06, 2002.
- WILLIS, A., N. DUNNETT, R. HUNT, and J. GRIME (1995, SEP). Does gulf-stream position affect vegetation dynamics in western-europe. *Oikos* 73(3), 408–410.

- 
- Wintle, B., M. McCarthy, C. Volinsky, and R. Kavanagh (2003). The use of bayesian model averaging to better represent uncertainty in ecological models. *Conservation Biology* 17(6), 1579–1590.
- Wold, S., M. Sjöström, and L. Eriksson (2001). Pls regression : a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* 58, 109–130.
- Wolman, M. and J. Miller (1960). Magnitude and frequency of forces in geomorphic processes. *The Journal of Geology* 68, 54–74.
- Wolter, C. (2007). Temperature influence on the fish assemblage structure in a large lowland river, the lower oder river, germany. *Ecology of Freshwater Fish* 16(4), 493–503.
- Woody, N. and S. Brown (2003, AUG 25). Partial least-squares modeling of continuous nodes in bayesian networks. *Analytica Chimica Acta* 490(1-2), 355–363. 8th International Conference on Chemometrics in Analytical Chemistry (CAC 2002), SEATTLE, WASHINGTON, SEP 22-26, 2002.
- Wu, J., K. Jones, H. Li, and O. Loucks (Eds.) (2006). *Scaling and uncertainty analysis in ecology, Methods and applications*. Springer.
- Zitek, A., S. Schmutz, and A. Ploner (2004). Fish drift in a danube sidearm-system : II. seasonal and diurnal patterns. *Journal of Fish Biology* 65(5), 1339–1357.
- Zitek, A., S. Schmutz, G. Unfer, and A. Ploner (2004). Fish drift in a danube sidearm-system : I. site-, inter-and intraspecific patterns. *Journal of Fish Biology* 65(5), 1319–1338.
- Zuur, A., E. Ieno, N. Walker, A. Saveliev, and G. Smith (2009). *Mixed effects models and extensions in Ecology with R*. Springer.



# Annexes



Annexe A

Publications scientifiques

Piffady, J., Souchon, Y., Capra, H. & Parent, E. (2010) Quantifying the effects of temperature and flow regime on the abundance of 0+ cyprinids in the upper River Rhone using Bayesian hierarchical modelling. *Freshwater Biology*, 55, 2359–2374  
doi:10.1111/j.1365-2427.2010.02453.x

## **Author Version**

**Quantifying the effects of temperature and flow regimes on 0+ cyprinid fish abundance in the upper River Rhone using Bayesian hierarchical modelling.**

JEREMY PIFFADY<sup>1</sup>, YVES SOUCHON<sup>1</sup>, HERVE CAPRA<sup>1</sup>, AND ERIC PARENT<sup>2</sup>

<sup>1</sup> Cemagref, UR MAEP, 3 bis Quai Chauveau CP 220, 69336 LYON cedex 09, France

<sup>2</sup> AgroParisTech-INRA UMR518, 16 rue Claude Bernard, 75231 PARIS Cedex 05, France

Correspondence: Jérémy Piffady, Cemagref, UR MAEP—LHQ, 3 bis Quai Chauveau CP 220, 69336 LYON cedex 09, France

E-mail: jeremy.piffady@cemagref.fr

Abbreviated title: Environmentally-induced 0+ cyprinid fish abundance

Keywords: environmental variations, 0+ fish, Bayesian hierarchical modelling, cyprinid fish, upper River Rhone.

## **SUMMARY**

1. Assuming that recruitment variation is one of the main sources of fish population and assemblage changes, it is necessary to understand how natural variations in the

environment influence 0+ fish abundance. Temperature regimes play an important role in enhancing both spawning activity and survival during early larval fish development. Flow regime variation, which is a powerful source of stream disturbance, is another factor to be taken into account.

2. Responses to these variables need to be assessed using long-term datasets, since standard statistical approaches fail to provide a causal structure or to quantify the different effects. We therefore used a 26-year dataset to evaluate the respective effects of seven derived independent variables describing the effects of temperature and flow regimes on the 0+ juvenile abundance of eight fish species in the River Rhone.
3. A clustering procedure using the Kendall tau rank correlation coefficient was implemented and identified three groups of fish according to their synchronic variations in juvenile abundance; i.e. varying with decreasing juvenile abundance, slightly increasing juvenile abundance, and increasing juvenile abundance. These clusters provided the basis for building hierarchical logPoisson Generalized Linear Models. The Bayesian paradigm gives a reliable framework for model selection, and the best model was determined using the Bayes Factor.
4. The posterior distribution of the regression parameters was coherent with what was expected based on knowledge of the biology of the different species. It indicates that temperature regime drives 0+ juvenile abundance but that flow regime also plays an important regulating role. The models thus detected evidence of the consequences of specific flow events such as larval drift and an increase of available habitat during higher flow discharges.
5. Our study illustrates the advantages of using a hierarchical modelling approach to quantify ecological effects by improving discrimination between the different sources of uncertainty, leading to better precision when estimating regression parameters.

## **Introduction:**

The theory of ecological filters has been well demonstrated for fish assemblages (Tonn *et al.*, 1990; Poff, 1997; Quist, Rahel & Hubert, 2005; Chessman, 2006). Regional filters such as temperature and flow (mostly latitude- and altitude-dependent) explain which species are to be found in local situations, where local constraints determine assemblage structure. These local factors can be natural, such as the interannual variations in temperature and flow regimes around general regional regimes, or anthropogenic, such as hydromorphological alterations and organic and chemical pollution (Bayley & Li, 1992).

Assessing the relative importance of these different constraints, determining the effects of anthropogenic pressures on freshwater bodies, and assessing the benefits to be expected from better management of these pressures are challenges that have shaped several national legislations (Clean Water Act USA, 1972; European Water Framework Directive 2000/60/EEC). There is consequently a need to conduct predictive studies, and statistical modelling is considered pivotal to achieving this objective since it can quantify the relationships between anthropogenic pressures and fish assemblages. In addition, climate change and its subsequent effects on flow and temperature regimes could exacerbate these problems (Firth & Fisher, 1992; Poff, 1992; Ficke, Myrick & Hansen, 2007; Schneider *et al.*, 2007).

Individual adult fish are generally more difficult to sample in large rivers than in small rivers, due to the deeper waters. Under these conditions, 0+ fish are probably easier to sample as they preferentially occupy the shallowest zones near banks. The quantitative and physiological assessment of 0+ fish has been shown to be a very useful way of assessing the ecological status of large rivers (Copp *et al.*, 1991; Schiemer *et al.*, 1991). Similarly, assessing the

number of young-of-the-year (YOY) is important as a meaningful indication of the reproductive success of different species, which in turn is a factor of primary importance for establishing self-sustaining populations (Schiemer, 2001). Furthermore, YOY sampling proves a more sensitive indicator of habitat suitability than sampling adult fish (Copp *et al.*, 1991).

Recruitment success depends on many factors: the condition of adult spawners, trophic resources and suitable temperature cycles, the availability of suitable substrate for spawning, the reproduction period during which the number of eggs deposited in the river is determined, the first few months of life that form an initial period of high mortality, and the first winter which is the second period of high mortality (Nunn *et al.*, 2003). In the River Rhone, even a small number of reproducing adult cyprinid individuals can lead to strong 0+ fish cohorts, as the reproductive strategy of these species consists of depositing a large number of eggs (Grenouillet *et al.*, 2001). The main driver of YOY numbers appears to be abiotic conditions (Larkin, 1978). As most fish species are ectothermic, water temperature over the year should be a major factor driving the ecology of species and the survival of YOY individuals, due to its preponderant influence on early growth rates (Nunn *et al.*, 2003). There is a large body of research into the effects of temperature on 0+ fish growth (Mills & Mann, 1985). The influence of flow regimes has also been highlighted (Schlosser, 1985), but it may play a more indirect role as it determines the quantity of available habitat (Souchon, 1994; Lamouroux *et al.*, 1999). However, major floods occurring during critical periods also play an important role. The larval period has been identified as the most vulnerable to floods, which can flush 0+ fish from cover and alter the substrate (Cattanéo *et al.*, 2001). The limited literature examining both thermal and flow effects on freshwater organisms suggests that both the direct and cumulative influences of these factors are critical (Olden & Naiman, 2010).

To a certain extent, the composition of adult fish assemblages results from different abiotic and biological phenomena, making their structures difficult to interpret. For riverine fish it has been shown that year-class-strength is largely determined during the first few months of life, and that a strong cohort of recruits from a given year remains a strong cohort of older fish in succeeding years (Mills & Mann, 1985). This indicates that abundance of YOYs is highly sensitive to the environment. The egg, larval and juvenile periods are the most critical stages in the life history of fish (Copp *et al.*, 1991), and it appears that recruitment variability is one of the main factors governing population variability (Houde, 1997). Many authors have underlined the need for a better understanding of recruitment variability mechanisms (Gulland, 1982; Houde, 1997), stressing the importance of environmental conditions. While Humphries *et al.* (2008) highlighted the need for datasets spanning over 20+ years, these remain rare. This scarcity of long-term datasets is mainly due to monitoring strategies and the structure of research funding (Bayley, 2002).

To quantify the direct effects of human activities on different bodies of water, it is first necessary to assess the effects of natural environmental variation on different species. The links between environmental variations and assemblage structures have been given increasing attention by researchers over the last decade, though mostly through general descriptive methods, such as Principal Components Analysis (PCA). These descriptive methods are based on correlative analysis of dependent and independent variables and are powerful tools for describing and understanding the qualitative structures of phenomena (Tonn *et al.*, 1990; Poff, 1997; Quist *et al.*, 2005; Chessman, 2006). However, they fail to give any formal quantification of the links between the target variables and the different independent variables. Regression models are better in this respect because they inherently assess the linkage strength between a predicted variable and its independent variables. Hierarchical modelling has recently improved our capacity to detect major links in data (Gelman *et al.*,

2004), and the advantages of this approach for analysing the effects of environmental trends on smelt abundance have already been demonstrated (Webb, Stewardson & Koster, 2010).

Furthermore, conditional reasoning appears to be a particularly interesting way of developing progressively complex models (Parent & Bernier, 2007).

In this paper we report the use of a Bayesian hierarchical model to perform a quantitative analysis of the effects of annual temperature and flow regimes on the juvenile abundance of eight River Rhone cyprinid fish species forming three functional groups. We then discuss the relative importance of interannual variations in temperature and flow regimes in determining these abundances.

## **Material and methods**

### *Study site*

The upper River Rhone has been transformed since the end of the 19<sup>th</sup> century by the construction of nine hydropower plants between Geneva (CH) and Lyon (FR), thus tightly regulating the flow regime of the river. The Bugey study site is a 13 km-long segment of the upper River Rhone (45.81N; 5.21E) in France, situated 12.5 km downstream from the Sault-Brénez hydropower plant (Fig. 1). The segment has a slope of 0.35%, a mean width of 120 m, a mean depth of 3 m and a mean annual flow of 500 m<sup>3</sup>.s<sup>-1</sup> (Daufresne *et al.*, 2003). A nuclear power plant (Bugey CNPE) was built on the segment in 1978. Despite the presence of these power plants, the hydromorphological characteristics of the study site have changed little over time (see the Cassini maps, 1815; <http://www.geoportail.fr/5069711/visu2D/afficher-en-2d.htm>). The study segment is a mainly morphologically unimpacted single stretch of channel, subjected to seasonal and daily variations of flow regime caused by the regulation of hydropower dams built upstream, and with generally good water quality (Agence de Bassin Rhône Méditerranée Corse, 1999, in Daufresne *et al.*, 2003).

