

MEMOIRE

PAR

Nicolas Casajus

Changement climatique et assemblages de
poissons des cours d'eau français : modification
des aires de distribution et incertitude des
projections

En vue de l'obtention du Diplôme de
Master 2 Recherche mention Ecologie spécialité
Biostatistiques et modélisation



Responsable du master : **Sovan Lek**
Maître de stage : **Gaël Grenouillet**

Soutenu le **26 Juin 2008**

SOMMAIRE

1. Introduction.....	2
1.1. Changements globaux et impacts sur les écosystèmes.....	2
1.2. Ecosystèmes aquatiques et communautés piscicoles	2
1.3. Modèles de distribution d'espèces - Modèles de Circulation Globale.....	3
1.4. Incertitudes dans les projections futures - « Ensemble forecasting »	3
1.5. Objectifs du stage	4
2. Matériels	5
2.1. Sites d'étude	5
2.2. Données faunistiques.....	5
2.3. Descripteurs climatiques	5
2.4. Descripteurs environnementaux	5
3. Méthodes.....	6
3.1. Modélisation de la distribution des espèces	6
3.2. Analyse et quantification de la variabilité sur les projections.....	9
3.3. Construction d'un modèle consensuel.....	9
3.4. Impacts du changement climatique sur les aires de répartition.....	10
4. Résultats.....	11
4.1. Comparaison des modèles.....	11
4.2. Quantification de la variabilité.....	13
4.3. Impacts du changement climatique sur les aires de répartition.....	14
5. Discussion.....	17
5.1. Modélisation de la distribution des espèces	17
5.2. Variabilité dans les projections futures	17
5.3. Modèle de consensus - Ensemble forecasting.....	19
5.4. Impact du changement climatique sur les aires de répartition	19
5.5. Limites des modèles de distribution d'espèce.....	20
6. Conclusion et perspectives	22
7. Références bibliographiques	23
8. Annexe : Description et paramétrage des SDM.....	27
8.1. Méthodes de regression	27
8.2. Méthodes de classification	27
8.3. Méthodes de machine-learning	28

1. Introduction

1.1. Changements globaux et impacts sur les écosystèmes

Depuis plusieurs années, l'impact des activités anthropiques sur le climat global est largement admis et reconnu par la communauté scientifique. Le Groupe d'experts Intergouvernemental sur l'évolution du Climat (GIEC, ou « IPCC » pour Intergovernmental Panel on Climate Change), dans son 4^{ème} rapport d'évaluation « Changement climatique 2007 » (IPCC, 2007) dresse un bilan complet de l'état des connaissances scientifiques sur les changements globaux. Ces experts prévoient pour les années à venir une accélération des phénomènes climatiques extrêmes. Ces perturbations climatiques (augmentation de la température moyenne de l'air et des océans, augmentation du niveau des mers, fonte des glaciers) ont des répercussions importantes sur les écosystèmes. Sala et al. (2000) ont montré que le changement climatique serait le deuxième moteur d'érosion de la biodiversité, après les modifications du paysage.

Les organismes sont constamment soumis à des pressions environnementales. Face à des modifications de leur environnement, ils sont contraints de s'adapter rapidement. Pour cela, ils peuvent modifier leur phénologie/physiologie, en décalant par exemple la période de reproduction (Dunn et Winkler, 1999) ou en avançant la période de migration (Taylor, 2008). Ils peuvent également modifier leurs aires de répartition (McCarty, 2001) afin de coloniser des habitats plus favorables. Cependant, un changement trop rapide des paramètres abiotiques peut avoir des conséquences écologiques parfois catastrophiques, notamment parce que les espèces les plus sensibles n'ont pas le temps de s'adapter à leur environnement changeant (Araújo, 2006 ; Thomas et al., 2003).

1.2. Ecosystèmes aquatiques et communautés piscicoles

Les écosystèmes aquatiques dulçaquicoles sont particulièrement sensibles aux changements globaux. Leur structure et surtout leur fonctionnement sont étroitement liés aux facteurs climatiques. Par exemple, les régimes hydrologiques, qui déterminent entre autres la structure des biotopes, sont essentiellement régis par les températures et les précipitations. Ainsi, les perturbations climatiques observées actuellement modifient de manière drastique le fonctionnement de ces écosystèmes, en affectant les communautés, et notamment les peuplements piscicoles.

En effet, les poissons sont particulièrement sensibles à des modifications thermiques. Chez ces organismes poïkilothermes, la température joue un rôle fondamental dans la biologie, comme par exemple lors du développement larvaire (Mann, 1991), de la reproduction (Lappalainen, 2003), ou de la migration (Hembre et al., 2001). Ce facteur

intervient également dans la répartition spatiale et la structuration des communautés (Wolter, 2007). Il devient donc urgent d'anticiper la réaction des espèces face à ces changements afin de pouvoir identifier les plus sensibles et de mettre en place des mesures de protection.

1.3. Modèles de distribution d'espèces - Modèles de Circulation Globale

De nombreuses études ont été menées sur les impacts du changement climatique sur la distribution d'une grande diversité d'organismes (Thuiller, 2004 ; Pearson et Dawson, 2003 ; Elith et al., 2006). La plupart d'entre elles sont basées sur la même technique de modélisation, les 'modèles de distribution d'espèces' (ou « Species Distribution Models », SDM). Ce sont des modèles statistiques (ou algorithmiques) qui permettent de modéliser l'occurrence (ou l'abondance) d'une espèce grâce à différents prédicteurs environnementaux (Guisan, 2005). Dans l'étude du changement climatique, ces prédicteurs sont des variables climatiques, souvent associées à d'autres descripteurs environnementaux, dont les valeurs futures peuvent être prédites, notamment grâce à des 'modèles de circulation globale (« General Circulation Model », GCM). Ces derniers sont des outils qui permettent de simuler l'évolution future de différents paramètres climatiques (températures de l'atmosphère et des océans, précipitations), sous différentes hypothèses d'évolution potentielle des sociétés humaines (démographie humaine, économie, développement technologique), appelées scénario d'émission de gaz à effet de serre (GES) (Nakicenovic et Swart, 2000). Ainsi, à partir de ces différents outils (SDM, GCM et scénarios), il est possible de simuler la répartition future potentielle d'une espèce.

1.4. Incertitudes dans les projections futures - « Ensemble forecasting »

Il existe un grand nombre de SDM (ex. modèles statistiques, modèles algorithmiques, modèles rule-based), de GCM (ex. CGCM2, HadCM3, CSIRO2) et de scénarios (ex. A1Fi, A2, B1, B2). Les prédictions de la distribution future seront différentes selon le choix de ces trois éléments. Le paramétrage du SDM a également une forte influence sur les prédictions (Thuiller et al., 2004). Ainsi, les sources de variabilité dans les projections futures sont diverses : jeu de données, SDM, paramétrage du SDM, GCM, scénario (Thuiller, 2004 ; Thuiller et al., 2004 ; Araújo et al., 2005 ; Pearson et al., 2006).

Au cours de ces dernières années, un nouveau courant s'est développé : l'« Ensemble Forecasting » (Araújo et New, 2007). Il consiste à réaliser plusieurs prédictions en utilisant différentes combinaisons 'SDM - GCM - scénario' pour différents jeux de données initiaux. Ces différents modèles conduisent à une multitude de prédictions futures, lesquelles seront combinées par une méthode de consensus (ex. médiane, moyenne, moyenne pondérée, ACP).

Cette technique permet : i) d'éviter de choisir un SDM, GCM ou scénario souvent de manière arbitraire, ii) de prendre en compte la variabilité dans les projections futures et ainsi, iii) d'apporter des éléments robustes quant aux réponses des espèces au changement climatique ce qui faciliterait leur conservation et leur gestion futures (Guisan et Thuiller, 2005 ; Rodríguez, 2007).

1.5. Objectifs du stage

Ce stage s'inscrit dans le cadre du programme européen Euro-Limpacs (6^{ème} PCRD) dont l'objectif principal est d'évaluer l'impact des changements globaux sur les écosystèmes aquatiques continentaux européens.

Les objectifs principaux de cette étude sont : i) quantifier les différentes sources de variabilité (SDM, GCM, scénarios) existant dans les projections futures, ii) évaluer l'impact du changement climatique sur les aires de distribution de plusieurs espèces de poissons. Pour cela, plusieurs SDM sont couplés avec différents GCM et scénarios afin de modéliser la réponse au changement climatique de ces espèces.

2. Matériels

2.1. Sites d'étude

Les données ont été extraites de la base de données de l'ONEMA (Office National de l'Eau et des Milieux Aquatiques). 1110 stations de référence (c'est-à-dire peu perturbées par l'action de l'homme) correspondant à des tronçons de rivières ont été sélectionnées. Celles-ci sont distribuées sur l'ensemble du territoire français, à l'exception de la Corse, et réparties au sein de 9 unités hydrographiques.

2.2. Données faunistiques

Pour chacune de ces stations, des pêches électriques ont été réalisées. Les espèces de poissons présentes ont été déterminées, permettant d'obtenir des données de présence-absence sur l'ensemble des espèces rencontrées. Cependant, pour des contraintes liées aux techniques de modélisation utilisées, seules les espèces rencontrées dans au moins 25 stations ont été conservées dans la suite de l'analyse, soit un total de 35 espèces.

2.3. Descripteurs climatiques

Quatre variables climatiques ont servi à cette étude : température moyenne de l'air du mois le plus chaud (MTW), température moyenne de l'air du mois le plus froid (MTC), température annuelle moyenne de l'air (TAN) et précipitation annuelle moyenne (PAN). Les variables MTC et MTW étant fortement corrélées à la variable TAN, une nouvelle variable, l'amplitude thermique (TAM), a été créée de la manière suivante :

$$TAM = MTW - MTC$$

Ainsi, après transformation, les descripteurs climatiques ne comprenaient plus que TAN, TAM et log(PAN).

2.4. Descripteurs environnementaux

Pour étudier les assemblages d'espèces à une échelle régionale, Pearson et Dawson (2003) ont proposé de prendre en compte des descripteurs environnementaux non climatiques afin d'améliorer la précision des modèles. Six variables environnementales décrivant localement l'habitat ont donc été sélectionnées pour les 1110 stations : distance à la source (DS), superficie du bassin versant (SDB), altitude (ELE), pente (SLO), largeur du lit (WID) et profondeur (DEP). Certaines de ces variables étant fortement corrélées entre elles, deux transformations ont été nécessaires.

Premièrement, à partir des variables SDB et DS (transformées en logarithme), une analyse en composantes principales (ACP) centrée et réduite a été réalisée. L'axe 1 résumant

93% de l'inertie totale, une nouvelle variable synthétique G a été construite. Celle-ci peut être définie comme un gradient longitudinal (ou gradient amont-aval) et correspond aux coordonnées de chaque station sur l'axe 1 de l'ACP.

Dans un second temps, les variables DEP, WID et SLO ont permis d'élaborer la variable V, représentant la vitesse locale du courant dérivée de la formule de Chezy (Oberdorff et al., 2001) et calculée de la manière suivante :

$$V = \log (WID) + \log (DEP) + \log (SLO) - \log (WID + 2DEP)$$

Ainsi, après transformation, les descripteurs environnementaux ne comprenaient plus que G, V et log (ELE).

Ces variables environnementales étant fortement corrélées aux variables climatiques, un modèle additif généralisé (GAM) d'ordre 4 (Leathwick et al, 2006) a été construit pour chacun des 3 descripteurs environnementaux en fonction des variables climatiques. Les résidus de ces 3 modèles (resG, resV et resELE) ont été utilisés comme prédicteurs environnementaux. Ainsi, l'effet de l'habitat a pu être distingué de l'effet du climat dans la modélisation des distributions d'espèces.

3. Méthodes

3.1. Modélisation de la distribution des espèces

3.1.1. Séparation du jeu de données

Araújo et al. (2005) proposent d'évaluer les qualités prédictives d'un modèle sur un jeu de données indépendant. Pour cela, le jeu de données initial comprenant les 1110 stations a été séparé en deux (Fig.1a) : un jeu de calibration (70% du jeu de données initial) et un jeu de validation (les 30% restant). L'échantillonnage aléatoire des stations a été réalisé au sein de chaque bassin hydrographique. Afin de minimiser l'impact de cet échantillonnage aléatoire sur la fiabilité des résultats, cette séparation a été répétée 100 fois. Ainsi, 100 jeux de calibration et 100 jeux de validation ont été construits.

3.1.2. Choix des modèles de distribution d'espèce (SDM)

Huit méthodes statistiques regroupées en 3 catégories ont été retenues :

- Méthodes de régression : modèle additif généralisé (GAM, (Hastie et Tibshirani, 1990)), modèle linéaire généralisé (GLM, (McCullagh et Nelder, 1997)) et *multivariate adaptive regression splines* (MARS, (Friedman, 1991)),

- Méthodes de « machine-learning » : réseaux de neurones artificiels avec algorithme de rétropropagation (ANN, (Ripley, 1996)), *random forest* (RF, (Breiman, 2001)), *aggregated boosted tree* (ABT, (Leathwick et al., 2006)),
- Méthodes de classification : analyse factorielle discriminante (AFD, Fisher, 1936), arbre de régression (CART, (Breiman et al., 1984)).

Le détail de ces différentes méthodes est présenté dans l'annexe.

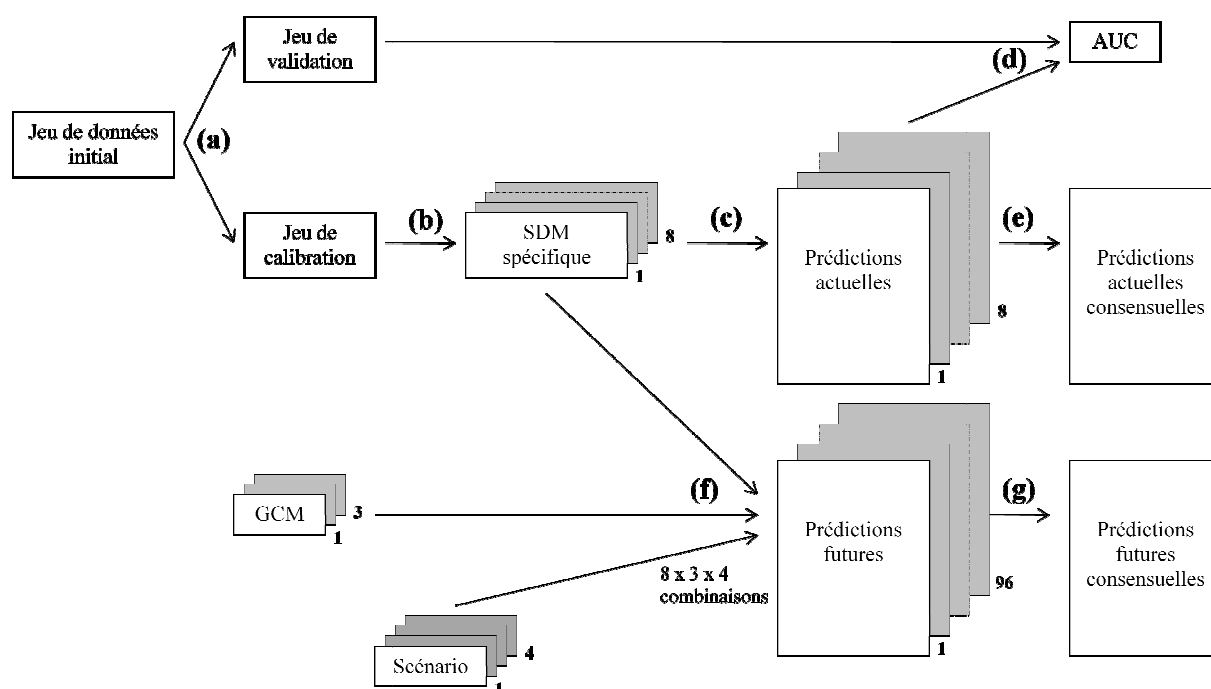


Figure 1 : Etapes de la modélisation pour une itération (un jeu de calibration et un jeu de validation) et une espèce. Ces différentes étapes conduisent à la création de 8 prédictions actuelles et 96 prédictions futures pour une itération, soit 800 prédictions actuelles et 9600 prédictions futures pour les 100 itérations. Les prédictions futures sont réalisées pour 2020, 2050 et 2080.

(a) Séparation du jeu de données initial, (b) Construction des modèles spécifiques, (c) Prédictions des probabilités d'occurrence actuelle, (d) Evaluation des modèles, (e) Construction des prédictions actuelles consensuelles, (f) Prédictions des probabilités d'occurrence future par couplage d'un modèle spécifique avec un GCM et un scénario, (g) Construction des projections futures consensuelles.

3.1.3. *Construction des modèles spécifiques*

Pour chacune des 35 espèces et à partir d'un jeu de calibration, 8 modèles statistiques ont été construits (Fig.1b), soit un total de 800 modèles par espèce (8 SDM x 100 jeux de calibration). Ces modèles reliaient la présence-absence de l'espèce considérée à 6 prédicteurs : les 3 variables environnementales (resG, resV et resELE) et les 3 variables climatiques (TAN, TAM et log(PAN)). Le paramétrage de chaque SDM est décrit dans l'annexe.

3.1.4. Prédictions des probabilités d'occurrence actuelles

Les modèles calibrés ont permis de calculer les probabilités d'occurrence actuelle (appelées par la suite *prédictions actuelles*) pour l'ensemble des 1110 stations (Fig.1c). Ainsi, pour chaque espèce, 800 probabilités d'occurrence actuelle furent obtenues pour chaque site.

Puis, elles ont été transformées en données binaires de présence-absence par application d'un seuil de transformation. Pour cela, une matrice de confusion regroupant les valeurs observées et les valeurs prédites (Tab.1) peut être construite pour différents seuils. Cette matrice permet de calculer différentes mesures de diagnostic (Tab.1). Le seuil retenu fut celui qui maximisait la somme 'sensibilité + spécificité' (Buisson et al., 2008b). En dessous ce seuil, l'espèce est considérée absente et au dessus, présente.

Tableau 1 : Matrice de confusion.

Nombres de sites	Absence observée	Présence observée
Absence prédite	a	b
Présence prédite	c	d

Pourcentage de bon classement

(« Accuracy ») : $(a+b) / (a+b+c+d)$

Sensibilité (mesure la capacité à prédire correctement les présences) : $d / (b+d)$

Spécificité (mesure la capacité à prédire correctement les absences) : $a / (a+c)$

3.1.5. Evaluation des modèles

L'évaluation des modèles se fait sur un jeu de données indépendant. Les jeux de validation ont donc été utilisés pour calculer 3 mesures du pouvoir prédictif obtenues à partir de la matrice de confusion : le pourcentage de bon classement (*accuracy*), la sensibilité et la spécificité (Tab.1). Cependant, ces mesures sont dépendantes du seuil de transformation choisi.