### *Fish data*

The Cemagref has been performing fish surveys by boat electrofishing four times a year since 1979. Fish are sampled at eight fishing sites by continuous electrofishing from a boat drifting downstream along the bank for 30 minutes, so that each site is around 500 metres long and 2-4 metres wide. Electrofishing is performed using a Dream Electronis apparatus (type: Heron DC; 300-400 V; 2-3 A, Pessac, France). Fish are identified to species, measured for total length, and released. For each species, juvenile abundance is expressed as Catch Per Unit Effort (CPUE), meaning the number of fish captured in 30 minutes. The eight sites are distributed from upstream to downstream of the nuclear power plant and between the left and right banks (Fig. 1).

This bank electrofishing method tends preferentially to sample younger fish, including YOY and one-year juveniles. Copp (1989) identified the late summer period as being particularly favourable for sampling YOY. It is a relatively stable biological phase between the larval phase and the winter period, both of which are characterized by high mortality rates. Thus, the fish dataset was limited to a summer (July–August) and autumn (October–November) fishing sessions subset spanning 1980 to 2005 (data for 1988 and 1995 are missing, as are the data for the 1992 and 2004 autumn samples and the 1981 summer sample). In the first trial performed for this study, the juvenile-fish CPUEs were annually and seasonally pooled over the eight sites for each species, so that two counts (summer and autumn) were obtained for each species and each year. By doing this, the effect of warming by water from the nuclear power plant becomes negligible for this study, and thus no nuclear plant effect has been included.

Age class size limits were derived from histograms of size distributions for the different fish species (Persat & Chessel, 1989). Fish measuring strictly less than the size class limits for 0+ fish were considered as 0+ fish for the eight species, each representing more than 5% of total

abundance: bleak (*Alburnus alburnus* Linnaeus, 8 cm), barbel (*Barbus barbus* Linnaeus, 9 cm), chub (*Leuciscus cephalus* Linnaeus, 8 cm), roach (*Rutilus rutilus* Linnaeus, 6 cm), gudgeon (*Gobio gobio* Linnaeus, 10 cm), nase (*Chondrostoma nasus* Linnaeus, 8 cm), stream bleak (*Alburnoides bipunctatus* Bloch, 7 cm) and dace (*Leuciscus leuciscus* Linnaeus, 8.5 cm). Autumnal fish sampling chronologies are given in Fig. 2.

### *Clustering procedure*

Fish species were clustered by the degree of synchrony of their juveniles' CPUEs. The Kendall tau rank correlation coefficient gives a correlation index between the ranks of two samples of fish species. It is computed as:

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)} \quad (1)$$

where  $n$  is the number of recorded years for the two species. Considering all the possible pairs of years,  $n_c$  is the number of concordant pairs (e.g. between years 3 and 7, the two samples either both increased or both decreased),  $n_d$  is the number of discordant pairs (during the period of time considered, one sample has increased while the other decreased), and the denominator is the total number of possible pairs.

The Kendall tau expresses the synchrony of two trajectories, and varies between -1 (the changes in reproductive success of the two species are always opposed) and 1 (the changes in reproductive success of the two species are fully synchronised). The Kendall tau was used as distance in a cluster procedure with the Ward method, which minimizes intra-cluster variance (Ward, 1963), to cluster the eight fish species into groups of synchronic interannual CPUE increases or decreases. The clusters are described and discussed below.

### *Temperature and flow regime data*

Mean daily water temperatures and mean daily flow have been recorded since 1 January 1980. In line with Fruget *et al.* (2006), nine synthetic variables describing the annual natural environment of the station between spawning and fishing were derived from these two datasets. Given that temperature is the most important factor in fish ecology management, we set three temperature thresholds (12°C, 18°C, and maximum water temperature; Mills & Mann, 1985) to determine two different time-windows in which to describe the natural variations (i.e. 12-18°C and 18°C-maximum temperature).  $C12$ ,  $C18$  and  $Cmx$  are the number of days between 1 January and the first day to record a 12°C water temperature, 18°C water temperature and maximum water temperature, respectively. Timing of minimum and maximum temperatures are important descriptors of thermal regime (Olden & Naiman, 2010).  $S1$  and  $S2$  are the sums of the differences between daytime temperature and the mean daily interannual temperature over each of the two periods, respectively. They express the difference in cumulative degree days between the considered year and the mean year. As temperature is known to play an important role in the early growth of 0+ fish, these two variables indicate whether the period is rather warm or cold compared to the 26-year interannual mean. They are computed using the following formulas, where  $T_{ij}$  is the temperature of the day  $j$  for the  $i^{\text{th}}$  year:

$$S_1 = \sum_{j=C12}^{C18} \left( T_j - \frac{\sum_{i=1}^I T_{ji}}{I} \right) \quad (2)$$

and

$$S_2 = \sum_{j=C18}^{Cmx} \left( T_j - \frac{\sum_{i=1}^I T_{ji}}{I} \right) \quad (3)$$

The other four variables describe the annual flow regime, divided between the two periods described earlier:  $Qm1$  and  $Qm2$  are the mean daily flows over periods 1 and 2 respectively,

while  $Qmx1$  and  $Qmx2$  represent the maximum daily flow discharge observed during periods 1 and 2 respectively.

After checking for correlations between independent variables, the C18 and Qmx1 variables were retrieved from the dataset as they exhibited strong correlations with variable S1 (i.e. Pearson's  $r = 0.8$  or  $r = -0.8$ ; Oberdorff *et al.*, 2002). A summary of these independent environmental variables is given in Table 1. In the following, we consider seven standardized vectors of length  $n=26$ :  $X^1 = C12$ ;  $X^2 = Cmx$ ;  $X^3 = S1$ ;  $X^4 = S2$ ;  $X^5 = Qm1$ ;  $X^6 = Qm2$ ;  $X^7 = Qmx2$ . The corresponding regression parameters are then annotated as  $\beta_{C12}$ ,  $\beta_{Cmx}$ ,  $\beta_{S1}$ ,  $\beta_{S2}$ ,  $\beta_{Qm1}$ ,  $\beta_{Qm2}$ ,  $\beta_{Qmx2}$ .

### *Hierarchical modelling and Bayesian inference*

The species groups previously obtained by the clustering method were used to write hierarchical models of Poisson regression (one model for each group) relating the two annual CPUEs to the annual environmental variables. The model's general structure is given as a directed acyclic graph (DAG; Fig. 3).  $X$  stands for the matrix containing the seven vectors of independent environmental variables. Thus  $X_j$  is the line vector of the environmental conditions in the  $j^{\text{th}}$  year. We define  $Z_{i,j,k}$  as a latent variable representing the number of YOY present in the zone the  $j^{\text{th}}$  year for the  $i^{\text{th}}$  species (belonging to the  $k^{\text{th}}$  group of species). A logPoisson was chosen, meaning that the logarithm of  $Z_{i,j,k}$  depends on the environmental variables under a linear combination of the independent variables:

$$\log(Z_{i,j,k}) = Z_i^0 + \beta_k \cdot X_{j,k} \quad (4)$$

where  $\beta_k$  is the vector of the coefficients of the regression for the  $k^{\text{th}}$  group of species and  $Z_i^0$  is the logarithm of the mean annual number of YOY for the  $i^{\text{th}}$  species. That means that the latent number of YOY for the  $i^{\text{th}}$  species (belonging to the  $k^{\text{th}}$  group of species) for the  $j^{\text{th}}$  year is obtained as the product of a mean interannual number of YOY by the effect of the

environment on the  $k^{\text{th}}$  group of species. The model is organized into a hierarchy, expressed by the common environmental effect on species belonging to the same group of species. Thus, the  $\beta_k$  parameters are the hyperparameters of the model, which describe (conditional upon environmental conditions  $X_{j,k}$ ) the environmental effects on the exchangeable fish species within each group. The number  $Y_{i,j,k,l}$  of YOY observed the  $j^{\text{th}}$  year in the  $i^{\text{th}}$  species during the  $l^{\text{th}}$  fishing event is considered to be a random variable according to a Poisson distribution as a distribution parameter.

$$Y_{i,j,k,l} \sim P(Z_{i,j,k}) \quad (5)$$

This Poisson distribution is justified by the fact that fish are assumed to be randomly distributed in the sample zone, which at first sight is both a simple yet rather realistic hypothesis.

The inference on the regression coefficient of this Generalized Linear Model was conducted using the Bayesian paradigm. Highly-dispersed Zellner Student priors were used on regression parameter priors and on  $Z_i^0$  as in Marin & Robert (2008). Priors are probability distributions describing the state of knowledge that scientists have on the possible range of values for their model parameters before carrying out any experiment. There is no single consensus definition of what “non-informative” means in the statistical community (see Bernardo & Smith (1994) for discussion), but using flat priors as we did is a practical and common way of providing vague prior knowledge. The knowledge acquired during the experiment is expressed via the model likelihood  $[Y|\theta]$  and is used to update the prior probability distributions  $[\theta]$  of parameters during the inference process to obtain the posterior distribution  $[\theta|Y]$  using Bayes’ formula:

$$[\theta|Y] = \frac{[Y|\theta] \cdot [\theta]}{[Y]} \quad (6)$$

where  $[Y]$  is the marginal probability of the data, used to normalize the posterior distribution.  $[Y]$  is generally approximated using a Monte Carlo Markov Chain (MCMC) algorithm. The models were written and implemented in the MCMC-based Bayesian analysis software WINBUGS 14 (Lunn *et al.*, 2000). Three independent Markov chains were run simultaneously. A burn-in period of 50,000 iterations was included to permit the Markov Chains to reach stationarity, after which their convergence was checked using the modified Gelman-Rubin diagnosis (Brooks & Gelman, 1998).

### *Model selection*

For each group of species, introducing  $I=7$  possible independent variables means that there are  $2^I$  different possible models (i.e. 128 models resulting from the different possible combinations of independent variables). To make the notations clearer, the model index  $M$  is made to appear explicitly in the probability distributions. The Bayes factor (BF) was used to determine which of these models best fitted the data (Kass & Raftery, 1995). BF compares the fit of two different models,  $M_1$  and  $M_2$ . It is defined as the ratio of the marginal distributions of the data for both competing models:

$$BF = \frac{[Y|M_1]}{[Y|M_2]} \quad (7)$$

Then, the marginal likelihoods  $[Y|M_i]$  of the data relative to a model  $M_i$  are computed for each of the  $2^7 = 128$  possible models using the importance sampling method described by Marin & Robert (2007). This method does not use WinBUGS parameter outputs, but instead employs samples generated from probability distributions whose mean and standard deviation are the classical logPoisson GLM estimators (see appendix 1 for mathematical development). For each group of species, the model with the highest mean marginal value is selected as the best-fitted model.

### *Model validation*

For each of the three selected models, the Poisson hypothesis was tested using an adjusted Kolmogorov-Smirnov test for discrete random variables. The means of the posterior distributions of the regression parameters of the best-fit models were considered.

A posterior predictive checking method (Gelman *et al.*, 2004) was then used to assess the extent to which the model reflects the natural synchrony between species of the same functional group. Relying on parameter values drawn at random from their posterior joint distribution, a sample containing 10,000 annual predicted catches is generated for each species. This sample is then used to check the synchrony level for each group, as follows. For each  $26 \times 25 / 2 = 276$  possible pairs of years, the variation signs (positive if the number of juveniles is higher for the latest year than it is for the earliest year; otherwise negative) are calculated for each species. Then, for a given functional group and each pair of years, a score of 1 is given if all the species in the group have the same variation sign, or a score of 0 otherwise. The test statistics associated with the synchrony level are then computed as the sum of the scores over 276. The 10,000 samples make it possible to assess the predictive density of probability of the test statistics for each group, and then to obtain the 0.025 and 0.975 quantiles forming the boundaries of a 95% tolerance interval. The observed synchrony levels are computed for each group of species using the observed number of juveniles, and are then compared to the 95% tolerance intervals previously computed. For a given group of species, if the computed value falls into the interval, then the fit is satisfactory and the model describes the natural synchrony level between species of the same group well. A smaller (resp. higher) observed value than the lower (resp. upper) tolerance interval boundary would indicate that the developed model oversynchronises the species trajectories (resp. little).

### **Results**

### *Species cluster*

The cluster procedure clearly distinguishes three groups of species according to their degree of synchrony (Fig. 4). The first separation ( $d=1.24$ ) discriminates an initial cluster composed of bleak and dace (hereafter referred to as “decreasing juvenile abundance”). The other branch then divides into two ( $d = 0.83$ ), isolating a second cluster composed of nase, barbel and gudgeon (hereafter referred to as “slightly increasing juvenile abundance”), while the last cluster encompasses roach, chub and stream bleak (hereafter referred to as “increasing juvenile abundance”).

### *Environmental effects*

The independent variables kept in the structure of the selected best-fit model for each group of species are shown in Fig. 5, along with the probability distributions of the regression parameters of these independent variables.

The results have to be interpreted in terms of variations in the values of the independent variables around their means and, logically, the unit of variation to be considered is the standard deviation (Table 1). Thus, the results reported here present the expected effects of one standard deviation unit change in the considered independent variable on 0+ juvenile abundance.

As a logPoisson model was used, the observed distributions have to be transformed by the exponential function in order to measure the expected effects of the independent variables. Considering one independent variable (and thus every other independent variable being taken as the mean), an increase in its mean value equivalent to one standard deviation will have no effect if the posterior predictive distribution of its corresponding regression parameter is centred on 0, but will double the juvenile abundance of the species if the posterior predictive distribution is centred on 0.7 (as  $e^{0.7} = 2$ ), and will divide this species population in two if it is

centred on -0.7 (as  $e^{-0.7} = 0.5$ ). Furthermore, the effect of an increase of  $n$  times the standard deviation will raise the exponential of the regression parameter to the power  $n$ .

The observed effects are different for the three groups of species. Each group shows a different pattern of response to the environmental variations, and these differences are responsible for the variations in fish assemblage. Nevertheless, some environmental factors seem to have similar effects on different groups of species, revealing possible subtle evolutions.

A late occurrence of the spring temperature threshold has a negative effect on the juvenile abundance of the three groups of species ( $\beta_{C12} = -0.275$ ;  $-0.138$  and  $-0.159$ , resp.). Similarly, a summer period warmer than the interannual average is slightly detrimental for the three groups of species ( $\beta_{S2} = -0.093$ ;  $-0.052$  and  $-0.044$ , resp.).

The two increasing juvenile abundance species groups show rather similar patterns of response to the environmental variations. Indeed, Cmx, S1 and Qm1 have positive effects on the juvenile abundance of these groups of species ( $\beta_{Cmx} = 0.110$ ,  $\beta_{S1} = 0.392$  and  $\beta_{Qm1} = 0.129$  for the slightly increasing juvenile abundance species group;  $\beta_{Cmx} = 0.372$ ,  $\beta_{S1} = 0.834$  and  $\beta_{Qm1} = 0.145$  for the increasing juvenile abundance species group) but have no influence on the group of decreasing juvenile abundance species.

The two environmental variables describing the summer flow conditions, Qm2 and Qmx2, have similar effects on the species of both the decreasing juvenile abundance group and the increasing juvenile abundance group, with Qm2 having a slightly positive effect ( $\beta_{Qm2} = 0.119$ ;  $0.043$ , resp.) while Qmx2 has a strongly negative effect ( $\beta_{Qmx2} = -0.514$ ;  $-0.498$ , resp.). Conversely, Qmx2 has no effect and Qm2 has a strong negative effect ( $\beta_{Qm2} = 0.681$ ) on the juvenile abundance of species in the slightly increasing juvenile abundance species group.

### *Model validation*

The posterior predicted levels of synchrony distributions for each group are given in Table 2. Synchronicity was overpredicted for each group of species (the observed values are lower than the lower 95% predictive tolerance interval bounds given by the model).

The Poisson hypothesis is too stringent. The data are more dispersed than allowed by a Poisson probability distribution function, and the Kolmogorov-Smirnov (KS) test of Poisson adequacy shows that the Poisson distribution fits poorly with the data (KS statistic value = 0.519; KS 95% quantile = 0.071). The implications of this test-of-fit result are discussed below.

## **Discussion**

### *Environmental effects*

Numerous studies have highlighted the role of water temperature on the growth of fish larvae (Schiemer *et al.*, 2004) and on the reproductive success of cyprinid fish species (Mills & Mann, 1985; Grenouillet *et al.*, 2001). It is generally recognized that warmer temperatures (in the range of physiological tolerance for the species) lead to better growth of cyprinids 0+ fish (Mills & Mann, 1985; Daufresne *et al.*, 2003; Wolter, 2007). However, our results indicate that flow conditions also play a major role in determining the juvenile abundances of the different flow-regulated large river species. Humphries & Lake (2000) suggested that flow regime variation and regulation had greater impacts on post-spawning factors responsible for recruitment than on preventing spawning, and it has been shown that in the first weeks of life YOY fish preferentially inhabit low-flow habitats such as slack waters (Moore & Thorp, 2008), where better temperature, flow, food production and light penetration conditions ensure better growth during critical life stages (Schiemer *et al.*, 2001; Moore & Thorp, 2008). However, there is evidence that spawning and YOY development do not occur only in the flood plain during high waters but also in the main channel (Humphries, Serafini & King,

2002). Our study site, with its regulated flow, contains very little slack water, except during high flows (over  $700 \text{ m}^3 \cdot \text{s}^{-1}$ ), meaning the YOY have to deal with the conditions of the main channel. The results observed for the early development period seem to confirm this. As mentioned previously, a higher mean flow regime could provide suitable conditions for spawning and growth habitats. Lytle & Poff (2004) highlighted that fish show behavioural adaptations to natural flow regimes, which mostly consist of finding favourable habitats during floods or droughts (Schiemer, Hein & Reckendorfer, 2007). However, Nunn *et al.* (2007) identified the larval period as most vulnerable to increases in river discharge, although the situation was complicated by interannual variations in hatching times and river discharge. Furthermore, several studies have shown that many cyprinid species undergo a larval drift period, especially during their few first weeks of life (Carter & Reader, 2000). This drift could be favoured or intensified by high flow discharges. Although studies have not detected any significant influence of river discharge on the seasonal pattern of juvenile drift densities (Reichard, Jurajda & Ondrakova, 2002; Zitek, Schmutz & Ploner, 2004; Zitek *et al.*, 2004), this kind of effect cannot be totally excluded. The results of this study are in agreement with these general statements, and a coherent hypothesis can be proposed to explain the juvenile abundances.

#### *Overall common effects*

Two environmental variables, C12 and S2, have similar effects on the eight species. The negative effect of C12 reveals that the earlier this temperature threshold is reached, the higher the annual juvenile abundance will be. The benefits of this earlier water warming have already been demonstrated for dace (Philippart, 1981; Mills & Mann, 1985). Dace is an early-spring spawner (Teletchea *et al.*, 2009), generally spawning when water temperature reaches  $10^\circ\text{C}$  (Tissot & Souchon 2008), and dace larvae benefit from the early water warming as it

enhances growth. Fractional spawning has been reported for bleak (Mackay & Mann, 1969) based on evidence of up to four size groups of eggs in the ovary prior to spawning (Barus & Prokes, 1993). Bleak spawns later than dace when water temperature reaches 15-16°C (Tissot & Souchon 2008), which normally occurs at around day 155 (first week of June). The early water warming may favour better growth of bleak larvae and thus enhance better survival of the successive batches. Teletchea *et al.* (2009) identified nase as an early-spring fractional spawning species, barbel as a late-spring possibly fractional spawning species, and gudgeon as a late-spring fractional spawning species (fractional spawning of gudgeon has also been reported by Mann, 1980). The earlier water warming could then lead to better growth and survival of the nase larvae, as for dace. Concerning the other two species, which are later spawners, earlier water warming probably enhances earlier first spawning and consequently promotes better growth and survival of the different successive batches of larvae. Roach usually spawn once within a 2-4 day period (Mackay & Mann, 1969) when water temperature reaches 15°C (Tissot & Souchon 2008), while chub is a late-spring spawner (Teletchea *et al.*, 2009) that is known to be capable of fractional spawning (Economou, Daoulas & Psarras, 1991; Vriese, Semmekrot & Raat, 1994), and its eggs can be distributed over up to three spawnings (Libosvsky, 1979), and stream bleak is probably capable of fractional spawning after water temperature reaches 14°C (Tissot and Souchon 2008). Globally, an early spring threshold probably enhances earlier spawning, allowing the larvae and subsequent juveniles better to survive the subsequent environmental conditions.

The overall negative effect of a hotter-than-average summer on juvenile abundance could be explained by both direct and indirect factors. It could reduce the available thermal juvenile niche and provoke increased juvenile mortality if temperature stays higher than the lethal temperature of the species for an excessively long period (Tissot & Souchon, 2008). Higher

water temperature could also be detrimental for aquatic invertebrates and algae, causing higher juvenile mortality due to starvation.

*Common environmental drivers for both increasing abundance groups*

The three environmental variables Cmx, S1 and Qm1 have positive influences on the species of both groups of increasing abundance but no effect on the species of declining abundance.

Thus, the specific responses to these three variables appears to be the main factor driving whether species will belong to the declining abundance group or to one of the two increasing abundance groups. At the same time, Cmx and S1 seem to be the two variables responsible for the particularly high increase in juvenile abundance of roach, chub and stream bleak, as they have a higher positive impact on these three species than on nase, barbel and gudgeon.

The very positive effect of S1, which describes the relative sum of degree-days in the spring period, seems to confirm the generally-accepted statement that warmer temperatures lead to better growth of cyprinid juveniles (Mills & Mann, 1985; Daufresne *et al.*, 2003; Wolter, 2007). A high S1 probably decreases the incubation duration of these species and promotes the early development of larvae, which in turn favours survival during the subsequent periods.

The positive effect of Cmx on the juvenile abundance of these species indicates the small benefit of a longer summer period. This effect is probably due to the fact that a later water temperature drop lets the early juvenile fish of these three species reach a higher fork length at the end of the summer, thus promoting better subsequent survival. The flow conditions seem to play an important role in determining the juvenile abundance of these groups of species.

Qm1 has a positive effect, confirming the importance of the availability of slack waters during spawning. Indeed, slack waters become available if the mean spring flow regime increases from  $565 \text{ m}^3 \cdot \text{s}^{-1}$  to  $741 \text{ m}^3 \cdot \text{s}^{-1}$ .

### *Benthic vs. pelagic species*

The two environmental variables describing the summer flow conditions ( $Q_{m2}$  and  $Q_{mx2}$ ) have different effects depending on whether they applied to benthic species in the slightly increasing abundance group or pelagic species in the two other groups.