Une quatrième mesure de performances des modèles, indépendante du seuil, a été calculée (Fig.1d) : l'AUC ou *Area Under the Curve of a Receiver Operating Characteristic plot* (Hanley et al., 1982). Une valeur de 0.5 indique des qualités prédictives médiocres alors qu'une valeur proche de 1 indique de très bonnes qualités prédictives (Swets, 1988).

3.1.6. Prédictions de la distribution future des espèces

Les prédictions futures ont été générées en attribuant aux variables climatiques (TAN, TAM et log (PAN)) les valeurs futures pour les années 2020, 2050 et 2080 dans les modèles calibrés sur la période actuelle (Fig.1f). Ces valeurs ont été obtenues à partir de 3 GCM : le CGCM2 (Coupled Global Climate Model, développé au Canada), le CSIRO2 (Commonwealth Scientific and Industrial Research Organization, développé en Australie) et le HadCM3 (Hadley Center for Climate Prediction and Research's General Circulation

Model, développé au Royaume-Uni). Chacun de ces 3 GCM a été couplé à 4 scénarios d'émission de gaz à effet de serre (Nakicenovic et Swart, 2000) : A1Fi (scénario le plus dramatique considérant un usage intensif de combustibles fossiles), A2, B1 et B2 (scénario le plus optimiste).

Ainsi, pour chaque variable climatique et pour une année future donnée, 12 valeurs différentes étaient disponibles par station, résultant de la combinaison 'GCM-Scénario'.

Grâce aux seuils de transformation déterminés pour les prédictions actuelles, les probabilités d'occurrence futures ont pu être converties en présence-absence.

3.2. Analyse et quantification de la variabilité sur les projections

3.2.1. Cohérence entre les projections futures

Dans le but d'évaluer la variabilité dans les projections futures issues des 96 combinaisons 'SDM - GCM - Scénario', une analyse de consensus a été réalisée (Thuiller, 2004). Pour cela, une analyse en composante principale (ACP consensus) a été effectuée sur les prédictions futures binaires (Thuiller, 2004). Dans cette analyse, l'axe 1 appelé 'axe de consensus' représente la similarité dans les projections tandis que tous les autres axes, appelés 'axes de variabilité' synthétisent les dissemblances. Ainsi, si l'axe de consensus explique 100% de l'information, cela voudrait dire qu'il n'existe aucune variabilité dans les prédictions (Thuiller, 2004). Cette méthode permet donc d'évaluer la cohérence entre modèles, mais pas de séparer l'effet de chacune des composantes de la variabilité.

3.2.2. Quantification de la variabilité

Pour déterminer la part de la variabilité attribuable à chacune des trois sources (SDM, GCM et scénario), un GLM reliant les projections futures aux 3 sources de variabilité a donc été construit. Ainsi, en décomposant la déviance expliquée par le modèle, il est possible de calculer la part de la variabilité due à chaque facteur (X_i) par la formule suivante :

$$\% \text{variabilité due à } X_i = \frac{\text{Déviance expliquée par } X_i}{\text{Déviance expliquée par le modèle}}$$

3.3. Construction d'un modèle consensuel

Un des objectifs de ce travail étant d'étudier l'impact du changement climatique sur les communautés de poissons d'eau douce, 9600 prédictions futures pour une seule espèce ne permettraient pas d'apporter des informations précises sur le devenir réel de cette espèce. Araújo et al. (2005, 2007) ont montré qu'un modèle consensuel (basé sur la moyenne des prédictions de plusieurs modèles) pouvait posséder de meilleures qualités prédictives que tous

les autres modèles pris séparément. De plus, Marmion et al. (2008) ont comparé les qualités prédictives de 5 méthodes de consensus et ont montré que la méthode basée sur la moyenne de toutes les prédictions possédait les meilleures qualités prédictives. Ainsi, une moyenne sur l'ensemble des prédictions fut réalisée afin de faciliter l'interprétation des résultats et de la rendre plus fiable.

Ces calculs de moyenne ont été réalisés pour les prédictions actuelles (moyenne sur 800 prédictions par espèce, Fig.1e) et sur les projections futures (moyenne sur 9600 prédictions par espèce et par an, Fig.1g). Puis, ces nouvelles prédictions ont été transformées en données binaires de présence-absence en calculant un nouveau seuil. Ainsi, les qualités prédictives de ce modèle moyen ont pu être évaluées en recalculant les valeurs d'AUC, sensibilité, spécificité et *accuracy*.

3.4. Impacts du changement climatique sur les aires de répartition

A partir des prédictions moyennes réalisées sur l'ensemble des combinaisons 'SDM - GCM - scénario', les espèces à fort potentiel d'extinction ou au contraire à fort potentiel d'expansion ont été identifiées. Pour cela, les pourcentages de sites où l'espèce devrait disparaître (% sites perdus) ou apparaître (% sites gagnés) ont été calculés de la manière suivante :

$$\% \text{ sites perdus} = \frac{\text{Nombre de sites perdus}}{\text{Nombre de sites où l'espèce était présente}}$$

$$\% \text{ sites gagnés} = \frac{\text{Nombre de sites gagnés}}{\text{Nombre de sites où l'espèce était présente}}$$

Enfin, afin d'évaluer spatialement les modifications des aires de distribution, cette étude a mis l'accent sur le devenir de trois espèces dont les exigences écologiques sont très contrastées : la truite commune (*Salmo trutta fario*), le brochet (*Esox lucius*) et le barbeau fluviatile (*Barbus barbus*).

4. Résultats

4.1. Comparaison des modèles

4.1.1. Qualités prédictives

Les valeurs moyennes d'AUC (Tab.2) vont de 0.72 pour les CART à 0.86 pour le modèle moyen (consensus, Tab.2). A l'exception des CART, tous les modèles présentent des valeurs moyennes d'AUC supérieure à 0.8. De manière générale, les écart-types sont faibles même si les CART présentent le plus fort écart-type (0.10), alors que tous les autres modèles ont des écart-types voisins de 0.06.

Tableau 2 : Performances prédictives des modèles. Les valeurs données sont les moyennes et les écart-types (entre parenthèses) calculées sur les 35 espèces et sur les 100 itérations.

Modèle	AUC (\pm Sd)	Accuracy (\pm Sd)	Sensibilité (\pm Sd)	Spécificité (\pm Sd)
ABT	0.83 (\pm 0.06)	76.10 (\pm 8.29)	75.48 (\pm 9.74)	81.88 (\pm 10.23)
AFD	0.82 (\pm 0.06)	75.02 (\pm 8.80)	74.29 (\pm 10.59)	80.13 (\pm 10.88)
ANN	0.83 (\pm 0.06)	74.22 (\pm 10.25)	73.15 (\pm 11.94)	80.86 (\pm 10.06)
CART	0.72 (\pm 0.10)	73.23 (\pm 20.92)	73.96 (\pm 23.22)	65.45 (\pm 20.84)
GAM	0.84 (\pm 0.06)	76.87 (\pm 8.09)	76.29 (\pm 10.04)	80.84 (\pm 10.52)
GLM	0.84 (\pm 0.06)	76.49 (\pm 8.25)	75.92 (\pm 10.17)	80.66 (\pm 10.61)
MARS	0.83 (\pm 0.07)	74.50 (\pm 8.91)	73.81 (\pm 10.42)	79.80 (\pm 11.26)
RF	0.85 (\pm 0.06)	77.80 (\pm 8.02)	77.39 (\pm 9.56)	82.35 (\pm 9.65)
CONSENSUS	0.86 (\pm 0.05)	77.86 (\pm 7.45)	77.17 (\pm 8.91)	83.97 (\pm 9.69)

En ce qui concerne le pourcentage de bon classement (accuracy), les valeurs moyennes s'échelonnent de 73% (CART) à 78% (RF et modèle moyen). Le modèle moyen a un faible écart-type sur les valeurs d'accuracy (7.45%) contrairement aux CART (21%). Les mêmes résultats s'observent pour les deux autres mesures. Les RF et le modèle moyen présentent les plus fortes valeurs moyennes de sensibilité (77%) et de spécificité (respectivement, 82% et 84%).

Quelque soit la mesure considérée, les méthodes RF et modèle moyen présentent les plus fortes valeurs moyennes avec un faible écart-type alors que les CART ont les plus mauvaises qualités prédictives (faibles valeurs moyennes et écarts-types importants), notamment pour la spécificité. Les autres modèles sont intermédiaires aux CART et RF (et modèle consensus) et présentent peu de différences entre eux.

4.1.2. Projections futures

Les différences observées dans les performances prédictives des modèles se retrouvent également dans les prédictions actuelles (résultats non présentés), mais aussi dans les projections futures. Pour une même espèce, l'aire de répartition future modélisée sous une

combinaison ‘SDM - GCM - Scénario’ est plus ou moins différente de l’aire prédite par une autre combinaison.

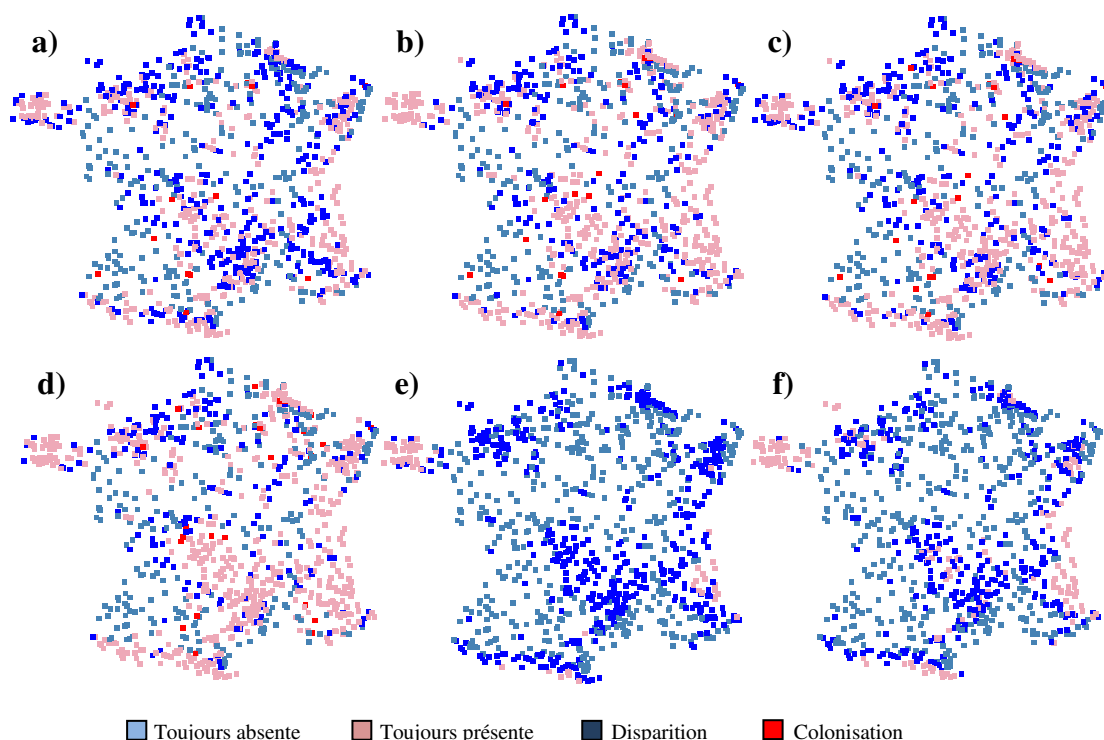


Figure 2 : Distributions futures potentielles de *Salmo trutta fario* sous différents SDM, GCM et scénario pour l’année 2080. **a)** Random Forest, HadCM3, scénario A1Fi, **b)** Random Forest, HadCM3, scénario B2, **c)** Random Forest, CGCM2, scénario A1Fi, **d)** Random Forest, CGCM2, scénario B2, **e)** GAM, HadCM3, scénario A1Fi, **f)** GAM, HadCM3, scénario B2.

Par exemple, la truite devrait modifier fortement sa distribution en 2080, mais l’intensité de ces modifications diffère selon les combinaisons (Fig.2). Tout d’abord, le scénario utilisé modifie l’intensité des extinctions locales puisque le nombre de sites où cette espèce serait absente en 2080 alors qu’elle y est aujourd’hui est plus élevé sous le scénario A1Fi, quelque soit le GCM et le SDM utilisés (comparaison Fig. 2a-b, Fig.2c-d, Fig.2e-f).

De la même façon, le GCM influence les changements dans la distribution de la truite (comparaison Fig. 2a-c, Fig.2b-d). La truite tendrait à disparaître d’un plus grand nombre de sites sous le GCM HadCM3 que sous le GCM CGCM2.

Enfin, le SDM utilisé provoque de grandes différences dans les prédictions de la distribution future de la truite (comparaison Fig. 2a-e, Fig.2b-f). D’une part, les GAM (Fig.2e-f) tendraient à prédire une plus forte disparition de la truite à l’échelle nationale que les RF (Fig.2a-b). De plus, les RF prédisent que quelques sites pourraient être colonisés par la truite alors que les GAM ne prédisent aucune colonisation potentielle.

Ainsi, l’exemple de la truite illustre bien les différentes sources de variabilité existant dans les projections futures.

4.2. Quantification de la variabilité

4.2.1. Cohérence entre les projections futures

L'ACP consensus réalisée sur les prédictions futures exprimées sous forme de données binaires (Fig.3) a permis de visualiser et de mesurer la cohérence entre les différents modèles pour les années 2020, 2050 et 2080. Deux résultats importants apparaissent.

Premièrement, toutes les combinaisons sont fortement corrélées entre elles et à l'axe 1 (axe de consensus), quelque soit l'année, à l'exception des combinaisons impliquant les ANN qui sont corrélées entre elles mais aussi à l'axe 2 (axe de variabilité).

Deuxièmement, il apparaît que la part consensuelle (c'est-à-dire, le pourcentage d'inertie apportée par l'axe 1) des prédictions futures diminue avec le temps : les prédictions tendent à s'accorder à 62.5% en 2020 (Fig.3a) alors que cette valeur de consensus n'est plus que de 52% en 2080 (Fig.3c).

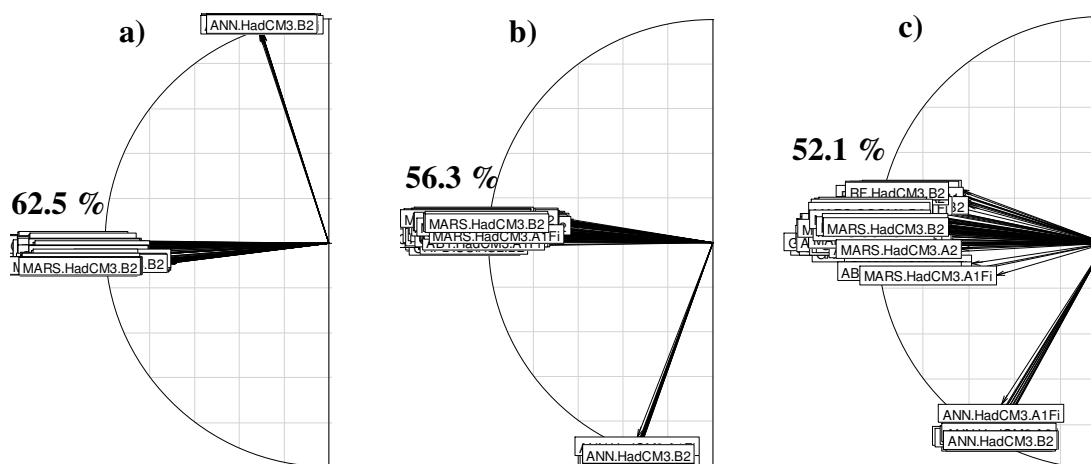


Figure 3 : Cercles de corrélation des variables des ACP consensus réalisées sur les projections futures a) 2020, b) 2050, c) 2080. Ces analyses ont été réalisées sur les prédictions futures exprimées en présence-absence, toutes espèces confondues, c'est-à-dire qu'une variable (combinaison 'SDM-GCM-scénario') contient 38850 valeurs binaires (35 espèces x 1110 stations). Les pourcentages indiquent la valeur de consensus.

4.2.2. Quantification de la variabilité

Afin de quantifier la part de variabilité attribuable aux 3 sources, un GLM par espèce et par année a été construit, reliant les prédictions futures sous forme de présence-absence aux 3 facteurs : SDM, GCM et scénario. Cependant, les ACP consensus ont montré que ces prédictions issues des combinaisons impliquant les ANN étaient fortement incohérentes avec celles issues des autres SDM. Pour cela, les prédictions issues des ANN ont été supprimées de cette analyse afin de limiter le biais lié à la variabilité uniquement due aux ANN.

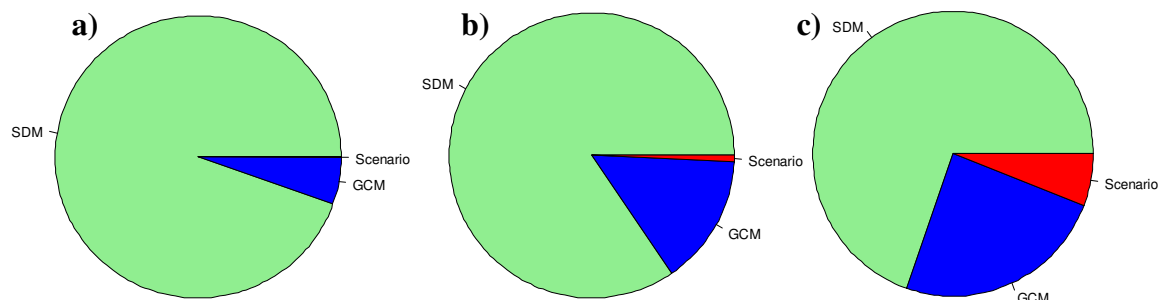


Figure 4 : Diagrammes sectoriels représentant la part de variabilité dans les projections futures expliquée par chaque facteur. a) 2020, b) 2050, c) 2080. La moyenne des pourcentages obtenus pour chacune des 35 espèces est représentée ici.

Quelque soit l'année, la part de variabilité due au SDM est majoritaire (> 69% pour les trois années, Fig.4) alors que celle due au scénario est très faible (< 6% pour les trois années). Le GCM explique une part intermédiaire de la variabilité dans les projections futures (de 5 à 25% selon l'année). On peut également remarquer que la part de variabilité due au SDM diminue progressivement au cours du temps, au profit de celle due au GCM et au scénario.

4.3. Impacts du changement climatique sur les aires de répartition

A partir des prédictions futures consensuelles, l'impact du changement climatique sur les aires de distribution des espèces peut être évalué. Globalement, le changement climatique devrait avoir un impact plus ou moins prononcé sur chacune des 35 espèces (Tab.3). Il est possible de regrouper les espèces en trois catégories.

Le premier groupe, contenant des espèces comme la truite commune (*Salmo trutta fario*), la lote de rivière (*Lota lota*), le chabot (*Cottus gobio*) ou encore l'épinoche (*Gasterosteus aculeatus*), est caractérisé par une forte perte d'habitat (47% en 2080 pour la truite et 84% en 2080 pour le chabot) et un gain d'habitat faible (5% en 2080 pour le chabot) voire nul (0% pour la truite). Ainsi, pour ces espèces, le changement climatique aurait un impact fortement négatif en diminuant leurs aires de distribution.