The negative effect of a high maximal flow discharge on pelagic species could be explained by three hypotheses. First, this high flow discharge could promote YOY drift. Carter & Reader (2000) suggested that drift of 0+ fish is probably part of a migratory cycle and has a function in allowing movement from the spawning habitat to nursery areas. They reported that drift of 0+ fish was mostly restricted to the first few weeks of life. In our case, during the defined summer period, YOYs have to be considered as early juveniles or larvae with poor swimming abilities. Secondly, Bolland, Cowx & Lucas (2008) showed that in lowland rivers, vast numbers of under-yearling coarse fish were found in temporary floodplain water bodies created by over-topping of floodbanks, as they were unable to return to the main channel when the floodwaters subsided. These young fish were mainly eurytopic species including bleak and roach, but also included rheophilic species such as dace and chub. Finally, concerning bleak, the summer floods could negatively impact the later reproductive activity of this fractional-spawning species, causing a drop in overall bleak juvenile abundance. In contrast, a high mean summer flow regime has a detrimental effect on the juvenile abundance of benthic species. This could be due to high water velocities leading to unfavourable benthic feeding conditions, or to higher water depth causing less algae development.

### *Community structure predictions and human management*

The mean annual water temperature at the Bugey study site has increased by 2°C over the last 20 years (Daufresne *et al.*, 2003), accompanied by an increase in the S1 and S2 variables. The

general trend indicates that a shift occurred in the early 1990s, and since 1995 S1 has constantly been higher than its interannual mean. This coincides with the sharp drop in the numbers of juvenile dace and bleak and the significant growth in numbers of juvenile roach, chub and stream bleak. This phenomenon seems to stem from warmer waters strongly favouring the latter three species. Furthermore, the increase in midsummer floods seems to be more detrimental for dace and bleak, probably because these two species do not benefit from water warming. Nevertheless, the system must be appreciated holistically, and certain compensatory mechanisms could be involved, such as juvenile or even adult recolonization from other source sites, thereby slowing down the decrease in some species and preventing their disappearance. Species mobility and spatial connectivity tend to reduce the risk of extinction (Lowe, 2002; Albanese, Angermeier & Peterson, 2009). A study by Phillips & Johnston (2004) suggested that patterns of recovery from disturbances could occur over distances of 20 km or more. The random occurrence of favourable years may act in the same way. Consequently, it is reasonable to think that, if water continues to become warmer, the fish community would continue to evolve toward the overall dominance of roach, chub and stream bleak, with nase, barbel and gudgeon continuing to accompany these species. Dace and bleak should probably remain present due to ecological adaptations and random environmental variations.

This work confirms that juvenile fish assemblages are shaped by the complex interactive effects of flow and thermal regimes (Olden & Naiman, 2010). Even if water temperature plays an important role in determining juvenile abundance, its effects can be significantly modulated by flow conditions. Since these are deeply impacted by human flow management for hydroelectric production, human flow management may be a major factor in structural variations in large river fish assemblages. It follows that readjusting watershed management

to decrease the amplitude of midsummer floods may help slow the decreases in dace and bleak numbers.

### *Hierarchical modelling*

Oberdorff & Porcher (1992) stressed the importance of describing, quantifying and modelling the relationships between assemblages and their habitats. Hierarchical modelling is particularly well-adapted in cases where the structure of the problem pleads for linking certain parameters, making it possible to build a joint probability model for these parameters that reflects their interdependence (Gelman *et al.*, 2004). Here, we assumed the existence of groups of species sharing similar responses to environmental constraints. Bayesian hierarchical modelling proved a reliable method of dealing with our complex system, making it possible to take these similarities into account and thus allowing a better assessment of environmental effects. In fact, by allowing the introduction of hidden variables (Clark, 2005), it becomes possible to separate uncertainty into process uncertainty (here, the environmental effects) and sample uncertainty (the Poissonian error). Furthermore, Bayesian hierarchical models are easy to write, as they directly reflect the way people conceptualize their system, and they are particularly flexible. This is an important advantage when using data from different sources, which leads to different source uncertainties, as is the case when using monitoring datasets (Webb *et al.*, 2010). Although our model has been developed using data from one study site, its structure is robust enough to be completed and then trialled on more general datasets.

### *Modelling hypothesis*

By using a Poisson model, we considered that fish are randomly distributed over the total sampled zone. However, as fish generally form shoals, they may show patchy distributions. It

appears in our study that the Poisson model is not well supported by the data, and that the latter are over-dispersed. This extra variation may, of course, first be imputed to the interannual variation stemming from the non-controlled conditions of the fishing protocol (change of boats, change of operators, changes in river flows between experiments).

Technically, but at the cost of estimating an additional parameter, the Poisson distribution could be replaced by a more dispersed alternative such as the negative binomial (a Gamma mixture of Poisson probability distribution functions). The extra variation feature may also be a consequence of pooling all the data from all the sites along the river, and much more refined models can be readily investigated in future research aimed at modelling these extra sources and variations by including fixed seasonal effects as well as random site effects. Nevertheless, it is very likely that the trends in environmental effects (meaning the means of the probability distributions of the regression parameters) will be conserved, despite loss of precision in the coefficient estimators.

Cressie *et al.* (2009) stressed the importance of assessing the model fit using posterior predictive distribution (Gelman *et al.*, 2004), but they also underlined the possible difficulties inherent in such an approach, especially in the case of high-dimensional distributions. Our study includes an example of a simple model test-of-fit, as the statistic used is relatively straightforward and can be directly computed from simulations using the posterior predictive distribution.

## **Conclusions**

Poff & Zimmerman (2010) summarized the latest literature in an attempt to quantify the effects of flow regime alterations on ecological assemblages. They pointed to the lack of studies on the effects of intermediate alterations. Our work combined a hierarchical modelling approach and a simple description of the annual flow regime to provide novel insight into how

0+ fish assemblages react to interannual flow variations in a flow-regulated river, and it points to the importance of flow management in managing fish assemblages. The study identified flow regime as an important driver, in addition to natural temperature regime variations (Daufresne *et al.*, 2003), in determining variations in fish assemblages. Furthermore, introducing the temperature regime confirmed the importance of interannual thermic variations in determining the abundance of species in the different assemblages, and illustrated the effects of water warming over the last two decades.

### **Acknowledgments**

We would like to thank J.A. Webb and an anonymous reviewer for their very useful comments and suggestions. We are grateful to Marie-Pierre Etienne and Etienne Rivot for their judicious initial suggestions and for their subsequent support. We also thank Bertrand Villeneuve for helping with statistical methods and Nolwenn Bougon for helping with the figures. Many thanks to Keith Hodson, from Accent Europe, for having corrected the English of this article. The data belong to Electricité De France.

### **References**

- Albanese, B., Angermeier, P.L. & Peterson, J.T. (2009) Does mobility explain variation in colonisation and population recovery among stream fish? *Freshwater Biology*, **54**, 1444-1460.
- Barus, V. & Prokes, M. (1993) Fecundity of the bleak (*Alburnus alburnus*) in the Vestonice reservoir. *Folia Zoologica*, **42**, 281-288.
- Bayley, P. & Li, H. (1992) Riverine fishes. In: *The Rivers Handbook, volume 2* (Eds P. Calow & G. Petts), pp. 251-281. Blackwell Scientific Publications, Oxford.

- Bayley, P.B. (2002) A review of studies on responses of salmon and trout to habitat change, with potential for application in the Pacific Northwest. Report to the Washington State Independent Science Panel (supplement), Washington.
- Bolland, J.D., Cowx, I.G. & Lucas, M.C. (2008) Factors affecting the dispersal of coarse Fish, Environment Agency, UK.
- Brooks, S.P. & Gelman, A. (1998) General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, **7**, 434-455.
- Carter, K.L. & Reader, J.P. (2000) Patterns of drift and power station entrainment of 0+ fish in the River Trent, England. *Fisheries Management and Ecology*, **7**, 447-464.
- Cattaneo, F., Carrel, G., Lamouroux, N. & Breil, P. (2001) Relationship between hydrology and cyprinid reproductive success in the lower Rhône at Montélimar, France. *Archiv für Hydrobiologie*, **151**, 427-450.
- Chessman, B.C. (2006) Prediction of riverine fish assemblages through the concept of environmental filters. *Marine and Freshwater Research*, **57**, 601-609.
- Clark, J.S. (2005) Why environmental scientists are becoming Bayesians. *Ecology Letters*, **8**, 2-14.
- Copp, G.H. (1989) The habitat diversity and fish reproductive function of floodplain ecosystems. *Environmental Biology of Fish*, **26**, 1-27.
- Copp, G.H., Olivier, J.M., Penaz, M. & Roux, A.L. (1991) Juvenile fish as functional descriptors of fluvial ecosystem dynamics: applications on the River Rhone, France. *Regulated Rivers : Research and Management*, **6**, 135-145.
- Cressie, N., Calder, C.A., Clark, J.S., Hoef, J.M.V. & Wikle, C.K. (2009) Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecological Applications*, **19**, 553-570.

- Daufresne, M., Roger, M.C., Capra, H. & Lamouroux, N. (2003) Long-term changes within the invertebrate and fish communities of the Upper Rhône River: effects of climatic factors. *Global Change Biology*, **10**, 124-140.
- Economou, A.N., Daoulas, C. & Psarras, T. (1991) Growth and morphological development of chub, *Leuciscus cephalus* (L.), during the first year of life. *Journal of Fish Biology*, **39**, 393-408.
- Ficke, A.D., Myrick, C.A. & Hansen, L.J. (2007) Potential impacts of global climate change on freshwater fisheries. *Reviews in Fish Biology and Fisheries*, **17**, 581-613.
- Firth, P. & Fisher, S.G. (1992) *Global climate change and freshwater ecosystems*, Springer Verlag, Paris.
- Fruget, J.F., Bady, P., Olivier, J.M., Carrel, G., Souchon, Y., Villeneuve, B., Daufresne, M. & Capra, H. (2006) Etude thermique globale du Rhône phase III, Lot 2: Etude à l'échelle du Rhône des compartiments biologiques. Cemagref.
- Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. (2004) *Bayesian Data Analysis*, Chapman & Hall/CRC, Boca Raton, FL.
- Grenouillet, G., Hugueny, B., Carrel, G.A., Olivier, J.M. & Pont, D. (2001) Large-scale synchrony and inter-annual variability in roach recruitment in the Rhône River: the relative role of climatic factors and density-dependent processes. *Freshwater Biology*, **46**, 11-26.
- Gulland, J.A. (1982) Why Do Fish Numbers Vary. *Journal of Theoretical Biology*, **97**, 69-75.
- Houde, E.D. (1997) Patterns and trends in larval-stage growth and mortality of teleost fish. *Journal of Fish Biology*, **51**, 52-83.
- Humphries, P., Brown, P., Douglas, J., Pickworth, A., Strongman, R., Hall, K. & Serafini, L. (2008) Flow-related patterns in abundance and composition of the fish fauna of a degraded Australian lowland river. *Freshwater Biology*, **53**, 789-813.