Le second groupe, incluant des espèces telles que le barbeau fluviatile (*Barbus barbus*), le blageon (*Leuciscus souffia*), le chevaine (*Leuciscus cephalus*), le goujon (*Gobio gobio*) ou encore le toxostome (*Chondrostoma toxostoma*) est caractérisé par un fort gain d'habitat (72% en 2080 pour le goujon et 41% pour le barbeau fluviatile) tandis que peu de sites seraient perdus (0.3% pour le barbeau fluviatile). Ces espèces ont un fort potentiel de colonisation et le changement climatique devrait donc avoir un effet positif sur leurs répartitions.

Enfin, le dernier groupe contient des espèces dont les modifications de l'aire de distribution sont peu prononcées. Parmi celles-ci figurent le saumon atlantique (*Salmo salar*), le hotu (*Chondrostoma nasus*), le brochet (*Esox lucius*) et la brème (*Abramis brama*). Pour

ces espèces, la perte d'habitat (9.6% pour le saumon en 2080) et le gain d'habitat (11.9% pour le saumon en 2080) se compensent globalement.

Tableau 3 : Nombre d'occurrences prédites actuellement et dans le futur (2020, 2050 et 2080), pourcentage de sites gagnés et pourcentage de sites perdus. Ces valeurs ont été calculées à partir des prédictions consensuelles.

Espèce	Nombre d'occurrences				% sites gagnés			% sites perdus		
	Actuel	2020	2050	2080	2020	2050	2080	2020	2050	2080
<i>Abramis brama</i>	341	373	402	386	5.6	10.9	9.8	3.2	6.7	8.8
<i>Alburnoides bipunctatus</i>	275	296	527	509	5.7	30.2	28	9.8	0	0
<i>Alburnus alburnus</i>	318	363	457	467	6.3	17.7	19.2	1.6	0.3	0.9
<i>Ameiurus melas</i>	185	259	486	459	10.2	33	30.4	10.8	2.2	3.8
<i>Anguilla anguilla</i>	450	545	582	691	14.4	20	36.5	0	0	0
<i>Barbatula barbatula</i>	613	675	749	735	15.1	31.8	35	2.1	3.6	8.5
<i>Barbus barbus</i>	305	348	545	634	5.7	29.9	41	1	0.3	0.3
<i>Barbus meridionalis</i>	244	327	454	666	10.6	24.7	48.8	3.7	1.6	0.4
<i>Blicca bjoerkna</i>	245	290	439	465	6.8	22.7	25.7	5.7	0.8	0.8
<i>Carassius carassius</i>	136	197	319	433	7	19.1	30.8	5.1	2.2	2.2
<i>Chondrostoma nasus</i>	186	181	305	247	2.1	13.0	7.8	12.9	0.5	5.9
<i>Chondrostoma toxostoma</i>	145	201	440	723	6.7	30.7	59.9	6.2	0.7	0
<i>Cottus gobio</i>	536	471	293	112	6.4	10.1	4.9	19.0	56.2	84.3
<i>Cyprinus carpio</i>	257	343	491	534	11.3	27.9	32.8	3.9	1.6	1.2
<i>Esox lucius</i>	348	356	414	379	3.5	11.4	9.7	5.5	6	12.4
<i>Gasterosteus aculeatus</i>	214	130	116	98	2.6	3.6	3.2	50	60.7	67.8
<i>Gobio gobio</i>	652	706	911	974	16.8	59.2	71.8	3.5	1.8	1.1
<i>Gymnocephalus cernuus</i>	167	175	188	153	2.2	3.8	1.6	7.8	9.0	17.4
<i>Lampetra planeri</i>	398	412	222	173	7.4	3.4	2.8	9.8	50.3	61.6
<i>Lepomis gibbosus</i>	252	382	576	606	16.3	38.1	41.4	4	1.2	0.4
<i>Leuciscus cephalus</i>	537	614	805	859	13.8	46.8	56.2	0.4	0	0
<i>Leuciscus leuciscus</i>	351	364	473	465	4.2	16.2	15.9	5.4	0.3	2
<i>Leuciscus souffia</i>	238	313	753	882	10.9	59.1	73.9	8.4	0	0
<i>Lota lota</i>	132	122	127	95	3	4.6	4.6	29.5	37.9	62.1
<i>Perca fluviatilis</i>	448	476	548	448	8.9	20.8	15.3	6.9	8.5	22.5
<i>Phoxinus phoxinus</i>	602	648	663	715	12.8	20.1	30.3	3.2	6.8	6.8
<i>Pungitius pungitius</i>	248	264	221	160	3.9	3.6	1.2	7.3	23.4	39.5
<i>Rhodeus amarus</i>	270	272	404	366	3	16.3	13	8.5	1.1	4.8
<i>Rutilus rutilus</i>	493	532	652	641	8.1	26.4	24.8	2.2	0.8	1
<i>Salmo salar</i>	116	109	155	257	0.7	6.8	15.3	12.1	25	9.5
<i>Salmo trutta fario</i>	776	693	534	408	0	0	0	10.7	31.2	47.4
<i>Sander lucioperca</i>	114	153	243	222	5.4	14.2	11.9	13.2	10.5	9.6
<i>Scardinius erythrophthalmus</i>	321	329	554	544	5.1	31.1	30.8	10	3.7	6.2
<i>Thymallus thymallus</i>	140	122	297	302	1.1	17	17.5	20.7	5.7	5.7
<i>Tinca tinca</i>	406	450	620	610	9.5	32.4	30.5	5.7	3.4	2.7

Afin d'étudier plus précisément la répartition spatiale des modifications dans les aires de distribution, trois espèces réagissant de façon contrastée au changement climatique ont été choisies.

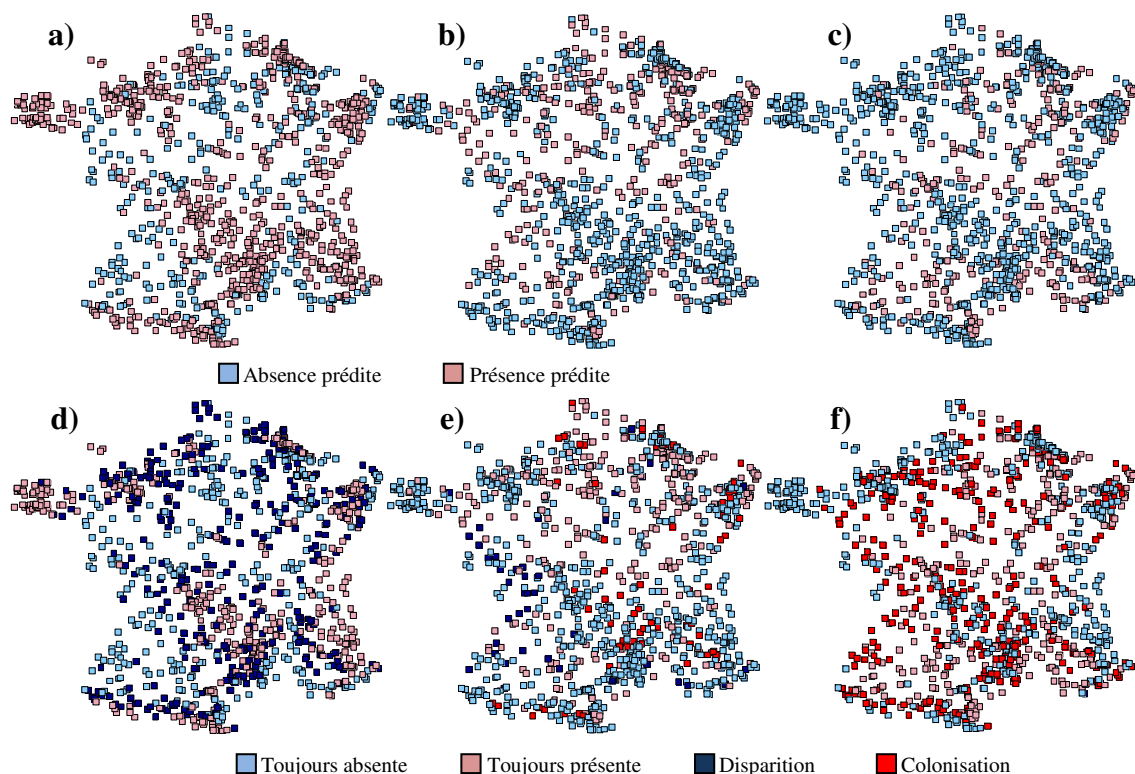


Figure 5 : Distributions prédites actuellement (a, b, c) et en 2080 (d, e, f) pour *Salmo trutta fario* (a, d), *Esox lucius* (b, e) et *Barbus barbus* (c, f). Ces cartes ont été obtenues à partir des prédictions actuelles consensuelles et des projections futures consensuelles transformées en données binaires de présence-absence.

La truite, espèce d'eau froide largement répartie sur l'ensemble du territoire national (Fig.5a), serait affectée de manière négative par le changement climatique. En 2080, elle devrait disparaître d'un grand nombre de sites, situés dans les parties aval des cours d'eau et devrait trouver refuge dans les tronçons de l'amont (Fig.5d). En effet, il semblerait que les cours d'eau des Pyrénées, des Alpes, du Jura et du Massif Central, ainsi que ceux de la Bretagne constituent des habitats qui resteraient potentiellement favorables pour la truite. Le modèle consensuel ne prévoit aucun phénomène de colonisation. A l'opposé, le barbeau fluviatile (Fig.5c-f), espèce plus tolérante que la truite quant aux conditions thermiques, devrait profiter du changement climatique pour étendre son aire de répartition. Le modèle prédit une colonisation des stations situées plus en amont que celles où il est présent actuellement (Fig.5f), comme dans les Pyrénées et le Massif Central. Il pourrait également coloniser des zones localisées plus en aval, comme par exemple certaines stations de la façade atlantique. Il semblerait que le barbeau ne disparaisse que d'un très faible nombre de sites. Enfin, le brochet (Fig.5b-e), poisson d'eau calme aurait un comportement intermédiaire aux deux espèces précédentes. Il serait susceptible de disparaître de quelques stations (Fig.5e), essentiellement situées à l'ouest du territoire, et de coloniser des stations plus en amont (piémont pyrénéen, Jura et Massif Central). Cependant, les phénomènes de disparition et de colonisation seraient moins intenses que dans le cas de la truite et du barbeau.

5. Discussion

5.1. Modélisation de la distribution des espèces

La réponse de 35 espèces de poissons communes dans les rivières françaises et confrontées au changement climatique a été modélisée en utilisant huit méthodes différentes. A l'exception des arbres de classification et de régression (CART), toutes les méthodes possèdent de bonnes qualités prédictives, avec un AUC moyen supérieur à 0.82.

D'autres études ont montré que les CART peuvent présenter de faibles performances prédictives (Thuiller, 2003 ; Austin, 2007 ; Marmion et al., 2008). Le sur-apprentissage en serait la principale raison. En effet, les CART sont fortement spécialistes du jeu de calibration et ils ne parviennent pas à prédire de façon correcte sur un jeu de données indépendant. Ce résultat n'est pas étonnant puisque cette technique sélectionne à chaque partitionnement des données, le prédicteur qui permet la meilleure discrimination des groupes créés, expliquant ainsi ce sur-apprentissage du jeu de calibration.

A l'opposé, les *random forest* (RF) présentent les meilleures qualités prédictives parmi les huit méthodes comparées. Ce résultat a été mis en évidence dans de nombreuses autres études (Prasad et al., 2006 ; Cutler et al., 2007 ; Peters et al., 2007). La forte capacité prédictive de cette technique de modélisation résulterait de leurs propriétés algorithmiques. En effet, le modèle final résulte de la combinaison d'un grand nombre de modèles construits à partir d'échantillons aléatoires du jeu de calibration (Prasad et al., 2006). Ainsi, à la fin du processus de construction des RF, les modèles obtenus sont des modèles consensuels. Or, nous avons montré que le modèle consensuel moyen construit à partir de tous les modèles possède de fortes capacités prédictives et qu'il est même meilleur que tous les autres SDM pris séparément. Le fait que le RF soit déjà un modèle consensuel explique vraisemblablement ses bonnes qualités prédictives.

Dans l'ensemble, les six autres méthodes (ABT, AFD, ANN, GAM, GLM et MARS) semblent être cohérentes en termes de qualités prédictives. Cependant, de légères différences entre modèles en termes d'AUC peuvent engendrer de grandes différences entre les projections futures issues de ces modèles (Thuiller, 2003).

5.2. Variabilité dans les projections futures

Dans cette étude, un des objectifs principaux est d'étudier la variabilité existant entre les prédictions futures issues de l'ensemble des modèles construits.

L'ensemble des combinaisons 'SDM - GCM - scénario' sont fortement cohérentes entre elles, à l'exception des combinaisons impliquant les réseaux de neurones artificiels (ANN). Ceci signifie que les prédictions futures issues des combinaisons impliquant les ANN

sont très différentes de celles issues des combinaisons impliquant les autres SDM. Néanmoins, elles ne sont pas opposées, car elles ne sont pas corrélées inversement aux prédictions issues des autres combinaisons. Le fait que seules les combinaisons utilisant les ANN divergent des autres permet d'avancer le problème du sur-apprentissage. Cependant, ce problème n'est pas apparu lors de la comparaison des qualités prédictives des modèles car, lors de la construction des ANN, une technique particulière d'optimisation a été utilisée. En effet, pour une espèce, plusieurs ANN avec différents paramétrages ont été construits et seul celui possédant le meilleur AUC fut conservé. Ceci pourrait expliquer le problème de sur-apprentissage rencontré, mais pas mis en évidence lors de l'évaluation de la qualité prédictive des SDM.

La valeur de consensus (pourcentage de similarité), calculée entre les prédictions futures est comprise entre 52.1% et 62.5%. Thuiller (2004) a procédé à la même analyse de cohérence entre les projections futures pour 1350 espèces européennes de plantes. Il trouve une valeur de consensus très proche avec un pourcentage de similarité de 56.1%. La part de variabilité existant entre les projections futures reste très importante. De plus, la valeur de consensus diminue au cours du temps, ce qui signifie que la variabilité augmente avec le temps. Ainsi, plus les projections seront réalisées dans un futur éloigné, et plus les prédictions issues des différentes combinaisons tendront à diverger entre elles. Cela peut être le résultat de la divergence croissante des conditions climatiques issues des 12 combinaisons « GCM - scénario » au cours du temps.

Les travaux actuels s'intéressent uniquement à la variabilité globale existant entre les prédictions. Ils ne cherchent pas à quantifier les différentes origines de cette incertitude. Cette étude est donc originale dans le sens qu'elle propose une analyse plus poussée de cette variabilité. En effet, la part relative de variabilité due à chaque facteur (SDM, GCM et scénario) a pu être quantifiée à l'aide de GLM. La variabilité due au SDM est majoritaire, quelque soit l'année, alors que celle due au scénario est minime. Le GCM explique une part intermédiaire de cette variabilité. La forte contribution des SDM à cette variabilité est probablement due à la grande diversité de techniques et de paramétrages utilisés. En effet, trois catégories de méthodes ont été utilisées : des méthodes de régression (GAM, GLM et MARS), des méthodes de machine-learning (ABT, ANN et RF) et des méthodes de classification (AFD et CART). Les techniques d'apprentissage sont extrêmement différentes d'une catégorie à l'autre, mais également au sein d'une catégorie. De plus, certains SDM ne font aucun présupposé sur la relation entre la réponse et les prédicteurs (par exemple, RF et ANN), tandis que d'autres font l'hypothèse d'une relation linéaire (GLM, AFD). Enfin, le nombre de paramètres à calibrer est très différent selon le SDM, ce qui peut conduire à des

prédictions plus différentes pour les SDM ayant de nombreux paramètres, comme par exemple, pour les MARS et ANN.

Enfin, il est apparu que la part de variabilité due au SDM diminue au cours du temps au profit des deux autres sources de variabilité. Comme pour la diminution de la valeur de consensus entre les prédictions futures, l'augmentation au cours du temps des différences entre les valeurs futures des descripteurs climatiques issues des différents GCM et scénario pourrait en être la raison.

5.3. Modèle de consensus - Ensemble forecasting

Afin d'évaluer l'impact du changement climatique sur la distribution des espèces de poissons, l'élaboration d'un modèle unique a été retenu pour simplifier les interprétations. Une solution classique aurait été de ne retenir que le SDM qui présente les meilleures qualités prédictives (Olden et Jackson, 2002 ; Segurado et Araújo, 2004), en l'occurrence les random forest dans cette étude. Cependant, il aurait encore fallu choisir de façon arbitraire un GCM et un scénario. Par exemple, l'étude aurait pu être centrée sur le scénario le plus pessimiste (A1Fi) et le GCM européen (HadCM3) comme cela est fait très communément (Thuiller et al., 2006 ; Araújo et Luoto, 2007).

Pour réduire l'incertitude liée aux projections futures, de nombreuses études (Thuiller, 2004 ; Araújo et al., 2005 ; Araújo et New, 2007 ; Marmion et al., 2008) ont développé des méthodes de consensus. Parmi ces méthodes, on retrouve la moyenne (ou la médiane) réalisée sur les prédictions issues de tous les modèles, ou simplement sur celles issues des modèles les plus performants. Une autre méthode, l'ACP consensus peut également permettre de diminuer cette incertitude. Marmion et al. (2008) ont montré qu'un modèle consensuel construit en moyennant les prédictions issues de l'ensemble des modèles présente des performances prédictives meilleures que celles des modèles pris séparément, mais aussi meilleures par rapport à d'autres méthodes de consensus.

Dans cette étude, deux méthodes de consensus ont été appliquées : la moyenne sur l'ensemble des prédictions et la moyenne pondérée par les valeurs d'AUC correspondantes. Entre ces deux méthodes, le modèle moyen simple possède les meilleures qualités prédictives. Cette méthode de consensus a donc été utilisée pour évaluer l'impact du changement climatique sur la distribution des espèces.

5.4. Impact du changement climatique sur les aires de répartition

Les résultats obtenus dans cette étude suggèrent que l'ensemble des 35 espèces de poissons seraient potentiellement affectées par le changement climatique. Certaines espèces,

notamment les espèces d'eau froide (truite, chabot, lamproie), verraient leurs aires de distribution se réduire et se maintiendraient uniquement dans les parties les plus apicales des cours d'eau. A l'opposé, quelques espèces thermophiles (barbeau, chevaine) tendraient à accroître leurs aires de distribution en colonisant des stations situées plus en amont.

Ces résultats sont assez cohérents avec des études menées en Amérique du Nord (Eaton et Scheller, 1996 ; Mohseni et al. 2003 ; Chu et al., 2005). En effet, Eaton et Scheller (1996) estiment que la grande majorité des poissons d'eau froide perdraient 50% de leurs habitats qui leur sont favorables actuellement. Pour les espèces thermophiles, toutes les études ne s'accordent pas entre elles. Certaines prédisent que les poissons d'eau chaude pourraient perdre 14% d'habitats favorables (Eaton et Scheller, 1996) alors que d'autres estiment un gain d'habitat pouvant aller jusqu'à 31% (Mohseni et al., 2003).