- Humphries, P. & Lake, P.S. (2000) Fish larvae and the management of regulated rivers. *Regulated Rivers-Research & Management*, **16**, 421-432.
- Humphries, P., Serafini, L.G. & King, A.J. (2002) River regulation and fish larvae: variation through space and time. *Freshwater Biology*, **47**, 1307-1331.
- Kass, R.E. & Raftery, A.E. (1995) Bayes Factors. *Journal of The American Statistical Association*, **90**, 773-795.
- Lamouroux, N., Olivier, J.M., Persat, H., Pouilly, M., Souchon, Y. & Statzner, B. (1999) Predicting community characteristics from habitat conditions: fluvial fish and hydraulics. *Freshwater Biology*, **42**, 275-299.
- Larkin, P.A. (1978) Fisheries Management - Essay for Ecologists. *Annual Review of Ecology and Systematics*, **9**, 57-73.
- Libosvsky, J. (1979) Gonad weight and egg numbers in chub, *Leuciscus cephalus*, from the Rokytina stream. *Folia Zoologica*, **28**, 35-42.
- Lowe, W.H. (2002) Landscape-scale spatial population dynamics in human-impacted stream systems. *Environmental Management*, **30**, 225-233.
- Lunn, D.J., Thomas, A., Best, N. & Spiegelhalter, D. (2000) WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, **10**, 325-337.
- Lytle, D.A. & Poff, N.L. (2004) Adaptation to natural flow regimes. *Trends in Ecology & Evolution*, **19**, 94-100.
- Mackay, I. & Mann, K.H. (1969) Fecundity of two cyprinid fish in the River Thames, Reading, England. *Journal of the Fisheries Research Board of Canada*, **26**, 2795-2805.
- Mann, R.H.K. (1980) The growth and reproductive strategy of the gudgeon, *Gobio gobio* (L.), in two hard-water rivers in southern England. *Journal of Fish Biology*, **17**, 163-176.

- Marin, J.-M. & Robert, C.P. (2007) *Bayesian core: a practical approach to computational Bayesian statistics*, Springer-Verlag, New York.
- Mills, C.A. & Mann, R.H.K. (1985) Environmentally-induced fluctuations in year-class strength and their implications for management. *Journal of Fish Biology*, **27**, 209-226.
- Moore, S.L. & Thorp, J.H. (2008) Coping with hydrogeomorphic variations in a prairie river: Resiliency in young-of-the-year fish. *River Research And Applications*, **24**, 267-278.
- Nunn, A.D., Cowx, I.G., Frear, P.A. & Harvey, J.P. (2003) Is water temperature an adequate predictor of recruitment success in cyprinid fish populations in lowland rivers? *Freshwater Biology*, **48**, 579-588.
- Nunn, A.D., Harvey, J.P., Britton, J.R., Frear, P.A. & Cowx, I.G. (2007) Fish, climate and the Gulf Stream: the influence of abiotic factors on the recruitment success of cyprinid fish in lowland rivers. *Freshwater Biology*, **52**, 1576-1586.
- Oberdorff, T., Pont, D., Hugueny, B. & Porcher, J.P. (2002) Development and validation of a fish-based index for the assessment of "river health" in France. *Freshwater Biology*, **47**, 1720-1734.
- Oberdorff, T. & Porcher, J.P. (1992) Fish assemblage structure in Brittany streams (France). *Aquatic Living Resources*, **5**, 215-223.
- Olden, J.D. & Naiman, R.J. (2010) Incorporating thermal regimes into environmental flows assessments: modifying dam operations to restore freshwater ecosystem integrity. *Freshwater Biology*, **55**, 86-107.
- Parent, E. & Bernier, J. (2007) *Le raisonnement bayésien: modélisation et inférence*, Springer, Paris.
- Persat, H. & Chessel, D. (1989) Size-Class Ordination - a Useful Method in the Study of Fish and Invertebrate Populations. *Acta Oecologica-Oecologia Generalis*, **10**, 175-195.

- Philippart, J.C. (1981) Ecologie d'une population de vandoise, *Leuciscus leuciscus* (L.) dans la rivière Ourthe (bassin de la Meuse, Belgique). *Annales de Limnologie*, **17**, 41-62.
- Phillips, B.W. & Johnston, C.E. (2004) Fish assemblage recovery and persistence. *Ecology of Freshwater Fish*, **13**, 145-153.
- Poff, N.L. (1992) Regional hydrologic response to climate change: an ecological perspective. In: *Global climate change and freshwater ecosystems* (Eds P. Firth & S.G. Fischer), pp. 88-115. Springer Verlag, New York.
- 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.
- Poff, N.L. & Zimmerman, J.K.H. (2010) Ecological responses to altered flow regimes: a literature review to inform the science and management of environmental flows. *Freshwater Biology*, **55**, 194-205
- Quist, M.C., Rahel, F.J. & Hubert, W.A. (2005) Hierarchical faunal filters: an approach to assessing effects of habitat and nonnative species on native fish. *Ecology of Freshwater Fish*, **14**, 24-39.
- Reichard, M., Jurajda, P. & Ondrakova, M. (2002) The effect of light intensity on the drift of young-of-the-year cyprinid fish. *Journal of Fish Biology*, **61**, 1063-1066.
- Schiemer, F., Hein, T. & Reckendorfer, W. (2007) Ecohydrology, key-concept for large river restoration. *Ecohydrology and Hydrobiology*, **7**, 101-111.
- Schiemer, F., Keckeis, H. & Flore, L. (2001) Ecotones and hydrology: Key conditions for fish in large rivers. *Ecohydrology and Hydrobiology*, **1**, 49-55.
- Schiemer, F., Keckeis, H., Nemeschkal, H., Schludermann, E., Winkler, G. & Zweimüller, I. (2004) Ontogenetic patterns in thermal adaptation of fish vs. long-term temperature trends in large rivers. *International Congress Series*, **1275**, 209-217.

- Schiemer, F., Spindler, T., Wintersberger, H., Schneider, A. & Chovanec, A. (1991) Fish fry associations: important indicators for the ecological status of large rivers. *Verhandlungen der Internationale Vereinigung für Theoretische und Angewandte Limnologie*, **24**, 2497-2500.
- Schlosser, I.J. (1985) Flow regime, juvenile abundance, and the assemblage structure of stream fish. *Ecology*, **66**, 1484-1490.
- Schneider, S.H., Semenov, S., Patwardhan, A., Burton, I., Magadza, C.H.D., Oppenheimer, M., Pittock, A.B., Rahman, A., Smith, J.B., Suarez, A. & Yamin, F. (2007) Assessing key vulnerabilities and the risk from climate change. *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden and C.E. Hanson, Eds., Cambridge University Press, Cambridge, UK, 779-810.
- Souchon, Y. (1994) Development of Fish Habitat Modelization in French Streams in Recent Years. *Bulletin Français De La Peche Et De La Pisciculture*, **67**, 57-71.
- Teletchea, F., Gardeur, J.N., Kamler, E. & Fontaine, P. (2009) The relationship of oocyte diameter and incubation temperature to incubation time in temperate freshwater fish species. *Journal of Fish Biology*, **74**, 652-668.
- Tissot, L. & Souchon, Y. (2008) Etat des connaissances sur les préférences thermiques des principales espèces de poissons des eaux continentales françaises. Rapport EDF LNHE/Cemagref.
- Tonn, W.M., Magnuson, J.J., Rask, M. & Toivonen, J. (1990) Intercontinental comparison of small-lake fish assemblage: the balance between local and regional processes. *American Naturalist*, **136**, 345-375.

- Vriese, F.T., Semmekrot, S. & Raat, A.J.P. (1994) Assessment of spawning and nursery areas in the river Meuse. *Water Science and Technology*, **29**, 297-299.
- Ward, J. (1963) Hierarchical Grouping to Optimize an Objective Function. *Journal of The American Statistical Association*, **58**, 236-244.
- Webb, J.A., Stewardson, M.J. & Koster, W.M. (2010) Detecting ecological responses to flow variation using Bayesian hierarchical models. *Freshwater Biology*, **55**, 108-126
- Wolter, C. (2007) Temperature influence on the fish assemblage structure in a large lowland river, the lower Oder River, Germany. *Ecology of Freshwater Fish*, **16**, 493-503.
- Zitek, A., Schmutz, S. & Ploner, A. (2004) Fish drift in a Danube sidearm-system: II. Seasonal and diurnal patterns. *Journal of Fish Biology*, **65**, 1339-1357.
- Zitek, A., Schmutz, S., Unfer, G. & Ploner, A. (2004) Fish drift in a Danube sidearm-system: I. Site-, inter-and intraspecific patterns. *Journal of Fish Biology*, **65**, 1319-1338.

## Tables

**Table 1** Summaries (means and standard deviations) of independent environmental variables used in the model.

Independent		
variable	mean	Sd
C12	115.4	11.4
Cmx	214.6	16.1
S1	-12.5	88.0
S2	26.4	71.5
Qm1	565.0	176.2
Qm2	580.9	152.7
Qmx2	880.2	221.1

**Table 2** Observed degrees of synchrony between species groups and posterior predicted degree of synchrony distributions

	Simulated quantiles			Observed values
	2.5%	50%	97.5%	
<hr/>				
Declining juvenile abundance species group	0.764	0.830	0.884	0.641
Slightly increasing juvenile abundance species group	0.721	0.783	0.837	0.652
Increasing juvenile abundance species group	0.754	0.815	0.866	0.558
<hr/>				

## Figure legends

Fig. 1 Study area showing the locations of the fish sampling sites (1–8) and the temperature (T) and discharge (Q) recording sites (adapted from Daufresne *et al.*, 2003).

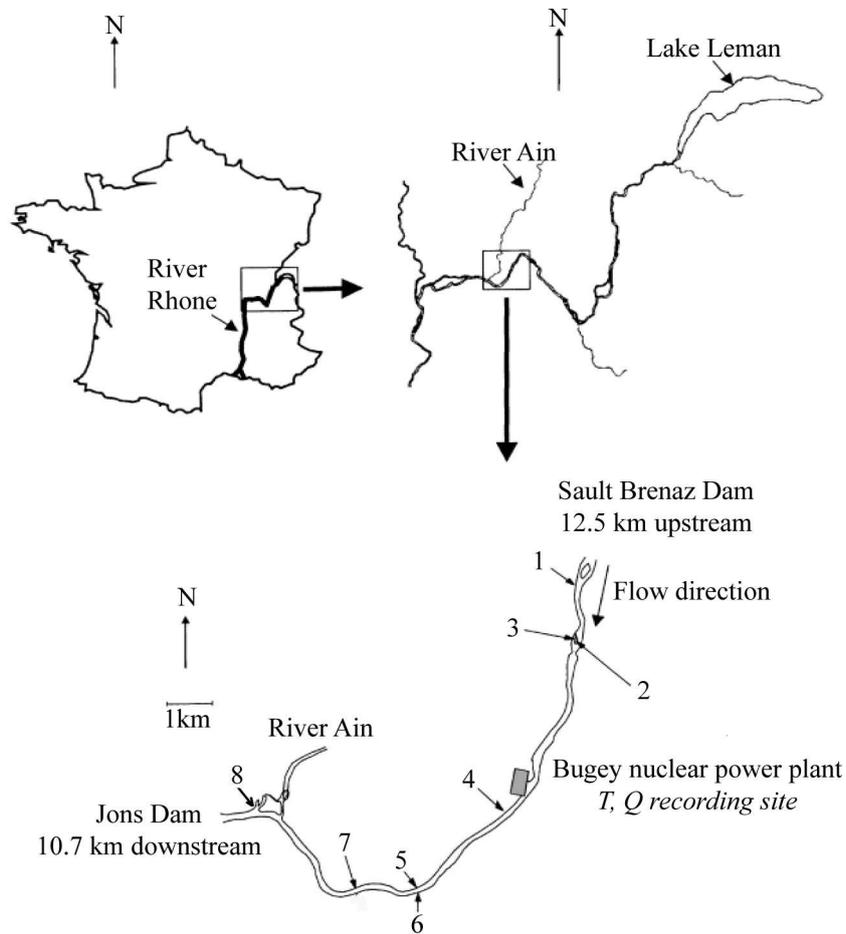


Fig. 2 Chronologies of autumn fish sampling results, expressed as Catch Per Unit Effort (CPUE, number of fish caught per 30 minutes), between 1980 and 2005 for the eight most abundant species.