Ainsi, contrairement aux espèces cryophiles, aucune tendance nette ne semble se dégager pour les espèces thermophiles. Nos résultats suggèrent que certaines espèces d'eau chaude (brème, perche) verraient leur probabilité d'occurrence diminuer, alors qu'on s'attendrait au contraire à une extension de leur aire de répartition. Ce phénomène peut être facilement expliqué. Ces espèces affectionnent particulièrement les eaux lentes voire stagnantes (Bruslé et Quignard, 2001) et seraient incapables de coloniser des cours d'eau plus rapides, même si les conditions thermiques leur sont favorables. De la même manière, elles ne colonisent pas les stations situées en aval, car les températures y sont trop élevées. Buisson et al. (2008a) ont montré que pour certaines espèces (dont la brème et la perche), la répartition spatiale pouvait être influencée en grande partie par les descripteurs environnementaux non climatiques, comme la position sur le gradient amont-aval.

5.5. Limites des modèles de distribution d'espèce

Dans l'ensemble, les résultats de cette étude sont assez cohérents avec les travaux déjà réalisés. Néanmoins, la démarche de modélisation utilisée dans cette étude peut être critiquable sur plusieurs points. En effet, les SDM impliquent plusieurs hypothèses sous-jacentes.

Tout d'abord, cette démarche considère que toutes les espèces ont le même pouvoir de colonisation. Or, les capacités de dispersion sont étroitement liées aux caractéristiques biologiques des espèces (Collingham et Huntley, 2000).

De plus, la notion de fragmentation des cours d'eau n'a pas été prise en compte dans la modélisation. Ainsi, une espèce serait susceptible de coloniser une station ne présentant aucune connexion avec celles où elle est présente actuellement.

Un autre élément à prendre en compte est le fait que d'autres facteurs que les températures et les précipitations vont être modifiés dans le futur. Les changements dans le régime des précipitations vont vraisemblablement modifier l'hydromorphologie des cours d'eau, alors que cette dernière joue un rôle important dans la structuration des communautés pisciaires. Or, dans cette étude, l'indice de vitesse du courant ne varie pas sous les scénarios de changement climatique. Il serait donc intéressant de formuler des hypothèses quant à l'évolution future de cette variable.

Enfin, les interactions biotiques (compétition, prédation) n'ont pas non plus été prises en compte dans cette étude. Même si la structuration des assemblages de poissons semble être principalement gouvernée par des facteurs physiques, et que les relations biotiques jouent un rôle minime dans cette structuration (Grossman et al., 1998), certains auteurs conseillent de les intégrer dans les modèles (Araújo et Luoto, 2007), afin d'en améliorer les performances et la précision.

6. Conclusion et perspectives

Cette étude a permis d'évaluer l'impact du changement climatique sur la répartition spatiale de plusieurs espèces de poissons. Elle propose dans un premier temps, une démarche permettant d'étudier de manière approfondie la variabilité existant entre les projections futures issues de différents modèles. L'innovation majeure de ce travail réside dans la quantification des différentes sources d'incertitude (SDM, GCM et scénario climatique).

Puis, à partir d'un modèle consensuel construit sur l'ensemble des projections futures, les modifications de la distribution des espèces ont pu être étudiées. Ainsi, cette étude a permis l'identification des espèces les plus sensibles au changement climatique (notamment les espèces cryophiles), mais aussi la localisation des milieux nécessitant une protection prioritaire, comme les cours d'eau d'amont des Pyrénées, des Alpes et du Massif Central.

Cette étude pourra servir de base de travail à de nombreuses autres études. Par exemple, l'évaluation de l'impact du changement climatique pourra être étudiée non plus à l'échelle spécifique mais pourra portée sur les assemblages de poissons. Il serait en effet intéressant de savoir si ces assemblages vont être profondément modifiés en conduisant à une homogénéisation.

Il conviendrait également d'étudier les modifications apportées sur les caractéristiques fonctionnelles des assemblages en s'intéressant aux traits fonctionnels des espèces. Ceci permettrait l'identification des traits d'histoire de vie les plus affectés par le changement climatique.

Enfin, les relations phylogénétiques entre espèces devraient être prises en compte dans les modèles. En effet, il a été montré que l'aspect phylogénétique, associés à des données écologiques permettaient de mieux expliquer la répartition actuelle des espèces (Wiens et Donoghue 2004 ; Jakob et al., 2007). Ainsi, la réponse des espèces au changement climatique serait mieux appréhendée.

7. Références bibliographiques

- Araújo MB, Whittaker RJ, Ladle RJ, Erhard M (2005) Reducing uncertainties in projections of extinction risk from climate change. *Global Ecology and Biogeography*. 14, 529-538.
- Araújo MB, Thuiller W, Pearson RG (2006) Climate warming and the decline of amphibians and reptiles in Europe. *Journal of Biogeography*. 33, 1712-1728.
- Araújo MB, Luoto M (2007) The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*. 16, 743-753.
- Araújo MB, New M (2007) Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*. 22-1, 42-47.
- Austin PC (2007) A comparison of regression trees, logistic regression, generalized additive models, and multivariate adaptive regression splines for predicting AMI mortality. *Statistics in Medicine*. 26, 2937-2957.
- Breiman LJ, Friedman J, Olshen R, Stone C (1984) Classification and Regression Trees. Wadsworth. Belmont, CA.
- Breiman LJ (2001) Random forests. *Machine Learning*. 45, 5-32.
- Bruslé J, Quignard JP (2001) Biologie des poissons d'eau douce européenne. *Collection Aquaculture-Pisciculture*. Editions TEC & DOC.
- Buisson L, Blanc L, Grenouillet G (2008a) Modelling stream fish species distribution in a river network: the relative effects of temperature versus physical factors. *Ecology of Freshwater Fish*. 17, 244-257.
- Buisson L, Thuiller W, Lek S, Puy L, Grenouillet G (2008b) Climate change hastens the turnover of stream fish assemblages. *Global Change Biology*. Sous presse.
- Chu C, Mandrak NE, Minns CK (2005) Potential impacts of climate change on the distributions of several common and rare freshwater fishes in Canada. *Diversity and Distributions*. 11, 299-310.
- Collingham YC, Huntley B (2000) Impacts of habitat fragmentation and patch size upon migration rates. *Ecological Applications*. 10, 131-144.
- Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ (2007) Random forest for classification in ecology. *Ecology*. 88, 2783-2792.
- Dunn OP, Winkler DW (1999) Climate Change Has Affected the Breeding Date of Tree Swallows Throughout North America. *Proceedings: Biological Sciences*. 266, 2487-2490.
- Eaton JG, Scheller RM (1995) Effects of climate warming on fish thermal habitat in streams of the United States. *Limnology and Oceanography*. 41, 109-1115.
- Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, McC. Overton J, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberon J, Williams S, Wisz MS, Zimmermann NE (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*. 29, 129-151.
- Elith J, Leathwick J (2007) Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. *Diversity and Distributions*. 13, 265-275.

- Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*. 24, 38-49.
- Fisher RA (1936) The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 7, 179-188.
- Friedman JH (1991) Multivariate adaptive regression splines. *Annals of Statistics*. 19, 1-141.
- Grossman DG, Ratajczak RE, Jr, Crawford M, Freeman MC (1998) Assemblage organization in stream fishes: Effects of environmental variation and specific interactions. *Ecological Monographs*. 63, 395-420.
- Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*. 8, 993-1009.
- Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 143, 29-36.
- Hastie T, Tibshirani RJ (1990) Generalized Additive Models. *Monographs on Statistics and Applied Probability*. vol. 43. Chapman and Hall, London.
- Hembre B, Arnekleiv JV, L'Abée-Lund JH (2001) Effects of water discharge and temperature on the seaward migration of anadromous brown trout, *Salmo trutta*, smolts. *Ecology of Freshwater Fish*. 10, 61-64.
- IPCC (2007) Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.
- Jakob SS, Ihlow A, Blattner RF (2007) Combined ecological niche modelling and molecular phylogeography revealed the evolutionary history of *Hordeum marinum* (Poaceae) - niche differentiation, loss of genetic diversity, and speciation in Mediterranean Quaternary refugia. *Molecular Ecology*. 16, 1713-1727.
- Lappalainen J, Dörner H, Wysujack K (2003) Reproduction biology of pikeperch (*Sander lucioperca* (L.)) – a review. *Ecology of Freshwater Fish*. 12, 95-106.
- Lawler JJ, White D, Neilson RP, Blaustein AR (2006) Predicting climate-induced range shifts: model differences and model reliability. *Global Change Biology*. 12, 1568-1584.
- Leathwick JR, Elith J, Francis MS, Hastie T, Taylor P (2006) Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine Ecology Progress Series*. 321, 267-281.
- McCarty JP (2001) Ecological consequences of recent climate change. *Conservation Biology*. 15, 320-331.
- McCullagh P, Nelder JA (1997) Generalized Linear Models. *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.
- Mann RHK (1991) Growth and production. In: Cyprinid Fishes - Systematics, Biology and Exploitation (eds IJ Winfield & JS Nelson). Chapman & Hall, London. 457-482.
- Marmion M, Parviainen M, Luoto M, Heikkinen RK, Thuiller W (2008) Evaluation of consensus methods in predictive species distribution modeling. *Diversity and Distribution*. doi: 10.1111/j.1472-4642.2008.00491.x.