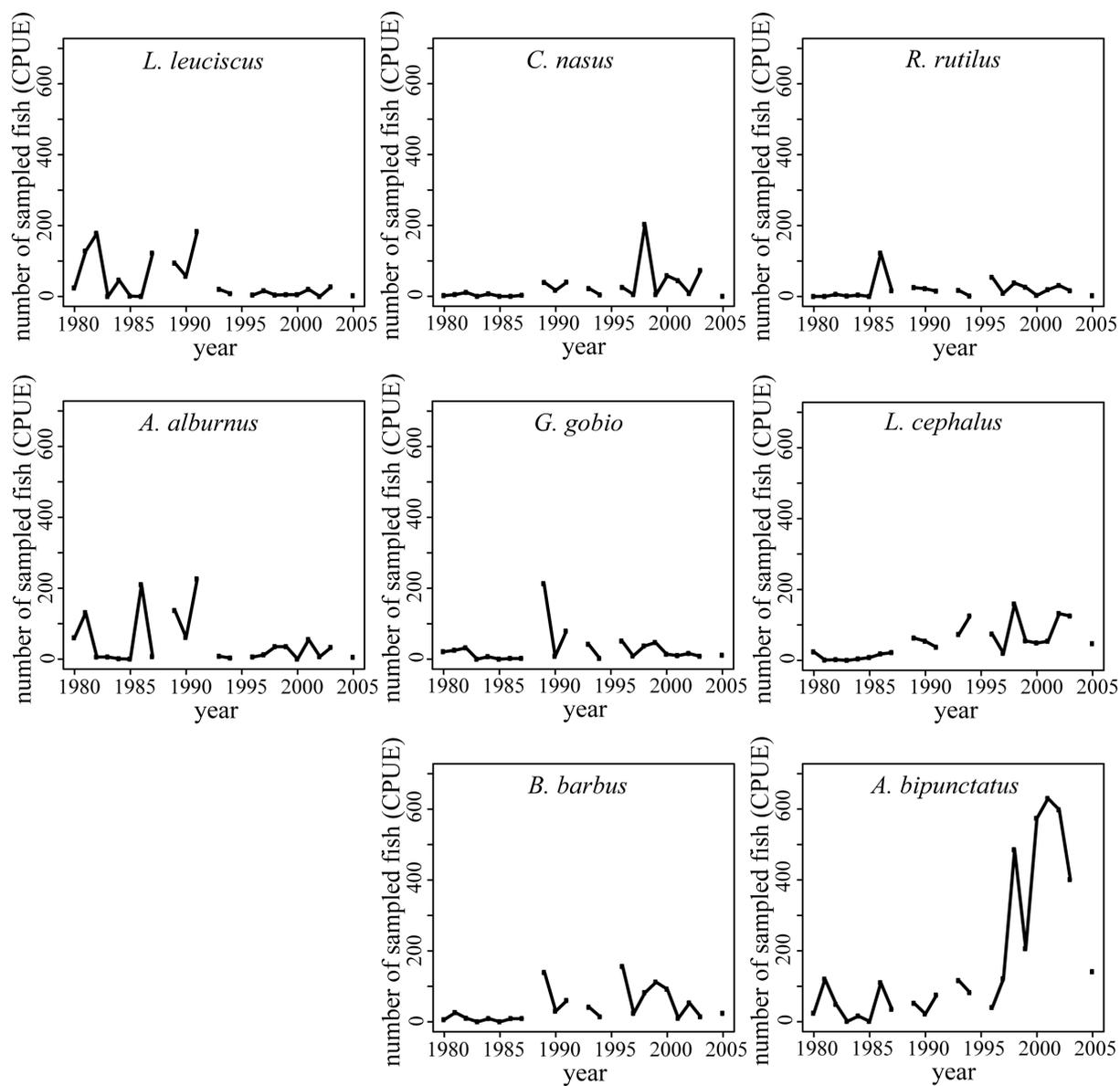


Fig. 3 Directed Acyclic Graph of the model relating the observed annual 0+ fish abundance to the environmental independent variables (square nodes are observed data, circle nodes are latent variables and parameters).

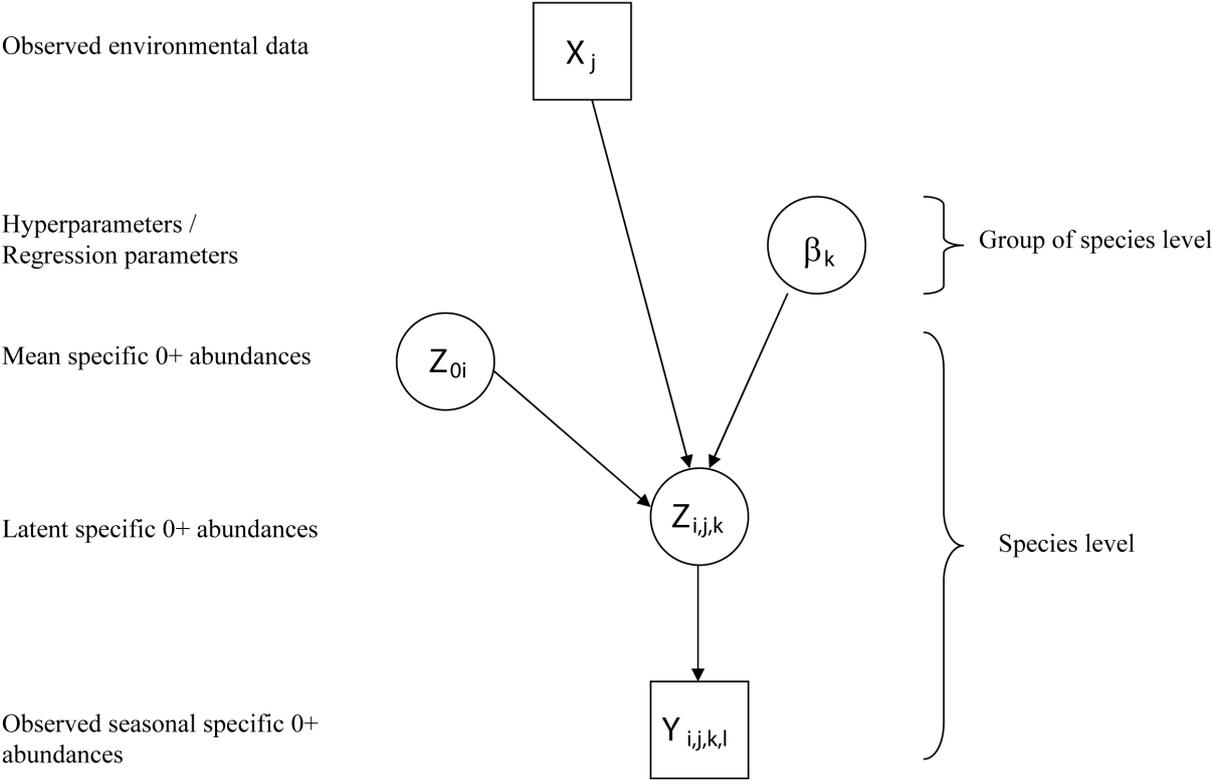


Fig. 4 Cluster dendrogram of fish species. The Kendall tau rank correlation coefficient is used as the distance unit and cluster are computed using Ward's (1963) method.

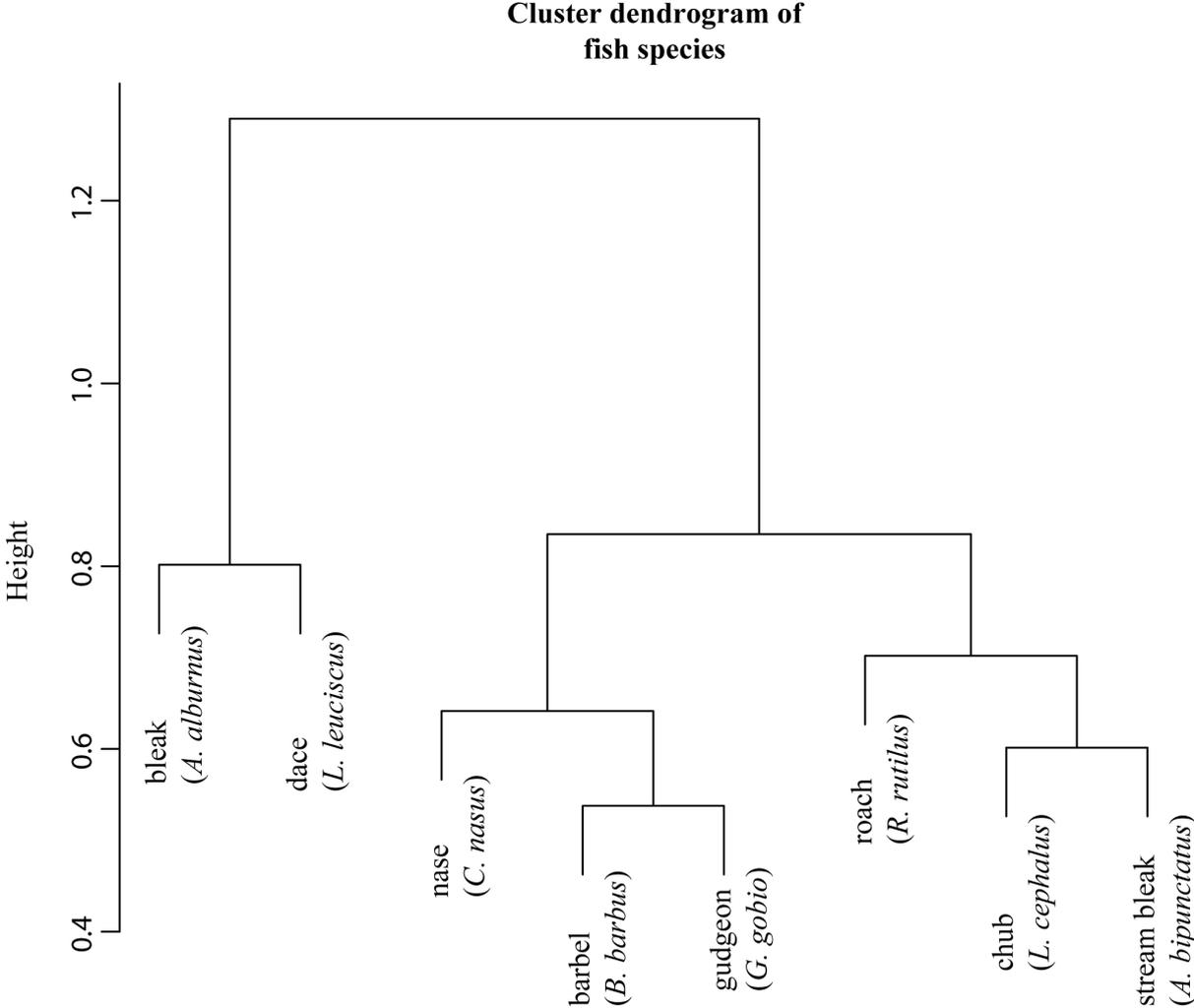
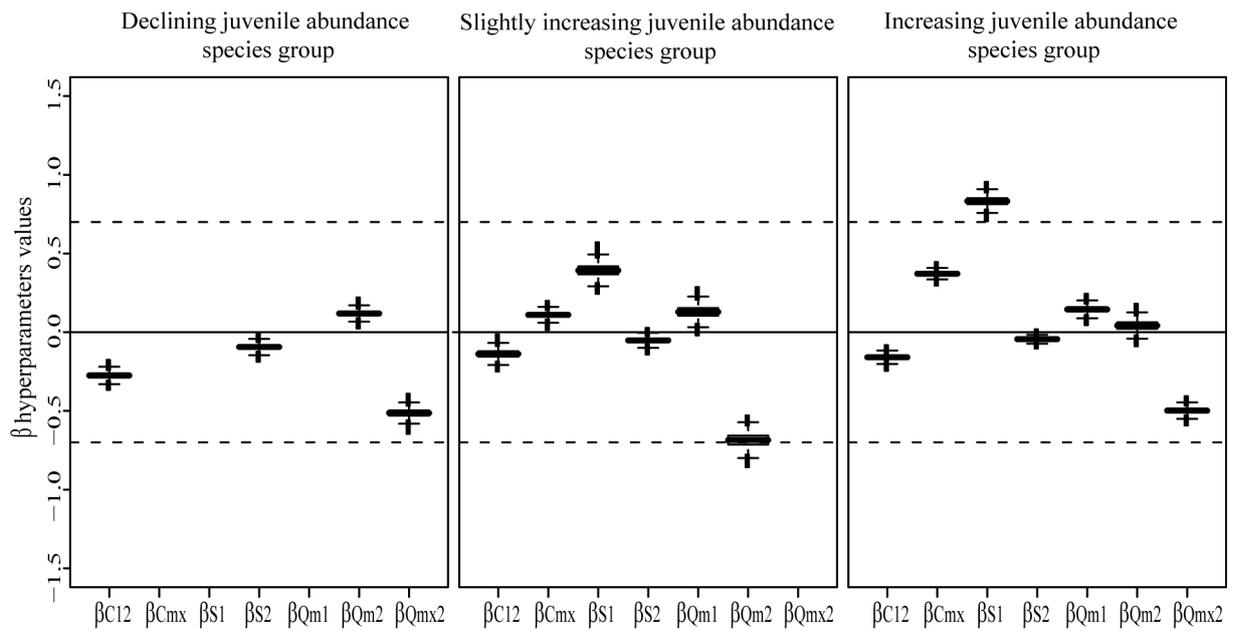


Fig. 5 Box plots of the predictive posterior probability distributions of the regression parameters  $\beta$  for the best selected models with a Bayes factor criteria (median, first and third quartiles, 95% credibility intervals). Lines are null effect (plain line), -0.7 and 0.7 effects (dashed lines).



## Appendix 1 Implementation of Bayes factor computation.

We denote  $\gamma$  a vector of size 8 with binary digits such as (0,1,1,0,1,0,1,0,1).  $\gamma_i = 0$  means that independent variable  $i$  is not considered as an explanatory variable. Knowing the dependency on the model  $\gamma$  is equivalent to know the structure of the independent variable matrix  $X$ . So, for each model  $\gamma$ , the predictive is expressed as:

$$[Y|X^\gamma] = \int \int_{\Sigma_B} [Y|\beta^\gamma, \sigma^{\gamma 2}|X^\gamma] \cdot d\beta \cdot d\sigma^2 \quad (1)$$

As we hypothesized an inverse gamma mixture of normal Zellner priors (i.e. a multivariate Student distribution), we can write

$$[Y|X^\gamma] = \int \int_{\Sigma_B} [Y|\beta^\gamma, X^\gamma] [\beta^\gamma|\sigma^{\gamma 2}, X^\gamma] [\sigma^{\gamma 2}|X^\gamma] \cdot d\beta \cdot d\sigma^2 \quad (2)$$

The successive conditional distributions are expressed as:

$$[Y|\beta^\gamma, X^\gamma] = \prod_i \frac{\exp(\beta^\gamma \cdot X_i^\gamma)}{Y^i!} \exp(-\exp(\beta^\gamma \cdot X_i^\gamma)) \quad (3)$$

$$[\beta^\gamma|\sigma^{\gamma 2}, X^\gamma] = (2\pi)^{\frac{p}{2}} \cdot \left(\frac{1}{\sigma^{\gamma 2}}\right)^{\frac{p}{2}} \cdot |X^{\gamma T} X^\gamma|^{\frac{1}{2}} \cdot \exp\left(-\frac{1}{2}\right) \cdot \left(\frac{(\beta^\gamma - \beta_0)^T X^{\gamma T} X^\gamma (\beta^\gamma - \beta_0)}{\sigma^{\gamma 2}}\right) \quad (4)$$

$$[\sigma^{\gamma 2}|X^\gamma] = \left(\frac{1}{\sigma^{\gamma 2}}\right)^{a-1} \cdot \frac{b^a}{\Gamma(a)} \exp\left(-\frac{b}{\sigma^{\gamma 2}}\right) \quad (5)$$

Once integrated upon the space of  $\sigma^{\gamma 2}$ , the prior distribution for the coefficients  $\beta^\gamma$  of  $X^\gamma$  is a multivariate Student distribution and the predictive becomes:

$$[Y|X^\gamma] = \int \prod_B \frac{\exp(\beta^\gamma \cdot X_i^\gamma)}{Y^i!} \exp(-\exp(\beta^\gamma \cdot X_i^\gamma)) \cdot \left(\frac{1}{\sqrt{2\pi}}\right)^p |X^{\gamma T} X^\gamma|^{\frac{1}{2}} \frac{b^a}{\Gamma(a)} \frac{\Gamma(a)}{b^{a^i}} \cdot d\beta \quad (6)$$

With:

$$\begin{aligned}
 a' &= \frac{p}{2} + a \\
 b' &= b + \frac{(\beta^\gamma - \beta_0) X^{\gamma T} X^\gamma (\beta^\gamma - \beta_0)}{2}
 \end{aligned} \tag{7}$$

As such, the integral (8) can not be analytically performed and one has to make recourse to a numerical implementation. The importance sampling method described in Marin & Robert (2007) was used. This consists of computing  $\hat{\beta}, \hat{\Sigma}$  the classical estimators of  $\beta^\gamma$  and the corresponding covariance matrix of the glm. Then generating a G-sample of  $\beta^{(g)}$  following the multivariate normal distribution  $N(\hat{\beta}, \hat{\Sigma})$ . Using this sample, the integral is estimated as via importance sampling:

$$[Y|X^\gamma] \approx \frac{1}{G} \sum_{g=1}^G \frac{\prod_i \frac{\exp(\beta^{(g)} \cdot X^{\gamma_i})}{Y^i!} \exp(-\exp(\beta^{(g)} \cdot X^{\gamma_i})) \cdot \left(\frac{1}{\sqrt{2\pi}}\right)^p |X^{\gamma T} X^\gamma|^{\frac{1}{2}} \frac{b^a}{\Gamma(a)} \frac{\Gamma(a')}{b^{a'}}}{[\beta^{(g)}|\hat{\beta}, \hat{\Sigma}]} \tag{8}$$

When computing Bayes factors for different models the determinant term  $|X^{\gamma T} X^\gamma|$  will compensate for the various model dimensions. One can see that we can even use an improper prior to compare all models by letting (a,b) going towards zero since the improper inverse gamma probability distribution function will be common to all models.



## Annexe B

Code WinBUGS : modèle  
logPoisson à "effet centrale"  
stationnarisé

```

model{
##### Priors
  for (clust in 1 : 3) {
    aC12[clust] ~ dnorm(0, 0.001)
    aCmx[clust] ~ dnorm(0, 0.001)
    aS1[clust] ~ dnorm(0, 0.001)
    aS2[clust] ~ dnorm(0, 0.001)
    aQm1[clust] ~ dnorm(0, 0.001)
    aQm2[clust] ~ dnorm(0, 0.001)
    aQmx2[clust] ~ dnorm(0, 0.001)
  }

##### Expression de l'effectif moyen attendu
  for (esp in 1 :8) {
    recru0[esp] ~ dnorm(0, 0.001)
    exprecru0[esp] <- exp(recru0[esp])
  }
  for (e in 1 : 360) {
##### Modèle statistique de l'échantillonnage
    recru[e] ~ dpois(true_recru[espece[e], annee[e]])
    recruprd[e] ~ dpois(true_recru[espece[e], annee[e]])

##### Expression des résidus, observés et prédits
    e.obs[e] <- (recru[e] - true_recru[espece[e], annee[e]]) / sqrt(true_recru[espece[e],
annee[e]])
    p2.obs[e] <- e.obs[e] * e.obs[e]
    e.prd[e] <- (recruprd[e] - true_recru[espece[e], annee[e]]) / sqrt(true_recru[espece[e],
annee[e]])
    p2.prd[e] <- e.prd[e] * e.prd[e]
  }

##### Calcul des sommes des résidus au carré
    SS <- sum(p2.obs[1 :360])
    SS.prd <- sum(p2.prd[1 :360])

##### Modèle écologique

```

---

```

for (an in 1 :24) {
for (esp in 1 : 8) {
  true_recru[esp, an] <- exp(recru0[esp] ) * exp(
    d[cluster[esp],1] * aC12[cluster[esp]] * C12[an] +
    d[cluster[esp],2] * aCmx[cluster[esp]] * Cmx[an] +
    d[cluster[esp],3] * aS1[cluster[esp]] * S1[an] +
    d[cluster[esp],4] * aS2[cluster[esp]] * S2[an] +
    d[cluster[esp],5] * aQm1[cluster[esp]] * Qm1[an] +
    d[cluster[esp],6] * aQm2[cluster[esp]] * Qm2[an] +
    d[cluster[esp],7] * aQmx2[cluster[esp]] * Qmx2[an] )
  } }
##### Utilisation des résultats de l'inférence pour réaliser les pré-
dictions
for (anpred in 1 :4) {
for (esppred in 1 : 8) {
  true_recru_pred[esppred, anpred] <- exp(recru0[esppred]) * exp(
    d[cluster[esppred],1] * aC12[cluster[esppred]] * C12pred[anpred]
+
    d[cluster[esppred],2] * aCmx[cluster[esppred]] * Cmxpred[anpred]
+
    d[cluster[esppred],3] * aS1[cluster[esppred]] * S1pred[anpred] +
    d[cluster[esppred],4] * aS2[cluster[esppred]] * S2pred[anpred] +
    d[cluster[esppred],5] * aQm1[cluster[esppred]] * Qm1pred[anpred]
+
    d[cluster[esppred],6] * aQm2[cluster[esppred]] * Qm2pred[anpred]
+
    d[cluster[esppred],7] * aQmx2[cluster[esppred]] * Qmx2pred[anpred]
  )
  recrupred[esppred, anpred] ~ dpois(true_recru_pred[esppred, anpred])
} }
}