- Mohseni O, Stefan HG, Eaton JG (2003) Global warming and potential changes in fish habitat in US streams. *Climatic Change*. 59, 389-409.
- Muñoz J, Felicísimo AM (2004) Comparison of statistical methods commonly used in predictive modeling. *Journal of Vegetation Science*. 15, 285-292.
- Nakicenovic N, Swart R (2000) Special Report on Emission Scenarios. Intergovernmental Panel on Climate Change. *Cambridge University Press*, Cambridge.
- Oberdorff T, Hugueny B, Chessel D (2001) A probabilistic model characterizing fish assemblages of French rivers: a framework for environmental assessment. *Freshwater Biology*. 46, 399-415.
- Olden JD, Jackson DA (2002) A comparison of statistical approaches for modeling fish species distributions. *Freshwater Biology*. 47, 1976-1995.
- Pearson RG, Dawson TP (2003) Predicting the impacts of climate change on the distribution of species: are bioclimatic envelope models useful? *Global Ecology and Biogeography*. 12, 361-371.
- Pearson RG, Thuiller W, Araújo MB, Martinez-Meyer E, Brotons L, McClean C, Miles L, Segurado P, Dawson TP, Lees DC (2006) Model-based uncertainty in species range prediction. *Journal of Biogeography*. 33, 1704-1711.
- Peters J, De Baets B, Verhoest NEC, Samsonc R, Degroeve S, De Becker P, Huybrechts W (2007) Random forests as a tool for ecohydrological distribution modeling. *Ecological Modelling*. 207, 304-318.
- Prasad AM, Iverson LR, Liaw A (2006) Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems*. 9, 181-199.
- R Development Core Team (2007) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Ripley BD (1996) Pattern Recognition and Neural Networks. Cambridge University Press. Cambridge.
- Rodríguez JP, Brotons L, Bustamante J, Seoane J (2007) The application of predictive modelling of species distribution to biodiversity conservation. *Diversity and Distributions*. 13, 243-251.
- Sala OE, Chapin FS III, Armesto JJ, Berlow E, Bloomfield J, Dirzo R, Huber-Sanwald E, Huenneke LF, Jackson RB, Kinzig A, Leemans R, Lodge DM, Mooney HA, Oesterheld M, LeRoy Poff N, Sykes MT, Walker BH, Walker M, Wall DH (2000) Global Biodiversity Scenarios for the Year 2100. *Science*. 287, 1770-1774.
- Segurado P, Araújo MB (2004) An evaluation of methods for modeling species distributions. *Journal of Biogeography*. 31, 1555-1568.
- Swets K (1988) Measuring the accuracy of diagnostic systems. *Science*. 240, 1285-1293.
- Taylor SG (2008) Climate warming causes phenological shift in Pink Salmon, *Oncorhynchus gorbuscha*, behavior at Auke Creek, Alaska. *Global Change Biology*. 14, 229-235.
- Thomas CD, Cameron A, Green RE, Bakkenes M, Beaumont LJ, Collingham YC, Erasmus BFN, Ferreira de Siqueira M, Grainger A, Hannah L, Hughes L, Huntley B, van Jaarsveld AS, Midgley GF, Miles L, Ortega-Huerta MA, Peterson AT, Phillips OL, Williams SE (2004) Extension risk from climate change. *Nature*. 427, 145-148.

- Thuiller W (2003) BIOMOD – optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology*. 9, 1353-1362.
- Thuiller W, Araújo MB, Lavorel S (2003) Generalized models vs. classification tree analysis: Predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science*. 14, 669-680.
- Thuiller W (2004) Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology*. 10, 2020-2027.
- Thuiller W, Araújo MB, Pearson RG, Whittaker RJ, Brotons L, Lavorel S (2004) Uncertainty in predictions of extinction risk. *Nature*. 430, 34.
- Thuiller W, Lavorel S, Sykes MT, Araújo MB (2006) Using niche-based modelling to assess the impact of climate change on tree functional diversity in Europe. *Diversity and Distributions*. 12, 49-60.
- Wiens JJ, Graham CH (2004) Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution*, 18, 639-644.
- 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.

8. Annexe : Description et paramétrage des SDM

8.1. Méthodes de regression

8.1.1. GLM (Generalized Linear Model)

Extension de la régression linéaire classique, cette technique permet de modéliser différents formes de distribution. Ici, la régression logistique (distribution binomiale) a été utilisée grâce à la fonction *glm* du package **MASS**. Les prédicteurs étaient sous forme linéaire ou quadratique et aucune interaction entre variables ne fut considérée. Les variables ont été sélectionnées (combinaison des méthodes « backward stepwise selection » et « forward stepwise selection ») selon le critère d'AIC (Akaike's Information Criterion) grâce à la fonction *stepAIC* du même package.

8.1.2. GAM (Generalized Additive Model)

Permet de modéliser des relations plus complexes que les GLM. Il s'agit d'une méthode non-paramétrique qui réalise un ajustement local des données grâce à des fonctions de lissage. Cependant, elle ne considère que les effets additifs des variables.

Ici, la fonction *gam* du package **gam** a été choisie avec une distribution binomiale de l'erreur (fonction de lien : logit). Comme pour les GLM, les variables ainsi que leurs formes (linéaire ou spline d'ordre 2) ont été sélectionnées selon le critère d'AIC.

8.1.3. MARS (Multivariate Adaptive Regression Splines)

Méthode d'ajustement adaptatif par régression non-linéaire. Elle est basée sur un partitionnement des données en sous-groupes pour lesquels un ajustement local par lissage est effectué grâce à des fonctions de base non paramétriques.

Le package **mda** a été utilisé. Les fonctions nécessaires sont issues du script de Elith et Leathwick (2007) « mars.public.functions.3.1.R ». La fonction *mars.glm* de ce script crée le modèle tout en donnant des probabilités d'occurrence comprise entre 0 et 1 (ce qui n'était pas le cas avec la fonction *mars* du package **mda**).

8.2. Méthodes de classification

8.2.1. AFD (Analyse Factorielle Discriminante)

Basée sur la construction de combinaisons linéaires des prédicteurs (fonctions discriminantes) permettant la discrimination de plusieurs groupes prédéfinis et l'affectation de nouveaux individus à un groupe (prédiction). Ces fonctions discriminantes maximisent la variance extra-groupe tout en minimisant la variance intra-groupe.

Dans cette étude, la fonction *lda* de la librairie **MASS** a été utilisée en incluant dans le modèle les formes linéaires et quadratiques des 6 prédicteurs. Les interactions entre prédicteurs n'ont pas été prises en compte.

8.2.2. CART (Classification And Regression Tree)

Technique de classification non-paramétrique basée sur un partitionnement récursif binaire des données. A chaque nœud de l'arbre est attribué le prédicteur le plus discriminant. Tout comme l'AFD, cette méthode tend à maximiser la variance extra-groupe et à minimiser la variance intra-groupe. Aucun présupposé sur la relation entre réponse et prédicteurs n'est fait.

Ce modèle a été réalisé grâce à la fonction *rpart* de la librairie **RPART**. La technique de post-élagage a permis de minimiser le sur-apprentissage. Différents élagages ont été réalisés et le modèle élagué possédant les meilleures qualités prédictives a été retenu.

8.3. Méthodes de machine-learning

8.3.1. ABT (Aggregated Boosted Tree)

Cette technique est basée sur la construction de séquences d'arbres de décision. Au fur et à mesure, les arbres deviennent de plus en plus performants grâce à une sur-pondération des objets mal classés et à une sous-pondération des objets bien classés (algorithme AdaBoost). Plusieurs séquences de tailles différentes sont construites, et les prédictions sont ensuite combinées par votes pondérés.

L'algorithme ABT se trouve dans le package **gbmplus** (fonction *gbm*). Le modèle a été construit avec une distribution de Bernoulli et sans interaction entre les variables. Les séquences initiales comportaient 3000 arbres.

8.3.2. ANN (Artificial Neural Network)

Méthode connexiste supervisée permettant d'établir des règles de classement pour un ensemble de groupes connus. Elle ne fait aucun présupposé sur la relation entre réponse et prédicteurs et permet de modéliser n'importe quel type de relation. Le perceptron multicouche avec algorithme de rétropropagation a été utilisé. Au cours des différentes phases d'apprentissage, les règles de classement sont améliorées afin de réduire l'erreur.

La fonction *newff* de la librairie **AMORE** permet de créer le réseau. Pour cette étude, le nombre de neurones a été fixé à 6 neurones en couche d'entrée, 12 neurones en couche cachée et 1 neurone en couche de sortie. La fonction sigmoïde a été choisie comme fonction de transfert. Le taux d'apprentissage a été fixé à 0.0001 ou à 0.01 et le momentum à 0.5 ou

0.95. Ainsi, 4 différents paramétrages ont été réalisés : celui offrant les meilleures qualités prédictives (AUC calculé sur le jeu de validation) fut retenu. Le nombre d'itérations (ou d'apprentissage), paramètre de la fonction *train* de cette même librairie, a été fixé à 3000.

8.3.3. RF (Random Forest)

Technique basée sur la construction d'un grand nombre d'arbre de classification (forêt). Chaque arbre est construit à partir d'un échantillon aléatoire des observations, et à chaque nœud, un échantillon de prédicteurs est tiré aléatoirement parmi l'ensemble des prédicteurs. Le prédicteur permettant le meilleur partitionnement des données est conservé. Les arbres de décision ne sont pas élagués (arbres les plus grands possibles) et sont agrégés par moyenne (« model-averaging approach »).

Cette technique a été réalisée grâce à la fonction *randomForest* du package **randomForest**. 300 arbres ont été construits et le nombre de variables choisies à chaque nœud fut de $\sqrt{6}$.