```



## Annexe C

# Code WinBUGS : modèle logPoisson à variabilité stationnelle

```
model{
##### Priors
  for (clust in 1 : 3) {
    aC12[clust] ~ dnorm(0, 0.01)
    aCmx[clust] ~ dnorm(0, 0.01)
    aS1[clust] ~ dnorm(0, 0.01)
    aS2[clust] ~ dnorm(0, 0.01)
    aQm1[clust] ~ dnorm(0, 0.01)
    aQm2[clust] ~ dnorm(0, 0.01)
    aQmx2[clust] ~ dnorm(0, 0.01)
  }
  for (esp in 1 :8) {
    effsta[esp,1] <- 0
  for (sta in 2 :8) {
    effsta[esp,sta] ~ dnorm(0, 0.001)
  } }
  effsais ~ dnorm(0, 0.001)
  for (esp in 1 :8) {
    effesp[esp] ~ dnorm(0, 0.01)
  }
##### Expression de l'effectif moyen attendu
  for (esp in 1 :8) {
  for (sta in 1 :8) {
  for (sais in 1 :2) {
    recru0[esp, sta, sais] <- effesp[esp] + effsta[esp,sta] + dsais[sais]* eff-
sais exprecru0[esp, sta, sais] <- exp(recru0[esp, sta, sais])
  } } }
  for (e in 1 : 3072) {
##### Modèle statistique de l'échantillonnage
    recru[e] ~ dpois(true_recru[espece[e], annee[e], station[e], saison[e]])
    recruprd[e] ~ dpois(true_recru[espece[e], annee[e], station[e], saison[e]])
  }
}
```

---

```
##### Expression des résidus, observés et prédits
  e.obs[e] <- (recru[e] - true_recru[espece[e], annee[e], station[e], saison[e]]) / sqrt(true_recru[espece[e], annee[e], station[e], saison[e]])
  p2.obs[e] <- e.obs[e] * e.obs[e]

  e.prd[e] <- (recruprd[e] - true_recru[espece[e], annee[e], station[e], saison[e]]) / sqrt(true_recru[espece[e], annee[e], station[e], saison[e]])
  p2.prd[e] <- e.prd[e] * e.prd[e]
}

##### Calcul des sommes des résidus au carré
  SS <- sum(p2.obs[1 :3072])
  SS.prd <- sum(p2.prd[1 :3072])

##### Modèle écologique
  for (an in 1 :24) {
  for (sta in 1 :8) {
  for (esp in 1 : 8) {
  for (sais in 1 :2) {

    true_recru[esp, an, sta, sais] <- exp(recru0[esp, sta, sais] ) * exp(
    d[cluster[esp],1] * aC12[cluster[esp]] * C12[an] +
    d[cluster[esp],2] * aCmx[cluster[esp]] * Cmx[an] +
    d[cluster[esp],3] * aS1[cluster[esp]] * S1[an] +
    d[cluster[esp],4] * aS2[cluster[esp]] * S2[an] +
    d[cluster[esp],5] * aQm1[cluster[esp]] * Qm1[an] +
    d[cluster[esp],6] * aQm2[cluster[esp]] * Qm2[an] +
    d[cluster[esp],7] * aQmx2[cluster[esp]] * Qmx2[an] )

  } } } }
}
```



## Annexe D

# Code WinBUGS : modèle logPoisson surdispersé

```
model{
##### Priors
  for (clust in 1 : 3) {
    aC12[clust] ~ dnorm(0, 0.01)
    aCmx[clust] ~ dnorm(0, 0.01)
    aS1[clust] ~ dnorm(0, 0.01)
    aS2[clust] ~ dnorm(0, 0.01)
    aQm1[clust] ~ dnorm(0, 0.01)
    aQm2[clust] ~ dnorm(0, 0.01)
    aQmx2[clust] ~ dnorm(0, 0.01)
  }
  for (esp in 1 :8) {
    effsta[esp,1] <- 0
    for (sta in 2 :8) {
      effsta[esp,sta] ~ dnorm(0, 0.01)
    }
    effsais ~ dnorm(0, 0.01)
    prec ~ dgamma(0.01, 0.01)
    V2 <- 1/prec
    for (esp in 1 :8) {
      effesp[esp] ~ dnorm(0, 0.01)
    }
##### Expression de l'effectif moyen attendu
    for (esp in 1 :8) {
      for (sta in 1 :8) {
        for (sais in 1 :2) {
          recru0[esp, sta, sais] <- effesp[esp] + effsta[esp,sta] + dsais[sais]* eff-
sais
          exprecru0[esp, sta, sais] <- exp(recru0[esp, sta, sais])
        } } }
    for (e in 1 : 3072) {
##### Modèle statistique de l'échantillonnage
```

```

recru[e] ~ dpois(true_recru[espece[e], annee[e], station[e], saison[e]])
recruprd[e] ~ dpois(true_recru[espece[e], annee[e], station[e], saison[e]])

##### Expression des résidus, observés et prédits

e.obs[e] <- (recru[e] - true_recru[espece[e], annee[e], station[e], saison[e]]) / sqrt(true_recru[espece[e], annee[e], station[e], saison[e]])
p2.obs[e] <- e.obs[e] * e.obs[e]

e.prd[e] <- (recruprd[e] - true_recru[espece[e], annee[e], station[e], saison[e]]) / sqrt(true_recru[espece[e], annee[e], station[e], saison[e]])
p2.prd[e] <- e.prd[e] * e.prd[e]
}

##### Calcul des sommes des résidus au carré

SS <- sum(p2.obs[1 :3072])
SS.prd <- sum(p2.prd[1 :3072])

##### Modèle écologique

for (an in 1 :24) {
for (sta in 1 :8) {
for (esp in 1 : 8) {
for (sais in 1 :2) {

true_recru[esp, an, sta, sais] <- exp(recru0[esp, sta, sais] ) * exp(
d[cluster[esp],1] * aC12[cluster[esp]] * C12[an] +
d[cluster[esp],2] * aCmx[cluster[esp]] * Cmx[an] +
d[cluster[esp],3] * aS1[cluster[esp]] * S1[an] +
d[cluster[esp],4] * aS2[cluster[esp]] * S2[an] +
d[cluster[esp],5] * aQm1[cluster[esp]] * Qm1[an] +
d[cluster[esp],6] * aQm2[cluster[esp]] * Qm2[an] +
d[cluster[esp],7] * aQmx2[cluster[esp]] * Qmx2[an] +
eps[esp, an, sta, sais] )

eps[esp, an, sta, sais] ~ dnorm(0, prec)

} } } }

##### Utilisation des résultats de l'inférence pour réaliser les prédictions

for (anpred in 1 :4) {
for (stapred in 1 :8) {

```

```
for (esppred in 1 : 8) {
for (saispred in 1 :2) {
  true_recrupred[esppred, anpred, stapred, saispred] <- exp(recru0[esppred,
stapred, saispred] ) * exp(
  d[cluster[esppred],1] * aC12[cluster[esppred]] * C12pred[anpred] +
  d[cluster[esppred],2] * aCmx[cluster[esppred]] * Cmxpred[anpred] +
  d[cluster[esppred],3] * aS1[cluster[esppred]] * S1pred[anpred] +
  d[cluster[esppred],4] * aS2[cluster[esppred]] * S2pred[anpred] +
  d[cluster[esppred],5] * aQm1[cluster[esppred]] * Qm1pred[anpred] +
  d[cluster[esppred],6] * aQm2[cluster[esppred]] * Qm2pred[anpred] +
  d[cluster[esppred],7] * aQmx2[cluster[esppred]] * Qmx2pred[anpred]
+
  epspred[esppred, anpred, stapred, saispred] )
  epspred[esppred, anpred, stapred, saispred] ~ dnorm(0, prec)
  recrupred[esppred, anpred, stapred, saispred] ~ dpois(true_recrupred[esppred,
anpred, stapred, saispred])
} } } }
}
```

## Annexe E

# Code WinBUGS : modèle à facteur latent partagé

```
model {  
  ##### Priors  
  for (t in 1 :24) {  
    Z[t] ~ dnorm(0, 1)  
  }  
  alpha1 ~ dunif(-1, 1)  
  alpha2 ~ dunif(-1, 1)  
  alpha3 ~ dunif(-1, 1)  
  alpha4 ~ dunif(-1, 1)  
  alpha5 ~ dunif(-1, 1)  
  alpha6 ~ dunif(-1, 1)  
  alpha7 ~ dunif(-1, 1)  
  alpha8 ~ dunif(-1, 1)  
  alpha9 ~ dunif(-1, 1)  
  
  sigma1 <- sqrt(1-pow(alpha1,2))  
  sigma2 <- sqrt(1-pow(alpha2,2))  
  sigma3 <- sqrt(1-pow(alpha3,2))  
  sigma4 <- sqrt(1-pow(alpha4,2))  
  sigma5 <- sqrt(1-pow(alpha5,2))  
  sigma6 <- sqrt(1-pow(alpha6,2))  
  sigma7 <- sqrt(1-pow(alpha7,2))  
  sigma8 <- sqrt(1-pow(alpha8,2))  
  sigma9 <- sqrt(1-pow(alpha9,2))  
  
  pX1 <- 1 / sigma1  
  pX2 <- 1 / sigma2  
  pX3 <- 1 / sigma3  
  pX4 <- 1 / sigma4  
  pX5 <- 1 / sigma5  
  pX6 <- 1 / sigma6  
  pX7 <- 1 / sigma7  
  pX8 <- 1 / sigma8  
  pX9 <- 1 / sigma9  
  
  a1 ~ dgamma(1, 1)  
  a2 ~ dnorm(0, 0.1)  
  
  b1 ~ dnorm(0, 0.1)  
  b2 ~ dnorm(0, 0.1)
```

```

c1[1] <- 0
c1[2] ~ dnorm(0, 0.1)
c2[1] <- 0
c2[2] ~ dnorm(0, 0.1)

d1[1] <- 0
d1[2] ~ dnorm(0, 0.1)
d2[1] <- 0
d2[2] ~ dnorm(0, 0.1)

##### Modèles statistique multinomial
for (e in 1 :360) {
  y[e, 1 :3] ~ dmulti(p[annee[e],saison[e],station[e],1 :3], N[annee[e]])
}

##### Expression des composantes principales sur l'hypersignal
for (an in 1 :24) {
  muX1[an] <- alpha1*Z[an] C12[an] ~ dnorm(muX1[an], pX1)
  muX2[an] <- alpha2*Z[an] C18[an] ~ dnorm(muX2[an], pX2)
  muX3[an] <- alpha3*Z[an] Cmx[an] ~ dnorm(muX3[an], pX3)
  muX4[an] <- alpha4*Z[an] S1[an] ~ dnorm(muX4[an], pX4)
  muX5[an] <- alpha5*Z[an] S2[an] ~ dnorm(muX5[an], pX5)
  muX6[an] <- alpha6*Z[an] Qm1[an] ~ dnorm(muX6[an], pX6)
  muX7[an] <- alpha7*Z[an] Qmx1[an] ~ dnorm(muX7[an], pX7)
  muX8[an] <- alpha8*Z[an] Qm2[an] ~ dnorm(muX8[an], pX8)
  muX9[an] <- alpha9*Z[an] Qmx2[an] ~ dnorm(muX9[an], pX9)

##### Expression des proportions relatives des groupes, via une
fonction de lien logit
for (sais in 1 :2) {
  for (sta in 1 :2) {
    p[an,sais,sta,3] <- exp(a1*Z[an] + b1 + c1[sais] + d1[sta]) /
      (1 + exp(a1*Z[an] + b1 + c1[sais] + d1[sta]))
    p[an,sais,sta,1] <- (1 - ((exp(a1*Z[an] + b1 + c1[sais] + d1[sta]) / (1
+ exp(a1*Z[an] + b1 + c1[sais] + d1[sta])))) /
      ((1 + (1/exp(a2*Z[an] + b2 + c2[sais] + d2[sta]))))
    p[an,sais,sta,2] <- 1 - p[an,sais,sta,3] - p[an,sais,sta,1]
  } } }
}

```



